Workplace Assistant Augmented Reality

Gabriel Camilleri

Supervisor(s): Dr. Vanessa Camilleri

Faculty of ICT

University of Malta

May 2020

Submitted in partial fulfillment of the requirements for the degree of B.Sc. ICT in Artificial Intelligence (Hons.)

L-Università ta' Malta

FACULTY/INSTITUTE/CENTRE/SCHOOL______________________

DECLARATIONS BY UNDERGRADUATE STUDENTS

Student's I.D. /Code _____________________________

Student's Name & Surname ____

Course ____

Title of Long Essay/Dissertation

____

____

____

Word Count ___________

(a) Authenticity of Long Essay/Dissertation

I hereby declare that I am the legitimate author of this Long Essay/Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) Research Code of Practice and Ethics Review Procedures

I declare that I have abided by the University's Research Ethics Review Procedures.

______________________ ______________________

Signature of Student Name of Student (in Caps)

____

Date

08.02.2018

Abstract:

Starting a new job in an office can be very stressful for an intern or a new employee, especially their rst day at the office. It takes time to adjust and learn what other employees' jobs are and how they can be benecial to them. It might additionally take some time for new members to learn the ropes and their purpose within the office building, while understanding and learning how to use certain equipment, for example, an automatic key lock or simply a coee machine. Therefore, the Workplace Assistant Augmented Reality tries to identify the user through user proling, while providing the necessary process for the user to learn and understand the information relevant to them.

Starting a new job in an office can be very stressful for an intern or a new employee, especially their rst day at the office. It takes time to adjust and learn what other employees' jobs are and how they can be benecial to them. It might additionally take some time for new members to learn the ropes and their purpose within the office building, while understanding and learning how to use certain equipment, for example, an automatic key lock or simply a coee machine. Therefore, the Workplace Assistant Augmented Reality will tries to identify the user through user proling, while providing the necessary process for the user to learn and understand the information relevant to them.

The conducted research and experimentation will ultimately determine whether using Vuforia's augmentation techniques is sufficient to complete certain augmented reality tasks. If not, better augmentation techniques will be compared with Vuforia's techniques and ultimately recommended.

Acknowledgements:

I would like to thank CCBill, for allowing me to carry out my research for WAAR on their behalf using their environment as a basis for my application.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Problem Definition

"Person-job fit is a substantial factor for decreasing job stress and the adjustment of employees to an organization is an important issue for eliminating stress"[11]. "New employees all bring expectations to their new jobs that are based on factors like their previous job experiences, their understandings of the profession, beliefs and experiences held by peers or family, promises made during recruitment, and their evaluation of the work situation during their interview"[25]. The rst month at the workplace might seem overwhelming for new employees. Therefore, during their first few months of settling and adjusting, the company may allow "a period of learning how to 'fit in' and adjusting to how things work in the new setting"[25] for the employee's benefit.

Providing an assistant augmented reality application to help speed up the process for the employee to adjust to their new workplace environment may offer several challenges. There are several Augmented Reality libraries which provide all the necessary techniques for one to build such applications, with no need to be highly skilled in any form of programming, especially where it involves Artificial Intelligence. When it comes to feature extraction, things can be challenging, especially if one is making use of traditional computer vision techniques, such as, SIFT and SURF alone. For instance, "[t]he SIFT algorithm deals with the problem that certain image features like edges and corners are not scale-invariant. In other words, there are times when a corner looks like a corner but looks like a completely different item when the image is blown up by a few factors"[22].

1.2 Motivation

"Whilst employees can be reasonably expected to adjust to changes in jobs over time, poor job or employee job t can result in increased stress and ineciency in organizations"[11].

A workplace is defined as the environment where people work. Adjusting to a new environment, especially one's workplace, can come with several challenges, such as, adapting to new people, finding certain offices within the environment, and using certain job equipment. "When humans feel a loss of control this causes physiological changes which can exacerbate feelings of stress"[11]. Job stress has become a common term in industry since several companies endeavour to sustain a healthy working environment for their employees. “Workload is one of the major factors which affect the employees’ productivity and efficiency. Job stress caused by high workload has become common in today’s scenario”[34].

Such level of stress can increase from certain necessary adjustments for the employee to settle within a company, such as, filling in papers and handing them to the right offices, and learning to use certain equipment around them. Therefore, proper training should always be provided, whether it is detailed or otherwise. “Application of training in the workplace and proper implementation of training can directly lead to improving the employees’ performance”[7].

There are two types of training, namely, on the job and o the job training. On the job training is a method of imparting knowledge and training directly while on the job. Conversely, off the job training is a method of imparting knowledge and training while not at the place of work, for example, through a site. The idea behind training is to minimise stress levels and allow the employee to improve without any pressure. “Training, which aims at empowerment, development, and qualifying employees through knowledge and skills, refers to end-oriented, organized, logical, on-going planned attempts to bring about the desired change in the knowledge, skills, capability and attitude of employees” [7].

1.3 Why the Problem is Non-Trivial

There have been previous attempts at making indoor augmented reality applications to guide users around a place. However, most attempts are usually made using ArCore and acquiring a 3D model of the building. ArCore is useful for catching movements and current positioning, as well as light detection. It further has the anchoring feature where a virtual object is given a marker to monitor its displacement. However, ArCore is incompatible with several devices, and it would thus be futile to apply it in real-life scenarios since not everyone will have the latest phone with the latest specs. Vuforia, on the other hand, is more user-friendly and can be used on several operating systems.

The second problem is that the augmented reality application can be fed a 3D model directly to anchor positions within the map and display the respective augmented information. This can be useful when applying indoor augmented reality navigation. However, creating a 3D model of the workplace can have several problems. Firstly, the company would not want to hand out freely a plot of its indoor workplace as this goes against its policy. Secondly, one would not be testing and experimenting with anything if a 3D model of the workplace were used. In this project, several features will be tested from Vuforia's library, such as, feature detection, and the library will be used to its full potential.

1.4 Approach

The proposed solution is to develop a workplace assistant augmented reality (WAAR) application to assist users by providing them with augmented reality information to guide them to offices, provide them with information about offices and rooms while walking down the corridor, and give instructions on how to use the office coffee machine. The application will make use of user profiling techniques to understand users' requirements, and will display relevant information related to the purpose for using the application. It will be necessary to fill in a form prior to using the application. The form will be quite short, and the collected data will not be stored anywhere and will only be used to display relevant markers on the augmented reality application. Once the application is closed, all data about the previous user will be forgotten, at least, for our testing purposes.

Augmented Reality development will be handled by Vuforia's libraries since Vuforia has some features which the application can well benet from. "It enables businesses and app developers to quickly spin-up high delity, mobile-centric, immersive AR experiences"[29]. For our research, image and object segmentation will be used to identify ofce workplace markers and Unity, and the proper content will be overlaid using game objects. There will be instances where model target and Vuforia's deep learning techniques are used to scan some objects in 3D. Vuforia is ideal because it can develop augmented reality application for Android and IOS devices.

Indoor navigation can be done in several ways. One can use GPS signals, beacons, RSS or WIFI signals, or simply Augmented Reality itself. Furthermore, Augmented reality can be either location-based or marker-based. Therefore, the proposed solution for our problem is to use Augmented Reality marker-based navigation by using several markers around the oce building to segment images or objects, while displaying the proper directions by recognising the markers in view thereby enabling the company to keep the application useful for oine use. For scenarios where WIFI or any other signals are down, users can still make good use of the application, for example, in case of an emergency to nd the re exit.

1.5 Aim and Objectives

The aim of this project is to research and develop a workplace assistant augmented reality application, using image and object detection provided by Vuforia, and filtered through user profiling.

The objectives of the final year project are:

- Perform image and object detection techniques using the Vuforia Library;.
- Use Augmented Reality techniques from the detected images and objects to overlay and augment information and navigation information.
- User profiling through a recommendation based system to filter out unnecessary information for augmentation.
- Applying and evaluating the developed and implemented artificially intelligent techniques through quality and quantity testing.

1.6 Report Layout

The layout of the report is as follows. Chapter 2 provides background information about the technologies used. Subsequently, Chapter 3 includes the literature review which was conducted while attempting to solve the problem at hand. Chapter 4 is a brief overview of the system and its design. Chapter 5 presents the implementation process, while Chapter 6 discusses the evaluation methods and approaches for the application, including both user and AI evaluation. The chapter further analyses the obtained results. Chapter 7 outlines the limitations and challenges encountered during the project, while offering recommendations for further development of the application and technologies used. Finally, the project is brought to an end with a conclusion.

2 Background Research

Besides providing some background research on the technologies applied in this project, this chapter will discuss thoroughly technical information about the workplace environment and applicable technologies in order to apply current Augmented Reality technologies in such environments, while outlining any difficulties which might arise.

2.1 User Profiling

"User Profiling is the process of Extracting, Integrating and Identifying the keyword-based information to generate a structured Profile and then visualizing the knowledge out of these findings" [19]. User profiling enables the system to tailor the required information for the user to see and use. It is rather annoying for users to have to go through irrelevant documents or data in an attempt to find information specific to what they require.

"User profile generation is done when we get user's complete information while he registers into our system. We have identified different user attributes for profiling him into our system" [19]. User profiling has taken the form of recommender systems, thus providing user specific and personalised recommendations. There are two forms of user profiling. The first is Explicit User Profiling, which is an approach where the "... user's behaviour is predicted by analysing the user's available data" [18]. This is also known as Static Profiling, where static and predictable user data is analysed. The second type is Implicit User Profiling, which "relies more on what we have known about user in future i.e. system tries to learn more about the user" [18]. This type is also referred to as Adaptive Profiling. After performing extraction, one might end up with redundant information.

To clean the information and see unique pieces of it, one must perform filtering. There are three filtering techniques for user profiling, namely, rule-based, collaborative, and content-based filtering. Rule-based filtering is the technique used to filter out content based on a set of rules, normally present using "if-then" statements. Content-based filtering "recommends items based on a comparison between the content of the items with a user profile and selects those items whose content best matches with the content of another item" [18]. On the other hand, collaborative filtering is the process of grouping users with a similar search criterion. Filtering is based on previously sought items as well and items which one is more likely to search for next.

2.2 Augmented Reality

“Augmented Reality (AR) is a new technology that involves the overlay of computer graphics on the real world” [32]. It is a term which refers to mixed reality, where the digital world and reality are combined and interwoven. Augmented Reality is a new form of technology that focuses on displaying realistic overlays on reality to provide extra information and content to what we see with our naked eyes.

There are different categories of Augmented Reality. The first category is marker-based AR, where the augmented overlay is only displayed once a marker is detected through a camera. It is also known as image recognition. The second category is markerless augmented, which makes use of an accelerometer, GPS, and velocity tracker to detect the location of the phone and display the AR overlay in that specific location, given its location is predefined. The third category is projection-based, which basically projects data in the form of light rays on objects, for example, an augmented-projected keyboard. The last category is superimposed AR, where AR replaces partially the real view with an augmented one of the object. IKEA use this application in their digital catalogues.

There are several Augmented Reality devices. The first device is Optical See-Through HMD. “Optical See-Through AR uses a transparent Head Mounted Display to show the virtual environment directly over the real world” [32]. HMD performs best when it fits perfectly to the users’ eyes and sits comfortable on their face, making it easy for them to move around when wearing it. The second type is Virtual Retinal Device, which “. . . projects a modulated beam of light (from an electronic source) directly onto the retina of the eye producing a rasterized image” [32]. The third device is Video See-Through HMD, the monitor-based Augmented Reality, which “. . . uses merged video streams but the display is a more conventional desktop monitor or a handheld display” [32]. Finally, projection display, which projects on surfaces and is useful for multiple user interaction. One such example of projection-based AR is Tilt Five.

2.3 Mobile Augmented Reality

Using Augmented Reality on mobile devices presents several challenges which are “related to context-awareness, usability, navigation, visualization and interaction design” [21]. Handheld devices are nowadays equipped with powerful processors, cameras, and sensors.

Smartphones use a "camera on the opposite side of the display [which] encourages the use of the 'magic lens' metaphor describing the fact that the users have to point and look 'through' the device to view the augmented representation of the real world"[21]. Although most cameras are equipped with high resolution, the screen and camera capture a limited range of field of view. Therefore, augmented information must clearly be placed on the smartphone screen and avoid obstructing the user from important views of the real world.

A mobile augmented reality framework is made up of three specific features [20]], namely, MAR Observer which obtains the target images or text from the augmented reality server, MAR Server which "serves as a bridge between the MAR customizer and MAR observer" [20], and MAR application customizer which defines interactions between the user and image targets. In this case, Vuforia serves as the MAR application customizer.

2.4 Augmented Reality Navigation

Outdoor navigation usually makes use of GPS localisation. However, this can be a problem for indoor navigation. There are several ways to provide indoor localisation. One can use beams either by Bluetooth signals or WIFI signals thereby obtaining continuous mapping, albeit with rather irregular results at times. The alternative is to use offline waypoints, where the user simply scans a marker to get a location or augment pre-programmed information within that location. However, "...the user needs to update his/her location by scanning another way-point on the way" [9].

One main challenge of augmented reality navigation is the process of registration, which "...is the process of correctly aligning the virtual information with the real world in order to preserve the illusion of coexistence" [9]. Although proper visual registrations must be met for the augmentation to be as realistic as possible, one must not forget that the user still needs to focus on what is in their path.

Improvement of AR can help to provide navigational information without distracting the user to look away to a secondary screen or view by, "[f]or example, showing navigation markers on the windshield of the car or augmenting the video camera output of a smartphone with the navigation path" [9]. To provide an augmented reality navigation system there are several steps one needs to take, namely, "1. Acquire the real-world view from the user's perspective. 2. Acquire the location information for tracking the user. 3. Generate the virtual world information based on the real-world view and the location information. 4. Register the virtual information generated with the real-world view" [9].

2.5 Traditional Computer Vision for Object Detection

Traditional computer vision is the "traditional approach. . . to use well-established CV techniques such as feature descriptors (SIFT, SURF, BRIEF, etc.) for object detection" [3]. Images contain several features which can be extracted using CV algorithms, such as, edge detection, corner detection, and threshold segmentation for improved detection of such features.

Image recognition works by detecting natural features such as edges and corners in an image. "[T]he feature tracking algorithm can determine what is a feature and map the positions of these features in the image" [14]. By shifting the positions of the image, features like edges are intensified, with even more corners as their position changes after shifting. Therefore, Vuforia makes use of pose feature detection techniques, where it takes into consideration the position and orientation of the natural features. It can make use of extended tracking, where the engine detects surrounding features as well. A proper image with high quality feature detection is an image that contains uniquely distinct features which are not repetitive. For example, a dark circle is difficult to recognise and establish as a unique feature.

"The difficulty with this traditional approach is that it is necessary to choose which features are important in each given image. As the number of classes to classify increases, feature extraction becomes more and more cumbersome. It is up to the CV engineer's judgment and a long trial and error process to decide which features best describe different classes of objects" [3]. There are several advantages when using traditional computer vision techniques. SIFT and SURF algorithms are generally used for applications such as image stitching, where classes do not need to be identified within the image. Traditional techniques make use of less processing power, and the problem at hand is simple enough to use such traditional computer vision techniques with little amount of data needed, unlike a deep learning model.

2.6 Deep Learning in Augmented Reality

The detection problem has been solved using camera-based tracking systems to apply them to Augmented Reality using deep learning techniques. The Vuforia Library has applied such techniques to scan 3D objects and create model targets for them in order to be easily recognisable within any developed AR app. This provides new advantages, such as, detectability from any angle of the recognisable real-world object. "Known model of the object can be used to determine the position and orientation of the object. Rendering of the virtual object follows easily" [5]. There are two ways how the object can be recognised. One can use traditional artificial vision techniques, or Convolutional Neural Networks for improved detection.

Model-based AR tracking is achievable in two steps. Firstly, one uses video tracking which "yields the pose of the camera with respect to the known target" [5]. Secondly, the pose is sent to an algorithm for tracking. Such algorithms as SIFT [24] and SURF [8] are commonly used for detection. The algorithm extracts a number of key points using a corner detection algorithm such as FAST [30]. In [5], a CNN implementation was trained using AlexNet to detect patches. FAST was used to detect features on a reference image, extracting 15 by 15 patches across each feature. HIPS [33] was used for 8 by 8 sparse sampled patches from the original set of patches. When comparing the overall performance of the CNN used in [5] with an algorithm such as ORB, the re-projection error shows that it was far improved in DeepAR. "DeepAR method produces consistently more inliers than HIPS. However, as can be seen in Figure 12 the percentage of inliers vs outliers are less for DeepAR" [5].

In their study, [5] concluded that "[t]he detector performance is very strong especially in the presence of error in feature localization" [5]. It is indeed comparable to one of the best feature detection algorithms to date.

2.7 Conclusion

This chapter provided and discussed background research and information on existing technologies and techniques which will be applied in FYP. The following chapter will present a literature review of the research published in the field of Augmented Reality.

3 Literature Review

The chapter serves as a review of the available research that was done in the area of Workplace Assistant Augmented Reality. I shall be discussing the two components involved in the Augmented Reality application as well as the research that has inspired my approach. The chapter is divided into three parts; the first part involves Image and Object Recognition techniques involved in Augmented Reality, the second is about applying user profiling methods with Augmented Reality, and the third is about different image and object recognition techniques involved in augmented reality technologies.

3.1 Workplace Augmented Reality

"Augmented Reality (AR) technology has rarely been discussed outside of the computer science world. It has taken years for this technology to become closer to a stable existence, and will most likely take several more years before it will be used by average citizens"[12]. Although the technology can be considered in its infancy it also has a wide variety of applications. One of its main applications in the 4.0 Industry is the use of AR in assisted learning. Every workplace needs adjustment and some form of training for employees to be adjusted to the process of the work they might be doing. Augmented Reality may help in assisting the employees by providing them with additional overlayed instructions to guide them through the whole process of adjustment as well as to provide them with training.

Workplace training normally comes in two forms; on the job training and off the job training. "OJT may be viewed as an apprenticeship where a novice AMT is mentored by an AMT who is an expert"[16]. It is a traditional form of training, especially for teaching maintenance. However "OJT may not be the best method for training because the feedback to learners may be infrequent and unmethodical"[16]. Off the job training maybe provided through face to face conversations or through use of multimedia. Augmented Reality is capable of combining the two aspects of training into one. Where the user is given on the job training through the use of multimedia, which is overlayed on top of the real world environment.

There are several useful application for Augmented Reality at a workplace. However, not every workplace might necessitate for AR. "there are situations where an AR system may be used to enhance the task completion process, or display and/or communication of information in conjunction with traditional technologies"[12]. As discussed in [12], the following are workplace conditions where AR is applicable; distance communication with 2D or 3D objects provided for visualization, training and education when making use of real life tools, recording of information obtained while training, and a collaborative design and interaction of 3D models is required.

An advantage which Augmented Reality provides is, for the workers and managers "the ability to author their own environment by embedding the relevant information needed for task completion"[12]. The common problem which is face during work training is for the expert individual to provide the respective information to the trainee in the most understandable way possible. Through AR technology the trainee is able to tailor how that information is presented. Therefore Augmented Reality may be capable of understanding its user's such that it may adapt to future possible users.

3.2 Recommendation Systems for Augmented Reality

Information during a particular job training is crucial for an employee to learn and adjust to the new environment. However, an overwhelming amount of information directed towards a new employee may demotivate them. Augmented Reality is a tool for providing interactive information towards the user as well as from the user. In spite of that, "The fact that the typical scene of these applications mix real and virtual elements can be a motivating factor for users. However, this feature may also make the interaction more complicated, which can affect the user experience in performing tasks within the application"[31]

"Recommender systems (RS) have proven to be a valuable tool for online users to cope with the information overload"[10]. Recommendation systems provide tailor made information to different users based on the users' preference. "Thus, it is important to offer the user a personal response, but also a context-dependent and constrained by the limited computing capacities of the mobile devices"[10, 4, 36, 28]. Therefore, the recommendation system should provide its user with information which might be of interest to them, but which also makes sense in the context and location they are using it.

Collaborative filtering techniques have been widely adapted in recommender systems. However, traditional recommender system in Augmented Reality cannot be easily adapted and deployed since they differentiate in the following areas; location, timing, first time use of the application, and immediate response from the AR application as discussed in [37]. In augmented reality distance-based filtering and visibility-based filtering are commonly used. In [37] a random walk algorithm was incorporated, which recommendations are based on user preferences, behavior patterns, history records and information from social media. However, in the latter research user-feedback was not evaluated, which would have helped in providing the efficiency and performance of the AR recommender system.

An alternative to using location or distance based recommendation, Augmented Reality applications can make use of time based recommendation systems. By time-based meaning, the amount of time one would generally spent on completing a particular task using the AR application. A task may take a lot of time to be completed by the user due to several factors, such as the task itself is complicated or simply the AR app is incapable of providing the user with the right instructions and guidance into solving the task, simply because it may lack different forms of interactive techniques. In [31], "a set of procedures to conduct experiments with users to identify how a set of aspects related to the user prole can be considered to improve mobile AR technology usage"[31] were defined. The result achieved from this research is that young groups of people spent less time in completing a task using AR, as they were accustomed to similar forms of technology. Users with little to no experience and of an advanced age took obviously longer to adjust to the technology. This was due to several factors other than being newly introduced to such forms of technology, one such example could be health issues which can hinder their overall performance, such as eyesight problems and motor coordination.

Therefore an augmented reality application cannot assume that the user will easily interpret what is being overlaid on the screen. They might need to be guided along, as to understand what different symbols, colour and size of the symbols being displayed might mean.

3.3 Computer Vision Approaches in Augmented Reality

Augmented Reality applications make use of several computer vision approaches to recognize images, objects and text. As previously discussed, Vuforia makes use of both traditional and deep learning approaches. Using deep neural networks will ensure highly accurate and efficient results. However, "it is well-known that training high capacity models such as deep neural networks requires huge amounts of labeled training data"[6]. Neural networks are data hungry architectures, that require huge amounts of data to train and test on, thus be capable of generalizing accurately.

As discussed in [23] marker based applications have been the main driving force of applying augmented reality in real life. "Most of the current approaches to 3D tracking are based on what can be called recursive tracking"[23]. Therefore the system must be initialized by hand and with a little bit of occlusion between the camera and the object being recognized the system fails to perform. However, a new computer vision approach has improved augmented reality, which is capable of registering the camera without camera pose introduction. This approach is called Tracking-by-Detection and in [23] it is tested to see its benefits. The approach works by extracting feature points from inputted frames during run-time. The features are then "matched against a database of feature points for which the 3D locations are known"[23]. However there were still a key number of limitations such as detecting reflective and shiny surfaces on the car, due to the fact that not many features could be extracted. Another limitation was dealing with occlusion especially if a person happened to be standing anywhere near the object being detected. One final challenge met was dealing with robustness, due to having to deal with different environments in which the object would be in in order to improve object recognition and be able to generalize.

According to the research conducted in [27], in order to solve the occlusion problem within current Augmented Reality technology a S-G Hybrid Recognition method was implemented. The approach takes "advantage of robustness of the SURF feature-based object identification and combine it with high reliability and effectiveness of the Golay error correction code detection"[27]. SURF along with SIFT are two traditional vision approaches commonly used for feature based detection. The advantage of SURF is scale and rotation in-variance. Golay error correction code on the other hand is a marker identification approach. "A marker based on the Golay error correction code (ECC) can be composed of a large white square in the top left corner and e.g. 24 black or white squares that encode a number. The large square provides information about the marker orientation"[27]. They tested the three main aspects which may hinder an AR application; distance variance, angle variance ,and occlusion. As a result, the S-G approach it was found that an object can be placed 2m away from the camera, the angles comparison was completely influenced by the SURF algorithm where it was able to detect under 55 degrees angle to the camera's axis, and that it could not be affected by up to 55% obstruction.

Another approach into solving the occlusion problem in augmented reality is by applying deep learning techniques as described in [13]. In this research, they "present a temporal 6-DOF tracking method which leverages deep learning to achieve state-of-the-art performance on challenging datasets of real world capture" [13]. Deep learning architectures can be trained on large amounts of data, and as a result this solves the occlusion, angle variance, and distance variance problem. Their approach involved getting a 3D model of the object and training the tracker for that specific object. Training involved two particular steps; first one was using a frame to capture the object in its predicted position, and secondly the frame of the object's actual position. "To encourage the network to be robust to a variety of situations, we synthesize both these frames by rendering a 3D model of the object and simulating realistic capture conditions including object positions, backgrounds, noise, and lighting" [13].

Deep learning architectures work well when making use of GPUs. The GPU is commonly used to run deep learning neural networks, hence the network takes less processing time to train and test. In [26], is presented "YOLO-LITE, a real-time object detection model developed to run on portable devices such as a laptop or cellphone lacking a Graphics Processing Unit (GPU)" [26]. YOLO-LITE primarily is designed to obtain a smaller, faster and more efficient model. "You Only Look Once (YOLO) was developed to create a one step process involving detection and classication. Bounding box and class predictions are made after one evaluation of the input image" [26]. The architecture developed runs at 10 frames per second, and its goal is to prove that shallow networks can run on non-gpu devices, and that shallow networks do not require batch normalization. The model had 18 trials achieving results of 33.77% mAP and 21 FPS, and 12.26% and 21 on PASCAL VOC and COCO dataset respectively.

3.4 Conclusion

Different approaches were defined and revised in this chapter. Traditional computer vision techniques, deep learning techniques, recommender systems, and augmented reality solutions were studied to obtain relevant information to acquire a state of the art Workplace Assistant Augmented Reality application. In the following chapters, the design and implementation of the proposed method shall be presented.

4 Design

This chapter will give an overview of the design of the system implemented. The components will be further discussed in detail in Chapter 5.

4.1 Overview

The augmented reality application is divided into four separate parts. The first part is the data extraction process, where data is collected for training both for the augmented reality side of the application, as well as for the recommendation process. The second part involves feature extraction, where relevant features are extracted and fed to the implemented or applied model for training. The third phase is building a suitable user-query model for user recommendation. The last part is implementing the trained data within the custom-built user interface to provide an user-friendly augmented reality experience.

4.2 Data Handling

The data extraction process is further divided into three phases. The first phase is gathering relevant images of the area around the workplace. The next phase is building 3D models of chosen markers for the Augmented Reality. The 3D models must capture as much detail as possible of the actual marker. The images and 3D models are then fed to the Vuforia's Library for training. The last phase is gathering data from a good number of previous users who rated the application when they performed a particular task. The reason behind this, is to perform collaborative filtering techniques using a set of machine learning algorithms and probabilistic methods to achieve a set of user preference recommendations. The results achieved are then combined with the item similarity based matrix.

4.3 Feature Extraction

As explained in [15], feature extraction can be decomposed in two steps: feature construction and feature selection. In this project feature extraction is done on images, 3D models and previous user ratings. Firstly, feature extraction on images is done using Vuforia's natural based feature selection technique [2] which is similar to the ones used in Sift[24] and Surf[8] algorithms. The next step is to pass the 3D models to Vuforia's model target generation [1] which makes use of the library deep learning techniques. For the collaborative filter techniques the SVD model makes use of the rating and the user-task id as features.

4.4 User Query Model

Two forms of user query models shall be created. The first one will be that of the intern. Every intern goes through a similar process of integration on their first day of their job. Since this is a prototype, it was decided to provide a step by step process as feature implementation for the intern, furthermore they will have the option of choosing whether they prefer the following for augmentation; the games room, the toilet rooms, and kitchen. Then the next part will be to allow the intern to first find the secretary, then the human resources' office and finally the manager's office. For each task, the system will be queried taking into account the preferred options the user had previously queried for and the office they wish to visit. The second form of user query model is that of the visitor querying the augmented reality system. To the contrary the visitor is presented with the top 3 recommendations according to which task they would need to accomplish when visiting the office. The system has to accommodate visitors with the following tasks: a delivery, an interview and a visitation. Once the user picks the relevant task and they choose the appropriate recommendation which falls under their preference, they are presented with relevant augmentation. For both the intern's and visitor's query, the system also takes into account the rooms and offices which the user will pass in front of, whilst visiting a particular office. Therefore, it not only takes into account what the user prefers based on previous users' preferences but also what they might require based on the location they are in. Hence, the user query model for the visitor makes use of both collaborative and item similarity based filtering techniques.

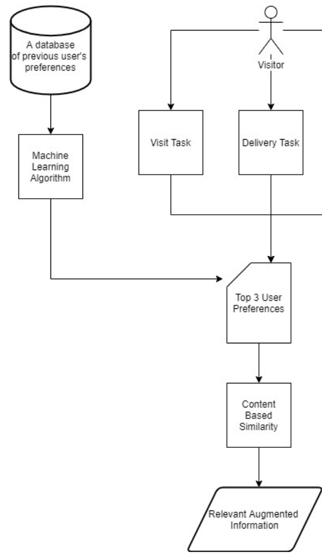

Figure 1: Visitor User-Query Model

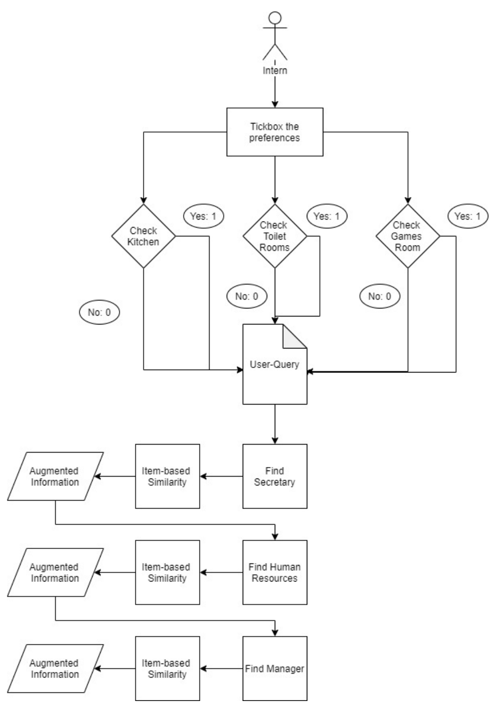

Figure 2: Intern User-Query Model

4.5 User Interface

The user interface has to be as user-friendly as possible. It needs to provide the user with several options for them of what information they are interested in. The user shall be presented with a main menu allowing them to either augment information about the coffee machine, the offices while wandering around or to locate an office in particular. The coffee machine interface is augmented once the coffee machine is recognized, allowing the user to learn how to make a cappuccino via an augmented video and text. The offices information interface is augmented once the user's phone recognizes the correct marker allowing them to view details about the office, or locate a particular office from where they are. Navigation is not provided through an artificially intelligent algorithm and it is not within this research's scope to implement it. The navigation provided is through a 3D sketched holographic map giving an idea to where the visitor or intern need to go to find the office within their interest.

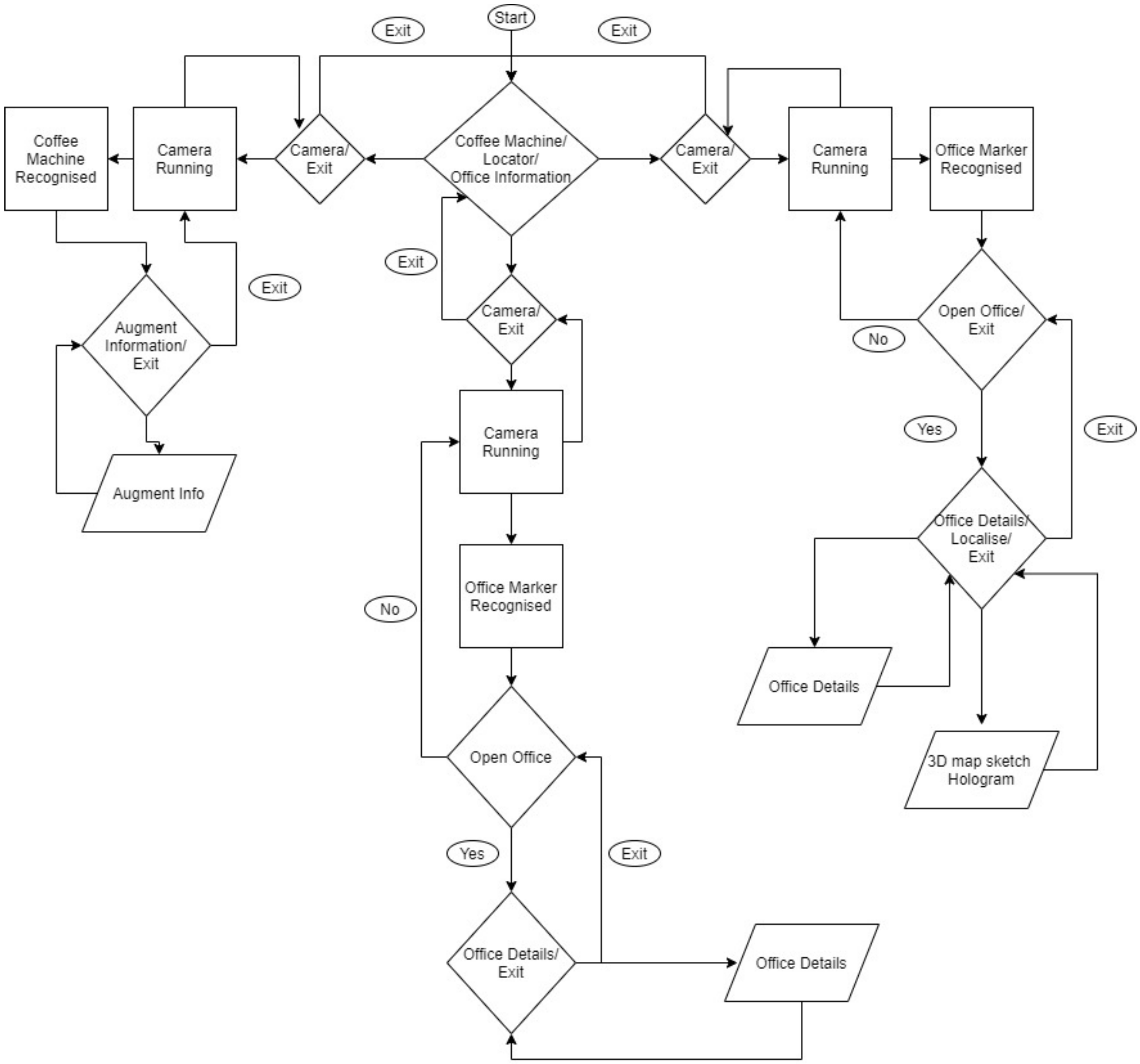

Figure 3: User-Interface

4.6 Conclusion

An overview of the implemented system in this chapter has been given. The next chapter will go into detail of the system as well as highlighting certain decisions taken and why they were taken.

5 Implementation

The components proposed in the previous chapter shall be further discussed in detail. The implementations used where done using the following applications and libraries: Vuforia [14], Unity Engine [14] and Python 3.7. Vuforia was mainly used to train image and model targets, Unity is a game engine/editor used in our case for building the augmented reality application architecture, and Python 3.7 was used for image sharpening as well as for training and testing the recommendation system. This chapter highlights all the necessary details about the data handling, feature extraction, user-query recommendation as well as the user interface and system architecture built accordingly. Vuforia was used for augmented reality, mainly because it is the "most popular SDK for developing AR-applications on a wide selection of devices"[14]. Similar to ARCore and ARKit, Vuforia can be used on multiple devices to recognise images, objects and text.

5.1 Data Handling

5.1.1 Images

Initially the project started by gathering a numerous amount of images of the workplace's first floor. A good number of images were taken of all the corridors, objects inside the corridors and doors. Images of the interior of the office rooms within the first floor were excluded, as it was not within the project's interest to capture them within the augmented reality application. Different variations of the same image were captured, capturing as possible different lighting and time scenarios. The reason behind this was mainly due to the fact that GPS signals are weak indoors. Therefore, the Augmented Reality had to be marker-based through image and model targets. The challenge in this particular case, was capturing the same images in different scenarios. For example if in one image a door was open, another image had to be captured where the door was closed, that way the system would later on manage to generalise.

5.1.2 Markers

The next step was choosing the markers, which Vuforia could use to recognise and overlay the augmentation on. Choosing the best markers is important, as that can entirely affect the user's experience within the augmented reality. Initially, it was thought to use images of the corridors themselves as a marker. However, it was soon realized that for one corridor, one needed to take several images capturing an infinitely amount of variations which could occur in that image. For example, as shown in figure 4 and in figure 5, one can see several clear sight variations of the same corridor. By finding differences in edges and contours computed in figure 6, a structured similarity index of 0.62 was found which is a very low similarity, hence making it difficult to augment corridors. Therefore, it was decided to use door signs as markers. The floor used for the application, had a good amount of doors, each had their own unique sign placed at the centre of them. Therefore, the signs were used as markers for their static looks, meaning, it was very unlikely for the marker to differ in different images taken at different times and angles and for their unique attributes as shown in figure 7.

5.1.3 3D Models

To build the model targets one has to provide a 3D object within the Vuforia Model Target Generator as explained in [14]. The 3D models generated were those of the door markers placed around the whole floor areas and of the coffee machine as shown in figure 8 and in figure 9. The 3D models for the door markers were generated through the use of an online library Selva3d which generates a 3D model from a given picture.

5.1.4 Recommendation Data Set

The recommendation system was made up of a combination of item-based similarity and collaborative filtering techniques. For the Collaborative filtering techniques, there was no existing dataset which one could make use of, since there were not any previous similar applications were users could rate the system. Therefore a prototype recommendation system was built through generated data, simply to analyse how the system works along with the augmented reality application. As previously mentioned three particular tasks were picked for a visitor profile, for which the collaborative techniques would be applied. The tasks were; an interview, a delivery and a visitation. For each task a dataset containing the user-tasks along with the rating of that particular tasks was created. Every entry had the following attributes; user id, task id, locations of interest, user-task id, and the rating. The task id represents the id of either an interview, a delivery or a visit. Locations of interest is a 7-bit binary code representing the accountant's office, the human resource's office, the manager's office, the secretary's office, the kitchen, the toilet rooms and the games room respectively. Each dataset has a respective smaller dataset containing a number of locations of interest which were activated by previous users. Their ids are represented by the user-task id.

5.2 Feature Extraction

5.2.1 Image Targets

The first step in the feature extraction process was feeding the images within the Vuforia library, for it to extract features from them as shown in [14]. The library applies natural feature extraction to identify contours and edges from within the image and as a result display a 5-star rating according to how augmentable the image is. Initially when corridors where tried as markers the library had a problem with identifying key features from plain white painted corridors, as they appeared to show no distinct edges and contours. However, then the door signs were used as markers, the images representing the markers were sharpened using OpenCV, taken in proper lighting scenes and fed into the library. The result is as shown in figure 10. The sharpening enhances the strength of certain edges and corners making the marker more detectable for the library.

5.2.2 Model Targets

The second step was feeding the 3D models within the model target generator. Vuforia's object recognition utilizes natural feature tracking by analyzing the object at 3 different axis. Vuforia does not specifically say what deep learning techniques they make use of. However, according to [14] they are most likely making use of Interest Point Detection, which makes use of a set of images of the same object at different scenarios, angles and lighting. The model target generator takes a maximum of 20 model targets. Their CNN is then trained on that one particular 3D object and meta data is outputted which can later be fed into Unity. The model target generator trained a total of ten 3D objects which consist of the coffee machine and the door markers.As shown in figure 11 the model's distance, angle and orientation are adjusted for training.

5.2.3 User-Query

The final step was extracting features from the Tasks dataset for the recommendation feature. The features extracted where the user id, the user-task id and the rating. This was done the same way for the three Tasks (Interview, Visit, and Delivery) separately, since there is no connection between each task and therefore the recommendation system is best trained on each one individually. In the next section the recommendation system shall be further explained in detail.

5.3 User-Query Recommendation

5.4 Overview

As previously explained in Chapter 4, along with the Augmented Reality it was thought to implement a recommender system depending on the profile of the user (Visitor or Intern). The system would incorporate functionalities of an item-item similarity based recommender system and functionalities of a collaborative filtering system.

5.4.1 Similarity Based

The item to item based similarity based recommendation consists of two main components. A user to item matrix which in our case is the user query, this changes depending on the user's preference and the item-item matrix. The latter matrix stays constant and represents the similarity one location has with another based on whether the user passes right in front of it, when trying to find an office in particular. The locations chosen were the accountant's office, the human resource's office, the manager's office, the secretary's office, the kitchen, the toilet rooms and the games room. Each row was computed using the term frequency approach. TF was chosen to enhance the relevant importance of a sought query, in this case the office one wants to go to, TF-IDF was not used as matrix scaling was done using one-fold normalization due to having a term (a location within the office) appearing in every task (a task is equivalent to a document). Below one can see the final normalized item-item similarity matrix.

$$
\begin{array}{cc}
\begin{array}{ccccccc} Accountant & HR & Manager & Secretary & Kitchen & Toilets & GamesRoom \end{array} & \\
\begin{pmatrix}
1.00 & 0.14 & 0.43 & 0.57 & 0.71 & 0.57 & 0.23 \\
0.23 & 1.00 & 0.86 & 0.57 & 0.43 & 0.57 & 0.43 \\
0.23 & 0.57 & 1.00 & 0.57 & 0.43 & 0.57 & 0.43 \\
0.23 & 0.14 & 0.43 & 1.00 & 0.43 & 0.86 & 0.43 \\
0.43 & 0.14 & 0.23 & 0.71 & 1.00 & 0.86 & 0.23 \\
0.23 & 0.14 & 0.23 & 0.43 & 0.23 & 1.00 & 0.23 \\
0.57 & 0.43 & 0.23 & 0.14 & 0.23 & 0.23 & 1.00
\end{pmatrix} &
\begin{array}{c} Accountant \\ HR \\ Manager \\ Secretary \\ Kitchen \\ Toilets \\ GamesRoom \end{array}
\end{array}
$$

5.4.2 Collaborative-Based Filtering Techniques

The collaborative filtering approach is a model built based on user's past experiences of the decisions, locations of interest picked and task they had in completing their own task. However, initially there were not any similar systems, where users could rate their experience. Therefore, the data was not gathered but was created solemnly to provide a prototype of the system and evaluate its performance. For building the collaborative filtering techniques the following python packages were used; pandas, numpy and sklearn. Each dataset contains 101 user ratings. Following a similar procedure as used in [35], the system uses a truncated Single Value Decomposition to generate a utility matrix. To avoid having empty fields within the utility matrix, add-k smoothing is applied where k is of 0.01. The matrix is then transposed and decomposed using TruncatedSVD. The decision to apply SVD based algorithms instead of PCA or CA-CF was to achieve a higher accuracy [35]. Finally a correlation matrix is generated and the final recommendation is based of it, where the results are then all stored in separate CSV files and fed into Unity Editor.

The three top recommendations are then presented within the AR application and whichever one the user chooses, the system then multiplies its item to item similarity matrix with that of the user's query matrix using dot product and the result matrix is finally obtained. The result matrix's elements represent the size of a location's gameobject within the 3D augmented map sketched. By applying a threshold value of 0.4 to a gameobject's size, if its size is below it, it simply does not appear within the 3D holographic map. The larger the size of an office pinpoint location is, the higher it is being recommended to the user. Therefore, that office in particular, has a higher relevance than the rest.

5.5 User Interface and System Architecture

5.5.1 System Architecture

There are three main components which make up the system architecture. The first most essential component is the user. The user is important as the system must serve as an essential tool for assisting them around the workplace. Their decision making process, drives the Augmented Reality application capabilities to its extent. The second component, is the recommendation system which as previously explained provides collaborative and similarity based filtering techniques depending on the user's profile. The third component is the Augmented Reality provided by the Vuforia. This component can be further subdivided into two other sub-components: the SDK library and the built architecture within unity tailor made for this project.

As explained in [17] the Vuforia AR SDK consists of the smartphone's camera and the target resources (the targets' database) communicating with the tracker. The tracker then detects the real world objects, converting each frame and snapshot to render augmented logic back on the user's smartphone. Figure 12 shows graphically how the Vuforia AR SDK works.

The built architecture within unity consists of three features. The first feature is called "Offices". This feature allows the user to wander around the workplace and view augmented information from the offices' marker. The second feature is "Locate" this offers office information and offline directions to the user, depending on their preferred recommendation. The third feature is "coffee machine", which provides employees and visitors with information of how to make use of the coffee machine. This information is provided via text and video instructions.

The main challenge encountered, was in the first two features. Vuforia can detect multiple image targets, yet can only detect one model target for every scene within unity. However, using model targets in certain scenarios can provide a more efficient user experience as the real world object is more recognizable due to deep learning techniques being utilised for recognition, rather than just traditional computer vision techniques. Therefore, a combination of image targets and model targets was used. The object is first detected using traditional computer vision methods, then once it is recognized the user if interested, is prompted to view relevant information according to the marker being recognized. Once the user prompts to view the respective augmented information the model target is activated. Once the user exits from the augmented information, the respective model target is deactivated. Therefore, through a combination of traditional computer vision and deep learning a more accurate system is provided, which as a result can offer efficiency and an immersive experience.

5.5.2 User Interface

When the application starts, the user is presented with three options: "Offices", "Location", and "Coffee Machine" as shown in figure 13. The "Offices" directs the user to the augmented reality system, where the user can wander around the workplace and view augmented information about the offices. "Location" directs the user to selecting a profile between an intern and a visitor. For the Visitor's profile the user picks a task from the following options: visit, interview, and delivery. As shown in figure 16 the user is provided with three recommendations. Each recommendation contains features which are recommended via an approval green sign, whilst the ones which are not recommended are represented by a X-sign in red. The approved signs for accountant, human resources, manager and secretary are clickable. Once any one of them is clicked, it will direct the user to the augmented reality system, where they can view augmented information about the offices through via the door markers and locate the respective office which was previously clicked. Each office augments its own respective main menu as shown in figure 17.

The details panel provides details of the room and the "Localize" button augments a 3D hologram of the workplace and provides the user with hard-coded directions towards the office they are interested in as shown in figure 19 and in figure 20. As one can observe the offices in the hologram are colour coded and are represented as a sphere. They are scale-wise recommended, this means that the larger the office's marker appears the more important and relevant it is to the user. The offices which do appear in the hologram, are the result of the collaborative and similarity based filtering techniques provided by the recommendation system.

The coffee machine functionality allows the user's phone camera to recognize the DeLonghi Esam2600 Coffee machine. The user can view text and video instructions on how to make a cappuccino, view a diagram of the machine's functionalities and a 3D view of the machine. This has further potential, such as controlling the coffee machine through the augmented reality system itself.

5.6 Conclusion

In this chapter we have seen in the detail how the workplace assistant augmented reality application was implemented and the decisions taken behind every feature which were provided. In the next chapter, tests and evaluations shall be provided to highlight the system's performance.

Appendices

A Chapter 5

Figure 4: Corridor Image Variation 1

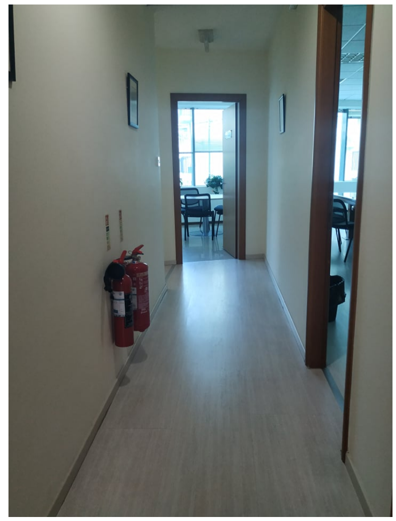

Figure 5: Corridor Image Variation 2

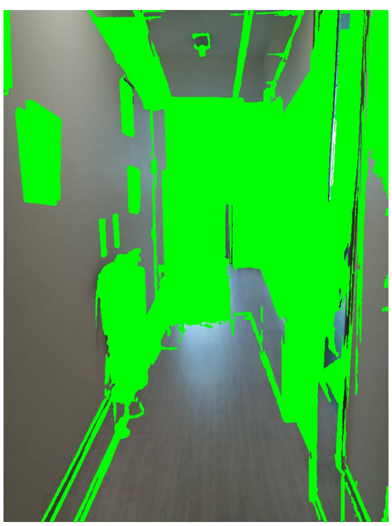

Figure 6: Result of differences using SSIM

Figure 7: Door Marker

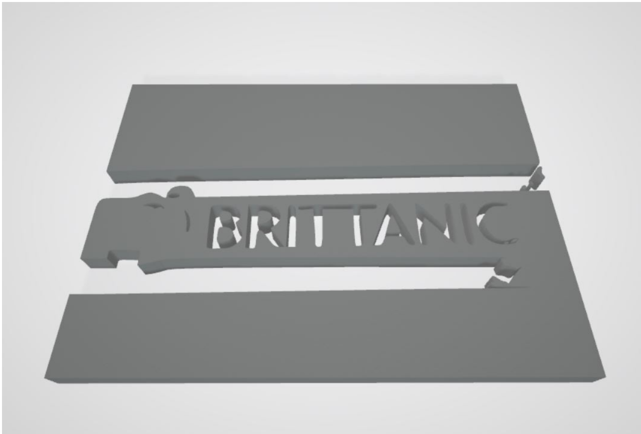

Figure 8: Door Marker 3D Object

Figure 9: Coffee Machine 3D Object

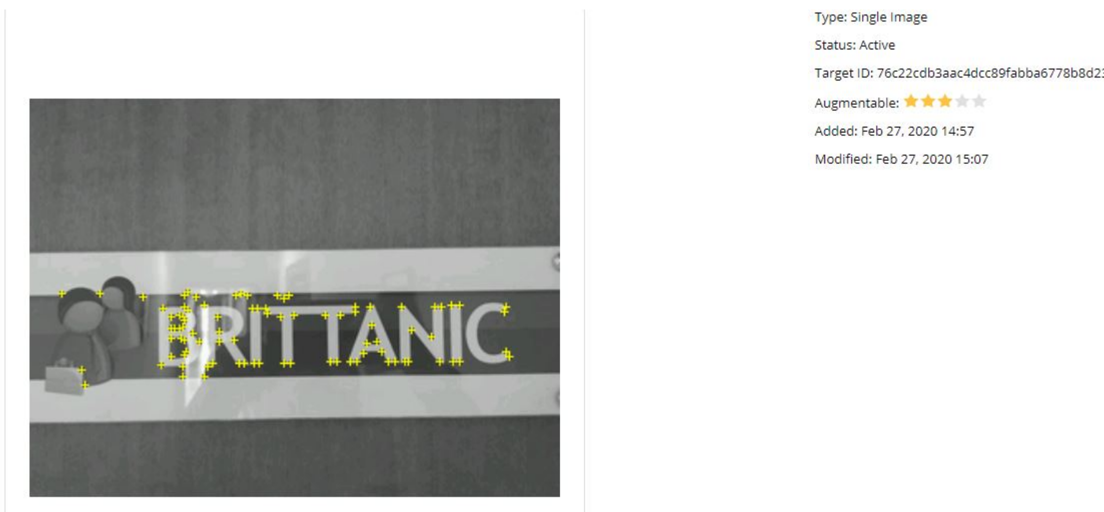

Figure 10: Image Target

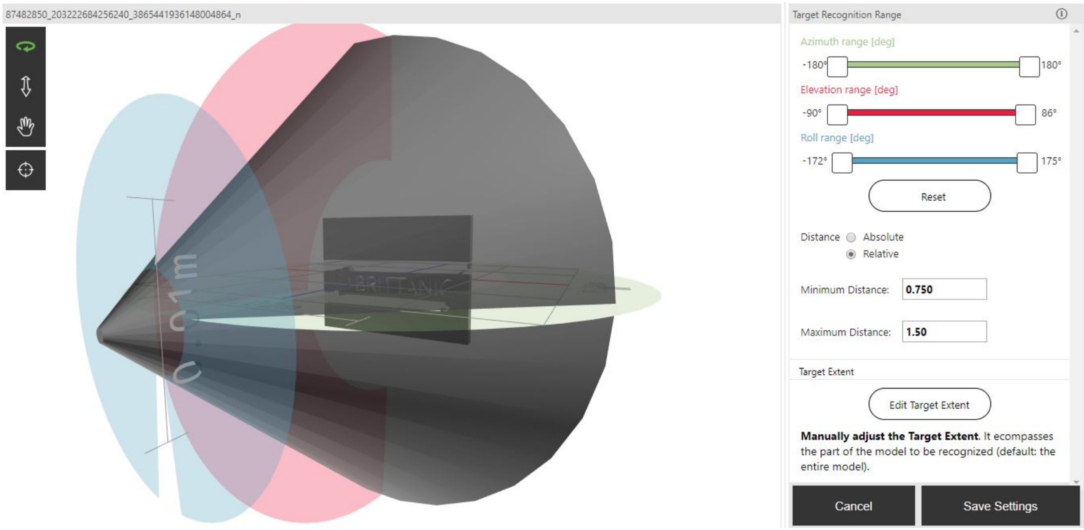

Figure 11: Model Target

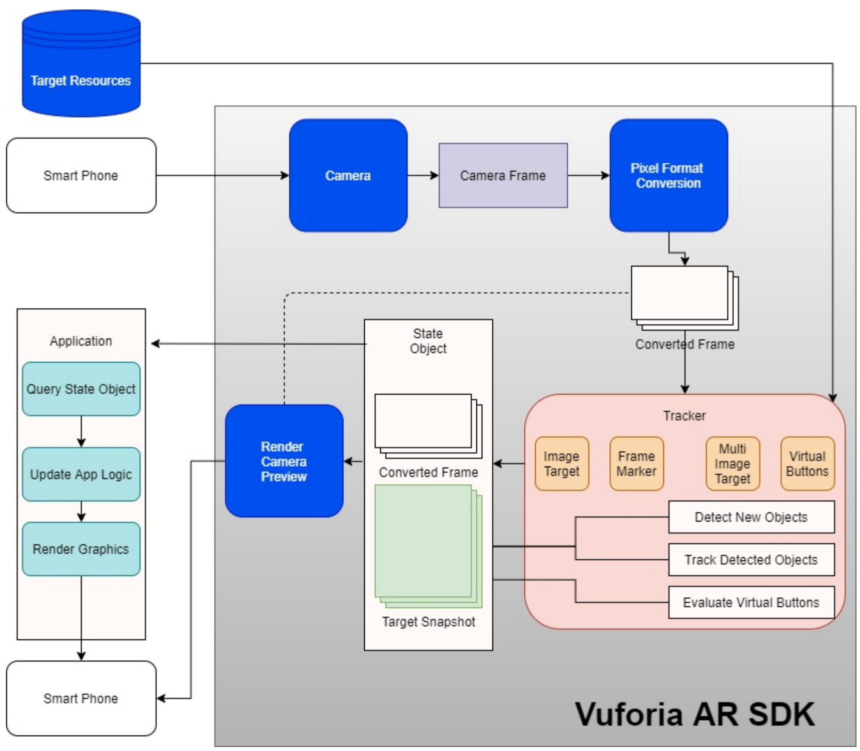

Figure 12: Vuforia AR SDK [17]

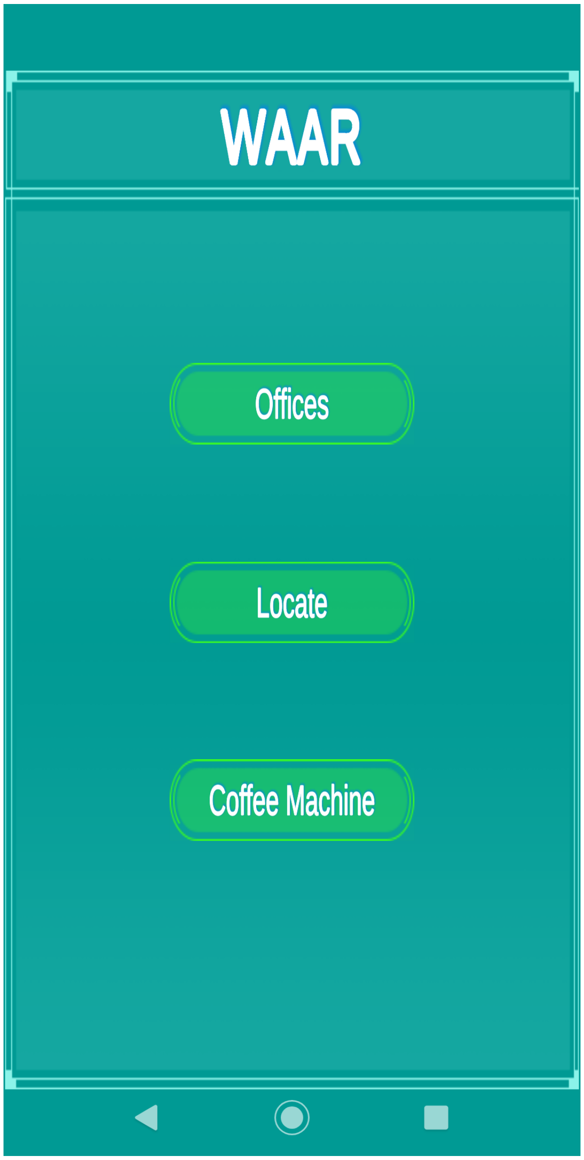

Figure 13: Main Menu

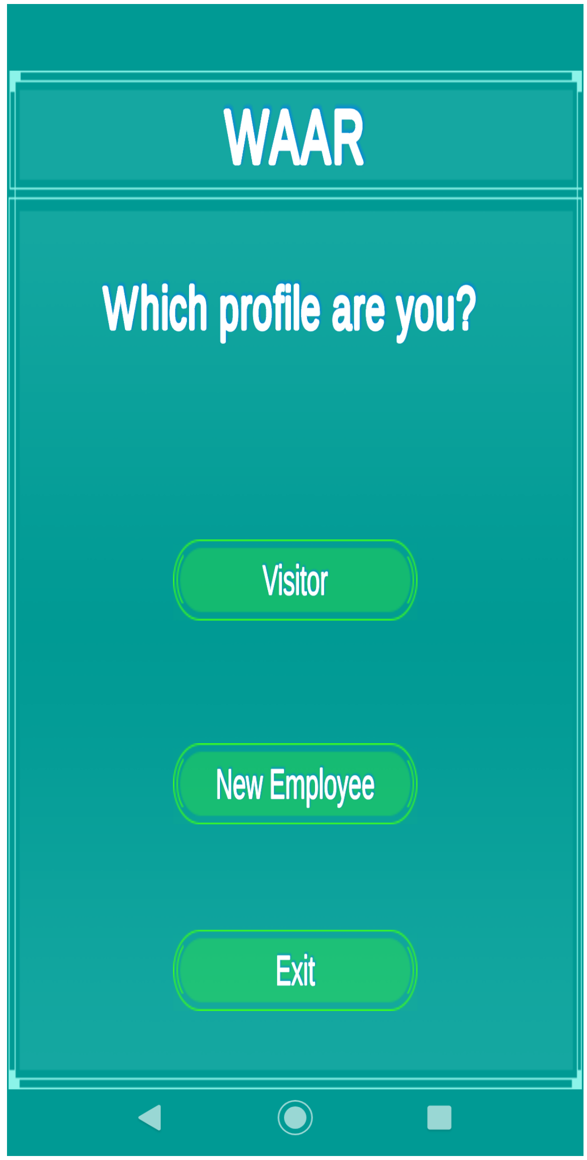

Figure 14: Choosing user profile

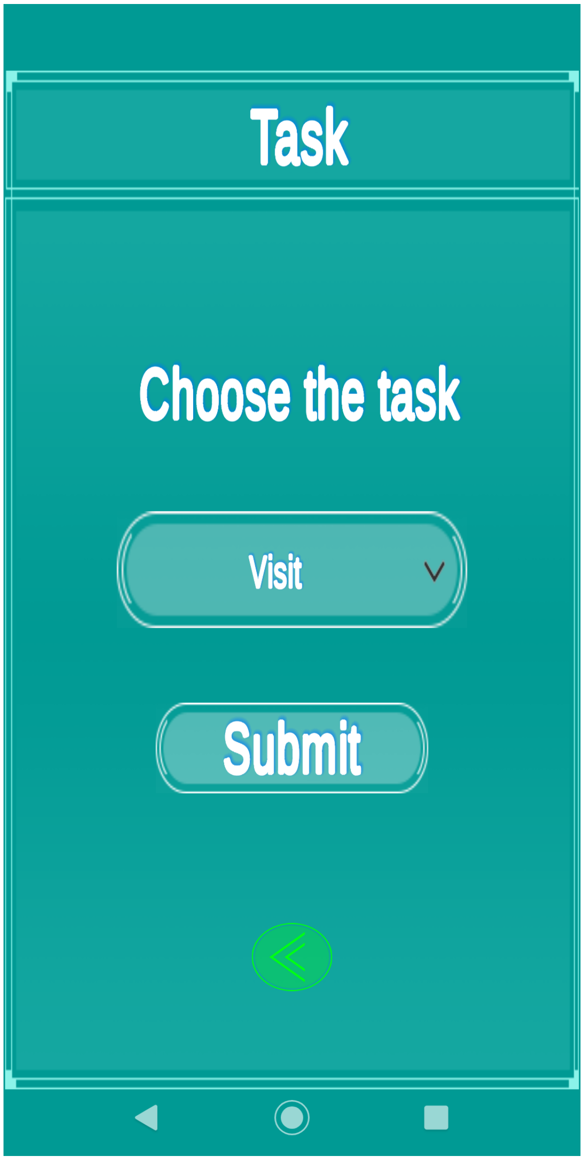

Figure 15: Choosing a visitor task

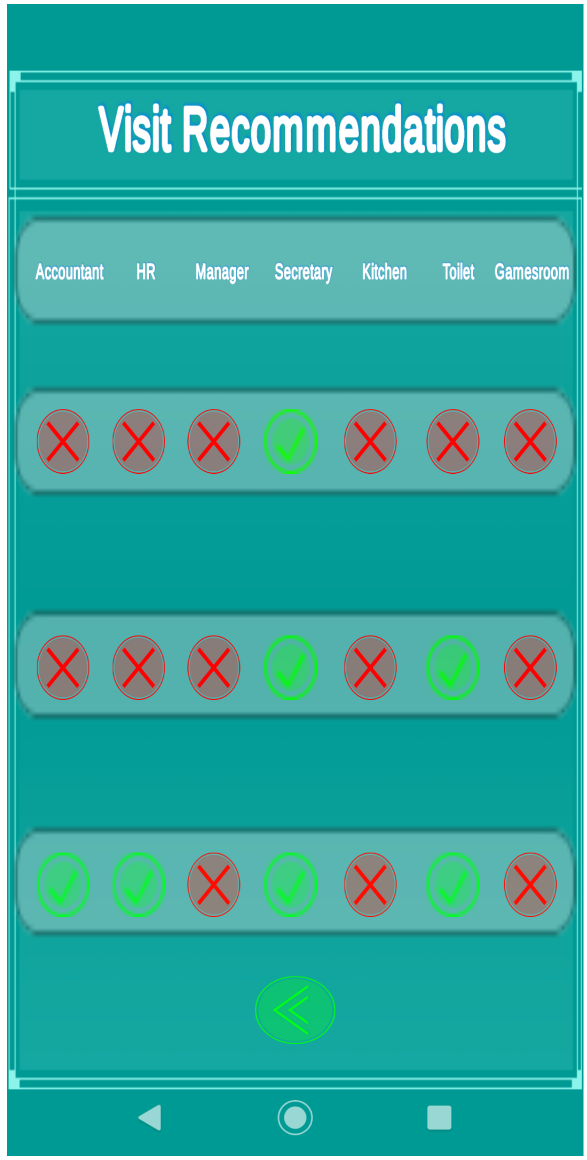

Figure 16: Visitor Recommendation for Visit

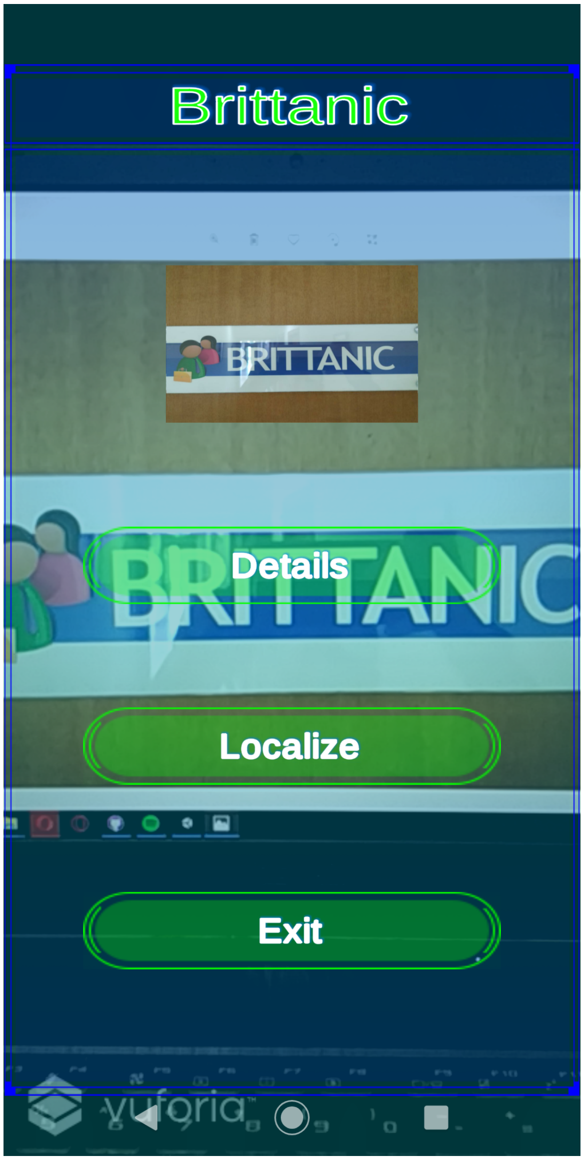

Figure 17: Marker Augmentation Menu

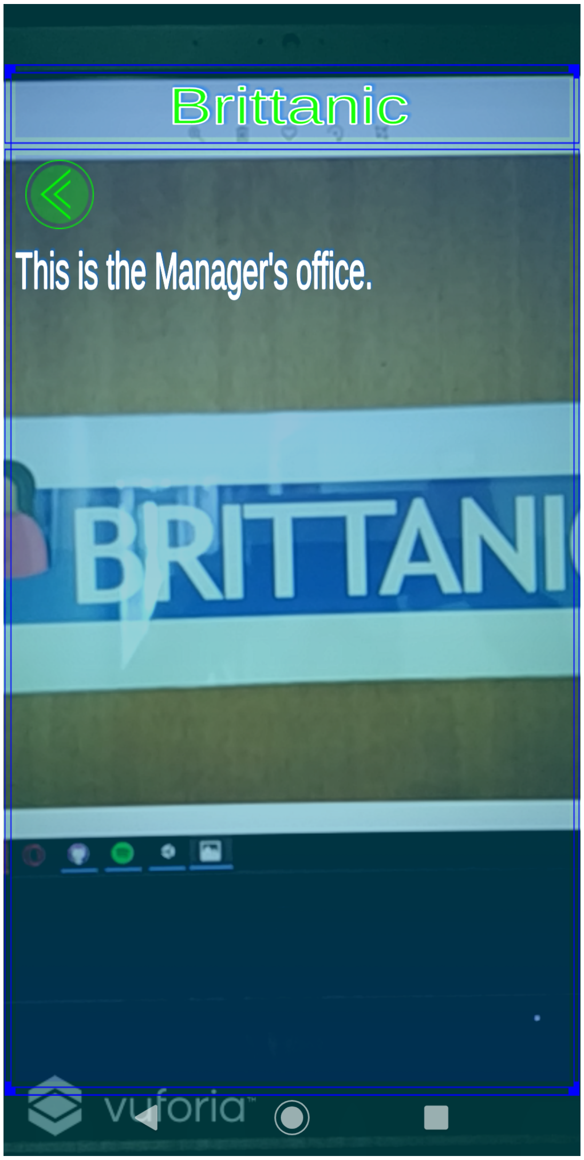

Figure 18: Office Details

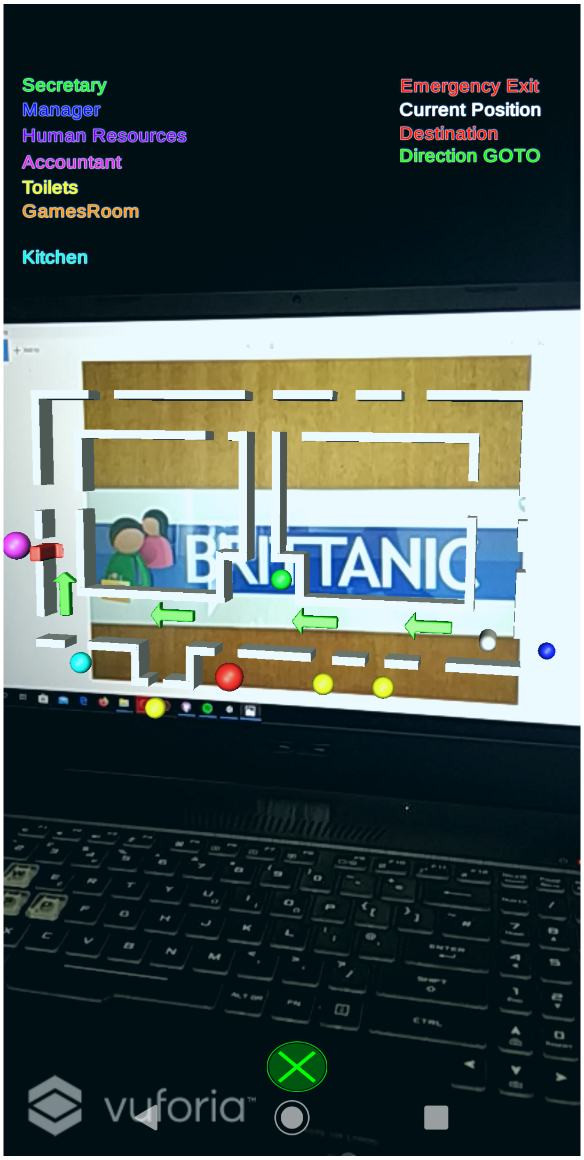

Figure 19: Holographic Sketch Map Front View

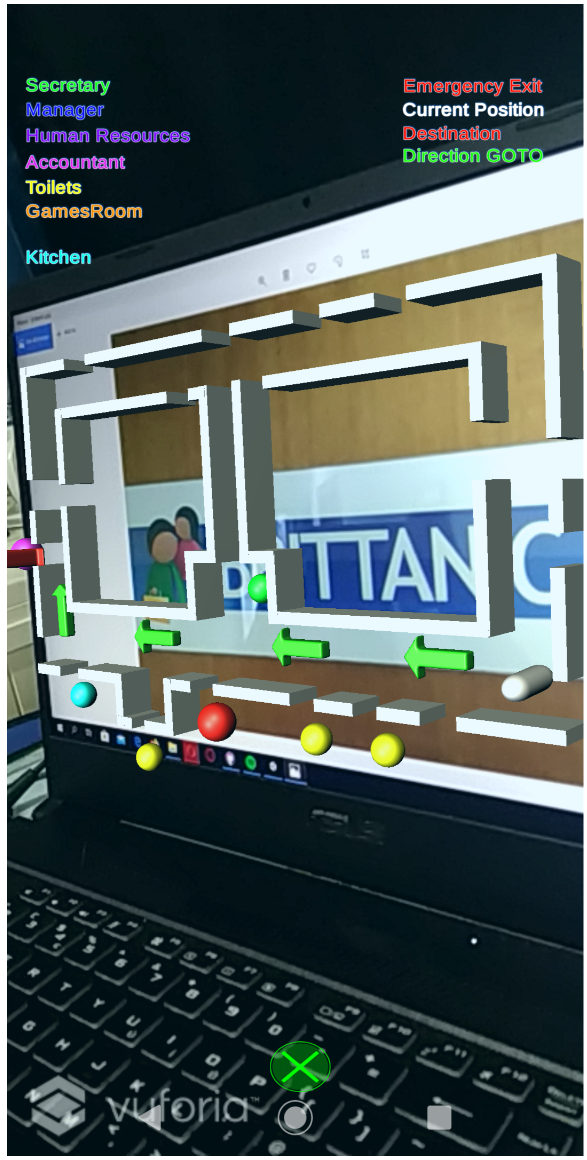

Figure 20: Holographic Sketch Map Side View

Figure 21: Coffee Machine Augmentation Menu

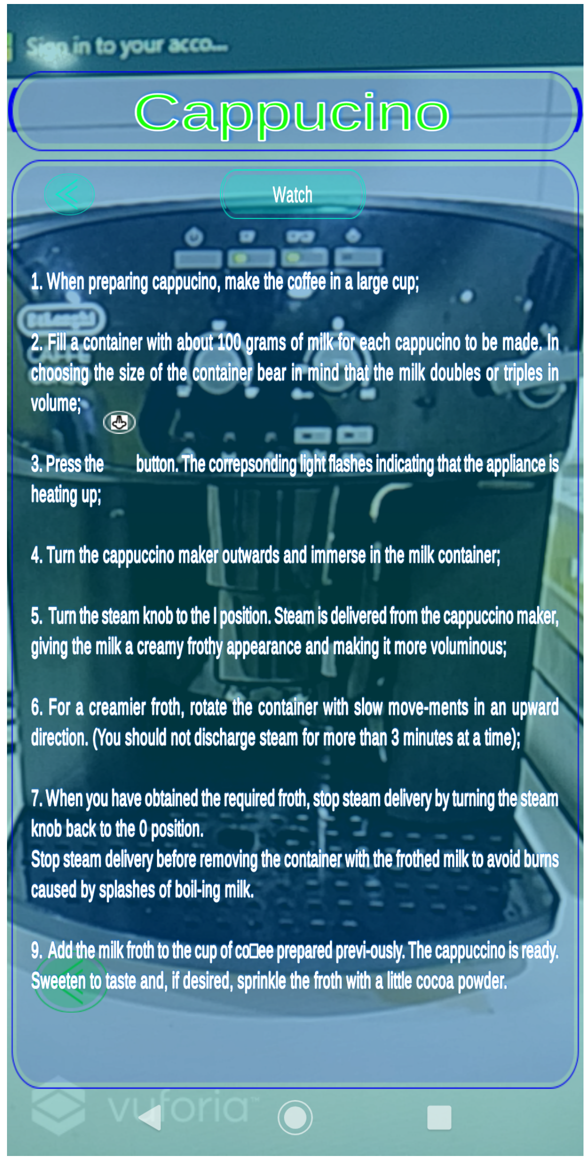

Figure 22: Cappuccino Details

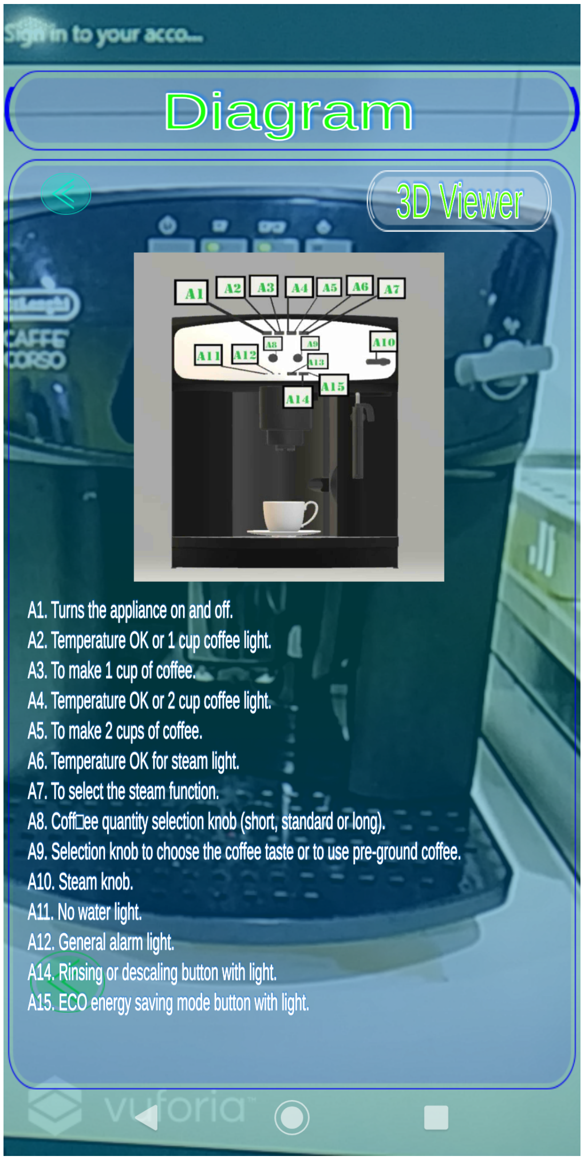

Figure 23: Coffee Machine Diagram

Figure 24: Coffee Machine 3D View

References

[1] Model target generator user guide.

[2] Optimizing target detection and tracking stability.

[3] Advances in computer vision. *Advances in Intelligent Systems and Computing*, 2020.

[4] Gediminas Adomavicius and Alexander Tuzhilin. *Context-Aware Recommender Systems*, pages 217–253. Springer US, Boston, MA, 2011.

[5] Omer Akgul, H. Ibrahim Penekli, and Yakup Genc. Applying deep learning in augmented reality tracking. *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pages 47–54, 2016.

[6] Hassan Alhaija, Siva Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets computer vision : Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, 08 2017.

[7] Nader Barzegar and Shahroz Farjad. A study on the impact of on the job training courses on the staff performance (a case study). *Procedia - Social and Behavioral Sciences*, 29:1942 – 1949, 2011. The 2nd International Conference on Education and Educational Psychology 2011.

[8] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. volume 3951, pages 404–417, 07 2006.

[9] Gaurav Bhorkar. A survey of augmented reality navigation. *ArXiv*, abs/1708.05006, 2017.

[10] Carlos Plaza de Miguel. Arlodge : Context-aware recommender system based on augmented reality to assist on the accommodation search process. 2014.

[11] Nevin Deniz, Aral Noyan, and Öznur Gülen Ertosun. Linking person-job fit to job stress: The mediating effect of perceived person-organization fit. *Procedia - Social and Behavioral Sciences*, 207:369 – 376, 2015. 11th International Strategic Management Conference.

[12] James Ford and Tobias Höllerer. Augmented reality: Information for workplace decision-makers, managers, workers and researchers. 02 2020.

[13] Mathieu Garon and Jean-François Lalonde. Deep 6-DOF tracking. *IEEE Transactions on Visualization and Computer Graphics*, 23(11), 2017.

[14] Ivar Grahn. The vuforia sdk and unity3d game engine : Evaluating performance on android devices. 2017.

[15] Isabelle Guyon and André Elisseeff. *An Introduction to Feature Extraction*, volume 207, pages 1–25. 11 2008.

[16] T. Haritos and N. D. Macchiarella. A mobile application of augmented reality for aerospace maintenance training. In *24th Digital Avionics Systems Conference*, volume 1, pages 5.B.3–5.1, Oct 2005.

[17] Alexandro Simonetti Ibáñez and Josep Paredes Figueras. Vuforia v1.5 sdk: Analysis and evaluation of capabilities. 2013.

[18] Sumitkumar Kanoje, Sheetal Girase, and Debajyoti Mukhopadhyay. User profiling trends, techniques and applications. *ArXiv*, abs/1503.07474, 2015.

[19] Sumitkumar Kanoje, Debajyoti Mukhopadhyay, and Sheetal Girase. User profiling for university recommender system using automatic information retrieval. *Procedia Computer Science*, 78:5 – 12, 2016. 1st International Conference on Information Security Privacy 2015.

[20] Sneha kasetty sudarshan. *AUGMENTED REALITY IN MOBILE DEVICES*. PhD thesis, 05 2017.

[21] Stan Kurkovsky, Ranjana Koshy, Vivian Novak, and Peter Szul. Current issues in handheld augmented reality. pages 68–72, 06 2012.

[22] Andy Lee. Comparing deep neural networks and traditional vision algorithms in mobile robotics. 2016.

[23] Vincent Lepetit. On computer vision for augmented reality. pages 13 – 16, 08 2008.

[24] David Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 11 2004.

[25] Paramjinang Moita. Adjustment to the work place by new recruits in libraries. Vol.5(2):71–85, 04 2015.

[26] Jonathan Pedoeem and Rachel Huang. YOLO-LITE: A real-time object detection algorithm optimized for non-gpu computers. *CoRR*, abs/1811.05588, 2018.

[27] David Procházka, Ondřej Popelka, Tomas Koubek, Jaromir Landa, and Jan Kolomaznik. Hybrid surf-golay marker detection method for augmented reality applications. *Journal of WSCG*, 20:197–204, 01 2012.

[28] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook*, volume 1-35, pages 1–35. 10 2010.

[29] M. Romilly. 12 Best Augmented Reality SDKs. (2019, Jan 25).

[30] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pages 430–443, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[31] S. R. R. Sanches, M. Oizumi, C. Oliveira, E. F. Damasceno, and A. C. Sementille. Aspects of user profiles that can improve mobile augmented reality usage. In *2017 19th Symposium on Virtual and Augmented Reality (SVR)*, pages 236–242, Nov 2017.

[32] Rodrigo Silva, Jauvane Oliveira, and G. Giraldi. Introduction to augmented reality. 01 2003.

[33] Simon Taylor and Tom Drummond. Binary histogrammed intensity patches for efficient and robust matching. *International Journal of Computer Vision*, 94(2):241–265, Sep 2011.

[34] Mathangi Vijayan. Impact of job stress on employees' job performance in aavin, coimbatore. 06 2018.

[35] Manolis Vozalis, Angelos Markos, and Konstantinos G. Margaritis. Evaluation of standard svd-based techniques for collaborative filtering. 01 2009.

[36] Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, and Ling Chen. Lcars: a location-content-aware recommender system. pages 221–229, 08 2013.

[37] Zhuo Zhang, Shang Shang, Sanjeev R. Kulkarni, and Pan Hui. Improving augmented reality using recommender systems. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, page 173–176, New York, NY, USA, 2013. Association for Computing Machinery.