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 first 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 coffee 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 first 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 coffee 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"[12]. "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"[26]. The first 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"[26] 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"[23].

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 inefficiency in organizations"[12].

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"[12]. 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”[36].

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 off 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"[31]. 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 office 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 offline 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" [20]. 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" [20]. 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" [19]. 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" [19]. 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" [19]. 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" [34]. 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" [34]. 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" [34]. 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" [34]. 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" [22]. 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"[22]. 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 [21]], 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" [21], 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" [15]. 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 [25] and SURF [8] are commonly used for detection. The algorithm extracts a number of key points using a corner detection algorithm such as FAST [32]. 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 [35] 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

This chapter reviews the available literature on Workplace Assistant Augmented Reality, while discussing the two components involved in the Augmented Reality application and research that has inspired the approach selected for this project. The chapter is divided into three parts, namely, Image and Object Recognition techniques involved in Augmented Reality, the application of user proling methods with Augmented Reality, and dierent image and object recognition techniques in augmented reality technologies.

3.1 Workplace Augmented Reality

> [A]ugmented 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.[13]

Although the technology can be considered still 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 necessitates adjustment and some form of training for employees to adapt to the process of the work they might be doing. Augmented Reality may assist employees by providing them with additional overlaid instructions to guide them through the whole process of adjustment while providing them with training.

Workplace training is usually provided in two forms, namely, on the job training (OJT) and off the job training. "[O]JT may be viewed as an apprenticeship where a novice AMT is mentored by an AMT who is an expert" [17]. This is the traditional form of training, especially when teaching maintenance. However, "[O]JT may not be the best method for training because the feedback to learners may be infrequent and unmethodical" [17]. Conversely, o the job training may be provided through face to face conversations or through multimedia. Augmented Reality can combine the two aspects of training, where the user is given on the job training through multimedia, which overlays the real-world environment.

There are several useful applications for Augmented Reality at the workplace. However, not every workplace might necessitate AR. "[T]here 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" [13]. As discussed in [13], the following are workplace conditions where AR is applicable, namely, distance communication with 2D or 3D objects provided for visualisation, training and education when using real-life tools, recording of information obtained while training, and a collaborative design and interaction of 3D models.

An advantage which Augmented Reality offers to workers and managers is "[t]he ability to author their own environment by embedding the relevant information needed for task completion" [13]. Nonetheless, a common problem during work training is that the expert individual needs to provide the respective information to the trainee in the most understandable way possible. Therefore, through AR technology, the trainee can tailor how that information is presented, and thus, Augmented Reality may be capable of understanding its users such that it may adapt to future possible users.

3.2 Recommendation Systems for Augmented Reality

Information during job training is crucial for an employee to learn and adapt to the new environment. However, an overwhelming amount of information provided to a new employee may be demotivating. Augmented Reality is a tool that provides interactive information to the user while also obtaining information from the user. Nevertheless,

> ...[t]he 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.[33]

"[R]ecommender systems (RS) have proven to be a valuable tool for online users to cope with the information overload" [11]. Recommendation systems provide tailored information to dierent users based on their preference. "[T]hus, it is important to oer the user a personal response, but also a context-dependent and constrained by the limited computing capacities of the mobile devices" [11, 4, 38, 30]. 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 systems in Augmented Reality cannot be easily adapted and deployed since they dierentiate in the following areas: location, timing, first time use of the application, and immediate response from the AR application, as discussed in [39].

Distance-based filtering and visibility-based filtering are commonly used in Augmented Reality. In [39], a random walk algorithm was incorporated, whose recommendations are based on user preferences, behaviour patterns, history records, and information from social media. However, in the latter research, user feedback was not evaluated, which would have helped to provide 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, that is, the amount of time one would generally take to complete a task using the AR application. A task may take extensive 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 to solve the task, simply because it may lack dierent forms of interactive techniques. The study by [33], defines “[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”. The result of this research is that young groups of people spent less time completing a task using AR since they were accustomed to similar forms of technology. On the other hand, users with little to no experience and those of an advanced age took obviously longer to adapt 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 AR application cannot assume that the user will interpret easily what is being overlaid on the screen. They might need to be guided throughout the process in order to understand the meaning of dierent symbols being displayed, as well as their colour and size.

3.3 Computer Vision Approaches in Augmented Reality

Augmented Reality applications make use of several computer vision approaches to recognise images, objects and text. As previously discussed, Vuforia makes use of both traditional and deep learning approaches. Using deep neural networks ensures highly accurate and efficient results. However, “[i]t is well-known that training high capacity models such as deep neural networks requires huge amounts of labelled training data”[6]. Neural networks are data-hungry architectures that require huge amounts of data to train and test on, thus being capable of generalising accurately.

As discussed in [24], marker-based applications have been the main driving force to apply AR in real life. "[M]ost of the current approaches to 3D tracking are based on what can be called recursive tracking" [24]. Therefore, the system must be initialised manually, and with some occlusion between the camera and the object being recognised, the system fails to perform. However, a new computer vision approach has improved Augmented Reality, and can register the camera without camera pose introduction. This approach is called Tracking-by-Detection, and in [24], it is tested to determine its benets. The approach works by extracting feature points from inputted frames during run-time. The features are then "[m]atched against a database of feature points for which the 3D locations are known"[24]. However, there were still key limitations, such as, detecting reective and shiny surfaces on the car, since not many features could be extracted. Another limitation was dealing with occlusion, especially if a person were standing anywhere near the object being detected. One nal challenge was dealing with robustness due to the dierent environments where the object would be in order to improve object recognition and generalise.

According to the research conducted by [29], a S-G Hybrid Recognition method was implemented in order to solve the occlusion problem within current Augmented Reality technology. The approach takes "[a]dvantage of robustness of the SURF feature-based object identication and combine it with high reliability and eectiveness of the Golay error correction code detection" [29]. SURF and 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 identication 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. [29]

The researchers tested the three main aspects which may hinder an AR application, namely, distance variance, angle variance, and occlusion. Consequently, through the S-G approach, it was found that an object can be placed 2m away from the camera, while the comparison of the angles was completely inuenced by the SURF algorithm which was able to detect under 55 degrees angle to the camera's axis, and that it could not be aected by up to 55% obstruction.

Another approach to solve the occlusion problem in AR is to apply deep learning techniques, as described in [14], where the researchers "[p]resent a temporal 6-DOF tracking method which leverages deep learning to achieve state-of-the-art performance on challenging datasets of real world capture" [14].Deep learning architectures can be trained on large amounts of data thereby solving occlusion, angle variance, and distance variance problems. Their approach involved getting a 3D model of the object and training the tracker for that specic object. Training involved two steps; firstly, using a frame to capture the object in its predicted position, and secondly, the frame of the object's actual position. "[T]o 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" [14].

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. The study by [27] presents "[Y]OLO-LITE, a real-time object detection model developed to run on portable devices such as a laptop or cell phone lacking a Graphics Processing Unit (GPU)". YOLO-LITE is primarily designed to obtain a smaller, faster, and more efficient model. "[Y]ou Only Look Once (YOLO) was developed to create a one step process involving detection and classification. Bounding box and class predictions are made after one evaluation of the input image"[27]. The developed architecture runs at 10 frames per second, and its goals are to prove that shallow networks can run on non-GPU devices, and that shallow networks do not require batch normalisation. The model had 18 trials, obtaining results of 33.77% mAP and 21 FPS, and 12.26% and 21 on PASCAL VOC and COCO dataset, respectively.

3.4 Conclusion

Dierent approaches were dened and reviewed in this chapter. Traditional computer vision techniques, deep learning techniques, recommender systems, and Augmented Reality solutions were analysed to acquire a state-of-the-art Workplace Assistant Augmented Reality application. The following chapters will present and discuss the design and implementation of the proposed method.

4 Design

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

4.1 Overview

The AR 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 aspect of the application, and 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 entails 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 a 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, while phase two entails building 3D models of chosen markers for the Augmented Reality. These 3D models must capture as much detail as possible of the actual marker. The images and 3D models are then fed into Vuforia's Library for training. The third phase is gathering data from a good number of previous users who rated the application when they performed a task. This third phase is necessary in order to perform collaborative filtering techniques using a set of machine learning algorithms and probabilistic methods to achieve a set of user preference recommendations. The yielded results are subsequently combined with the item similarity-based matrix.

1. Image extraction process of the workplace;
2. Building 3D models of the markers chosen from the extracted images; and
3. Building a user-task rating dataset of the previously used system.

4.3 Feature Extraction

As explained in [16], feature extraction follows two steps, namely, 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 [25] and Surf [8] algorithms. The next step involves passing the 3D models to Vuforia's model target generation [1], making use of the library deep learning techniques. In this project for the collaborative filter techniques, the Truncated SVD (Singular Value Decomposition) model performs feature extraction using the rating and user-task ID features.

4.4 User Query Model

Two user-query models will be created. The first one is related to the intern, where every intern goes through a similar process of integration on their first day on the 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 to choose whether they prefer the following for augmentation: the games room, restrooms, or kitchen. Then, the next part is 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 according to the preferred options the user would have previously queried and the office they wish to visit. Figure 2 further explains the flow of logic.

The second user-query model is related to the visitor querying the augmented reality system. Here, the visitor is presented with the top 3 recommendations according to which task they would need to accomplish when visiting the office. The system must accommodate visitors with the following tasks: a delivery, an interview, and a visitation. Once the user chooses the relevant task and selects 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 further considers the rooms and offices which the user will walk past while visiting an office. Therefore, it not only considers what the user prefers based on previous users' preferences, but also what they might require based on the location they are in. Therefore, the user-query model for the visitor makes use of both collaborative and item similarity-based filtering techniques. Figure 3 further details the visitor's query model.

4.5 User Interface

The user interface must be as user-friendly as possible, and needs to provide the user with several options about what information they are interested in. The user will be presented with a main menu, allowing them to augment information about the coffee machine, or the offices while wandering around, or to locate an office. The coffee machine interface is augmented once the coffee machine is recognised, 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 recognises 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 the scope of this research to implement it. Navigation is provided through a 3D sketched holographic map which gives an idea of where the visitor or intern needs to go to find the office.

4.6 Conclusion

This chapter discussed the custom design of the methods to apply the workplace assistant augmented reality, while taking into consideration previously applied implementations and methods using Augmented Reality in similar areas. The following chapter will delve further into the implementation and application of the designed methods.

5 Implementation

This chapter discusses in further detail the components proposed in the previous chapter. The implementations used a number of applications and libraries, namely, Vuforia [15], Unity Engine [15] and Python 3.7. Vuforia was mainly used to train image and model targets, while Unity is a game engine/editor used in this project to build the augmented reality application architecture, and Python 3.7 was used for image sharpening and for training and testing the recommendation system. This chapter highlights all the necessary details about data handling, feature extraction, user-query recommendation, user interface, and system architecture which was 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" [15]. Similarly 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

The initial step of the project entailed gathering many images of the first floor of the workplace. Indeed, a good number of images were taken of all the corridors and objects inside them, as well as the doors. Images of the interior of the office rooms on the first oor were excluded since it is not within the scope of this project to capture them within the augmented reality application. Dierent variations of the same image were captured, ensuring to capture possible dierent lighting and time scenarios. The reason behind this was mainly because GPS signals are weak indoors and Vuforia does not support location-based tracking. Therefore, the Augmented Reality had to be marker-based through image and model targets. The challenge in this case was capturing the same images in dierent scenarios. For example, if one door was open in one image, another image had to be captured where the door was closed thereby training the system in the various settings and contexts.

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 it can affect entirely the user's experience within the augmented reality. Initially, it was thought to use the images of the corridors themselves that were going to be used as markers. However, this meant that for one corridor, one needed to take several images capturing an innitely amount of variations which could occur in that image. For example, in figures 5 and 6, one can see several clear sight variations of the same corridor. By finding differences in edges and contours computed in figure 7, a structured similarity index of 0.62 was found, which is very low similarity, hence making it dicult to augment corridors. Therefore, it was decided to use door signs as markers. The oor used for the application had a number of doors, each with its unique sign placed at the centre. Therefore, the signs were used as markers for their static looks, and thus, it was very unlikely for the marker to dier in dierent images taken at dierent times and angles, especially considering their unique attributes (figure 8).

5.1.3 3D Models

To build the model targets, one must provide a 3D object within the Vuforia Model Target Generator, as explained in [15]. The 3D models generated were those of the door markers placed around the whole oor area as well as the coffee machine (figures 9 and 10). 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 no previous similar applications for users to rate the system. Therefore, a prototype recommendation system was built through generated data, simply to analyse how the system works with the augmented reality application. As previously mentioned, three tasks were picked for a visitor prole which the collaborative techniques would be applied to, namely, an interview, delivery, and site-visitation. For each task, a dataset containing the user-tasks and the rating of that task was created. Every entry had the following attributes: user id, locations of interest, user-task id, and rating. Locations of interest is a 7-bit binary code representing the accountant's office, the human resources office, the manager's office, the secretary's office, the kitchen, the toilet rooms, and the games room respectively (0 meaning not interested 1 meaning interested). Each dataset has a respective smaller dataset containing several locations of interest which were activated by previous users. The primary key of the tasks' dataset is the foreign key within the user-task dataset called 'user-task id'.

5.2 Feature Extraction

5.2.1 Image Targets

The first step in the feature extraction process was feeding the images into the Vuforia library which extracts features from them, as shown in [15]. The library applies natural feature extraction to identify contours and edges from within the image, and consequently, display a 5-star rating according to how augmentable the image is. Initially, when corridors were 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, and due to lighting, the images had a glossy texture which was harder for Vuforia to track. However, the door signs were eventually used as markers, and the images representing the markers were sharpened using OpenCV, taken in proper lighting scenes and fed into the library. Figure 11 illustrates the result. 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 into the model target generator. Vuforia's object recognition utilises natural feature tracking by analysing the object at 3 dierent axes. Vuforia does not say specically what deep learning techniques they use. However, according to [15]], they are most likely making use of Interest Point Detection, which uses a set of images of the same object at dierent scenarios, angles, and lighting. The model target generator takes a maximum of 20 model targets, whose CNN is then trained on that one 3D object, and meta data is outputted and later fed into Unity. The model target generator trained a total of 10 3D objects, which consist of the coffee machine and the door markers. As shown in Figure 12 the model's distance, angle and orientation were adjusted for training, hence making the object easy to track.

5.2.3 User-Query

The nal step was extracting features from the tasks' dataset for the recommendation feature. The features extracted were the user id, user-task id, and rating. All the three tasks (i.e. interview, visit, and delivery) underwent the same process separately since there is no connection between each task, and thus, the recommendation system is best trained on each one individually. The recommendation system will be explained further in the next section.

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 recommendation system, depending on the user's prole (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 similarity-based recommendation consists of two main components, namely, a user to item matrix which, in our case, is the user query, that changes according to the user's preference, and the item-item matrix. The latter matrix stays constant and represents the similarity between one location and another, based on whether the user passes right in front of it when trying to nd an office. The selected locations were the accountant's office, the human resources 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 since matrix scaling was done using one-fold normalisation since a term (a location within the office) appears in every task (a task is equivalent to a document). Below is the nal normalised item-item similarity matrix.

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

5.4.2 Collaborative-Based Filtering Techniques

The collaborative filtering approach is a model based on the user's past experiences of the decisions taken, locations of interest picked, and task they had in order to complete their own task. However, initially, there were no similar systems where users could rate their experience. Therefore, it was not possible to gather data, but instead, it was created solely to provide a prototype of the system and evaluate its performance. To build the collaborative filtering techniques, the following python packages were used, namely, pandas, NumPy, and Surprise. Each dataset contains 1000 user ratings. Following a similar procedure as used in [37], the system uses a Single Value Decomposition Plus Plus provided by the Surprise library. The decision to apply SVD based algorithms instead of PCA or CA-CF was to achieve a higher accuracy [37]. An SVD++ model works similar to an SVD model but has the capability to infer implicit data from explicit ratings. The SVD model generates a utility matrix, where from that matrix the model generates a number of predicted ratings along with the task id. The top three predicted ratings' 'taskId' are then stored in a csv file and fed into Unity.

The three top recommendations are then presented within the AR application. Whichever of them the user chooses, the system multiplies its item to item similarity matrix by that of the user's query matrix using dot product, finally obtaining the result matrix, whose 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 this value, it simply does not appear within the 3D holographic map. The larger the size of an office pinpoint location is, the more it is recommended to the user. Therefore, that office 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, who is important as the system must serve as an essential tool to assist them around the workplace. Their decision-making process drives the capabilities of the Augmented Reality application to their extent. The second component is the recommendation system which, as previously explained, provides collaborative and similarity based filtering techniques, depending on the user's prole. The third component is the Augmented Reality provided by Vuforia. This component can be further subdivided into two other sub-components, namely, the SDK library and the built architecture within Unity tailored for this project.

As explained in [18], 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 13 shows graphically how Vuforia's AR SDK works.

The built architecture within Unity consists of three features. The first feature is called 'Offices', which allows the user to wander around the workplace and view augmented information from the office markers. The second feature is 'Locate', which 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 on how to make use of the coffee machine. This information is provided through text, images, and video instructions.

The main challenge encountered was in the first two features. Vuforia can detect multiple image targets, but 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 recognisable since deep learning techniques are 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, and once it is recognised that the user if interested, can prompt to view relevant information according to the marker being recognised by clicking the open button. The model target is activated once the user prompts to view the respective augmented information, and is deactivated once the user exits the augmented information. Therefore, through a combination of traditional computer vision and deep learning, a more accurate system is provided thereby offering efficiency and an immersive experience. It is further a solution to one of Vuforia's lacking features which still does not enable the users to use multiple model targets in one. Therefore, the aforementioned technique served as a workaround to use multiple model targets.

5.5.2 User Interface

When the application starts, the user is presented with three options: 'Offices', 'Locate', and 'Coffee Machine' (figure 14). 'Offices' directs the user to the augmented reality system, where the user can wander around the workplace and view augmented information about the offices. 'Locate' directs the user to select a prole between an intern and a visitor. For the visitor's prole, the user chooses a task from the following options via a dropdown button: visit, interview, and delivery. As shown in Figure 17, the user is provided with three recommendations, each containing features which are recommended via an approval green sign, whereas those which are not recommended are represented by an 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 via the door markers and locate the respective office which was previously clicked. Each office augments its own respective main menu (figure 18).

The details panel provides details of the room, while the ‘Locate’ button augments a 3D hologram of the workplace and provides the user with hard-coded directions to the oce they are interested in (figures 20 and 21). It can be observed that the offices in the hologram are colour-coded and are represented as a sphere. They are scale-wise recommended, that is, the larger the office 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 ltering techniques provided by the recommendation system.

The ‘Coffee Machine’ functionality allows the user’s phone camera to recognise 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 view a 3D model of the machine.

5.6 Conclusion

This chapter discussed thoroughly the implementation of the workplace assistant AR application, as well as the decisions taken for every feature provided. The next chapter will provide the tests and evaluations to highlight the system’s performance.

6 Testing and Evaluation

6.1 Methods of Evaluation

The workplace assistant augmented reality system was evaluated using the images of the markers at different angles, distances, lighting and scenarios to ensure that no bias was developed towards the tested markers. The system was tested both quantitatively and qualitatively. The two components which were tested were the augmented reality and the recommendation system.

6.2 Augmented Reality Quantitative Testing and Evaluation

Ideally, the system would be tested by evaluating the precision and recall of the traditional computer vision and deep learning techniques utilised. However, Vuforia does not allow its users full access to its API. One can also assume that since Vuforia is a commonly used commercial library, therefore it is most certainly making use of efficient and accurate models for training and testing. However, as seen in [28], quantity testing on the AR application was also achieved by testing its competency in recognising and augmenting the markers used in different scenarios, that enable the application to be tested to its limits. The evaluations used were: colour variance, distance variance, rotation variance and occlusion variance. The tests were performed using a Xiaomi Mi A2 smartphone. Tests were done on the model target and image target separately. Image targets amount

Device	**Specifications**
Xiaomi Mi A2	octa-core (4x2.2GHz + 4x1.8GHz) processor 4GB Ram, 64GB internal storage 12MP + 20MP dual Camera 20MP front camera

Table 1: Smartphone Specifications

up to 10 whilst model targets amount up to 11 which includes the coffee machine. Hence, in total combining the image and model targets, one test is repeated 21 times. Vuforia also provides a 5-star rating to every image target which represents the detection quality.The higher the rating is the better the image target is detected.

Image Target	**Rating**	**Figure**
Abessinia	4	Figure 26
Ayanami	3	Figure 27
Brittanic	3	Figure 28
Enola	3	Figure 29
Hellespont	3	Figure 30
Pomeranian	2	Figure 31
Secretary Office	5	Figure 32
Sirenia	2	Figure 33
Takanami	2	Figure 34
Titanic	1	Figure 35

Table 2: Image Target Rating

6.2.1 Color Variance

For color detection the images were turned to grayscale and each image was successfully detected via the AR application. Firstly as presented in [15], this is due to the fact that

Vuforia uses natural features such as edges and contours to detect images both in the traditional computer vision techniques and the pose estimation deep learning techniques. Secondly, the images when analysed by the library are analysed in grayscale to be able to generalise rather than form a bias for images or objects with a specific color.

6.2.2 Distance Variance

The distance variance tests the maximum range the smartphone can be away from the object or image being recognised. Each image and model target was tested and the mean variance, standard deviation, and mean standard error were calculated as shown in table 3. As observed the model targets were on average more detectable than the image targets. This further emphasises the strength of deep learning techniques in comparison to traditional computer vision approaches. The average maximum distance test also shows how stable the AR application is in recognising an object from far away. Figures 40 and 41 show results highlighting the difference between each image and model target respectively.

Targets	**Mean distance**	**Standard Deviation**	**Mean Standard Error**	**Count (N)**
Image	130cm	26.46cm	8.37	10
Model	150	0cm	0	11

Table 3: The maximum distance detection test

6.2.3 Orientation Variance

The maximum orientation test, tests the maximum orientation angles of the image or model target the AR app can recognise. The results in figures 38 and 39 show that the AR app can handle any type of orientation. Therefore, the application is capable of generalising between orientations due to the fact that natural features were used in both modern and tradition computer vision approaches.

Targets	**Mean Orientation**	**Standard Deviation**	**Mean Standard Error**	**Count (N)**
Image	270°	0°	0	10
Model	270°	0°	0	11

Table 4: The maximum target orientation detection test

6.2.4 Occlusion Variance

The occlusion variance tested the AR application's capability in recognising an object while it is being occluded. This is necessary as the more occlusion the AR can handle the more applicable it is in real life scenarios, hence the application is capable of recognising the targets while an object or a person is partially occluding them. Each image was tested by occluding them by 25%, 50%, and 75% as shown in figure 1. In contrary to the procedure used in [28] in evaluating occlusion, a different form of occlusion was tested. The procedure used within this project aimed to mimic real life occurrences with regards to an occluded object. As in reality the markers (the door signs in this case) may be obstructed horizontally by another person in view. Table 5 shows that the AR application was capable of recognising on average more than 50% of an obstructed object both using model and image targets. Therefore, the results show that the application is more applicable and efficient to use in real life scenarios, where noise in data can be of an issue due to the environment one is using the application in. Figures 36 and 37 provide in detail the maximum occlusion an object could take to be detected efficiently by the augmented reality.

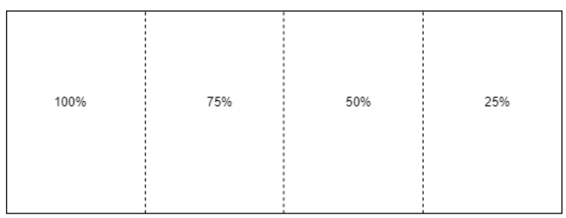

Figure 1: Marker's Occlusion Percentages

Targets	Mean Occlusion	Standard Deviation	Mean Standard Error	Count(N)
Image	0.55	0.22	0.07	10
Model	0.64	0.20	0.06	11

Table 5: The Occlusion Variance test

6.3 Recommendation System Quantitative Testing and Evaluation

The main challenge behind this project was choosing the correct collaborative filtering algorithm. According to [10], an SVD++ algorithm was the one that gave the best results when applied to the MovieLens dataset. The datasets built in this FYP project were similar in structure to the MovieLens dataset. Therefore, a SVD++ algorithm was applied on the three datasets separately due to the fact that each task is not related to another and the algorithm is compared to other algorithms on the RMSE and MAE values.

The first step of evaluation was evaluating and analysing the data which the SVD++ algorithm was going to be applied to. Three distribution barcharts per task type were plotted. Figures 42, 43, and 44 show how the ratings were distributed. One can notice that the distribution of the ratings is similarly balanced, thus showing no bias towards one specific rating. The second form of plotted distributions were of distribution of number of ratings per task as shown in figures 45, 46, and 47. As previously explained, for a delivery, a visit or an interview there are multiple tasks depending on the locations of interest. The plots clearly show that some tasks received more ratings than others, thus highlighting the users' preference for specific locations they were interested in. The third and last form of data distribution analysis is that of the distribution of a delivery, an interview or a visit per user as shown in figures 48, 49, and 50 respectively. As each distribution shows each user has an equivalent amount of ratings made, thus having no bias towards one specific user's preference.

The second step of evaluation was baseline comparison of SVD++ along with other algorithms. The algorithms were compared with each other on Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values. RMSE compares a predicted value with the actual vale and measures the amount of error, whilst MAE is the absolute distance between points in a scatter plot. The other algorithms compared with the SVD++ were the ones in table 6. The SVD++ and SVD are matrix factorization based algorithms, the SlopeOne and CoClustering are collaborative filtering algorithms, whilst the KNN Baseline, KNN Means and KNN Basic are different types of classification and regression methods which use neighbours.

A 5-fold cross-validation was carried out to compare the algorithms' accuracy. The cross validation was carried out on the whole the three separate datasets (Visit,Interview and Delivery Tasks). On average the SVDpp performed the best across the three visitor's tasks as shown in figures 54 and 55. The SVDpp achieved the lowest RMSE and MAE values on average. It measured the least amount of average magnitude errors during prediction. Therefore, the algorithm was the most accurate. On average the KNN based algorithms, the collaborative filtering algorithms, and the SVD Vanilla algorithm were close in accuracy as shown in figures 53, 51, and 52. However, this could be due to the fact that they were not applied to larger-scales of datasets, were the difference in accuracy might be magnified.

Algorithm	Mean RMSE	Mean MAE	Mean Fit Time	Mean Test Time
SVDPP	3.1226	2.6866	0.1053	0.0023
KNNBaseline	3.1281	2.6878	0.0242	0.0013
Baseline	3.1278	2.6884	0.0023	0.0013
SVD	3.1286	2.6893	0.0620	0.0016
KNNMeans	3.1292	2.6904	0.0333	0.0017
KNNBasic	3.1275	2.6904	0.0237	0.0015
SlopeOne	3.1279	2.6909	0.0141	0.0021
CoClustering	3.1300	2.6914	0.1094	0.0015
KNNZscore	3.1286	2.6915	0.0602	0.0021
NormalPredictor	4.1130	3.3603	0.0018	0.0019

Table 6: Average MAE,RMSE,Fit time and Test time from Delivery, Interview and Visit Tasks

Finally, the hyperparameters for SVD++ is tuned using the grid search method. The GridSearchCV() method provided by Surprise library calculate the RMSE and MAE values for every combination of hyperparameters through in this case a 5-fold cross validated dataset. Finally, it returns with a set of correct parameters that improve the model's accuracy. Table 7 provides the hyperparameters to get the best RMSE and MAE values for each dataset SVD++ was applied to.

Dataset	RMSE	MAE	n epochs	lr-all	reg-all
Delivery	3.1282	2.6849	25	0.005	0.6
Interview	3.1429	2.7248	25	0.002	0.4
Visit	3.0906	2.6395	25	0.0015	0.6

Table 7: Hyper-parameters for model tuning for Delivery, Interview and Visit tasks

6.4 Qualitative Testing and Evaluation

Appendices

A Chapter 4

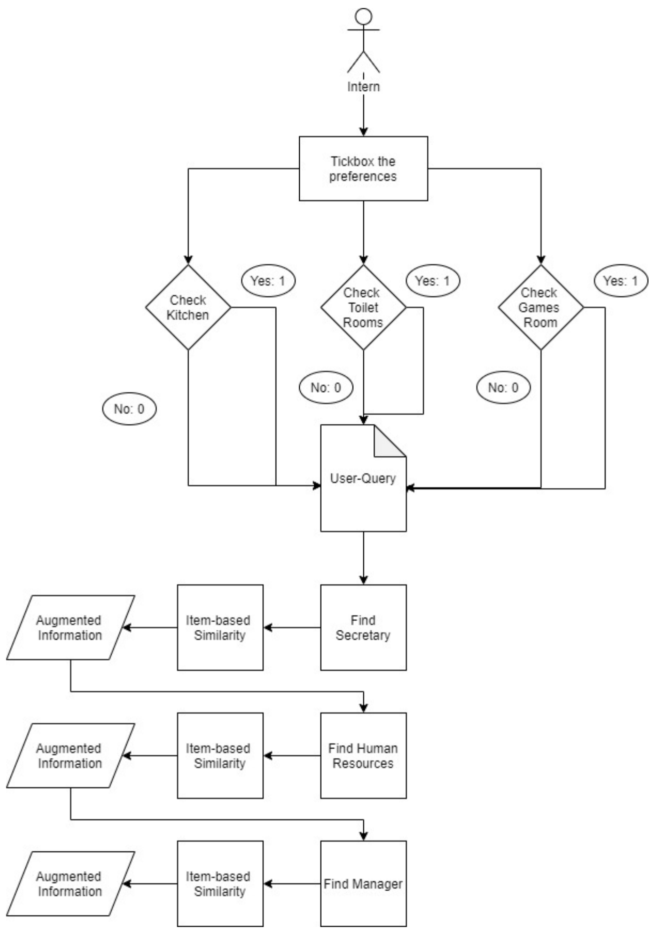

Figure 2: Intern User-Query Model

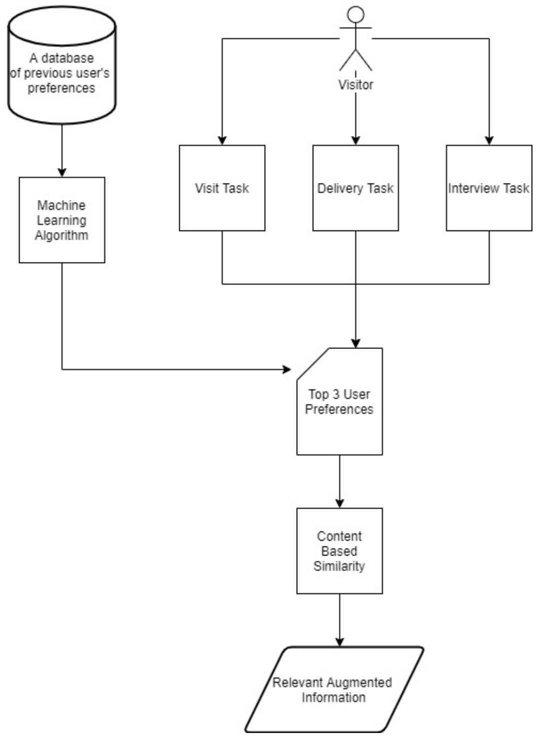

Figure 3: Visitor User-Query Model

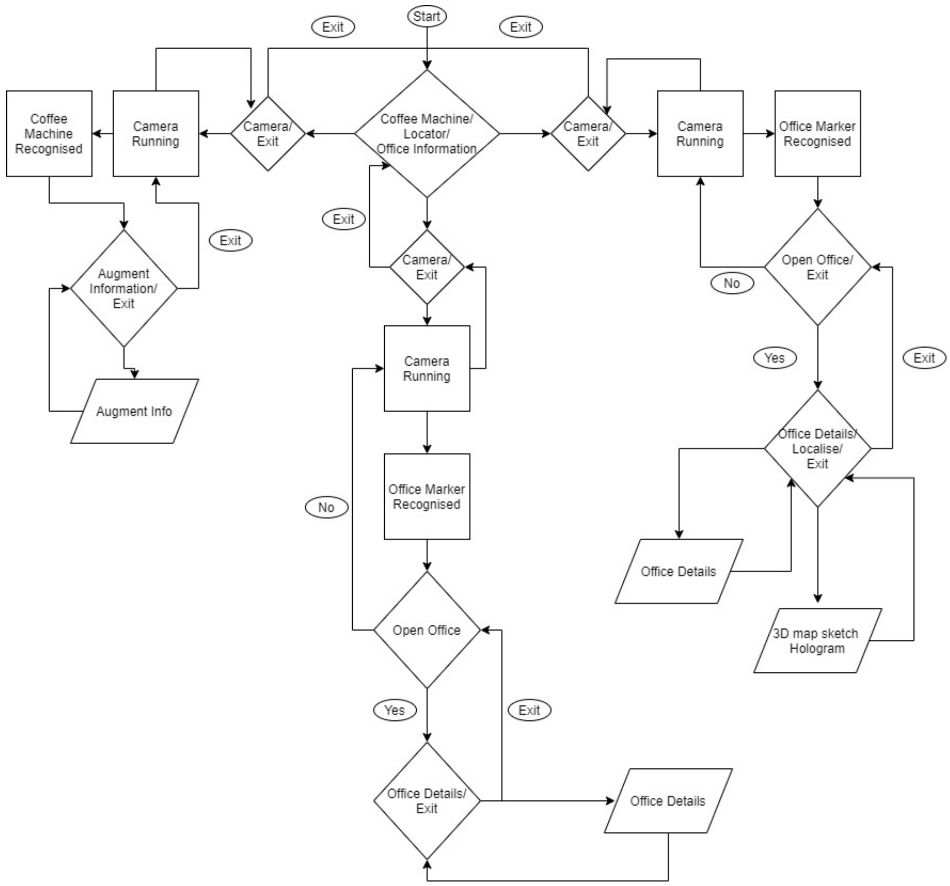

Figure 4: User-Interface and System Architecture

B Chapter 5

Figure 5: Corridor Image Variation 1

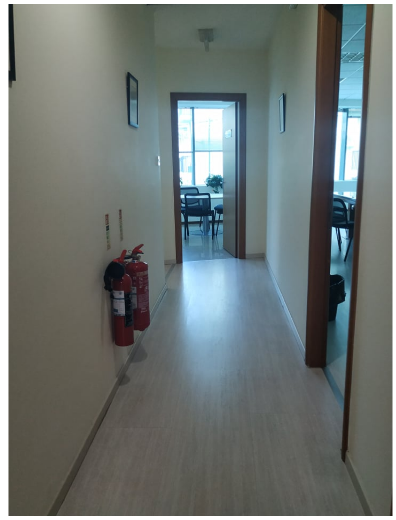

Figure 6: Corridor Image Variation 2

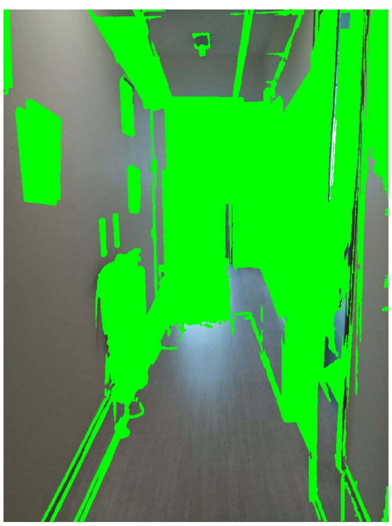

Figure 7: Result of differences using SSIM

Figure 8: Door Marker

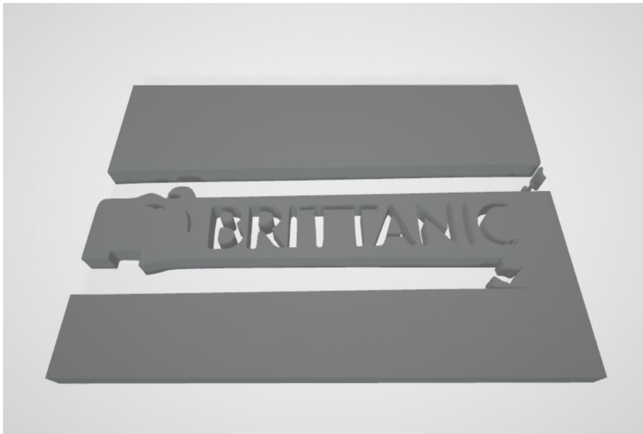

Figure 9: Door Marker 3D Object

Figure 10: Coffee Machine 3D Object

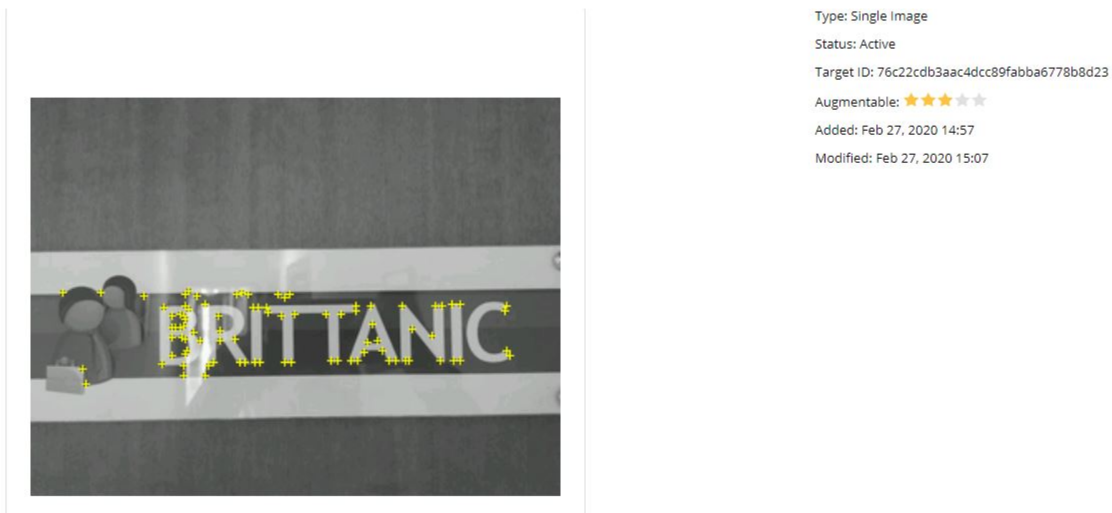

Figure 11: Image Target

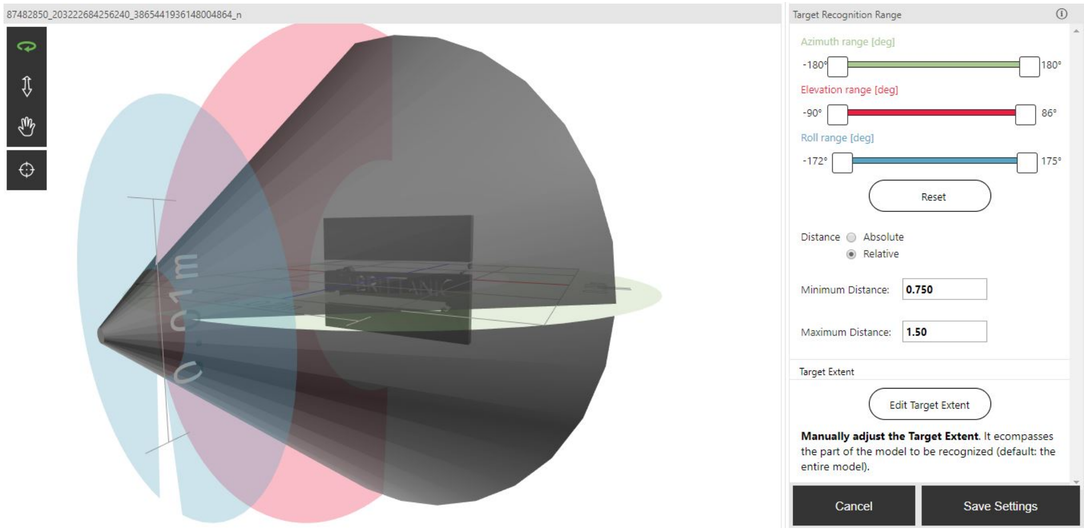

Figure 12: Model Target

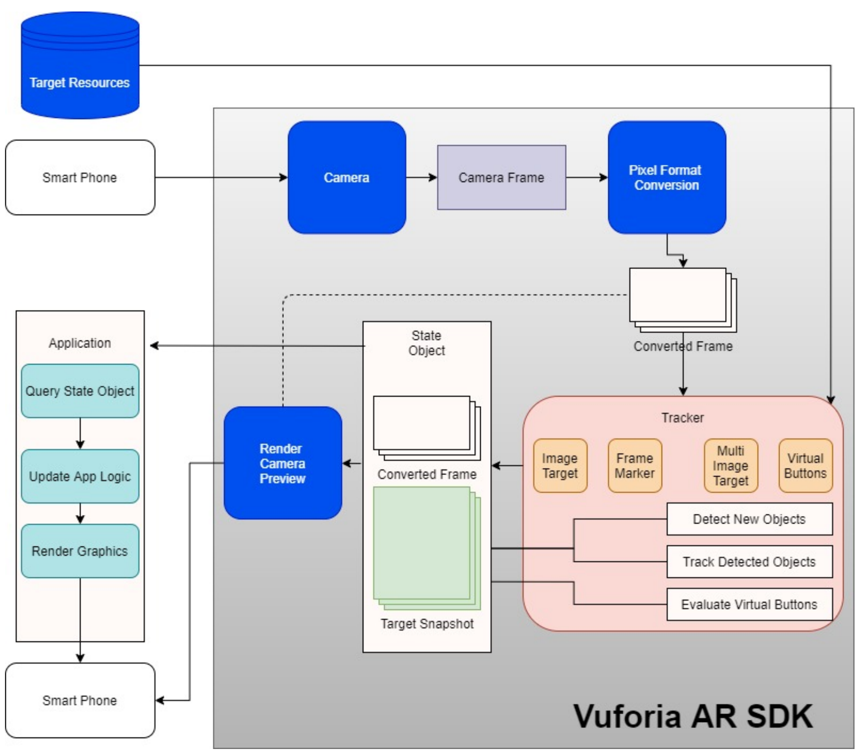

Figure 13: Vuforia AR SDK [18]

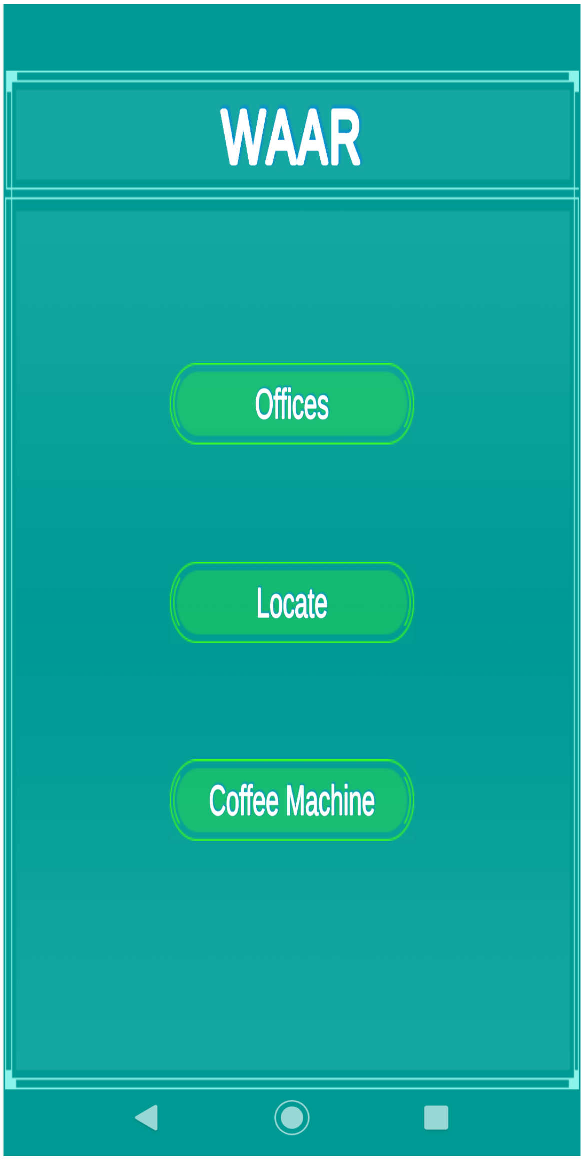

Figure 14: Main Menu

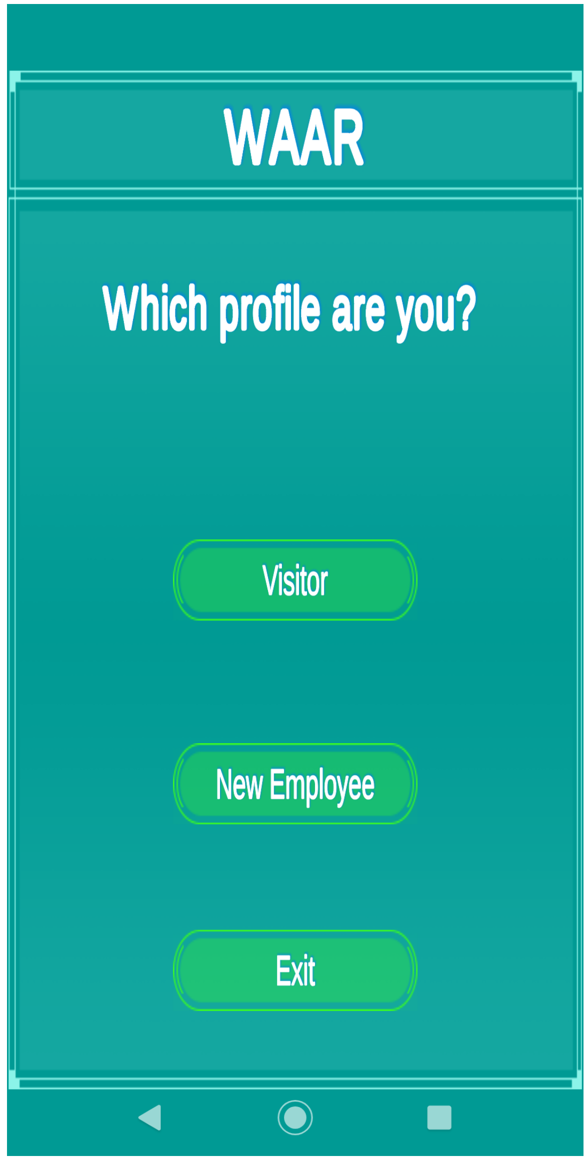

Figure 15: Choosing user profile

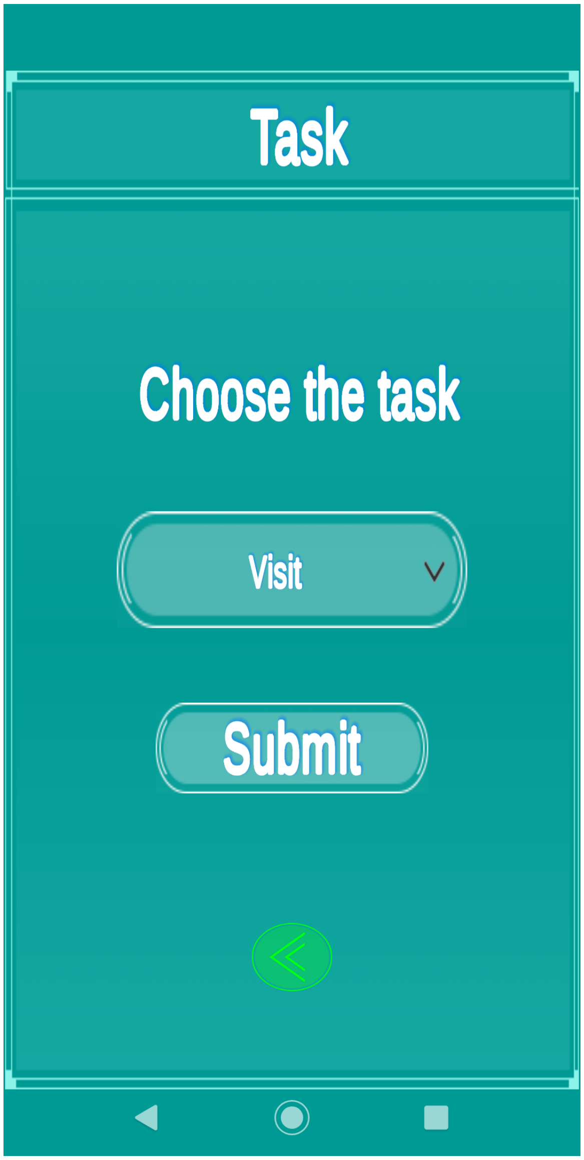

Figure 16: Choosing a visitor task

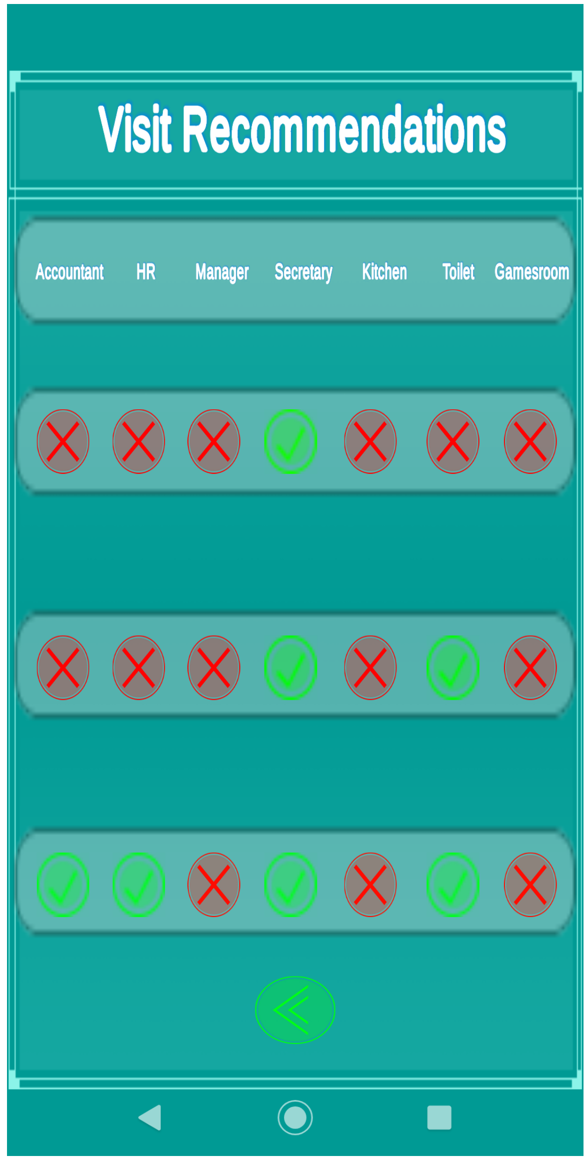

Figure 17: Visitor Recommendation for Visit

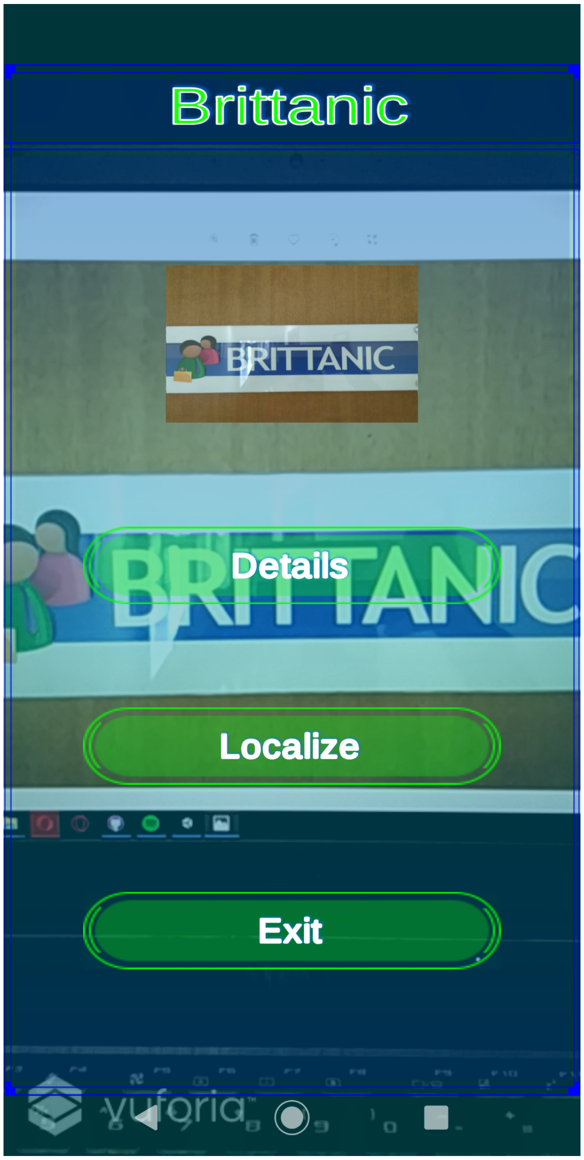

Figure 18: Marker Augmentation Menu

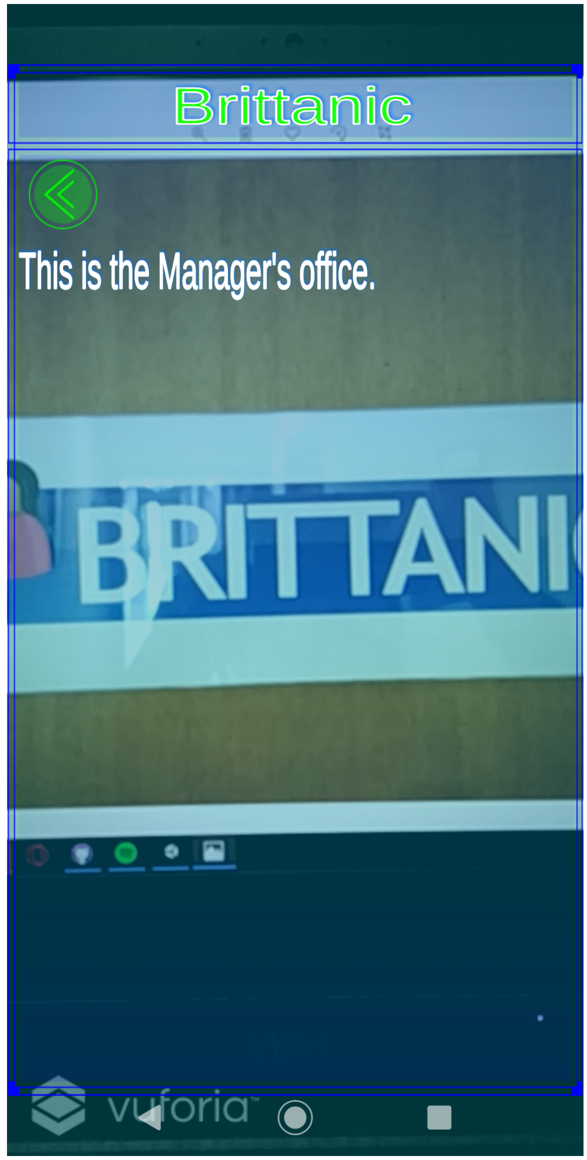

Figure 19: Office Details

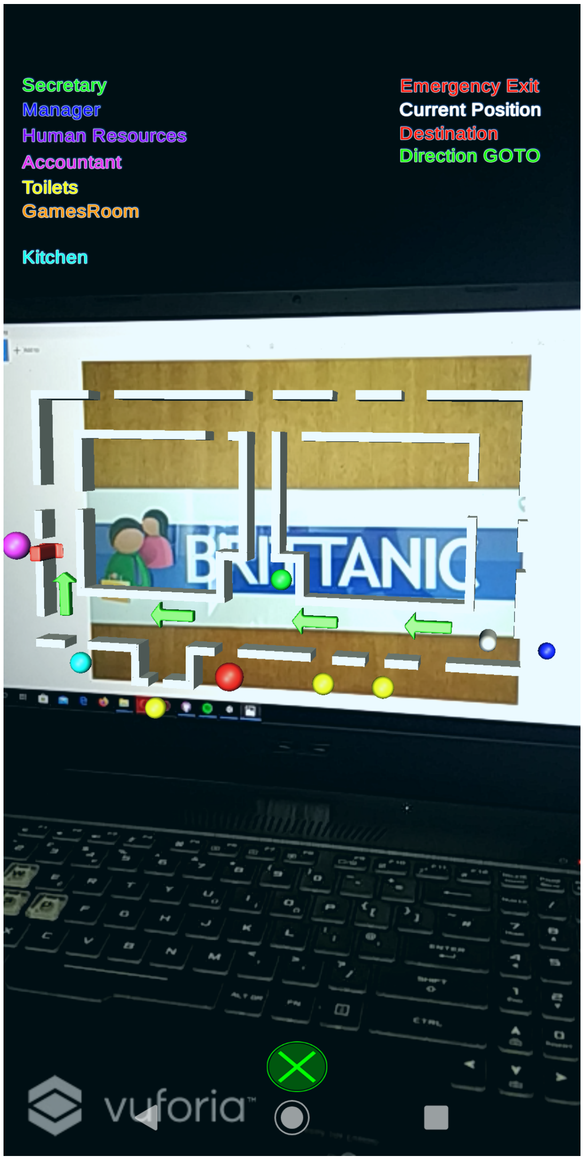

Figure 20: Holographic Sketch Map Front View

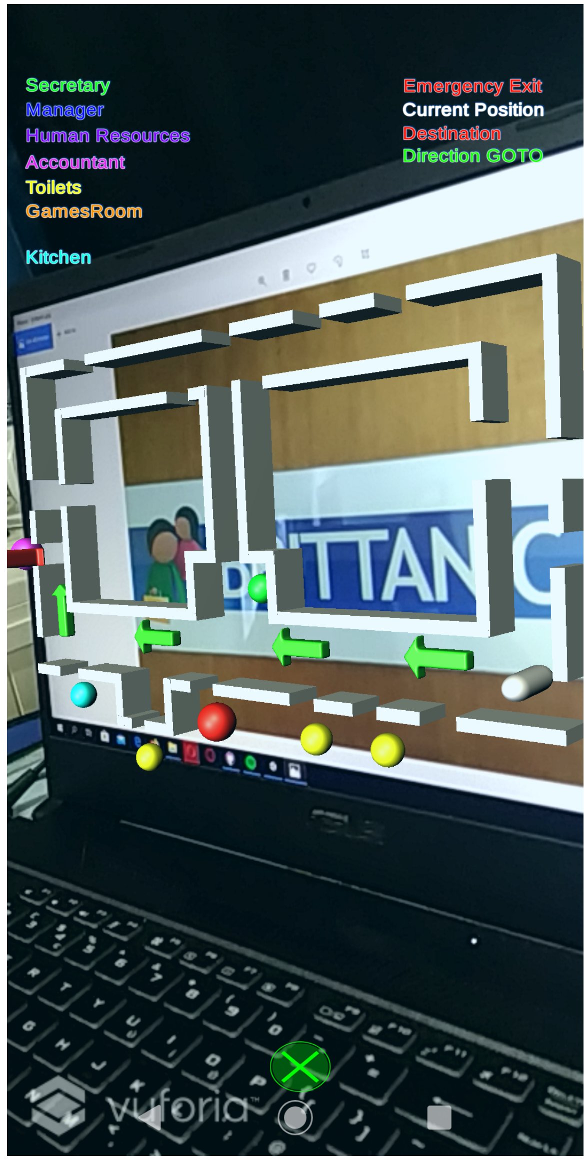

Figure 21: Holographic Sketch Map Side View

Figure 22: Coffee Machine Augmentation Menu

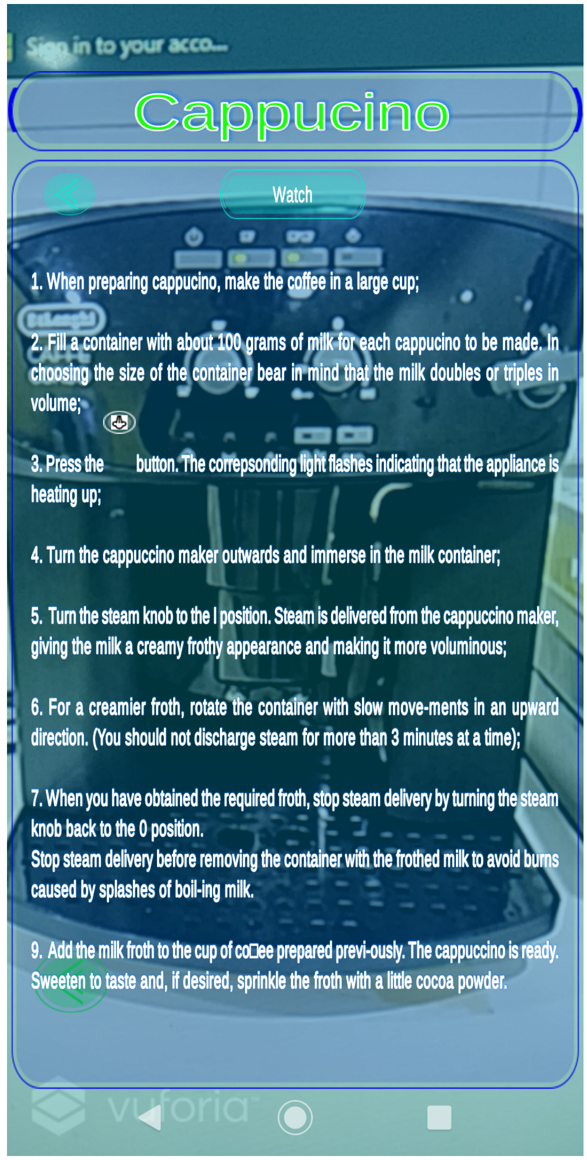

Figure 23: Cappuccino Details

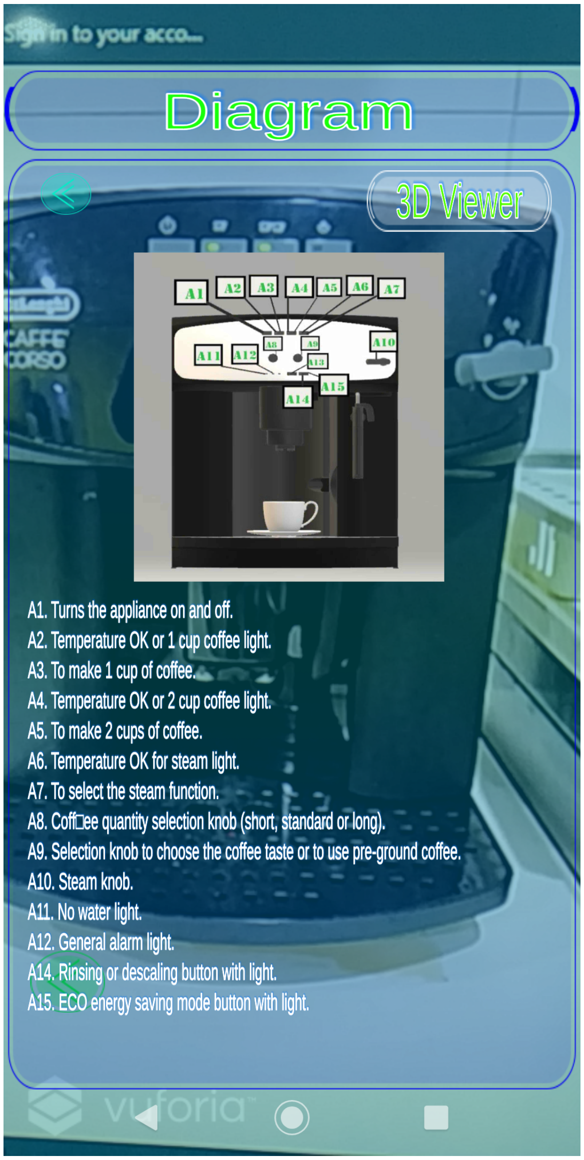

Figure 24: Coffee Machine Diagram

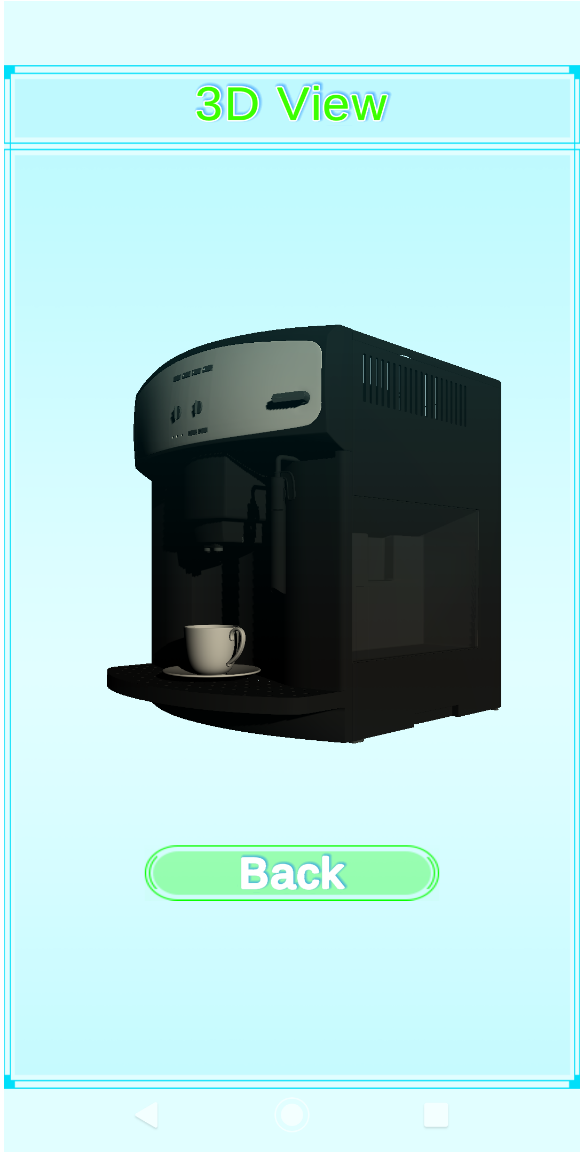

Figure 25: Coffee Machine 3D View

C Chapter 6

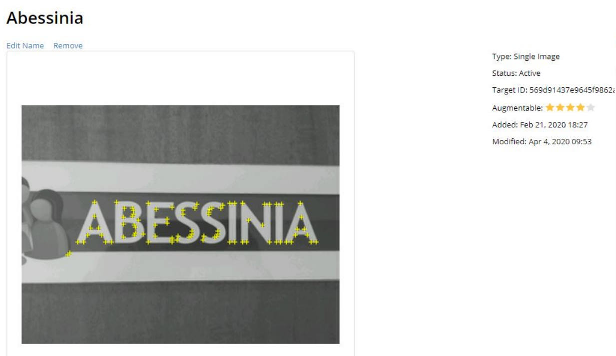

Figure 26: Abessinia Image Target

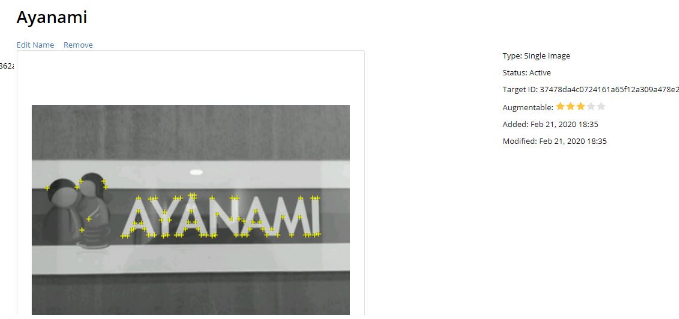

Figure 27: Ayanami Image Target

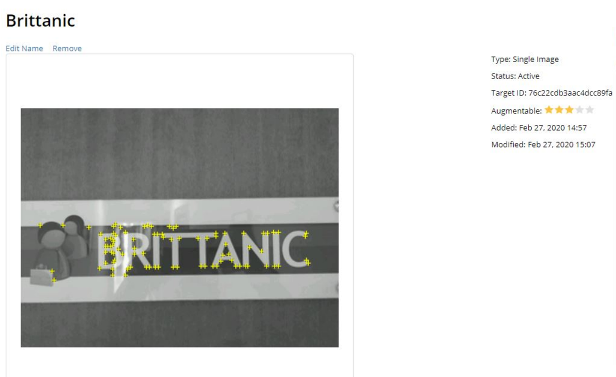

Figure 28: Brittanic Image Target

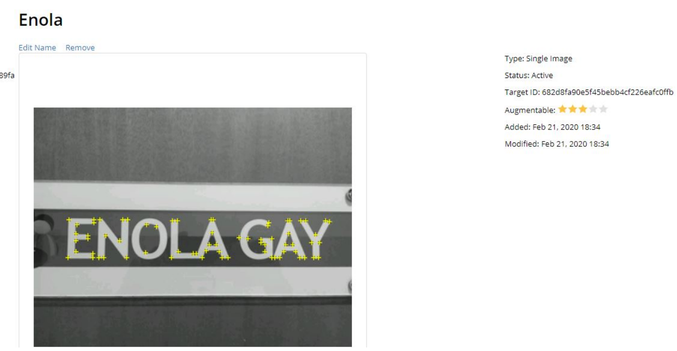

Figure 29: Enola Image Target

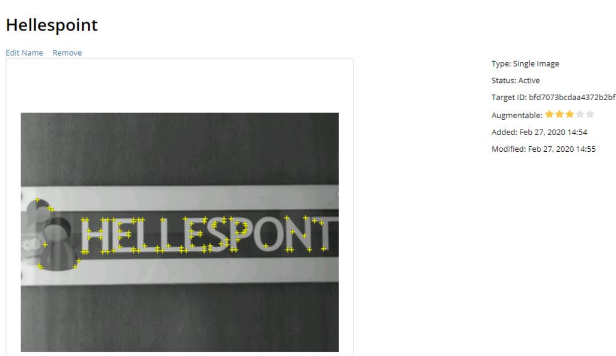

Figure 30: Hellespont Image Target

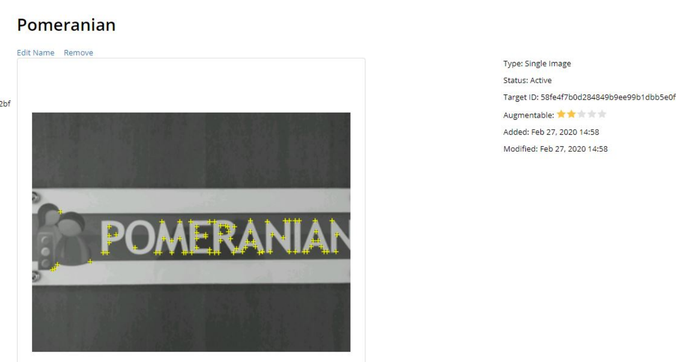

Figure 31: Pomeranian Image Target

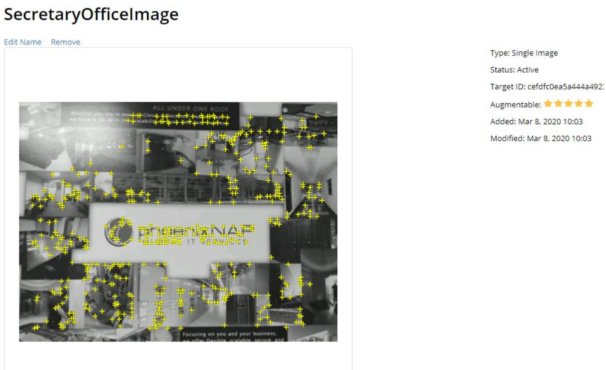

Figure 32: Secretary Image Target

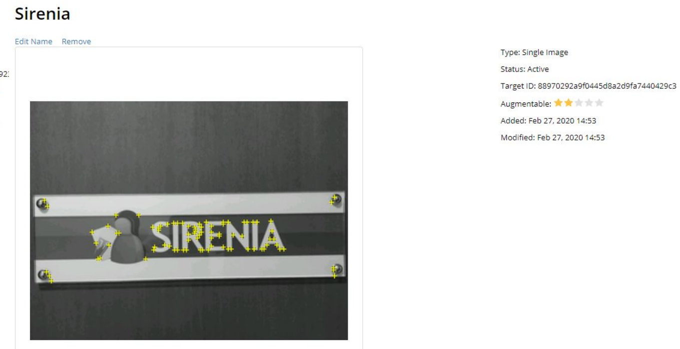

Figure 33: Sirenia Image Target

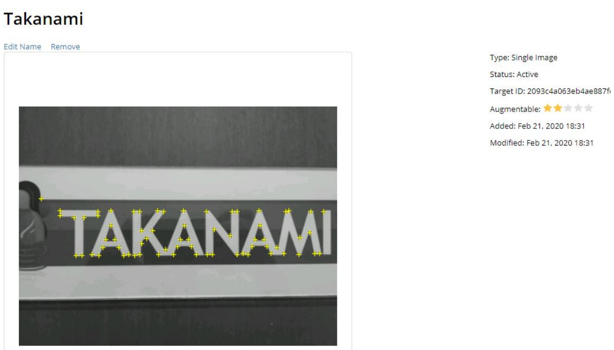

Figure 34: Takanami Image Target

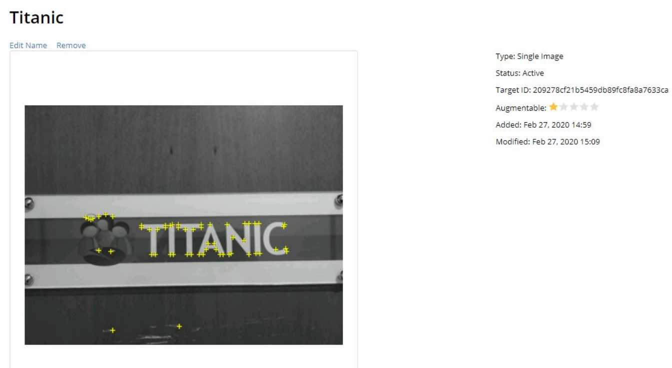

Figure 35: Titanic Image Target

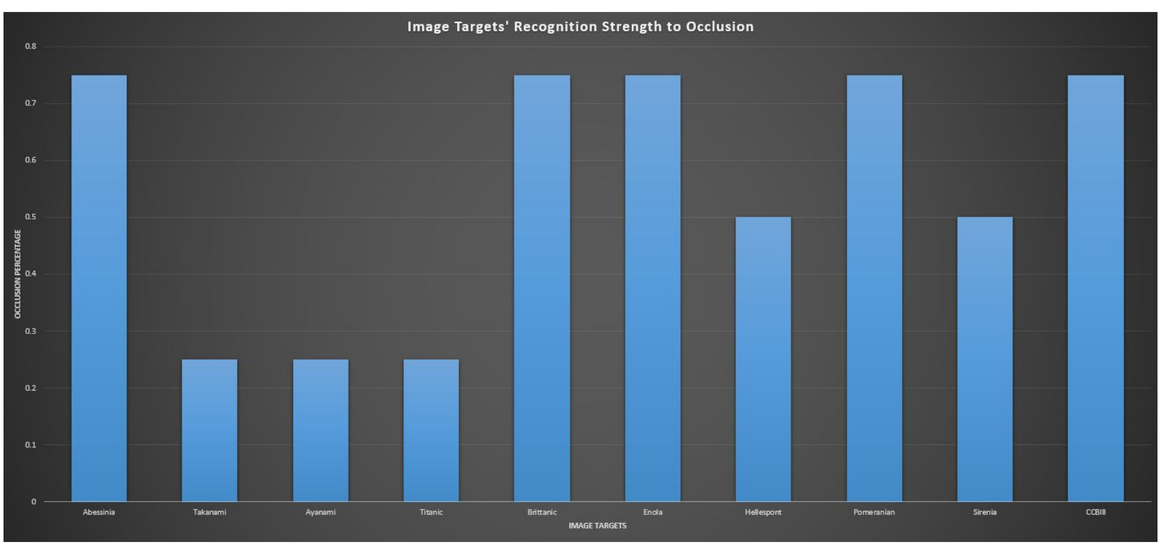

Figure 36: Image Targets' Recognition Strength to Occlusion Bar Chart

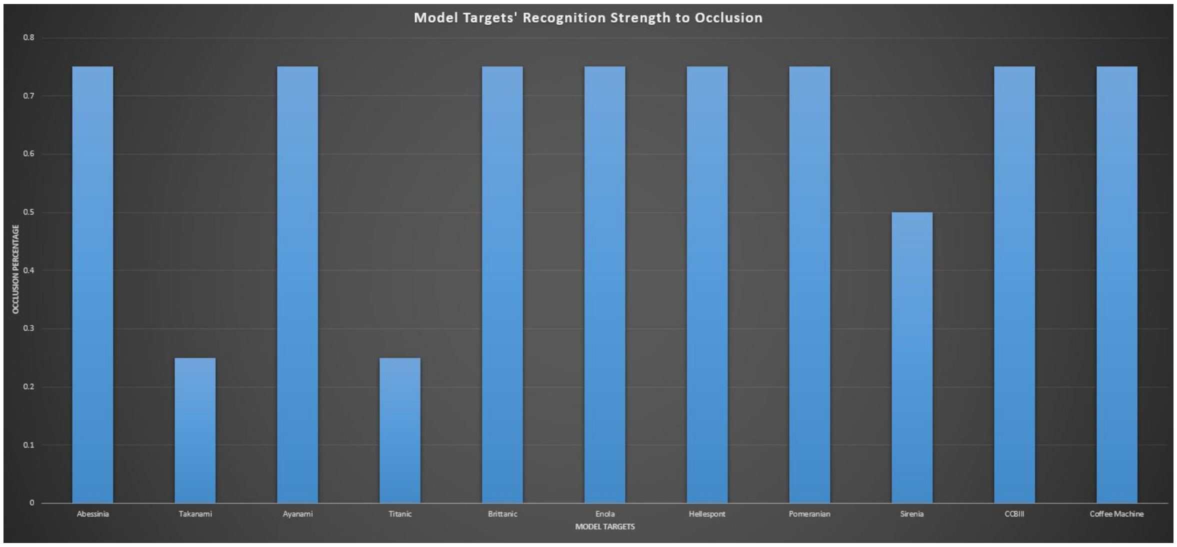

Figure 37: Model Targets' Recognition Strength to Occlusion Bar Chart

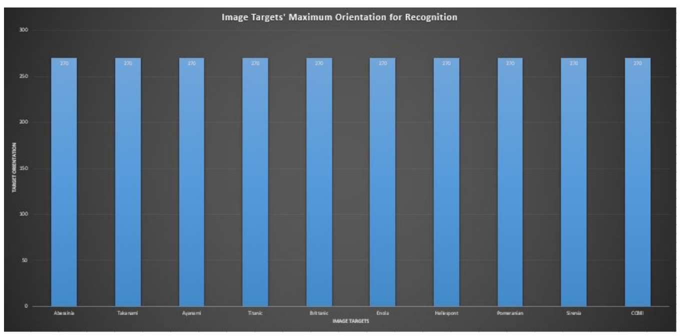

Figure 38: Image Targets' Recognition Strength to Orientation Variance Bar Chart

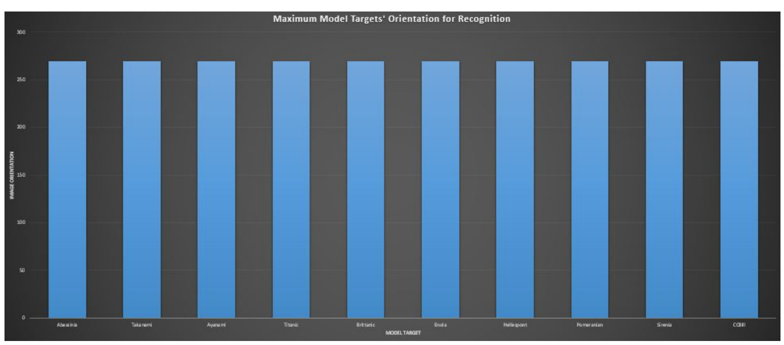

Figure 39: Model Targets' Recognition Strength to Orientation Variance Bar Chart

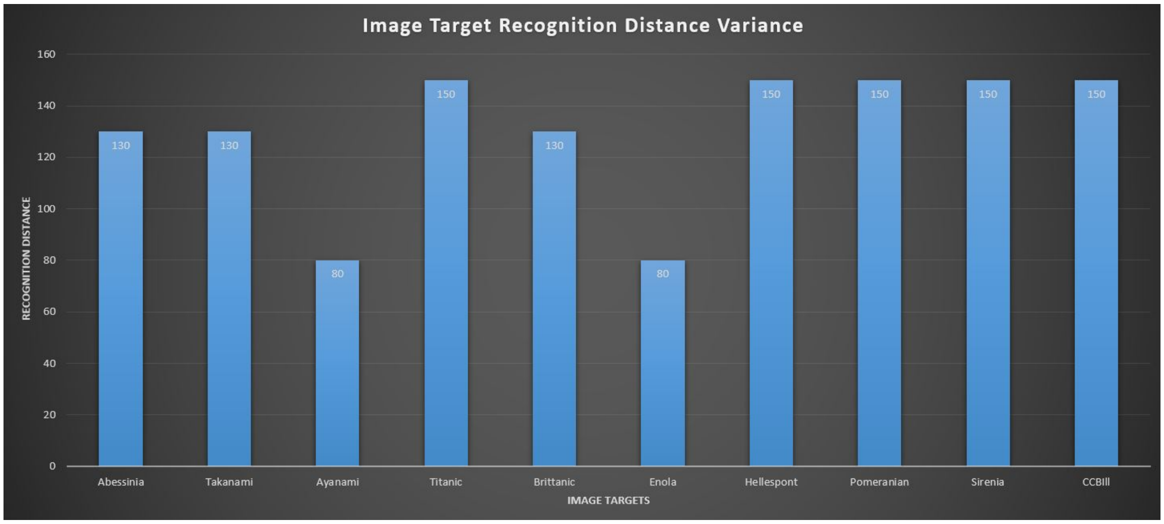

Figure 40: Image Targets' Recognition Strength to Distance Variance Bar Chart

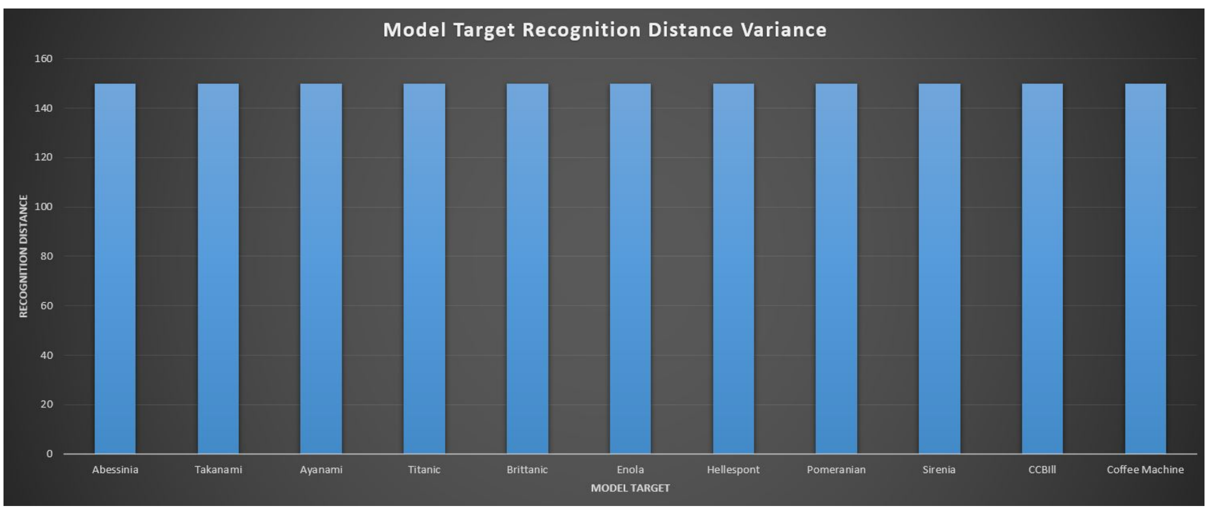

Figure 41: Model Targets' Recognition Strength to Distance Variance Bar Chart

Figure 42: Distribution of 1000 Delivery Task-Ratings Barchart

Figure 43: Distribution of 1000 Interview Task-Ratings Barchart

Figure 44: Distribution of 1000 Visit Task-Ratings Barchart

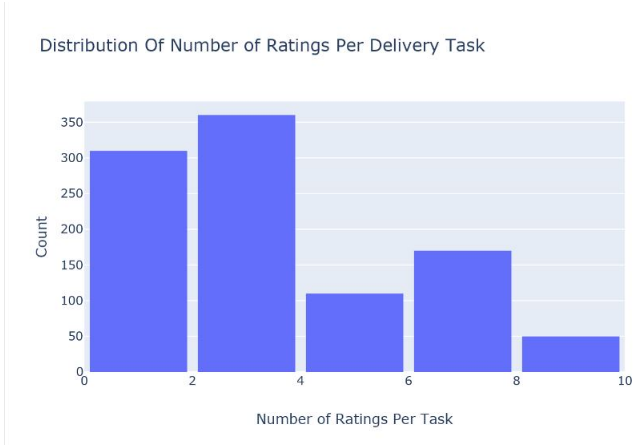

Figure 45: Distribution of Number of Ratings Per Delivery Task Barchart

Figure 46: Distribution of Number of Ratings Per Interview Task Barchart

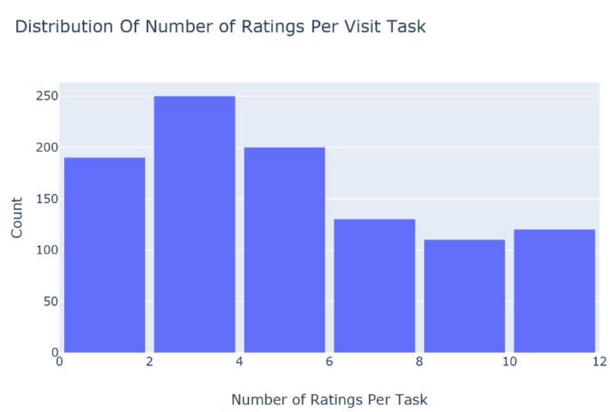

Figure 47: Distribution of Number of Ratings Per Visit Task Barchart

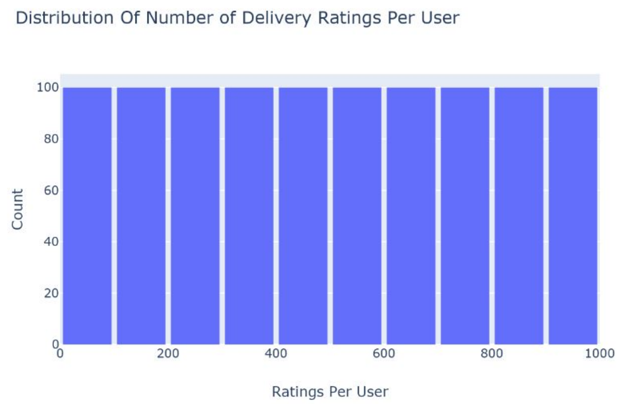

Figure 48: Distribution of Delivery Ratings Per User Barchart

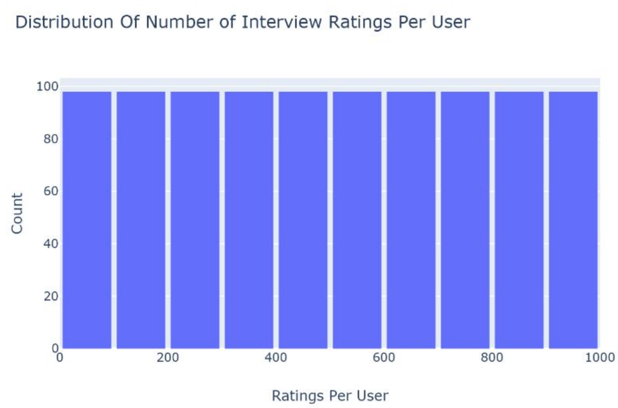

Figure 49: Distribution of Interview Ratings Per User Barchart

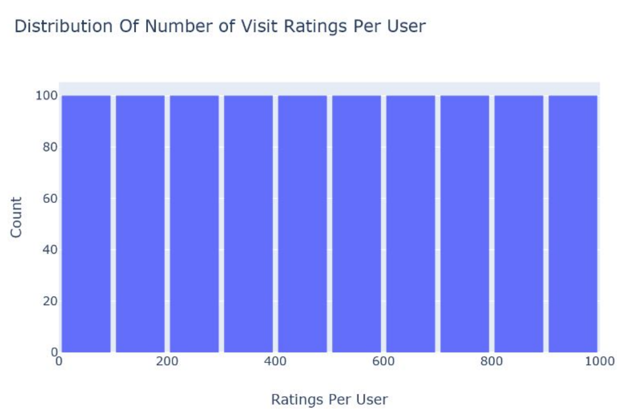

Figure 50: Distribution of Visit Ratings Per User Barchart

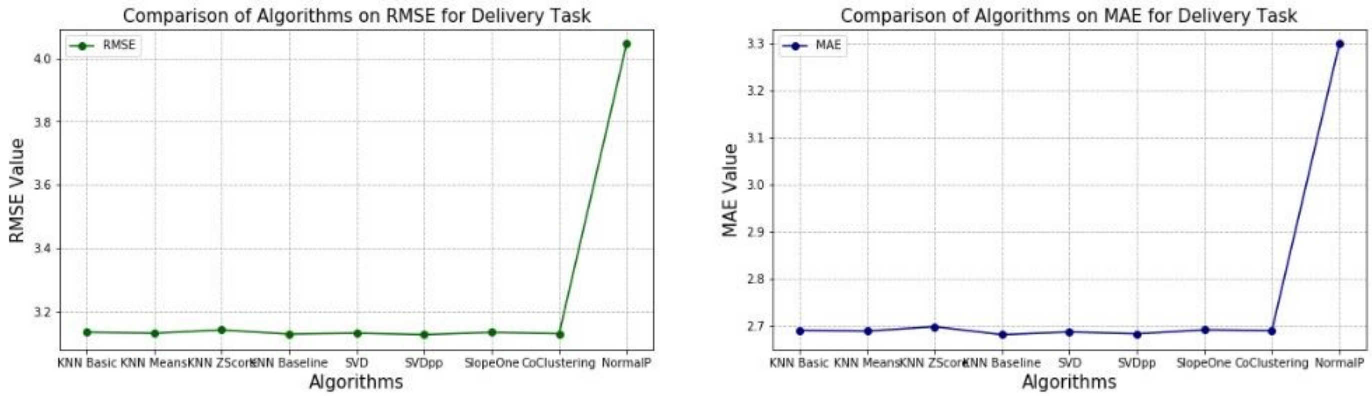

Figure 51: Comparison of Algorithms on RMSE and MAE for Delivery

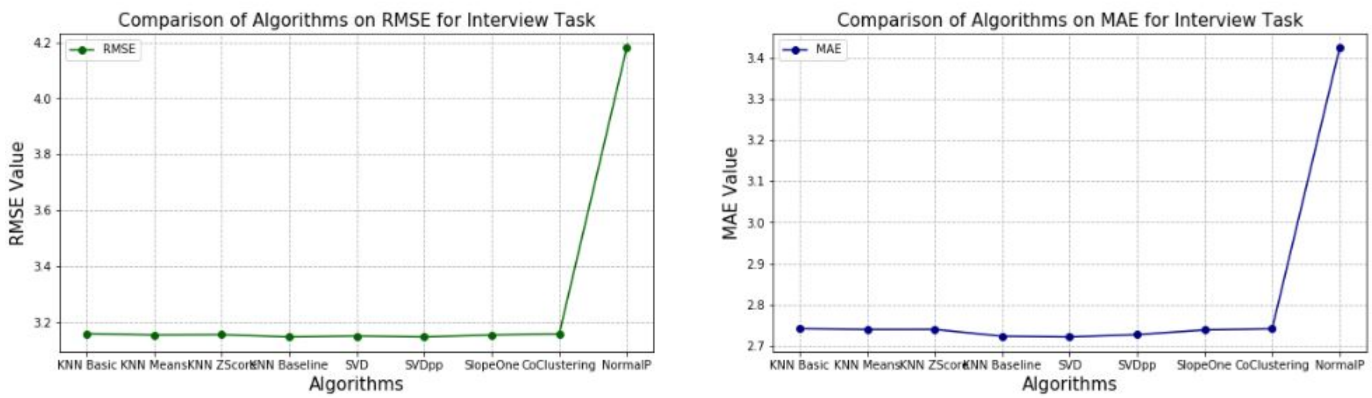

Figure 52: Comparison of Algorithms on RMSE and MAE for Interview

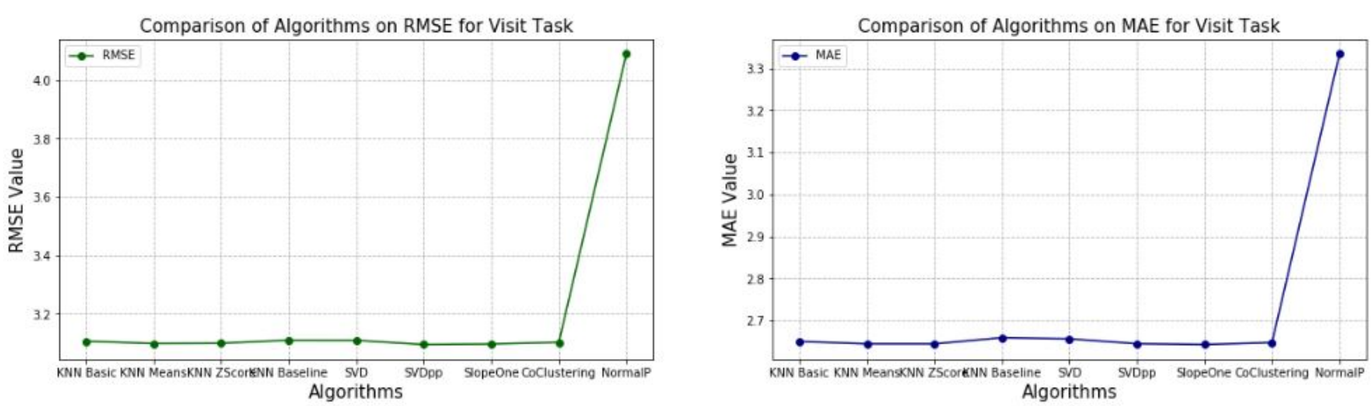

Figure 53: Comparison of Algorithms on RMSE and MAE for Visit

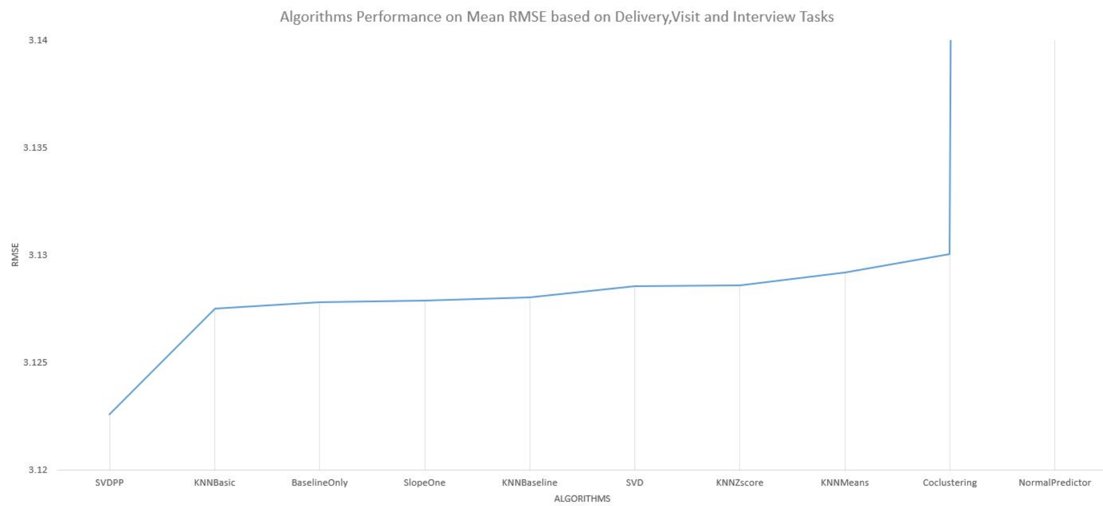

Figure 54: Comparison of Algorithms on Average RMSE for Visit, Interview and Delivery

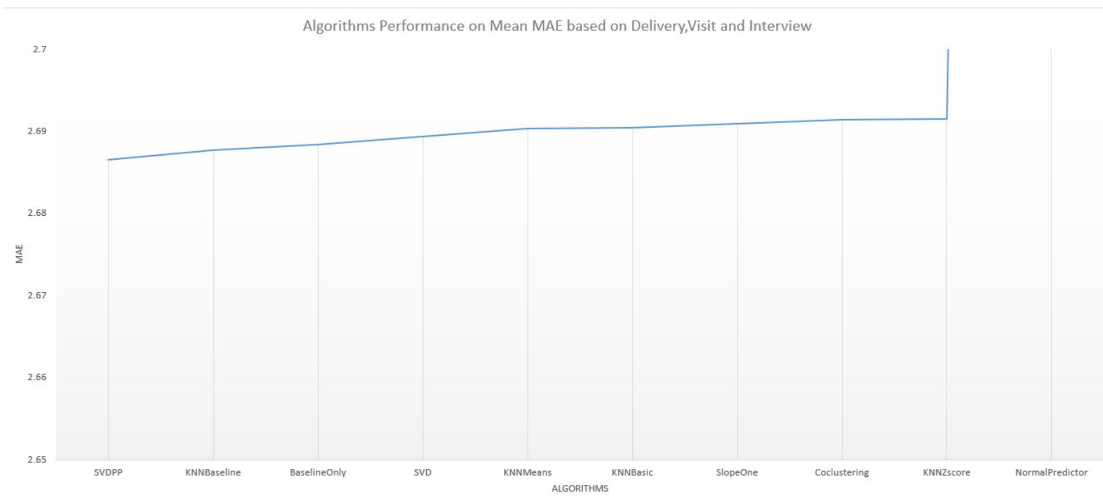

Figure 55: Comparison of Algorithms on Average MAE for Visit, Interview and Delivery

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] Fidel Cacheda, Victor Carneiro, Diego Fernández, and Vreixo Formoso. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *TWEB*, 5:2, 01 2011.

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

[12] 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.

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

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

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

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

[17] 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.

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

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

[20] 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.

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

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

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

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

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

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

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

[28] Hendra Pradibta, Wijaya Indra Dharma, Ferdian Ronilaya, and Usman Nurhasan. The evaluation on a mobile augmented reality application as therapy media. *MATEC Web of Conferences*, 197:15004, 01 2018.

[29] 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.

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

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

[32] 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.

[33] 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.

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

[35] 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.

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

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

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

[39] 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.