

Plane Detection Based Object Recognition for Augmented Reality

Aparicio Carranza¹, Juan Estrella¹, Syed R. Zaidi², Harrison Carranza³

¹New York City College of Technology of the City University of New York
300 Jay Street, Brooklyn, NY, USA

acarranza@citytech.cuny.edu, juan.estrella@mail.citytech.cuny.edu

²Bronx Community College of the City University of New York
2155 University Avenue, Bronx, NY, USA
syed.zaidi@bcc.cuny.edu

³Vaughn College of Aeronautics and Technology
8601 23rd Avenue, East Elmhurst, NY, USA
harrison.carranza@vaughn.edu

Abstract - Machine Learning and Deep Learning; subfields of Artificial Intelligence have been implemented in applications that can detect and label objects and figures such as a chair, a person or even some kinds of animals. We have developed an application that makes use of Augmented Reality, a more modest goal of Plane Detection. The purpose of this app is to label detected objects with tags and relay the information to people who are visually impaired. A significant first step towards our goal is to concentrate on Plane Detection. This technology can benefit individuals in public places by helping them reach their destination safely. The app will guide individuals in need to prevent from colliding with walls and objects by recognizing and navigating plane surfaces and physical objects.

Keywords: Augmented Reality, Artificial Intelligence, Deep Learning, Machine Learning, Object Recognition

1. Introduction

Augmented Reality (AR) and Artificial Intelligence (AI) are two of the most important frontiers of technological development. However, as it stands, these two fields are not as integrated as they're supposed to be. However, with future research and development, the potential for more effective integration is attainable. AR is a technology that superimposes digital content, generated by computers, over real-world environments; it provides an enhanced view of the real world. AR can provide an aid in visualization by simulating 3D objects that can be manipulated [1]. For example, as shown in Fig.1, the Object Recognition (OR) technology can recognize objects and label them with tags, combine that with Augmented Reality and the visually impaired will be more aware of their visual surroundings. Furthermore, AR can portray features of large real-world facilities on a smaller scale [2].

AI aims to make computer programs or machines to learn and make decisions. The most widely used approach to AI is called Narrow AI, which uses a “narrowly “defined skilled set to achieve high level outcomes with limited problems (such as IBM’s chess playing computer, “Big “Blue”). It uses algorithms and statistical models to learn from data sets to classify things and to make predictions; this is known as Machine Learning (ML) [3]. A subfield of AI is Computer Vision (CV), which deals with the capabilities of machines to recognize and interpret the visual world from images on videos and real-world examples. Modern CV techniques are based on Deep Learning (DL), a modern approach to ML based on Neural Networks, to accurately identify and classify objects. The versatility of AI significantly benefits society in various realms, such as marketing, manufacturing, security, and healthcare. It is particularly advantageous in the healthcare industry, enabling faster diagnosis in contrast to humans [4]. In addition, it may one day provide self-driving-services for the disabled. AI plays a big role in building of intelligent, adaptive interfaces, identifying gestural inputs, object recognition and tracking. Manipulation of the virtual environment through voice commands will be explored in the future. For example, Visually Impaired people have difficulty getting to a specified location in an unknown environment. While human helpers may not always be available, a Holographic Guide, with real-time destination routing and voice capabilities, increases the likelihood that they will arrive at their desired destinations efficiently [5].

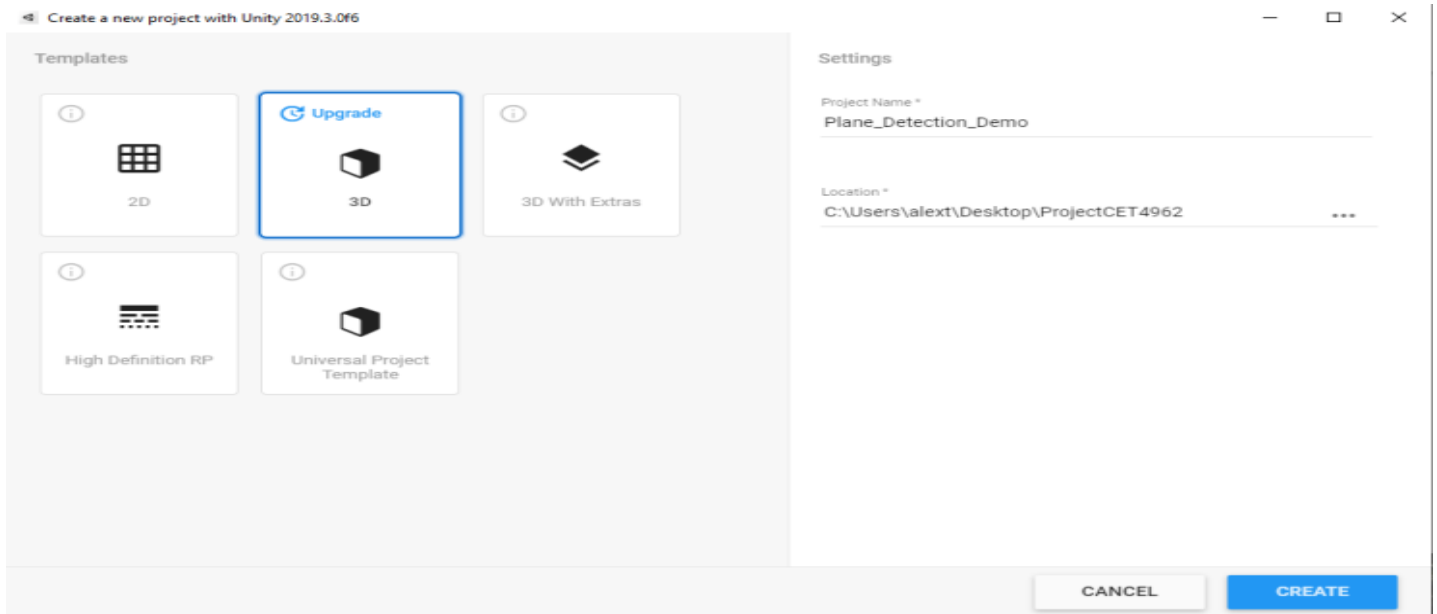


Fig. 1: 3D project selection in Unity

Artificial Intelligence and Augmented Reality are ideally suited to one another. In fact, AR relies on AI to be effective. Computer vision is used on AR for recognition of gestures, such as eye tracking, hand gestures and body movements to control applications. However, there is no standardized approach for developing AR applications that involves AI techniques such as Machine Learning and Object Recognition. Developers must explore alternatives to integrate OR and AI. Developers find it difficult to integrate AI with AR because there is a lack of proper practices, resources, and tools to develop their projects [6]. There are many ways to associate AI with AR, but developers must find effective ways to combine tools and approaches to craft their projects. However, it is not clear which one integrates better with the AR development platforms [1].

Plane Detection is a subset of Object Recognition and is part of its foundational elements. Plane detection scans and uses the depth in an environment. When processing is complete it visually displays the distance between objects and people. With this technology featured in Augmented Reality, virtual objects can and will be anchored onto real world environments to be displayed for all to see. However, at the early stages of the plane detection method, texture information is mainly used but this approach may fail when a plane has inconsistent color or texture. The technology has been around for quite some time as a remedy to devices with limitations such as depth cameras have; “depth maps offer one clue and turns out to be effective in texture challenging situations” [7]. There is an increase use in consumer depth cameras, e.g. Microsoft Kinect, Intel Realsense d400, and the depth-base plane segmentation and plane detection have become popular. Depth maps represent spatial information of each point within a scene and these points connect as clustering points to refine their distance estimation from the origin. Indoor and outdoor structures crafted by man are generally dominated by different shapes of planes. Such planes carry orientation and size data of the 3D objects within the scene. Further, 3D reconstruction can be simplified by detecting these planes and setting up the piecewise planar model of the indoor and outdoor structures. Plane detection techniques have been widely used in robot navigation systems and computer vision for object recognition [8].

The remainder of this paper is organized as follows: Section 2, introduces the Advantages of Plane Detection Based Object Recognition; Section 3, discusses the Disadvantages of Plane Detection Based Object Recognition; Section 4, discusses Object Recognition and Augmented Reality Combined; Section 5, presents the Procedure and Detailed Implementation of the plane detection app. Finally, Section 6 ends with concluding statements and future work.

2. Advantages of Plane Detection Based Object Recognition

Computer Vision provides machines with the ability to see and visually sense the world around them, which is similar to how humans use their eyes. It uses automatic extraction, analysis and understanding of useful information from one or several images. An important aspect of Computer Vision is Plane Detection. It has become a crucial feature in the process of building markerless augmented reality [5]. It helps developers understand better the geometry of the environment around them and place 3D content into the world [5]. Both Computer Vision and Plane Detection are the foundations of Augmented Reality. In order to use Augmented Reality effectively, the user must interact with computer generated and real objects and surfaces. One of the principle challenges of Augment Reality is to keep track of physical environments such as walls, floors, objects and people; and effectively in real time compute their relation to digital objects and surfaces [8].

Object Detection (OD) or Recognition has been around for years but is becoming more apparent across a range of industries now more than ever before. Object Detection is emerging in a wide range of industries, with use cases ranging from personal security to productivity in the workplace. Facial detection is one form of OD - it can be utilized as a security measure to allow only certain people into a highly classified area of a government building. It can be used to count the number of people present within a business meeting to automatically adjust other technical tools that will help streamline the time dedicated for that meeting. In such collaboration, business leaders will be able to count the number of participants within a meeting to help them automate the process further and monitor room usage to ensure spaces are being used properly. The possibilities are endless when it comes to future use cases for object detection. For instance, sports broadcasting will utilize this technology in instances such as detecting when a football team is about to make a touchdown and notify fans via their mobile phone or at home. In crowded places, a relatively new “people counting” method that detects heads rather than bodies and motion will allow for more accurate detection to take place, which will enable even more applications for the security industry. The future of object detection has massive potential across a wide range of industries. Markers to trigger augmented reality may be complex as well as simple [9].

3. Disadvantages of Plane Based Object Recognition

By checking the points’ normal vectors, depth-based plane detection can be done in real-time, but it has a problem with poor precision and low accuracy. In addition, the normal vector can be affected by noise points when it is developed using a small number of fixed points (e.g., 2). The normal vector for boundary points may also be inaccurate if the number is too large. In the condition of robotic navigation, accuracy is not as crucial as detection speed. Under these circumstances, normal vector use is adequate. However, in the implementation of guiding visually impaired people with a human navigational system, the most important feature is detection accuracy. Algorithms must be advanced to provide almost accurate results when recognizing objects. A lot of calibration research and time is poured into object recognizing devices and their software to achieve precise results [10].

Due mainly to developments advances in Convolutional Neural Networks (CNN) and Deep Learning, computer vision has significantly improved. Building on pre-trained neural nets and adjusting the last throughput layer, computers label and classify images. It was only a few years ago that the recognition and classification of an unknown objects inside an image was considered a major challenge. Such recognition of objects is not only possible but has been turned into products by IBM and Google. However, Object detection still presents other major challenges beyond simple image classification. Algorithms for Object detection require speed as well as accuracy for real-time video processing. Several important improvements have increased algorithms speed from the 0.02 frames per second (fps) of R-CNN to the impressive 155 fps of Fast You Only Look Once (YOLO) [4].

Yet another important kind of object detection system is based on a unified one-state approach. Such single-shot detectors completely locate and then classify objects with just a single pass. This substantially decreases test time. YOLO is a kind of single-shot detector. It starts by creating out a grid over the image and permits each cell to detect a limited number of varying sized objects. Every time YOLO decides a true object is present, the grid cell related to the center of that object is held responsible for predicting this object’s center. To ensure that all classification and localization happens inside a single process a complex, multi-term loss function is used. Fast YOLO is a variety of this method that has obtained

speeds of 155 fps (frames per second). However, at higher speeds accuracy for localization and classification drops off sharply [4].

The final objective of modern object recognition algorithms is to balance accuracy and speed. There are various design schemes beyond the basic detection framework to achieve this outcome. 164 layers (very deep network) are used in Inception-ResNet-V2. The accuracy is very good, but it is exceeded by VGG-16 frameworks in speed. In addition, YOLOv3 permits varying image resolutions (low resolution gives higher speed while high resolution improves accuracy but decreases speed) [4].

Objects can appear in many sizes and aspect ratios in real-world applications. In order to capture such variable objects, a number of different techniques are used. Multiple Region of Interests (MRoI) is evaluated in relation to anchor boxes in addition to their prediction at each position. Anchor boxes sizes and shapes are variable across range of scales and aspect ratios. In this way, bounding box coordinates do not have to be adjusted that much during the task of localization. In addition, the faster R-CNN updated region proposal network uses a sliding window that moves across the targets convolutional feature map in order to create good potential RoI. Finally, in order to initialize regions of interest, single shot detectors utilize anchor boxes [4].

Another hurdle for good object recognition is the lack of annotated data sets. Object detection and object recognition have different demands. For example, while ground truth examples for object detection data sets may have up to 100 classes, good image classification data sets need upwards to 100,000 classes. One way to meet this demand is through crowd-sourcing image classification tags. Even so, the process of building a working data set is very arduous. We have personally experienced the COCO data set and although it can be precise, overtime it becomes very slow due to so much image classification [4].

Several challenges remain which relate to limited data, multiple scales, speed, class imbalance and dual priorities. The balance between accuracy and speed remains a problem when doing a top-level classification with real-time detection design decisions. Another problem is picking out small objects that are bunched together and partially occluded. Using three dimensional bounding cubes instead of 2-dimensional bounding boxes is yet another enhancement. Finally, video tracking can be improved by assuming frame continuity rather than processing every frame by itself [4].

4. Object Recognition and Augmented Reality Combined

In order to make augmented reality work good object recognition is needed. We need to identify form, shape, position, and location for a variety of objects. The final product is an AR enhancement of the normal world with digital overlays. One possible application of augmented reality would be to assist in New York City to Con Edison's street repair crews by showing a 3D enhanced version of all the conduits, heat, and steam pipes and electrical lines on a superimposed street image. Another example is a holographic avatar that would be a guide for those who are visually impaired, which can guide them to locations safely with directional indicators.

A special programming language is used with augmented reality to integrate digital information with real world objects. A variety of applications present themselves which include the following: GPS based driving, education, medicine, and military applications. Automobile driving and pedestrian street navigation can be made faster and safer when GPS based AR routes are overlaid on streets and roads. Using the phone's camera in combination with the GPS, the users see the selected route over the live view of what is in front of them. Education would benefit when audio, video, text, and graphics can be overlaid on the student's real-world environment. Medicine can also be improved by combining a surgeon's real-world experience with X-rays, MRI's and relevant diagnostic information. Military training and flying high performance aircraft are enhanced by combining AR object recognition along with gesture recognition [11].

5. Procedure and Detailed Implementations

In this section we present detail steps of our implementations in a sequential order:

1. Get right Project Setting (Player) as shown in Fig. 2 - Remove Vulcan from graphic API's, Check off Multithreaded Rendering, Change Package Name, Change Minimum API Level to the closest version of Android used in the project

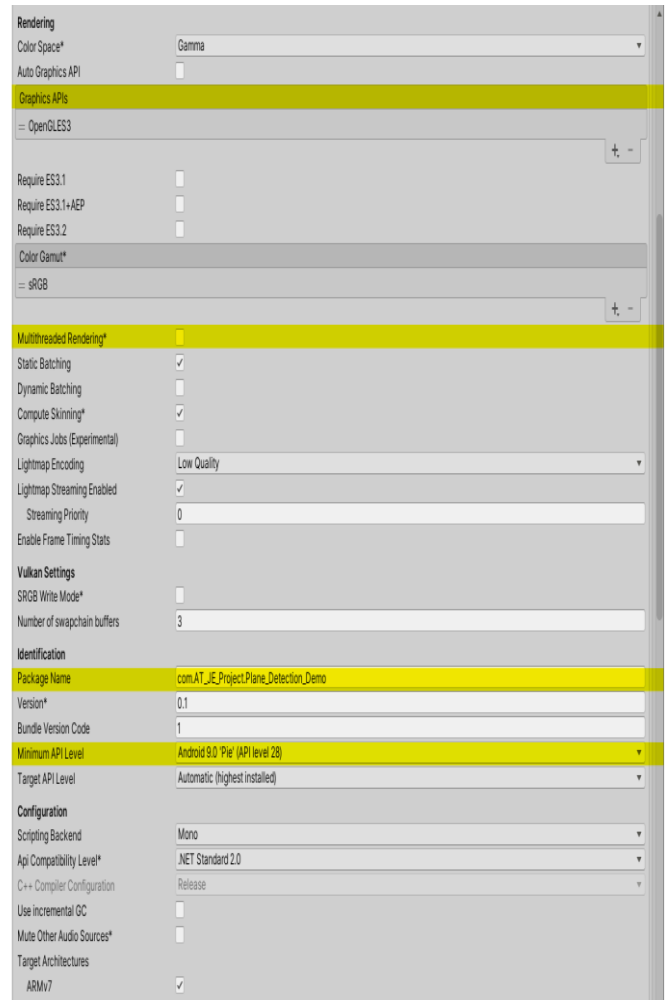
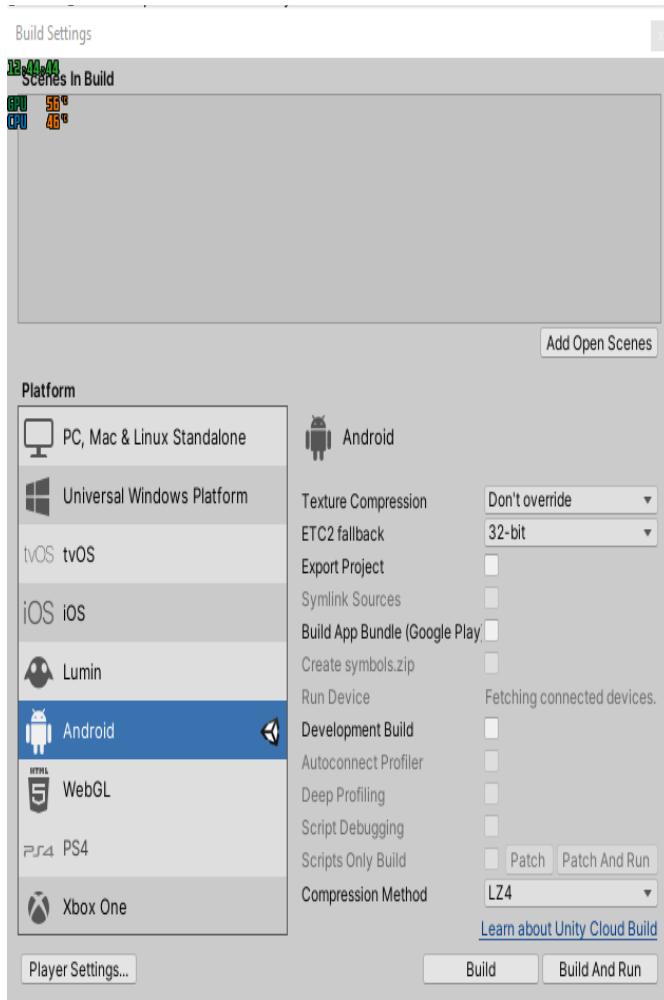


Fig. 2: Modify Build settings for the platform that the project is going to be built on

2. Select packages needed for Android program development as shown in Fig. 3: AR Foundation, AR Subsystems, ARCore XR Plugin, AR Session Origin and AR Session.

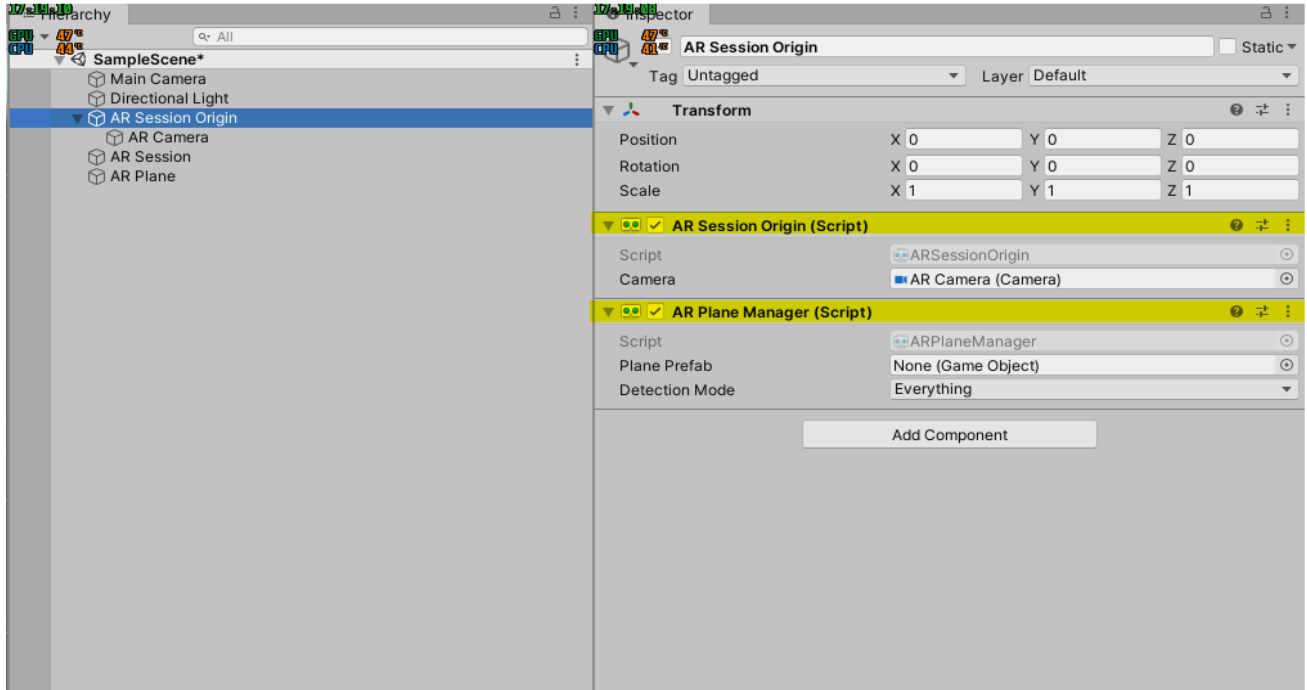


Fig. 3: Add AR Session and AR Plane Manager in AR Session Origin

3. Add following components to the AR Plane, the exercise is shown in Fig. 4: AR Plane, AR Plane Mesh Visualizer, Mesh Collider, Mesh Filter, Mesh Renderer, Line Renderer.

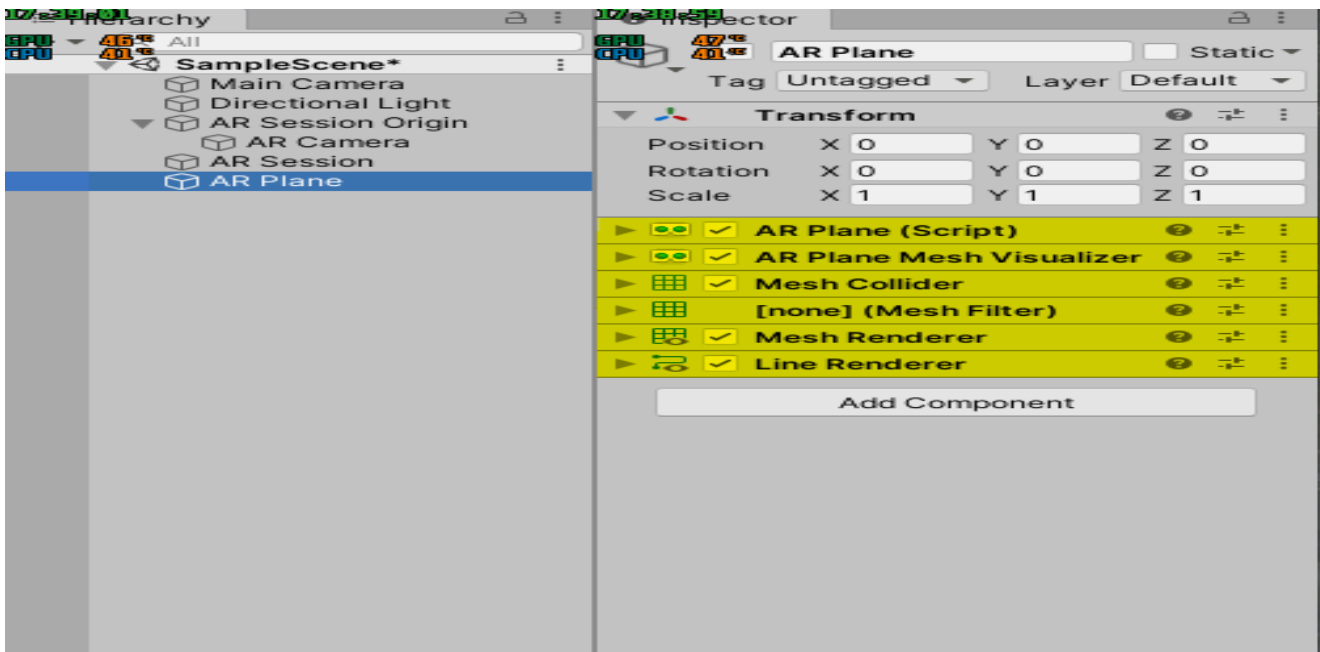


Fig. 4: Setup outline that will show the borders

4. Set up outline that will show the borders, shown in Fig. 5 and Fig. 6: Setup Width, Setup Color, Setup Corner Vertices, End Cap Vertices, Turn Off Cast Shadows, Check off Receive Shadows.

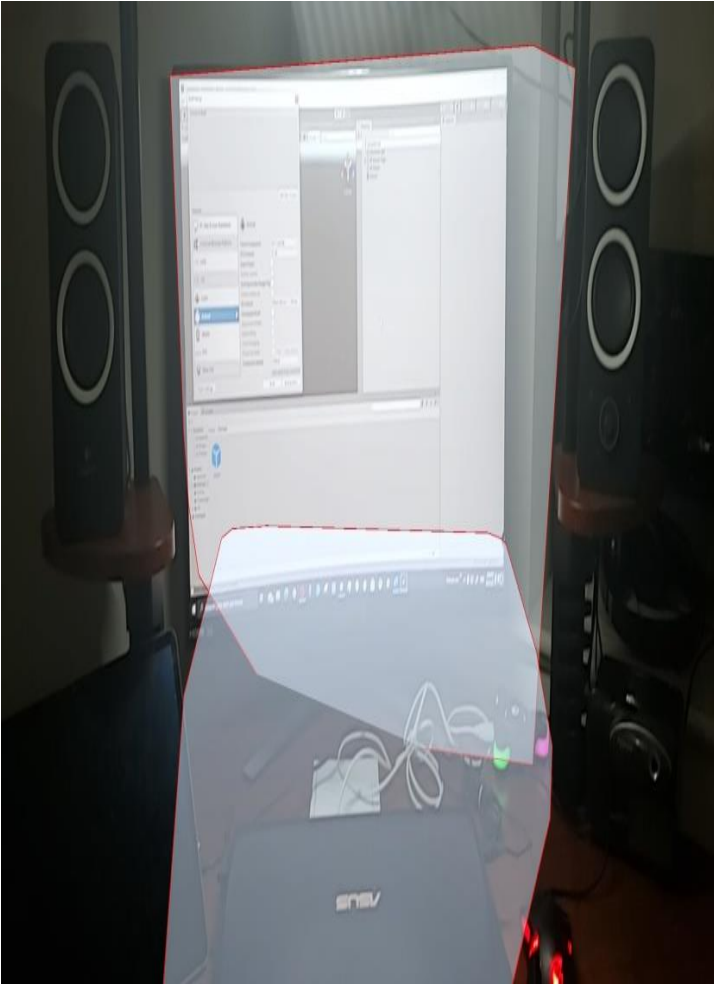


Fig. 5: App can scan Computer Monitor and Table



Fig. 6: App can scan surface of beds.

6. Conclusion and Future Work

In recent years, augmented reality technology has received extensive attention from researchers. Driven by Computer Vision and Artificial Intelligence, Augmented Reality has shown strong momentum of development. The tracking registration accuracy, display equipment performance and the nature of human-computer interaction have been greatly improved. However, there are still many problems to be solved in augmented reality. In terms of tracking registration, the current method can only make use of a small amount of information in the scene, such as the feature point information, which leads to incomplete understanding of the system to the environment. In terms of display technology, the size and price of augmented reality glasses that can provide users with a high sense of submergence cannot meet the demand of the public. For the interaction mode, the more natural and multi-user augmented reality interaction technology remains to be studied. Our future project plan is to continue with the implementation in smart phones as a way of creating object recognizing smart glasses with augmented reality.

References

- [1] Ballard, B. (2018, July 3) "Giving voice to augmented reality." Aerospace Manufacturing and Design. [Online]. Available: <https://www.aerospacemanufacturinganddesign.com/article/giving-voice-to-augmented-reality/>
- [2] T. Palladino, (2018, September, 13) "Mobile AR Apps Can Now Track Any Surface Using Plane Detection via Wikitude SDK," Next Reality. [Online]. Available: <https://next.reality.news/news/mobile-ar-apps-can-now-track-any-surface-using-plane-detection-via-wikitude-sdk-0187402/>
- [3] Dickson, B. (2020, September). "What Is Artificial Narrow Intelligence (Narrow AI)?" [Online]. Available: <https://bdtechtalks.com/2020/04/09/what-is-narrow-artificial-intelligence-ani/>.
- [4] K. Fessel, (2020, September, 21) "5 Significant Object Detection Challenges and Solutions," Medium. [Online]. Available: <https://towardsdatascience.com/5-significant-object-detection-challenges-and-solutions-924cb09de9dd>
- [5] Davenport, T. and Kalakota, R. (2020, December, 10). *The Potential For Artificial Intelligence In Healthcare*. [online]. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/>
- [6] ActiveWizards (2018, October) "Comparison of the Top Speech Processing APIs." KDnuggets. [Online]. Available: <https://www.kdnuggets.com/2018/12/activewizards-comparison-speech-processing-apis.html>
- [7] Jin, Z., Tillo, T., Zou (2018, July 02). Depth image-based plane detection. *Big Data Anal 3*. [Online]. Available: <https://doi.org/10.1186/s41044-018-0035-y>
- [8] Gargalakos, M., Rogalas, D. (2012) "The EXPLOAR project: Visualizing the invisible." Proceedings of Science Center to Go Workshop in Augmented Reality in Education. Athen, Greece, October 27 - 29, 2011.
- [9] Aouam, D., Benbelkacem, S., Zenati, N., Zakaria, S., Meftah, Z. (2018) "Voice-based Augmented Reality Interactive System for Car's Components Assembly." Proceedings of the 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS). Tebessa, Algeria. 24-25 Oct. 2018
- [10] A. Bobeshko. (2020, December 15). "Object Recognition in Augmented Real. [Online]. Available: <https://virtualrealitypop.com/object-recognition-in-augmented-reality-8f7f17127a7a>
- [11] (December, 16, 2020). "Augmented reality (AR), virtual reality (VR) different areas of application, advantages and disadvantages," Panono 360 Camera 16K. [Online]. Available: <https://www.panono.com/en/blog/blog-posts/augmented-reality-ar-virtual-reality-vr-different-areas-of-application-advantages-and-isadvantages/index.html>.