# 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 proﬁling methods with Augmented Reality, and diﬀerent image and object recognition techniques in augmented reality technologies.

## 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. [12]

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” [16]. 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” [16]. Conversely, oﬀ 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” [12]. As discussed in [12], 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” [12]. 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.

## Recommendation System 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. [31]

## “[R]ecommender systems (RS) have proven to be a valuable tool for online users to cope with the information overload” [10]. Recommendation systems provide tailored information to diﬀerent users based on their preference. “[T]hus, 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 systems 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]. Distance-based ﬁltering and visibility-based ﬁltering are commonly used in Augmented Reality. In [37], 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 diﬀerent forms of interactive techniques. The study by [31] 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 diﬀerent symbols being displayed, as well as their colour and size.

## 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 [23], 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” [23]. 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 [23], it is tested to determine its beneﬁts. 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” [23]. However, there were still key limitations, such as, detecting reﬂective 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 ﬁnal challenge was dealing with robustness due to the diﬀerent environments where the object would be in order to improve object recognition and generalise.

## According to the research conducted by [27], 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 identiﬁcation and combine it with high reliability and eﬀectiveness of the Golay error correction code detection” [27]. 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 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]

## 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 inﬂuenced by the SURF algorithm which 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 to solve the occlusion problem in AR is to apply deep learning techniques, as described in [13], 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” [13]. 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 speciﬁc 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” [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. The study by [26] 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” [26]. 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.

## Conclusion

Diﬀerent approaches were deﬁned 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.

# 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 ¨Oznur Gu¨len 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 Ho¨llerer. Augmented reality: Information for workplace decision-makers, managers, workers and researchers. 02 2020.

[13] Mathieu Garon and Jean-Fran¸cois 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 Andr´e 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 Ib´an˜ez 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 Proch´azka, Ondˇrej 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 Aleˇs 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 locationcontent-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.