

Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence

JACKY CAO, University of Oulu, Finland

KIT-YUNG LAM, The Hong Kong University of Science and Technology, Hong Kong

LIK-HANG LEE, KAIST, Republic of Korea

XIAOLI LIU, University of Helsinki, Finland

PAN HUI, The Hong Kong University of Science and Technology, Hong Kong; University of Helsinki, Finland

XIANG SU, Norwegian University of Science and Technology, Norway; University of Oulu, Finland

Mobile Augmented Reality (MAR) integrates computer-generated virtual objects with physical environments for mobile devices. MAR systems enable users to interact with MAR devices, such as smartphones and head-worn wearables, and performs seamless transitions from the physical world to a mixed world with digital entities. These MAR systems support user experiences by using MAR devices to provide universal accessibility to digital contents. Over the past 20 years, a number of MAR systems have been developed, however, the studies and design of MAR frameworks have not yet been systematically reviewed from the perspective of user-centric design. This article presents the first effort of surveying existing MAR frameworks (count: 37) and further discusses the latest studies on MAR through a top-down approach: (1) MAR applications; (2) MAR visualisation techniques adaptive to user mobility and contexts; (3) systematic evaluation of MAR frameworks including supported platforms and corresponding features such as tracking, feature extraction plus sensing capabilities; and (4) underlying machine learning approaches supporting intelligent operations within MAR systems. Finally, we summarise the development of emerging research fields, current state-of-the-art, and discuss the important open challenges and possible theoretical and technical directions. This survey aims to benefit both researchers and MAR system developers alike.

CCS Concepts: • **Human-centered computing** → **Mixed / augmented reality**; *Graphical user interfaces*; • **Software and its engineering** → *Development frameworks and environments*; • **Computing methodologies** → *Machine learning approaches*.

Additional Key Words and Phrases: mobile augmented reality, user interactions, development framework, artificial intelligence

1 INTRODUCTION

Over the past several decades, Augmented Reality (AR) has evolved from interactive infrastructures in indoors or fixed locations to various mobile devices for ubiquitous access to digital entities [133]. We consider the first research prototype of MAR to be the *touring machine system* [18] that offers cues of road navigation on a school campus. After 20 years of development, MAR devices have shrunk from the size of huge backpacks to lightweight and head-worn devices. The interfaces on wearable head-worn computers have shifted from micro-interactions displaying swift (e.g., Google Glass) and small-volume contents to enriched holographic environments (e.g., Microsoft HoloLens) [129, 134]. MAR offers interactive user experiences on mobile devices by overlaying digital contents, e.g., computer-generated texts, images, and audio, on top of physical environments. Numerous MAR frameworks have been developed to support the displaying of such enriched real-world environments by managing different sensors and components for tracking physical objects and various user interaction features (e.g., anchors and ray-casting), as well as offloading computational

Authors' addresses: Jacky Cao, University of Oulu, Oulu, Finland, jacky.cao@oulu.fi; Kit-Yung Lam, The Hong Kong University of Science and Technology, Hong Kong, kylambd@connect.ust.hk; Lik-Hang Lee, KAIST, Daejeon, Republic of Korea, likhang.lee@kaist.ac.kr; Xiaoli Liu, University of Helsinki, Helsinki, Finland, xiaoli.liu@helsinki.fi; Pan Hui, The Hong Kong University of Science and Technology, Hong Kong, pan.hui@helsinki.fi; Xiang Su, Norwegian University of Science and Technology, Norway, xiang.su@ntnu.no.

demanding tasks to remote servers. Nowadays, MAR, albeit primarily on smartphones, has been employed in various sectors, for instance, facilitating teaching and learning [109, 138, 146, 154, 210], surgery training and reducing human errors during surgical operations [32, 118, 136], as well as virtual tour guides during sight-seeing [68, 69, 82, 249]. The multitude of MAR applications has proven its practicality in becoming a ubiquitous interface type in the real world. With the recent advancements of Artificial Intelligence (AI) and Internet of Things (IoT) devices, numerous mobile equipment and IoT devices construct smart and responsive environments. Users leverage head-worn computers to access a multitude of intelligent services through MAR. Accordingly, we advocate that MAR has to develop adaptability, or *adaptive MAR frameworks*, to address the on-demand user interactions with various IoT devices in the aforementioned smart environments, such as drones, robotics, smart homes, and sensor networks.

One critical research challenge to realise the MAR vision is the provisioning of seamless interactions and supporting end-users to access various MAR services conveniently, for instance, through wearable head-worn computers and head-mounted displays (HMDs). HMDs are technological enablers for user interactions within augmented environments. That is, these headsets project digital contents in the form of windows, icons, or more complex 3D objects (e.g., virtual agents/avatars) on top of views of the physical world. Users with such wearable head-worn computers can access various AR services ubiquitously and new forms of services and user experiences will spread and penetrate into our daily routines, such as purchasing tickets from transport service providers, setting privacy preferences and control on small-size smart cameras and speakers, and buying snacks and drinks from convenience stores. The interfaces of MAR experiences are regarded as ubiquitous service points to end users.

AI is crucial to support features of high awareness to user physical surroundings and user context to support MAR experiences, and in order to achieve user-oriented services. Considerations of user performance, user acceptance, and qualitative feedback in such MAR systems make adaptive MAR user interfaces (UIs) a necessity in rich yet complex physical environments. To support the creation of MAR experiences, several frameworks are available where developers can deploy pre-built systems, which already contain features such as displaying augmented objects, performing environment analysis, and supporting collaborative AR with multiple users. From environment analysis found in MAR frameworks to adaptive MAR UIs, and to enable low-latency and good user experiences, AI methods are generally required to execute these tasks. Nowadays, more commonly, machine learning (ML) methods are typically used due to their ability to analyse and extract several layers of information from data (e.g., geographical data, user data, image frames taken from physical environments) in an efficient manner.

Thus, this survey examines the recent works of MAR with focusing on developments and synthesis of MAR applications, visualisations, interfaces, frameworks, and applied ML for MAR. We also strive to move beyond the individual applications of MAR and seek research efforts towards highly intelligent and user-oriented MAR frameworks that potentially connects with not only digital and virtual entities, but also physical daily objects. We outline key challenges and major research opportunities for user-oriented MAR UIs and the AI methods for supporting such UIs. The contributions of this survey paper are as follows:

- (1) Provisioning of a comprehensive review of the latest AR applications and framework-supported interface designs in AR in both hand-held and head-worn AR system scenarios;
- (2) Provisioning of a comprehensive review of existing AR frameworks and SDKs, which explores the supported hardware platforms and features of the MAR frameworks;
- (3) Provisioning of a comprehensive review of relevant ML methods to be applied in key AR components of MAR frameworks;

- (4) A research agenda for the development of future adaptive AR user interactions concerning systematic MAR frameworks and ML-assisted pipelines.

Among our main calls to action in the agenda are issues related to AR experiences, such as investigating the feasibility of AR interfaces in user-orientated and high adaptive manners, developing ML methods for MAR experiences in dynamic and complicated environments, and advancing MAR systems for two-way and seamless interactions with all intelligent objects in evolving MAR landscapes.

1.1 Overview of Mobile Augmented Reality

Figure 1 presents the typical pipeline of an MAR system. The components in MAR systems can be approximately grouped into *MAR Device* and *MAR Tasks*, where the tasks may be executed on device, or on an external server. Generally, hardware sensors (i.e., cameras, GPS and IMU sensors, and LiDAR modules) on the MAR device first captures images and sensor data which are then fed into the MAR tasks. The first of which is frame pre-processing, where data is cleaned and prepared to be inputted into the analysis tasks to then retrieve information on which objects are in the environment, e.g., object detection, feature extraction, object recognition, and template matching tasks. These tasks can be classed as the *ML Tasks*, where ML techniques are used to complete the tasks in a more efficient manner due to their complexity. After data analysis, successful results and associated annotations are returned to the MAR device for object tracking and annotation rendering on the client. Together with ML methods, this can form an overall *AR Framework* where developers can employ pre-built frameworks to create their own MAR applications. The annotation rendering and display components of the MAR device can be grouped together as part of *Adaptive UI*, the contents and interaction with the rendered augmentations are important for users as this allows them to experience MAR. Therefore, the scope of this survey primarily aims to address the following questions, which are missing in the existing surveys (detailed in Section 1.2): (1) What features and functions are available in MAR frameworks nowadays? (2) What are the latest efforts of AR research and accordingly, the prominent gap between these efforts and the identified frameworks? (3) Then based on the above questions, how MAR engineers and practitioners design the next generation of MAR frameworks that bridge the user requirements?

1.2 Related Surveys and Selection of Articles

This subsection briefly outlines previous relevant surveys in various domains, i.e. user interfaces, frameworks, and intelligence. One of the earliest AR surveys, focusing on AR systems, applications, and supporting technologies, is proposed by Carmigniani et al. [40]. Later on, Lee and Hui [133] focus on the user interaction techniques on MAR devices (i.e., smartglasses). A recent survey by Chatzopoulos et al. [41] present a comprehensive state-of-the-art survey of MAR application fields, how MAR user interfaces can be developed, what key components MAR systems consists of, and how data is managed in MAR systems. Zollmann et al. [271] re-organise the recent works of different components in the visualisation pipelines of AR systems and presents the findings in three main areas, including filtering, mapping, and rendering. Marneanu et al. [152] provide an early evaluation of AR frameworks for Android developments. Recent AR development platforms on both Android and iOS platforms have been generally discussed in some development communities, but have never been evaluated in a systematic and scientific way. Herpich et al. [91] perform a comparative analysis of AR frameworks, albeit for the scenario of developing educational applications. Finally, utilisation of ML in different core components of MAR applications are discussed in several surveys. For example, Krebs et al. [125] present a survey of leveraging deep neural networks for object tracking. Liu et al. [141] focus on deep learning for object detection. Yilmaz et al. [262] survey

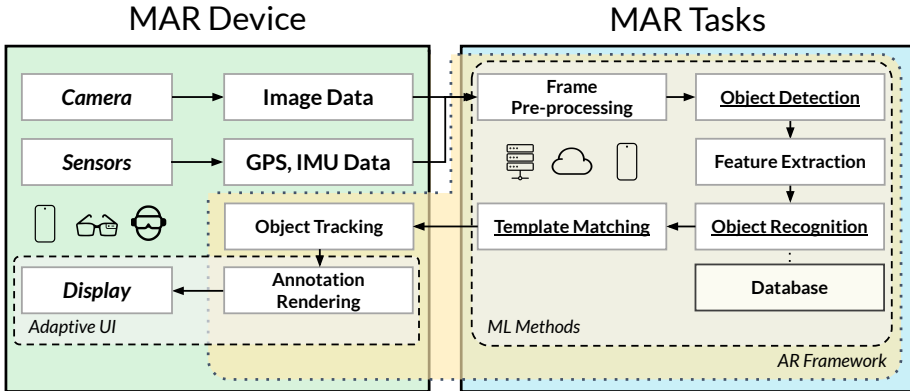


Fig. 1. A typical MAR system pipeline, splitting into MAR device and MAR tasks. ML methods can be utilised in the majority of required MAR tasks.

object tracking. Sahu *et al.* [207] conduct a comprehensive study on fundamental ML techniques for computational pipelines of AR systems (e.g., camera calibration, detection, tracking, camera pose estimation, rendering, virtual object creation, and registration), in AR-assisted manufacturing scenarios. Our article addresses the gaps in these existing surveys and provides a comprehensive survey on MAR interface design, frameworks and SDKs, and relevant ML methods utilised in key AR components.

The articles are initially selected by searching recent research articles using a small number of keyword seeds on Google Scholar, IEEE Xplore, ACM Digital Library, Scopus, Web of knowledge, etc. The queries combine keywords from the set *MAR*, *ubiquitous interaction*, *user interaction*, *AR adaptive UI*, *AR visualisation techniques*, *AR camera calibration*, *camera pose estimation*, *AR remote collaboration*, *AR user collaboration*, *MAR frameworks*, *object detection*, *object tracking*, *object recognition*, *rendering*, etc. The initial article set was expanded by taking into consideration articles that cite or are cited by the articles within this set. We further complement the set of articles by adding prominent research featured in media to cover a comprehensive set of important publications in this area. This process was continued until no new articles were found. We then discussed the papers among the authors and analysed the most relevant and important of the selected articles by reading the abstract and the main findings of the papers. Papers that were less relevant for the scope of the survey were filtered out during this process. The selected papers form the core of this survey and we have performed continuous updates during the survey writing process to cover papers published since the start of our process.

1.3 Structure of the survey

The remainder of this article is organised as follows. Section 2 summarises significant application areas which MAR can fulfil. We present detailed discussion and analysis on the development and recent advances of adaptive UI for MAR, MAR frameworks, and ML for MAR, in Section 3, Section 4, and Section 5, respectively. Finally, we discuss research challenges and future directions in Section 6 and conclude the paper in Section 7.

2 MAR APPLICATIONS

Over the past few years, AR has gained wide spread media attention and recognition due to easy access through modern smartphone applications. As an emerging technology, research and

development of MAR have been seen in several different fields. The versatility and ease-to-use of MAR highlight the potential for the technology to become a ubiquitous part of everyday life. In 2017 alone, the revenue of the AR market was worth \$3.5 billion [142] and by 2023 this value could reach up to \$70–\$75 billion [56]. A study conducted by SuperData and Accenture Interactive found that 77% of AR users are likely to use AR as a content viewer (e.g., for file viewing), 75% are likely to use AR in shopping scenarios, and 67% are likely to use AR for gaming entertainment. It reveals the importance of considering AR user interfaces, the underlying AR system, and how users can continually be encouraged to use AR. Among the different AR scenarios, the requirements of user interface are different between each of scenarios. In this section, we study several particular AR scenarios and applications with an emphasis on the types of user interfaces employed. This leads to a greater understanding as to why AR has been such a prominent technology and why it continues to attract research attention.

2.1 Entertainment & Tourism

AR technologies push the boundaries of entertainment and tourism. Pokémon GO [197] is a mobile game which uses AR to bring fictional Pokémon characters into the real world. Released in July 2016, at the height of the game's popularity, there were 45 million users and generated a first month of revenue of \$207 million [101]. This has led to the creation of similar concept AR gaming apps, such as Harry Potter: Wizards Unite [248]. AR gaming is not restricted to enhancing optical camera feeds. Audio AR experiences is another type of augmentation, for example, as found in the running app, Zombies, Run! [214]. Social media applications, such as Snapchat and Instagram allow users to apply “face filters” or “lenses” to themselves using AR [216]. In addition, Snapchat and social media application TikTok, have utilised the LiDAR sensor in the 2020 generation of iPhones to create environmental AR effects [72].

AR could act as digital tour guides for landmarks or cultural sites in tourism applications. Several projects create interesting MAR applications, which are offered to visitors to Tuscany, Italy [68] and Basel, Switzerland [69] with city navigation guidance, sight-seeing destinations, and general information. The design of such applications is traveller-centric so information is tailored to that specific audience, for example, providing multi-language functionality [82], offering alternative haptic or audio AR experiences [249], or supplying information which is suited to an individual's knowledge level [124]. Wide-spread adoption of AR has yet to emerge. However, with continual exposure and improving technologies, there is no doubt that AR will become a prevalent part of everyday life in several entertainment and tourism sectors.

2.2 Education

Over the past 30 years, classrooms have been evolving through the introduction of various technologies. From computers in classrooms [59] to lecture theatres with projection systems [165], classroom-based learning is now incorporating AR technology as an educational tool. Educational AR can be categorised as subject specific AR applications and separate AR teaching tools. Subject specific AR applications are developed for promoting a specific set of knowledge, and the user interfaces for these applications are well designed for presenting interactive contents. AR teaching tools apply AR technologies for improving the efficiency of teaching. The user interface for AR teaching tools is designed for facilitating the communications between users during education activities. Advantages are shown to arise from the use of AR in educational settings, for example, the individual skills of students can be improved, such as an increased ability to understand and handle information [109, 210], enhanced confidence [146], and a development in spatial abilities [117, 138, 154].

AR could enhance student-teacher communications within classrooms. Zarraonandia *et al.* [265] suggest providing teachers with HMDs and students with devices which can alert the teacher if a student is struggling. Holstein *et al.* [93] design intelligent tutoring systems as a rich formative assessment tool, which can augment teachers' perceptions of student learning and behaviour in real time. For students with disabilities and learning difficulties, the teaching can be tailored to individual specific needs. For example, pupils with attention deficit hyperactive disorder could use AR to help emphasise and improve engagement, motivation, and content visualisation [2]. Martín-Gutiérrez *et al.* [153] present a collaborative and autonomous 3D AR learning experiment, which combines every learning process from an electrical machines course in an electrical engineering degree. Their AR tool allows interactive and autonomous studying as well as collaborative laboratory experiments with other students and without assistance from teachers.

2.3 Healthcare

AR applications in the medical field offer new approaches for patient-doctor relationships, treatments, and medical education. For example, AR is categorised as an assisting technology to support individuals who have Alzheimer's disease to help identify objects and people, provide reminders to take the correct medication, and aid caregivers in locating patients [113]. Alternatively, current medical imaging modalities, such as ultrasound and CAT scans, can be greatly enhanced through scan virtualisation into 3D models and their superposition onto physical body parts [178]. With a scenario such as a pregnant mother receiving an ultrasound, the scan could be casted in real-time to several different observers wearing HMDs within the same room. All participants could receive the same ultrasound hologram on top of the patient, but midwives and other medical staff could be viewing additional information, such as past medical history and other vital signs [44, 151]. Novarad OpenSight presents one particular Microsoft HoloLens application product for CT images [75], allowing a user to view virtual trajectory guides and CT images which are superimposed on physical objects as two or three dimensional augmentations.

One of the most significant areas of AR for medical research pertains to utilising AR for surgery in various different stages of treatment. Firstly, a surgery can be explored and planned before execution, which allows surgeons to develop the most optimal treatment strategy [37]. Once the surgery procedure begins, image guidance can be used to provide surgeons with navigation information so their attention is not drawn away from the operative field [57, 66]. In addition, mixed reality image-guided surgery (IGS) has been frequently used for surgery training [32, 118] and minimally invasive surgeries. In IGS, surgical instruments are tracked and visualised with respect to patient-specific data sets to guide surgeons while in the operating theatre. Several surgical fields have already begun to apply AR, for example, neurosurgery, general surgery, and orthopedic surgery [213].

Adaptive user interface techniques are critical for surgery scenarios. To support this, several techniques, including label placement, occlusion representation, and registration error adaptation, are extensively studied. For example, Fuchs *et al.* [66] use depth cues of occlusion to aid in the perception of the order of objects. Specifically, a pixel is painted only if it lies above the surface of the closest object. In the AR system of Wieczorek *et al.* [252], occlusion handling of the surgeon's hand gives a more realistic perception and blending of realities. Pratt *et al.* [187] demonstrate that AR can assist accurate identification, dissection, and execution of vascular pedunculated flaps during reconstructive surgery.

3 COMPOSING AR VISUALISATION AND USER ENVIRONMENTS

Among all the AR applications (Section 2), the general AR experience is to display virtual entities overlaid on top of the physical world. For producing immersive AR experiences, information

management and rendering are key processes for delivering virtual entities. Collaborative AR system enables users to collaborate with each other within the physical and virtual environments. Numerous MAR frameworks are designated for the aforementioned purposes. Before we describe the existing MAR frameworks (Section 4), this section provides an overview of the relevant concepts and existing works in UI adaptability as well as device and user collaborations. We also explain the existing metrics of measuring the effectiveness of AR.

3.1 Adaptive UI for MAR

After data is prepared for visualisation, there are three main steps in the visualisation pipeline [271], including (1) filtering, (2) mapping, and (3) rendering. Zollmann *et al.* [271] perform a comprehensive review organised by this visualisation pipeline. We adopt their approach for classifying the characteristics of visualisation techniques and we further investigate the context-aware adaptive interface techniques which adapt to a person, task, or context environment. Several algorithms, such as ML-based algorithms, have been applied in adaptive user interfaces for improving human-interface interactions. They are designed to aid the user in reaching their goals efficiently, more easily, or to a higher level of satisfaction. We evaluate these works by considering the user experience measurement perspectives, and summarise four adaptive user interface techniques [108, 271] commonly found in the existing frameworks, with the following requirements for the implementations of each set of interfaces:

Information Filtering and Clustering (InfoF): Overload and clutter occur when a large amount of information is rendered to users. Complex visualisation of augmented contents negatively affects visual search and other measures of visual performance. In order to produce better user interfaces, there are two approaches to reduce the complexity of interfaces by decreasing the amount of displayed contents. The first approach is information filtering which can be implemented by developing a filtering algorithm that utilises two steps, namely, a culling step followed by a detailed refinement step. Another approach is information clustering [226], which groups together and represents the information by their classification.

Occlusion Representation and Depth Cues (OcclR): Occlusion representation provides depth cues to determine the ordering of objects. Users can identify the 3D locations of other physical objects when large structures occlude over or under each other. Julier *et al.* [107] suggest three basic requirements for depth cues, including (1) the ability to identify and classify occluding contours; (2) the ability to calculate the level of occlusion of target objects, and parameterising the occlusion levels if different parts of the object require this; and (3) the ability to use perceptually identified encodings to draw objects at different levels of occlusion.

Illumination Estimation (IEsti): Light estimation is another source of information depicting the spatial relations in depth perception, which enhances visual coherence in AR applications by providing accurate and temporally coherent estimates of real illumination. The common methods for the illumination estimation can be classified into auxiliary information or non-auxiliary information.

Registration Error Adaptation (RegEA): Another mapping step of the rendering pipeline. Trackers are imprecise and time-varying registration often exists. Therefore, the correct calibration of devices and displays is difficult. This subsequently leads to graphical contents not always aligning perfectly with their physical counterparts. Accordingly, the UI should be able to dynamically adapt while visualising the information. Ambiguity arises when the virtual content is not clearly interacting with the context environment around users.

Adaptive Content Placement (AdaCP): AR annotations are dynamic 3D objects, which can be rendered on top of physical environments. Labels have to be drawn with respect to the visible part of each object, in order to avoid confusing and ambiguous interactions.

InfoF: For some context-aware AR applications, information clutter is prevalent in outdoor environments due to the complexity of them. Exploring and searching for information on AR screens become indispensable tasks for users. Information filtering is a necessary technique in modern AR systems [229], especially in large and complicated environments, where information overload is significant. Without intelligent automation of filtering and selection tools, the display would always lead to difficulty in reading information [119]. There are three main methods for filtering information, including (1) spatial filters, (2) knowledge-based filters, and (3) location-based filters [229]. Spatial filters select information displayed on screens or in the object space based on physical dimension rules. These filters require user interactions to investigate the entirety of virtual contents, for example, users have to move their MAR devices to view large 3D models. But the method only works locally in a small region. The immersive AR application always applies the spatial filters to exclude the information that is out of the user's view frustum. Knowledge-based filters enable user preferences to be the filtering criteria [229]. Expert knowledge filters embedded behaviour and knowledge in the coding, regulating a series of data structures used by the system to infer recommendations and output the items satisfying user requirements. Such knowledge coding can be done in different ways, such as rules in a rule-based systems [183]. Finally, location-based filters are a hybrid method combining spatial and knowledge-based filters. As new sensors are embedded in modern AR headsets, such as gaze sensors, the user bio-information and context information can be used for filtering [6].

OcclR: Comprehensive AR systems not only track sparse geometric features but also compute depth maps for all pixels when visualising occluded objects or floating objects in AR [108]. Depth maps provide depth values for each pixel in captured scenes, they are essential for generating depth cues, and they help users to understand their environment and to aid interactions with occluded or hidden objects. Recent AR frameworks, such as Google ARCore [77] and Apple ARKit [15], provide depth map data for enabling depth cue features in MAR applications [58]. Physical and virtual cues are two options for producing depth cues in AR applications to support depth perception [271]. Physical cues can be used to rebuild natural pictorial depth cues [256], such as occlusion or shadows. Integrating depth maps and RGB camera images together can provide the necessary natural pictorial depth cues [271]. Subsequently, virtual cues and graphical aids are generated by applications to provide similar depth cues as physical alternatives. "X-ray vision" is a technique for virtual cues frequently used for perceiving graphics as being located behind opaque surfaces. DepthLab [58] is an application using ARCore's Depth API, enabling both physical and virtual cues, helping application developers to effortlessly integrate depth into their AR experiences. DepthLab implements the depth map and depth cues for at least six kind of interactions in AR, including (1) oriented reticles and splats, (2) ray-marching-based scene relighting, (3) depth visualisation and particles, (4) geometry-aware collisions, (5) 3D-anchored focus and aperture effect, and (6) occlusion and path planning [58]. Illumination estimation is typically achieved with two traditional approaches, including: (1) Methods utilising auxiliary information which leverages RGB-D data or information acquired from light probes and the methods can be an active method like the fisheye camera used by Kán *et al.* [111] or a passive method like reflective spheres used by Debevec [55], and (2) Estimating the illumination using an image from the main AR camera without the need of having an arbitrary known object in the scene. The auxiliary information can also be assumptions of some image features which are known to be directly affected by illumination, or simpler models like Lambertian illumination. Shadows, gradient of image brightness [35, 36, 115] and shading [80, 81, 105, 145, 203, 251] are the typical image features used for estimating illumination direction.

RegEA: Addressing registration and sensing errors is a fundamental problem in building effective AR systems [23]. Serious errors result in visual-kinesthetic and visual-proprioceptive conflicts.

Such conflicts between different human senses may be a source of motion sickness. Therefore, the user interface must automatically adapt to changing registration errors. MacIntyre B. *et al.* [148] suggest using Level Of-Error (LOE) object filtering for different representations of an augmentation to be automatically used as registration error changes. This approach requires the identification of a target object and a set of confusers [149]. Afterwards, their approach calculates the registration errors for the target and all confusers. The delegation error convex hulls are used to bound the geometry of the objects. The hulls are constructed for two disjointed objects in the presence of substantial yaw error. The hull surrounding each object together with a suitable label is sufficient to direct the user to the correct object edges.

Registration error adaptation is a critical issue in safety-critical AR systems, such as for surgical or military applications. Recent AR frameworks provide real-time registration error adaption with precise IMU tracking data and camera image fusion algorithms, which minimises the registration error. Robertson and MacIntyre [204] describes a set of AR visualisation techniques for augmentations that can adapt to changing registration errors. The first and traditional technique is providing general visual context of an augmentation in the physical world, helping users to realise the intended target of an augmentation. It can be achieved by highlighting features of the parent object, and showing more feature details as the registration error estimate increases. The second technique is presenting the detailed visual relationships between an augmentation and nearby objects in the physical world, by highlighting a unique collection of objects near the target of the augmentation in the physical world, and allowing the user to differentiate between the augmentation target and other parts of the physical world that are similar.

AdaCP: Major label placement solutions include greedy algorithms, cluster-based methods, and screen subdivision methods [24]. Other methods include making the links between objects and their annotations more intuitive, alleviating the depth ambiguity problem, and maintaining depth separation [156]. Labels have to be drawn with respect to each visible part of the object. Otherwise, the results are confusing, ambiguous, or even wrong. The implementation is simple computed axially-aligned approximations of the object projections, the visibility is then determined with simple depth ordering algorithms [29]. Several research works focus on providing appropriate moving label dynamics to ensure that the temporal behaviour of the moving labels facilitates legibility. From these works, certain requirements arise, such as the ability to determine visible objects, the parameterisation of free and open spaces in the view plane to determine where and how content should be placed, and labels should be animated in real-time because the drawing characteristics should be updated on a per-frame basis [108]. The requirements for adaptive content placement solutions are targeting to produce better user experiences.

View management algorithms address label placement [24]. Wither *et al.* [254, 255] provide an in-depth taxonomy of annotations, especially regarding the location and permanence of annotations. Tatzgern *et al.* [229] propose a cluster hierarchy-based view management system. Labels are clustered to create a hierarchical representation of the data and visualised based on the user's 3D viewpoint. Their label placement employs the "hedgehog labeling" technique, which places annotations in real world space to achieve stable layouts. McNamara *et al.* [157] illustrate an egocentric view-based management system that arranges and displays AR content based on user attention. Their solution uses a combination of screen position and eye-movement tracking to ensure that label placement does not become distracting.

For a comprehensive MAR system, current interfaces do not consider walking scenarios. Lages *et al.* [126] explore different information layout adaptation strategies in immersive AR environments. A desirable property of adaptation-based interface techniques is developed in their study. Adaptive content management is implemented in an MAR system, where the behaviours function in a

modular system can combine individual behaviours, and a final minimal set of useful behaviours is proposed that can be easily controlled by the user in a variety of mobile and stationary tasks [88, 104, 176, 191].

3.2 Collaborative UIs in Multi-user and Multi-device AR

Adaptive AR UIs can serve as ubiquitous displays of virtual objects that can be shown anywhere in our physical surroundings. That is, virtual objects can be floating in the air on any physical background, which can be reached out to or manipulated by multiple users with their egocentric views [123]. Users engaged in their AR-mediated physical surroundings are encouraged to accomplish tasks in co-facilitated environments with shared and collaborative AR experiences amongst multiple users [144]. Multiple dimensions of AR collaborative UIs are discussed throughout various applications, for instance, working [144, 218] and playful [61, 123] contents, local/co-located [170] and remote [196] users, sole AR [144, 170] and a mixture of AR and VR [61, 182], co-creation/co-editing by multiple users (i.e., multiple users at the front of AR scenes) [144, 170], supported and guided instruction (i.e., multiple users connecting to one user at the front of AR scenes) [47, 61], human-to-human interactions [123, 144, 170], and the interaction between human and AR bots, such as tangible robots [174] and digital agent representatives [17].

Multi-user collaborative environments has been a research topic in the domain of human-computer interaction, which has evolved from sedentary desktop computers [221] to mobile devices and head-worn computers [61, 144, 196, 218]. The success of such collaborative environments needs to cope with several design challenges, including (1) high awareness of others' actions and intentions, (2) high control over the interface, (3) high availability of background information, (4) information transparency amongst collaborators, and (5) information transparency whilst not affecting other users [38, 263]. These design challenges of awareness, control, availability, transparency, and privacy serve as fundamental issues enabling multiple users to smoothly interact with others in collaborative and shared environments. When the collaborative and shared environments are deployed to AR, features such as enriched reality-based interaction and high levels of adaptability are introduced. The additional design challenges extend from resolving multi-user experiences, as discussed previously, to reality-based interaction across various AR devices including AR/VR headsets, smartwatches, smartphones, tablets, large-screen displays, and projectors. These challenges place a priority on exploring the management of multiple devices and their platform restrictions, unifying the device-specific sensing and their interaction modalities, connecting the collaborative AR environments with physical coordinate systems in shared views [218]. The complexity of managing diverse devices leads to the need of an AR framework to systematically and automatically enable user collaborations in co-aligned AR UIs. It is important to note that the majority of evaluation frameworks focus on small numbers of quantitative metrics, such as completion time and error rate in an example of visual communication cues between an on-site operator and a remote-supporting expert [121], and they neglect the multi-user responses to the physical environments.

3.3 Metric Evaluations for AR UIs

AR interfaces were originally considered for industrial applications, and the goodness or effectiveness of such augmentations were limited to work-oriented metrics [54]. A very early example refers to augmented information, such as working instructions for front-line labourers on assembly lines, where productivity, work quality, and work consistency are regarded as evaluation metrics [67]. However, these task-orientated metrics are not equivalent to user experience (e.g., the easiness of handling the augmentation), and neglects the critical aspects especially user-centric metrics. The multitudinous user-centric metrics are generally inherited from traditional UX design issues, which can be categorised into four aspects: (1) user perception of information (e.g., whether the

information is comprehensible, understandable, or can be easily learnt), (2) manipulability (i.e., usability, operability), (3) task-oriented outcome (e.g., efficiency, effectiveness, task success), and (4) other subjective metrics (e.g., attractiveness, engagement, satisfaction in use, social presence, user control) [19]. When augmentations are displayed on handheld devices, such as smartphones and tablets, ergonomic issues, such as ease in AR content manipulation with two-handed and one-handed operations are further considered [209]. Nowadays, AR is deployed in real-world scenarios, primarily acting as marketing tools, and hence business-orientated metrics are further examined, such as utility, aesthetic, enjoyment, and brand perception [194, 237]. Additionally, multi-user collaborative environments require remote connections in AR [267], and lately, the quality of experience (QoE) through computation offloading mechanisms to cloud or edge servers encounters new design challenges of AR UIs [208, 232].

The user perception of AR information which leads to the comprehensibility of AR cues and the learnability of AR operations in reality-based interactions, has been further investigated as a problem of multi-modal cues, such as audio, video, and haptics in various enriched situations driven by AR [73]. The intelligent selection of AR information and adaptive management of information display and multi-modal cues are crucial to users' perception to AR environments [42, 129]. This can be considered as a fundamental issue of interface plasticity. In the mixed contents between the digital and physical realities [19], the plasticity of AR interfaces refers to the compatibility of the information to physical surroundings as well as situations of users (i.e., context-awareness) [74, 140].

After defining evaluation metrics, AR UI practitioners (e.g., software engineers and designers) often examine the dynamic user experience in interactive environments while satisfying the aforementioned metrics. There are attempts to assess the AR experience by building mini-size studio interactive spaces to emulate AR environments [120]. However, such physical setups and iterative assessments are costly and time-consuming, especially when AR is implemented on large-scales (i.e., ubiquitously in our living spaces). Moreover, the increasing number of evaluation metrics calls for systematic evaluations of AR UIs [102]. It is therefore preferable to assess AR UI design metrics through systematic approaches and even automation, with prominent features of real-time monitoring of system performance, direct information collection via user-device interaction in AR, and more proactive responses to improve user perceptions to AR information [102]. However, to the best of our knowledge, the number of existing evaluation frameworks is very limited and their scopes are limited to specific contexts and scenarios, such as disaster management [167]. More generic evaluation frameworks with a high selectivity of AR evaluation metrics pose research opportunities in the domain of AR interface designs [73, 102].

4 MAR FRAMEWORKS

MAR systems can either be deployed individually on client devices (e.g., smartphones and HMDs) without the need for separate hardware, or client devices and external servers together. This section investigates existing MAR frameworks, a large number of which are designed for the creation of MAR applications. A general hardware benchmark for these frameworks can be found by exploring the minimum requirements for MAR frameworks, such as ARCore and ARKit. These software development kits (SDKs) should be most optimised for the devices released by Google and Apple, respectively. According to the device-specification for both SDKs, the oldest devices still actively supported were released in 2015, i.e., the Nexus 5X [78] and the iPhone 6S [14]. However, older devices can still be supported by commercial and open-source solutions, such as ARToolKit which supports Android devices released in 2013 [52], and Vuforia which supports the 2012 released iPhone 5 [189]. Through software optimisation, MAR support for a variety of mobile hardware is possible. However, hardware backwards-compatibility can be reduced as more features are developed for the SDKs, and those features may require better processing capabilities or additional

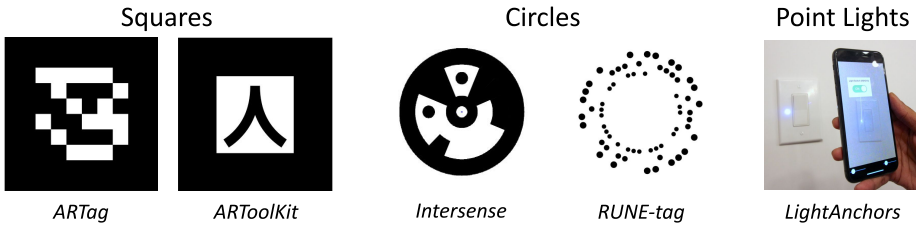


Fig. 2. Examples of square, circular, and point light fiducial markers used in MAR.

hardware, which may not always be available from older devices. Therefore, features may be omitted from version releases, or device support may be removed completely.

4.1 Evaluation Metrics

MAR frameworks have various features as they aim to support different platforms and constraints. We compare these frameworks by evaluating the availability of the features, capabilities, and the devices which the frameworks can be deployed on as evaluation metrics.

- (1) **Platform support:** MAR is not necessarily restricted to only smartphones running Android (🤖) or iOS (🍏) operating systems (OSs). Wearables and HMDs are other MAR devices, which run alternative OSs, such as Linux (🐧) [65, 180] and Windows Holographic OS/Windows 10 (🪟) [159]. The frameworks are typically deployable on several different hardware types. Cross-platform support is also found through web browser-based MAR (🌐) and using the cross-platform Unity Real-Time Development Platform (🎮) [122].
- (2) **Tracking:** To provide MAR experiences, virtual annotations are placed and tracked within physical environments. The two most commonly used techniques to enable and produce these experiences are (a) marker- or fiducial-based, and (b) markerless tracking.
 - Marker or fiducial tracking uses 2D patterns of predefined shapes and sizes, which are placed within environments, these are recognised from optical images and accompanying information associated with the patterns are returned to users [64]. Among the different types of planar markers used in AR, Figure 2 presents typical examples, including shaped-based markers such as squares (ARTag [64], ARToolKit [130]), circles (Intersense [162], RUNE-tag [30]), and pervasive point lights (e.g., LEDs or light bulbs) to spatially-anchor data (LightAnchors [5]). Planar patterns, such as QR codes [112] are good at encoding information. However, they do not function well in certain MAR situations such as when there are large fields of view and when perspectives are distorted [64].
 - Markerless tracking utilises local environments of users to generate and fix virtual markers for 3D tracking [63, 168]. Also known as natural feature tracking (NFT), the natural features are points and regions of interest which are detected and selected, then motion and pose estimation are performed to track those features, and augmented content can be fixed to physical environments [168]. The most popular techniques for this environment analysis is simultaneous localisation and mapping (SLAM) [201]. SLAM methods generally reconstruct an environment into a 3D spatial map (3D point cloud) to build a global reference map while simultaneously tracking the subject's position [26, 43]. Markerless tracking has advantages such as not requiring prior knowledge of a user's locale, and not needing to place additional objects (e.g., fiducial markers) in an environment to create MAR experiences.

Some frameworks also support a variety of other tracking techniques and behaviours which have varying degrees of support across the different frameworks [235], including (a) Device

tracking: the device's position and orientation are tracked with respect to physical space. (b) Plane tracking: flat, horizontal and vertical surfaces are detected and tracked. (c) Hand tracking: human hands are recognised and tracked. (d) Body tracking: the entire human body is recognised in relation to physical space. (e) Facial tracking: the human face is detected and tracked.

- (3) **Features:** Frameworks and SDKs support various features, which enhance the overall MAR experience for users, whether through quality of service, or quality of experience improvements [235]: (a) Point clouds: feature point maps from NFT, frameworks sometimes allow the maps to be accessible to developers and users. (b) Anchors: these are points and locations of arbitrary position and orientation which the system tracks. (c) Light estimation: the average colour temperature and brightness within a physical space is estimated. (d) Environment probes: probes which are used to generate a cube map which represents a particular area of a physical environment. (e) Meshing: a physical space is virtually segmented and masked using triangular meshes. (f) Collaboration: positions and orientations of other devices are tracked and shared to create a collaborative MAR experience. (g) Occlusion: objects in the physical world have known or calculated distances to aid in the rendering of 3D content, which allows for realistic blending of physical and virtual worlds. (h) Raycasting: the physical surroundings are queried for detected planes and feature points. (i) Pass-through video: the captured camera feed is rendered onto the device touchscreen as the background for MAR content. (j) Session management: the platform-level configuration is changed automatically when AR features are enabled or disabled.
- (4) **Sensors:** The hardware providing MAR experiences are now equipped with a wide range of sensors. In addition, they contain high resolution displays and cameras; inertial measurement unit (IMU) sensors such as gyroscopes, accelerometers, and compasses; and Global Positioning System (GPS) sensors are now a standard inclusion in smartphones and HMDs. They benefit MAR frameworks with estimating user location and orientation to display location-relevant information [40]. Ranging technologies, such as LiDAR (light detection and ranging) sensors are now used to enhance current arrays of optical cameras on mobile smartphones for better spatial mapping, which is available from recent Apple iPhone and iPad devices [99].
- (5) **Architecture:** With every iteration release of mobile hardware, these devices are more capable with increasingly powerful CPUs and GPUs to perform onboard tasks (*offline* processing). However, not every user has access to the latest and most powerful devices, therefore in these scenarios, a framework could employ external servers to offload computations (*online* processing). Frameworks could then support either offline and online recognition of objects, or both.

Other considerations for selecting an MAR framework include, whether the framework is available for free, or requires the purchasing of a commercial software license; whether the framework is open source for developers to expand and to build upon; whether the framework is for general "all-purpose" usage, i.e., not just for one use-case like facial-tracking; and whether there are studio tools to simplify and aid in the development of MAR experiences.

4.2 Frameworks

We perform a comprehensive summary of existing MAR frameworks and SDKs, with considering the evaluation metrics specified in Section 4.1. We tabulate our results in Table 1. To provide a basic filter on the frameworks and SDKs listed, we decide to exclude frameworks which have not been updated within a four year time span. This represents the age of the iPhone SE (1st generation, 2016) which is supported by the latest version of iOS 14 from Apple (i.e., the oldest supported

Table 1. Comparisons of several available features in MAR frameworks and SDKs.

Framework/SDK	Platform support					Tracking					Features										Sensors			Others								
	Android	iOS	ARCore	ARKit	WebXR	Markers	NFT	Device	Plane	Hand	2D & 3D body	Facial	Point clouds	Anchors	Light estimation	Environment probes	Meshing	Collaboration	Occlusion	Raycasting	Pass-through video	Session management	Camera	LiDAR	IMU	GPS	Architecture	Price	Open source	General	Studio tool	
A-Frame (1.2.0) [1]	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Online	Free	✓	✓	X
ALVAR (0.7.2) [244]	✓	✓	✓	✓	X	✓	✓	✓	✓	X	X	X	✓	✓	X	X	X	X	X	X	X	X	X	✓	X	✓	✓	Offline	Free	X	✓	X
Amazon Sumerian (N/A) [9]	✓	X	X	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Online	Free, Paid	X	✓	✓
ApertusVR (0.9*) [224]	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Offline	Free	✓	✓	X
ARCore (1.23.0) [77]	✓	✓	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Offline	Free	✓	✓	X	
ARKit (4) [15]	X	✓	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Offline	Free	X	✓	✓	
ARMedia SDK (2.1.0*) [100]	✓	✓	X	✓	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	✓	Offline	Paid	X	✓	X	
ARToolKit (5.4*) [116, 130]	✓	✓	✓	✓	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Offline	Free	✓	✓	X	
artoolkitX (1.0.6.1) [198]	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Offline	Free	✓	✓	X	
ArUco (3.1.12) [13, 205]	✓	X	✓	✓	X	X	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Offline	Free	✓	✓	X		
AR.js (3.3.1) [62]	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Online	Free	✓	✓	X	
Augment (4.0.6*) [20]	✓	✓	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Online	Paid	X	✓	✓	
Augmented Pixels (N/A) [21]	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	✓	Offline	Paid	X	✓	✓	
AugmentedPro (2.4.3) [22]	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	✓	Offline	Paid	X	✓	✓	
Banuba (0.35.0) [27]	✓	✓	X	X	✓	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Offline	Paid	X	X	X	
Blippar (N/A) [33]	✓	✓	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Online	Paid	X	✓	✓	
CraftAR (5.2.1*) [49]	✓	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	Both	Free, Paid	X	✓	✓	
DeepAR (2.3.1) [98]	✓	✓	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Offline	Free, Paid	X	✓	✓	
EasyAR (4.2.0) [241]	✓	✓	X	✓	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Offline	Free, Paid	X	✓	✓	
HERE SDK (3.17) [90]	✓	✓	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Offline	Paid	X	X	X	
Kudan AR SDK (1.6.0) [260]	✓	✓	X	X	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Offline	Free, Paid	X	✓	X	
Lumin SDK (0.25.0) [150]	X	X	X	X	X	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Offline	Free	X	✓	X	
MAXST AR SDK (5.0.3) [155]	✓	✓	X	X	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Both	Free, Paid	X	✓	X	
Minsar (2.0) [172]	✓	✓	X	X	✓	X	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Both	Free, Paid	X	✓	✓	
MRTK (2.6.0) [160]	✓	✓	X	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Offline	Free	✓	✓	X	
NyARToolkit (5.0.8*) [190]	✓	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Offline	Free	✓	✓	X	
Onirix (N/A) [171]	✓	✓	X	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Online	Paid	X	✓	✓	
Pikkart AR SDK (3.5.8*) [181]	✓	X	X	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Both	Free, Paid	X	✓	X	
PlugXR (1.0.0) [185]	✓	✓	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Online	Paid	X	✓	✓		
Universal AR SDK (N/A) [264]	✓	✓	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Online	Free	X	✓	X	
Vectary (N/A) [238]	✓	✓	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Online	Paid	X	✓	✓	
Vidioti SDK (N/A) [239]	✓	✓	X	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	✓	Both	Paid	X	✓	✓	
ViewAR SDK (N/A) [240]	✓	✓	X	X	X	X	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Offline	Paid	X	✓	X	
Vuforia (9.7) [188]	✓	✓	X	X	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Both	Free, Paid	X	✓	X	
Wikitude (9.6) [253]	✓	✓	X	X	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	✓	✓	✓	Both	Paid	X	✓	✓	
WebXR (N/A) [28]	✓	✓	X	✓	✓	X	✓	✓	○	○	○	○	○	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	○	○	○	Offline	Free	✓	✓	X	
XZIMG (2.0.2*) [261]	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	X	✓	Offline	Free, Paid	X	✓	X	

Frameworks and SDKs marked with an asterisk (*) have not been updated in 1+ years. Features or functions optionally supported in different platform are marked with ○.

iPhone) [16], and similarly, Google supports the Pixel 2 (2017) in their latest Android 11 OS update [79]. Frameworks which last had updates four years ago may not function on the wide range of new hardware and software now available to users. Certainly, this may also be true for frameworks which have not had updates since one year ago, these are therefore marked in Table 1 with an asterisk (*).

- **A-Frame** [1] is a framework primarily dedicated for web-based virtual reality (VR). Built using JavaScript, A-Frame has additional capabilities for displaying AR content. However, greater MAR functionality is provided when using the community component, AR.js.

- **ALVAR** [244] is a suite of SDKs developed by the VTT Technical Research Centre in Finland. Functionality is split into three distinct C++ libraries: (1) *ALVAR Core* is a low-level AR toolkit; (2) *ALVAR Platform* abstracts the camera, IMU, and GPS sensors and a data processing pipeline; and (3) *ALVAR Tracker* provides 3D point cloud, planar image, multi-marker, and panorama tracking.
- **Amazon Sumerian** [9] is an Amazon-managed service for building and deploying AR, VR, and 3D applications. Sumerian applications are accessible through web browsers and is based on JavaScript API and WebGL 2. Additionally, hybrid experiences can be achieved with iOS ARKit and Android ARCore.
- **ApertusVR** [224] is primarily an open source VR software library composed of a main core (ApertusCore) and a set of plugins (ApertusVR) which extends the core to provide both AR and VR functionality. Their architecture is based on a “distributed plugin mechanism” which allows for IoT devices and other smart hardware to interface with the system.
- **ARCore** [77] is an SDK developed by Google for mobile platforms, with the main development environment being Android smartphones. However, there are APIs which enable cross-platform AR experiences. One notable service is ARCore Cloud Anchor which allows for anchors and feature points to be sent to a cloud host, and subsequently shared with multiple user devices to create collaborative and multiplayer apps.
- **ARKit** [15] is Apple’s solution to MAR for their products, i.e., iPhone, iPad, and iPod devices. Visual-inertial odometry is used to analyse notable features from a scene taken by a device’s camera to provide physical world tracking, and there are additional classes to track people’s movement using body anchors, and using alpha and depth information from an image frame to occlude virtual content. The framework can additionally utilise LiDAR (light detection and ranging) data from an iPhone and iPad sensor to enhance spatial environment mapping.
- **ARMedia SDK** [100] is developed by Inglobe Technologies which provides object, planar image, geolocation-based and motion tracking. Additionally, a simple object rendering engine is included to visualise 3D contents.
- **ARToolKit** [116, 130] was a software library developed by the Human Interface Technology Laboratory at the University of Washington. The framework uses camera position and orientation relative to physical fiducial markers to determine scene geometry. The last update to the original ARToolKit SDK was made in May 2007 [131]. Android, iOS and JavaScript support was later introduced in ARToolKit in 2015 when version 5 was released as an open-source framework by DAQRI [53]. This project was developed until November 2017, when the developers’ focus shifted to a new open-source project, artoolkitX.
- **artoolkitX** [198] is a framework supported by Realmax Inc. and has four distinct libraries: (1) camera and video source processing, (2) feature point tracking with OpenCV, (3) fiducial marker tracking, and (4) virtual content rendering with OpenGL/GLES. All of these components are managed by a central controller class.
- **ArUco** [13, 205] is a minimal open source library for camera pose estimation using square markers. Discrimination correlation filters are used for the tracking of markers and their corners, and achieves continuous detection under extreme blurring and scaling conditions.
- **AR.js** [62] is an MAR library built for mobile smartphone web browsers. Adapting the 2015 JavaScript version of ARToolKit, this framework is a lightweight AR solution which uses fiducial markers and on-device GPS sensors to localise and track 3D content. The only requirement client devices must have, is the web browser should contain support for WebGL and WebRTC JavaScript APIs. Users do not need to install a device application, rather, just load a web app. AR.js can be expanded with additional JavaScript frameworks such as A-Frame [1] and three.js [231] to provide alternative methods for rendering virtual content.

- **Augment** [20] is a paid platform for 3D and AR product visualisation for mobile devices and web browsers. The platform has several products for viewing and managing 3D models, and an SDK to embed the model player in custom applications.
- **Augmented Pixels** [21] is an SDK which provides a custom SLAM platform for developers to fuse data from multiple sensor inputs to calculate positioning. The primary deployment scenarios are drones and robotics navigation, and providing real-time SLAM for mobile smartphones and AR/VR glasses.
- **AugmentedPro** [22] is a two component AR solution for industry. An *AugmentedPro Creator* program is used to modify and create virtual content, this can then be viewed and displayed as an AR experience through a companion *AugmentedPro Player*.
- **Banuba** [27] provides several augmented face AR solutions. They have two main SDKs, the first of which is tracking and augmenting facial features in a Face AR SDK, the second is an AI Video Editor SDK which allows users to alter their video streams through the inclusion of face filters, modifying video backgrounds, and applying AR video overlays.
- **Blippar** [33] provides several products for commercial and educational AR applications. Developers can develop their own experiences in custom applications using the Blippar AR SDK. Alternatively, their Blippar app and their creation tools allows for a simplified development process, as their application functions as an AR browser to which users can obtain multiple AR experiences for different products and services. The Blippar SDK is also used for other frameworks such as Layar [132].
- **CraftAR** [49] is a marker-based AR SDK providing content rendering on-device or through their cloud service. There is an image recognition solution which is also marketed under the same name [48], and provides on-device or cloud-based recognition. Both SDKs are combined and utilised as part of their Pro SDK product.
- **DeepAR** [98] is an AR platform mainly focusing on providing Snapchat- and Instagram-style face augmentations. Utilising neural networks and deep learning models, DeepAR is able to perform face detection, face tracking, and emotion recognition. Face shape heatmaps and facial feature points are generated and used to track facial features.
- **EasyAR** [241] is an SDK supporting surface, 3D object, and planar image tracking; as well as motion tracking by utilising visual inertial odometry, and on-device 3D mesh generation to create sparse and dense environment spatial maps. This allows for features such as persistent AR content, collaborative multi-user AR, collision detection, and object occlusion effects.
- **HERE SDK** [90] is primarily a collection of SDKs and APIs which are used to build location-aware apps such as navigation applications. The main AR component of both SDKs for Android and iOS is the LiveSight feature which enables geospatial location markers in the real world.
- **Kudan AR SDK** [260] is developed by Kudan, and is supported by KudanSLAM, Kudan's own SLAM solution. Their implementation uses hardware sensors such as stereo and visual-inertial depth cameras to create MAR experiences.
- **Lumin SDK** [150] is the SDK by Magic Leap for producing native apps which are deployed on their AR smartglasses product, the Magic Leap 1. The SDK is available for several development platforms, and allows developers to create their self-described "spatial computing" experiences.
- **MAXST AR SDK** [155] is a cross-platform AR engine for developing MAR applications. Visual SLAM is used to obtain a map of the virtual environment, and subsequent coordinate systems are created with the *MAXST AR Fusion Tracker* which uses technology from ARKit and ARCore.

- **Minsar**, or Minsar Studio [172] is a “no-code” tool for creating both AR and VR content for mobile devices, and HMDs (i.e., HoloLens, Magic Leap, and Oculus Quest). MAR experiences are made using the Minsar Studio application found on both iOS and Android, and are exported using either the companion *XR View* application, or through a web browser using WebXR.
- **MRTK**, or Mixed Reality Toolkit [160] is a Microsoft-driven framework for cross-platform mixed reality (MR) application development. MRTK primarily supports and drives applications on Microsoft HoloLens devices, but supports OpenVR headsets, and mobile devices running iOS and Android. On HoloLens 2 devices, MRTK provides additional features such as hand and eye tracking.
- **NyARToolkit** [190] is a collection of AR libraries based on ARToolKit. Several platforms are supported including Android and Unity, where both marker-based and NFT tracking is available to serve augmented content.
- **Onirix** [171] is composed of three SDKs: (1) *Onirix Places* utilises user location and viewing direction along with Google Maps to provide geo-spatial AR content, (2) *Onirix Spaces* employs SLAM algorithms to provide markerless AR in indoor environments, and (3) *Onirix Targets* uses 2D marker objects and images for marker-based AR. Onirix uses ARCore and ARKit to deliver the SLAM and light estimation features within the Onirix Spaces SDK.
- **Pikkart AR SDK** [181] utilises marker images and geolocation data to track virtually drawn content. A notable feature is Pikkart AR Logo, which uses printable steganography to produce micro-variations on markers so different virtual content can be returned using the “same” marker.
- **PlugXR** [185] is a cloud-based MAR solution containing a *PlugXR Creator* tool to create MAR experiences which are deployable on several platforms, and a *PlugXR App* to view those experiences. The primary focus is providing tools so users with no coding knowledge can create AR applications.
- **Universal AR SDK** [264] by ZapWorks is composed of several SDKs for different platforms to enable MAR experiences, including web browser-based MAR through compatibility with JavaScript, A-Frame, and ThreeJS. The main features across the frameworks include image, world, and face tracking.
- **Vectary** [238] is a web-based studio tool to create 3D and AR experiences. Those experiences are then accessible through mobile or desktop web browsers using their Vectary Web AR system.
- **Vidinoti SDK** [239] is an SDK which delivers MAR experiences using a combination of on-device image recognition and tracking, as well as their cloud recognition service, however, their SDK is able to function completely offline.
- **ViewAR SDK** [240] allows developers to create both AR and VR applications using a JavaScript-based API. Both ARCore and ARKit frameworks are integrated within the SDK and are automatically selected depending on the device type.
- **Vuforia** [188] is an SDK which can use several different types of images and objects as targets: pre-existing 3D models, planar images, cylindrical objects (e.g., columns, coffee cups, beverage bottles/cans), and targets can be created through scanning objects to produce 3D models. Additionally, video data from external non-client cameras can be used to enhance the recognition and tracking capabilities of the platform.
- **Wikitude** [253] is an MAR orientated SDK. Offering their own SLAM technology which can connect with ARCore and ARKit backends, and supports object fixing onto a scene even after the target has left the field of view. Traditional marker-based recognition is also available in

the SDK, along with location-awareness and model occlusion. Wikitude is also deployable and optimised for smart glasses such as Epson Moverio and HoloLens.

- **WebXR** [28] is a specification developed by the World Wide Web Consortium, W3C. WebXR API is an implementation of the WebXR feature set. Since the WebXR API serves as an interface between XR Web content and the devices on which they run, the functions and features are not the same across different platforms or web browsers.
- **XZIMG** [261] produces two main AR products, *Augmented Vision* and *Augmented Face*. Both are deployable on standard MAR hardware, as well as through web browsers. Augmented Vision is a typical fiducial marker recognition system, while Augmented Face has face tracking and augmentation technology.

From the 37 listed frameworks, Android and iOS are the two most consistently supported platforms. Support for Unity is also often available as it simplifies the MAR application development process, e.g., easily exporting produced applications to multiple platforms. In terms of support for different tracking technologies: (1) 30 of the 37 frameworks support 2D image marker tracking, (2) 27 support natural feature tracking, (3) only 6 support hand tracking, (4) facial tracking is found in 8 of the frameworks, and (5) only 2 frameworks supports 2D and 3D body tracking. This indicates that there is still a primary focus on environment-based MAR, and human body-based tracking is still of limited interest to framework developers.

By comparison, there is varied supported for the different features in frameworks. Both AR-Core and ARKit support the majority of the described features, they therefore form the baselines upon which other frameworks can be compared to and built upon. Among the 10 listed features, the following are organised as the most-to-least supported: (1) pass-through video, (2) session management, (3) anchors, (4) point clouds and occlusion, (5) raycasting, (6) meshing, (7) light estimation, and (8) collaboration and environmental probing. While the overall average number of supported features is approximately 3 per framework, there are 10 frameworks which only support 1, i.e., pass-through video. In relation to the type of architecture in use, offline and on-device-based processing is the most popular with 28 frameworks (21 offline-based, and 7 supporting both offline and online), and 16 frameworks use purely an online-based processing architecture. These online frameworks are often obtainable as paid services which are offered alongside a free trial variant, while the offline frameworks are generally free, albeit not always open source.

5 MACHINE LEARNING METHODS FOR MAR

Different MAR frameworks and implementations contain approximately the same basic components, irrespective of the hardware used, or whether the system is self-contained on one device or distributed between client and an external server. Figure 1 in Section 1 presents the typical pipeline of MAR systems where different tasks function together to create the MAR experience. Several of these tasks often require the use of ML algorithms to be successfully completed, namely, machine vision-based tasks. ML is a set of methods and techniques used for modelling or building programs that employ knowledge gathered from solving previous problems to answer unseen examples of similar problems, i.e., learning from past experiences [158, 161]. ML technologies are crucial for MAR adoption, because they make the processing of visual, audio, and other sensor data to be more intelligent while simultaneously protecting privacy and security. A notable subfield of ML is deep learning, where multiple layers are used to represent the abstractions of data to build computational models [186]. Deep learning methods are useful in different domains [3].

In this section, we explore several MAR tasks which require the usage of ML methods. More specifically, we provide the requirements of the tasks, the ML methods used to fulfil the tasks, and the advantages and disadvantages of those methods. Table 2 provides a collection of most

Table 2. List of example AR works which use ML methods for completing AR tasks or as part of the system.

Device	Overview	Task	ML method
Server	Deep learning for AR tracking [7]	Object detection	CNN
	Real-time moving object detection for AR [51]	Object detection	Background-foreground non-parametric-based Mask R-CNN
	Deep-learning-based smart task assistance for wearable AR (HoloLens) [173]	Object detection and instance segmentation	(2D) CNN
	Interacting with IoT devices in AR environment[222]	Real-time hand gesture recognition	DNN
	Edge-assisted distributed DNN for mobile WebAR [202]	Object recognition	History of Oriented Gradientsm, Haar-like features
	Object detection and tracking for face and eyes AR [84]	Object detection, object recognition	Random forest
	AR surgical scene understanding improved with ML [175]	Object identification	CNN, XGBoost
	Dynamic image recognition methods for AR [46]	Feature extraction, object recognition	Manifold learning
	AR platform for interactive aerodynamic design and analysis [25]	Object recognition	SSD
	AR retail product identification [236]	Object detection	ResNet50 (CNN)
	Improving retail shopping experience with AR [50]	Object recognition	Faster R-CNN
	AR instructional system for mechanical assembly [127]	Object detection	CNN
	Deep learning for AR [128]	Object tracking, light estimation	CNN
	AR for radiology [233]	Image segmentation	CNN
	AR design personalisation for facial accessory products [97]	Facial tracking	AdaBoost
Low cost AR for automotive industry [192]	Feature extraction, object classification	Linear SVM, CNN	
AR training framework for neonatal endotracheal intubation [269]	Assessing task performance	CNN	
AR video calling with WebRTC API [106]	Semantic segmentation	CNN	
AR gustatory manipulation [163]	Food-to-food translation	GAN	
Edge	Edge-based inference for ML at Facebook [259]	Image classification	DNN
	Supporting vehicle-to-edge for vehicle AR [270]	Object detection	Deep CNN (YOLO)
	AR platform for operators in production environments [234]	Object detection	SSD
	Spatial AR with single IR camera [83]	3D pose estimation, image classification	Hough Forests, Random Ferns
	Federated learning for low-latency object detection and classification [45]	Object classification modelling	Federated learning
Optimising learning accuracy in MAR systems [87]	Delay and energy modelling	CNN	
Client	Application to support users in everyday grocery shopping [245]	Object classification	Random forests
	Outdoor AR application for geovisualisation [195]	Object detection	SSD
	MAR object detection [135]	Object detection	DNN
	Reducing energy drain for MAR [12]	Object detection	DNN
	HoloLens surgical navigation [243]	Object detection and pose estimation	CNN
	AR inspection framework for industry [177]	Object detection	R-CNN
	Learning Egyptian hieroglyphs with AR [184]	Object detection	SSD MobileNets
	Interest point detection [219]	Object detection	CNN
	Edge-assisted distributed DNN for mobile WebAR [202]	Object recognition	DNN
	AR navigation for landmark-based navigation [10]	Prediction of speed of movement	Penalised linear regression
	Enhancing STEM education with AR [11]	Object detection	MobileNets
	Campus navigation with AR [137]	3D model placement determination	CNN
	AR assisted process guidance on HoloLens [200]	Predicting process quality metrics	Decision tree classification
	Indoor AR for Industry 4.0 smart factories [220]	Object detection	MobileNets
	AR application for science education of nervous systems [95]	Image classification	CNN

significant work of the aforementioned AR tasks as well as the ML methods used to fulfil them based on where ML algorithms are deployed on, Server, Edge, or Client. Subsequently, Figure 3 provides a summary mapping of these tabulated works, linking the AR tasks to the ML methods, this figure demonstrates the popularity of certain methods, such as CNN and SSD (more details in Section 5.1), and the roles they play in accomplishing AR tasks.

5.1 ML for Object Detection and Object Recognition

Several MAR systems include object detection and/or recognition as the primary pipeline task which uses ML methods during execution. Object detection supports MAR experiences by allowing the system to understand and detect which objects are in a particular scene at a given moment, and then return the spatial location and extent of each object instance [141]. For MAR in particular, object detection is used to provide environmental context-awareness and system knowledge of what supplemental augmentations should be provided users and where to place them on UIs.

Deep learning are used to achieve object detection, where feature representations are automatically learned from data. The most commonly used type of deep learning models for object detection and object recognition are Convolutional Neural Networks (CNN) [8], Region-based Convolutional Neural Networks (R-CNN) [76], Faster Region-based Convolutional Neural Networks (Faster

R-CNN [266], You Only Look Once (YOLO) [199], Single Shot MultiBox Detectors (SSD) [143], ResNet [86], Neural Architecture Search Net (NASNet) [272], Mask Region-based Convolutional Neural Networks (Mask R-CNN) [85], DenseNet [96], RetinaNet [139], and EfficientNet [225]. These networks can be used to create algorithms which are able to produce the same detection or recognition results. The major differences between each neural network are related to the structure and approach of the network itself when processing input data (e.g., camera images for optical object detection/recognition). For example, R-CNN, which tries to find rectangular regions which could contain objects in an image, based on the features of each rectangular region extracted by CNN, is one of the most important approaches for objective detection. Fast R-CNN has been proposed by applying the CNN to the whole image to find a rectangular region with fully connected layers to reduce computation cost caused by R-CNN. Fast R-CNN is proposed by merging R-CNN with region proposal network (RPN) to share full image convolutional features. Mask R-CNN is usually used in segmentation and could be regarded as a mixture of Faster R-CNN and fully convolutional networks for object detection. CNNs and the subsequent variants are therefore powerful components in the AR pipeline for producing AR experiences, without which, the system would require more time to analyse the environment which comes at the cost of user QoE.

Both object detection and object recognition tasks in MAR systems are typically deployed on servers, namely, due to the limitation of client devices and their lack of computation and battery resources to sufficiently sustain acceptable MAR experiences. Examples of server-based deployment include usage of CNNs for object detection to support AR tracking [7], Mask R-CNN for object detection and instance segmentation to support smart task assistance for HoloLens-deployed AR [173], and CNN for object recognition in improving retail AR shopping experiences [50]. Additionally, some works explicitly state their usage of edge servers, citing the processing acceleration gained by the object detection and recognition tasks when a GPU is used to execute the functions, for example, using YOLO to accomplish object detection to support vehicle-to-edge AR [270], and SSD for supporting an edge-based AR platform for operators in production environments [234]. Comparatively, there are several research works which deploy object detection and recognition in an enclosed client system, i.e., without needing to offload computation components to an external server or device. Amongst these works, both CNNs and DNNs are the most commonly attributed ML models. DNN models have been used to fulfil MAR object detection [135] and support the reduction in energy drain for MAR applications [12]. Alternatively, CNNs are used for object detection and pose estimation for HoloLens surgical navigation [243], interest point detection for an AR application [219], as well as for object detection in an industrial AR inspection framework [177], albeit the model for this latter work is R-CNN. The deployment of these models on the client is supported by increasing efforts to produce more efficient neural networks, as well as reducing their implementation size through using ML libraries which are specifically designed for deployment on mobile devices, i.e., TensorFlow Lite [230].

5.2 ML for Object Tracking

Object tracking in the MAR pipeline allows for objects in the physical world to be located and tracked across the MAR system, which subsequently supports the placement and tracking of virtual annotations derived from the analysis of the surrounding user environments in object detection and object recognition routines. As discussed in Section 4, two common used object tracking techniques are marker-based tracking and markerless tracking. Similar to object detection and object recognition, object tracking can be achieved through analysis of optical camera data [250]. However, object tracking differs from object detection as objects must be tracked over time while the MAR experience is occurring, whereas object detection is more concerned about detecting objects in every frame or when they first appear [262].

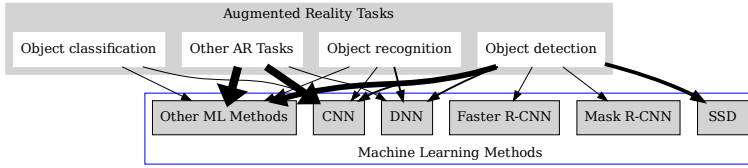


Fig. 3. Mapping of AR tasks found in the works of Table 2 to the ML methods used to fulfil the tasks. The width thickness of the arrow lines denotes the amount of works associated with the tasks-to-ML method.

Some ML-based approaches for object tracking are Deep Regression Networks [89], Recurrent YOLO (ROLO) [169], Multi-Domain Networks (MDNet) [164], Deep SORT [257], SiamMask [246], TrackR-CNN [242], Deep Regression Learning [70], Adversial Learning [217], Tractor++ [31], and Joint Detection and Embedding (JDE) [247]. As evident by some of their names, these object tracking methods are based on deep learning models which are used for object detection and object recognition. For example, ROLO combines the object detection utility of a YOLO network with a Long Short-Term Memory (LSTM) network for acquiring the trajectory of an object. The combined ROLO network is able to track in both spatial and temporal domains, while dealing with occlusion and motion blur effects. TrackR-CNN uses Mask R-CNN for initial object detection, and then applies 3D convolutions to incorporate temporal context from the input data streams to form the object tracker, as well as detection and tracking. Meanwhile, TrackR-CNN is also able to perform segmentation within the single convolutional network. These techniques are not typically found in works which contain object tracking for an MAR system. Rather, custom computer vision tracking algorithms are combined with deep learning-based object detection and recognition models to track objects across MAR scenes.

One such AR system which uses deep learning detection and a non-deep learning tracker is proposed by Ahmadyan *et al.* [4], where they use a CNN to initialise the pose of an object and then a planar tracker using sensor data to track the object in 9 degrees-of-freedom (DoF). The 3D tracking is performed on the client device, but utilises the mobile GPU to enhance the tracking performance and increase the rate the tracker is able to run at. Alternatively, a computer vision algorithm which is often used in MAR to map and track environments is SLAM, where sensor data from mobile devices such as gyroscopes and accelerometers are used to achieve reliable tracking [179]. Sernani *et al.* [212] use SLAM for a self-contained tourism AR application, where SLAM localisation allows the system to locate the user’s orientation and assists in the superposition of arrows and icons to guide user’s attention in an environment. On the other hand, Rambach *et al.* [193] investigate the usage of model-based tracking as an alternative to SLAM tracking as extracting real-world scale information and displaying object specific AR content may not always be possible with SLAM. Model-based tracking is achieved by using a predefined model or 3D reconstructed textured model of the tracking object, and then matching with a live view of the object during tracking to uncover the 6DOF pose of the object. Similar to previously mentioned MAR systems, they deploys the computation heavy object recognition and tracking tasks to a high performance edge server which is enabled with a GPU to reduce the total system latency, and the client is then free to solely perform data capturing and annotation rendering tasks.

Compared with purely deep learning-based object trackers, these latter systems fuse together a deep learning object detection routine with a non-deep learning tracking system. The usage of ML-based methods for tracking is less common when compared to object detection and recognition

as there is generally a lack of tracking training data suitable for the creation of deep learning tracking models [125]. However, exploratory works are now using recurrent neural network (RNN) for implicit learning of temporal dependencies during tracking from data.

5.3 ML for Adaptive UIs

In this section, we describe the ML methods applied to adaptive MAR UIs.

InfoF: The limitations of traditional information filtering methods are twofold [229], (1) the methods can be destructive and cause data loss for data visualisation, and (2) the methods may fail when the amount of data grows tremendously and overloads. To avoid these limitations, using ML methods in layout generation and data clustering is the recent trend. The advantages of ML methods are summarised by Sahu *et al.* [207], which is more efficient and effective rendering with ML-based layout creation, and secondly, better performance of transforming data into clustering-friendly nonlinear representations with deep learning-based data clustering. The information filtering, clustering, and rendering stages in MAR could be integrated as a context-aware recommendation system, with the ML methods for generating relevant recommendations for users based on the contextual situation of them. Recent works apply deep learning architectures for collaborative filtering in recommendation systems [34]. Their results show strong improvements in both prediction and recommendation of information for users as a result of incorporating context-awareness, user customisation, expert knowledge, and flexibility into recommendation systems [60]. Context-awareness modules in AR can utilise the 5W1H context model to determine what kind of contextual information to be adopted [94]. Adopting necessary context information within ML models and subsequently deciding which models to deploy, could reflect the logical thinking of the user, this is then a critical issue in AR recommendation systems. Jacucci *et al.* [103] present a recommendation system for a mixed reality-based urban exploration application. They consider using user profiles, and weightings from both crowd-sourced ratings and personal preferences. Their ML technique for personalised search and recommendations is based on two main components, (1) a data model that defines the representation of multiple data sources (e.g., content, social, and personal) as a set of overlaid graphs, and (2) a relevance-estimation model that performs random walks with restarts on the graph overlays and computes a relevance score for information items. However, their study does not present any experimental results or performance data. Zhang *et al.* [268] present an aggregated random walk algorithm incorporating personal preferences, location information, and temporal information in a layered graph to improve recommendations in MAR ecosystems. Readme [42] is a real-time recommendation system for MAR with an online algorithm and by comparing this system with the following baselines: (i) k-Nearest Neighbours (kNN) which is based purely on physical distance, and (ii) focus-based recommendation which infers user's interests based on their viewpoint and attempts to find the most similar objects within the current focus. The evaluation results show that ReadMe outperforms these baselines. Alternatively, contextual sequence modelling with RNN is another method to produce recommendation systems in MAR [215].

RegEA: Recent works show that applying ML algorithms in safety-critical AR systems could improve the tracking accuracy and lead to minimisations in the registration error. To adapt to real-time registration errors, safety-critical AR systems should provide some hints on the user interface to inform users of any potential harmful or dangerous situations. U-Net [206] is a modified full CNN initially built for precise medical image segmentation. Brunet *et al.* [39] propose a physics-based DNN with U-Net, for providing AR visualisation during hepatic surgery. Physics-based DNN solves the deformed state of the organ by using only sparse partial surface displacement data and achieves a similar accuracy as a finite element method solution. They achieve the registration in only 3 ms with a mean target registration error (TRE) of 2.9 mm in the ex vivo liver AR surgery. Their

physics-based DNN is $500\times$ faster than a reference finite element method solution. For the RegEA visualisation, this work obtains the surface data of the liver with an RGB-D camera and ground truth data acquired at different stages of deformation using a CT scan. Markers were embedded in the liver to compute TRE. By knowing the TRE values, this work can highlight the region around the 3D model of the virtual liver to alert the user for the error minimisation.

Occlusion Representation and Depth Cues: Traditional methods for calculating the depth of a scene are ineffective in recovering both depth and scene information whenever perceptual cues are used. This is due to the heavy and under-constrained nature of the problem and the methods themselves, as well as the use of limited cues. The advantage of ML methods is their ability to exploit almost all cues simultaneously, thus offering better depth inference, as well as the capability to estimate depth information from RGBE images in a similar concurrent manner for the completion of other tasks, such as, object detection, tracking, and pose estimation [207]. Park *et al.* [173] present a comprehensive smart task assistance system for wearable AR, which adopts Mask R-CNN for several adaptive UI techniques. Their work supports occlusion representation calculation, registration error adaption, and adaptive content placement. Alternatively, Tang *et al.* [227] present GrabAR, a system using a custom compact DNN for generating occlusion masks to support real-time grabbing of virtual objects in AR. The model is able to calculate and segment the hand to provide visually plausible interactions of objects in the virtual AR environment.

IEsti: The major limitation of classical methods for illumination estimation is the low accuracy problem due to insufficient contextual information, especially the methods using auxiliary information, the methods are computationally expensive and not scalable, which disturbs real-time performance of AR scene rendering. ML-based illumination estimation resolves the low accuracy and performance issue in AR, which does not require definitive models of the geometric or photometric attributes [207]. DNN approaches are based on the assumption that prior information about lighting can be learned from a large dataset of images with known light sources. Large datasets of panoramas are used to train an illumination predictor, they use the concept of finding similarities between an input image and one of the projections of individual panoramas [114]. Gardner *et al.* [71] present an illumination estimation solution by calculating diffuse lighting in the form of an omnidirectional image using a neural network. DeepLight [110] is a CNN-based illumination estimation for calculating a dominant light direction from a single RGB-D image. DeepLight uses a CNN to encode a relation between the input image and a dominant light direction. DeepLight applies outlier removal and temporal filtering to achieve temporal coherence of light source estimation. The results of DeepLight are evaluated with a comparison to the Gardner *et al.* [71] approach. DeepLight achieves a higher accuracy for the estimated light direction than the approach of Gardner *et al.* [71] by calculating the angular error to the ground-truth light direction and by rendering virtual objects in AR using an estimated light source. LightNet [166] is the latest illumination estimation using a dense network (DenseNet) architecture, which trained with two softmax outputs for colour temperature and lighting direction prediction.

6 RESEARCH CHALLENGES AND FURTHER DIRECTIONS

We discuss some prominent challenges for MAR, including seamless frameworks for interactable immersive environments, audio AR, user interactions, user interactions with environments, and federated learning. It is crucial to identify and analyse these challenges and seek for novel theoretical and technical solutions.

6.1 Towards seamless frameworks for interactable immersive environments

Existing frameworks and SDKs offer various features to understand physical environments. *ARKit* remarkably achieves a broad coverage of feature recognition, including point clouds, anchors, light

estimation, and occlusion. In general, the AR content annotations in existing frameworks already go beyond simple overlays in mid-air. Feature recognition enables these augmented annotations to become a part of the physical world, albeit users can distinguish unnatural positioning and some rendering issues. Interactive AR on human faces, such as with Banuba and XZIMG, demonstrates a real-life use case that requires AR overlays to merge with the physical world. We see an obvious trend for the merging of AR augmentations in the physical world for different application scenarios, e.g., for medical surgery [39]. One potential research direction could be to extend the different sensor types and employ multi-sensor information to construct user context intelligently.

Alternatively, ML approaches can serve as an automated tool to automatically generate lively and animated digital entities [92] as well as imitating real world creatures [211, 228]. In addition to this, the collection of ML-generated objects can potentially become a library of animated objects for AR. However, “*how to merge the digital overlaid content (i.e., AR content) interactions with the physical world (e.g., everyday objects and humans)*” are in the nascent stage. The merged AR overlays will eventually co-exist with both human users and their surroundings. This leaves us with an unexplored research challenge of designing novel spatial environments. Sample research directions for the next generation of MAR frameworks could be: (1) user issues such as immersive spatial environments and their relationship and interaction among human users, virtual entities, and tangible objects; (2) AI becomes a key actor in spatial environments and how objects (or agents) take actions in an automated fashion; and (3) how to evaluate the interaction of AI-supported animated objects with physical surroundings (i.e., human users and physical objects) without active participation of designers and software engineers (analogous to the concept of *evaluation probe* [147]).

6.2 Audio augmented reality

While the majority of existing frameworks have made significant efforts in visual-based AR cues, another neglected and important aspect is audio-based interaction. Audio can be regarded as an ineluctable design material for audio-based interactions, where individualisation, context-awareness, and diversification are the primary strategies of designing such audio-based interaction [223]. Among three strategies, context awareness (i.e., voices to be tailored to the user’s context) aligns with the primary aspect of AR. In addition, audio’s sociophonetic aspect should reflect the user’s social quality. For example, the scenario of paying a credit card bill requires audio that brings the user an impression of trustworthiness, honesty, and efficiency [223]. Audio AR will be an interesting topic in examining the alignment of audio and the properties of virtual visuals (e.g., appearance and other contexts) in such audio-driven AR.

6.3 User interaction and user workload

Section 4.1 summarises 37 MAR frameworks and SDKs. Generally, these frameworks have not prioritised natural user interfaces (NUIs) [133] and very limited numbers of existing frameworks and SDKs support on-body interfaces such as hand tracking, facial tracking, and body tracking. Although such feature tracking on user bodies enables alternative input modals on AR devices, especially head-worn AR headsets, such discussed frameworks and SDKs do not offer sufficient considerations into the user workload across various types of user interaction tasks, ranging from clicking a button on a 2D menu to manipulating a 3D sphere with rotation movements. More importantly, there exists no integration of user workload measurements into existing frameworks.

As natural user interfaces rely on body movements, user interaction with such interfaces cannot be prolonged, due to user fatigue [133, 134]. It is necessary to include computational approaches to understand how the user interacts with AR interfaces to reduce user workload (i.e., collecting user interaction traces). The user interaction trace will be collected for the foundation of data-driven user

interaction design, which further offers systematic evaluation modules inside MAR frameworks and SDKs to optimise the user interaction design. The systematic evaluation modules should also acquire the ability to assist the iterative design process that allows for natural user interfaces to improve the user workloads with constantly changing AR environments.

6.4 Users and their environments

As mentioned above, with user interaction and the potential architecture of collecting data user traces, as these facilities are becoming mature, such frameworks can further consider the user interaction with their tangible environments. This then supports hyper-personalisation of user experiences, i.e., *letting audiences shape their user interaction designs*. Users with AR devices, e.g., smartphones (as-is) and smartglasses (to-be), encounter digital entities that are highly customised to the user preference. It is worthwhile mentioning that digital entities will merge with the nearby tangible environments and user interaction in AR should work on both physical and digital worlds.

With such premises, the next generation of AR frameworks will encounter two primary challenges. On the one hand, the tangible environments are not static objects. The information and property registration of such objects will not remain static forever, and any changes in such objects will largely impact the user experience. On the other hand, we cannot guarantee that users will act in the same identical way when interacting with AR augmentations on tangible environments, as initially planned by software engineers and product managers when they designed and implemented the AR physical environments, which presumably considered the world to be a static property.

The second challenge is more challenging to be addressed compared to the first one, because the first will eventually be solved by the enhanced sensing capability of our tangible environments and the crowd-based mechanisms that can constantly monitor and reflect the changes in the tangible environments. Regarding the second challenge, we cannot blame such user misalignment between the real user behaviours and the intended design as human error. We have the following example, derived from a classic design example of “*the desired path*”, to reinforce the above statement. In a school campus, a user is finding a way to reach a certain building, and there exists two official paths (left or right) and a lawn in front of the user. The user with an AR device receives navigational information of either ‘going onto the path on the left-hand side’ or ‘going along the path on the right-hand side’. Such navigation information is contextualised with the school campus map and registered in the design and implementation phases. However, due to the incentive of the shortest path, the user eventually takes the diagonal (unofficial) path across the lawn, instead of taking any of the official paths. Meanwhile, the AR system keeps reminding the user about the erroneous path being taken, according to such pre-defined AR information (referring to the latter challenge). In addition, if such a diagonal path has been repetitively taken, the tangible environment has changed, but the updates of AR information lags behind (referring to the first challenge).

Such uncertainties from dynamic tangible environments and user behaviours give the motivation to building adaptive information management in AR frameworks. Although we have no definitive solutions to address the above challenges at the time being, we identify a research gap in fulfilling the orientation towards user interaction and their environments in adaptive and personalised manners.

6.5 Deepening the understanding of users’ situations

Nowadays, the frameworks and SDKs provide well-established connections to common sensors such as cameras, inertial measurement units (IMUs), and GPS sensors. The sensing capability allows AR frameworks or systems to detect objects and people within the user’s in-situ environments, regardless of marker-based (Fig. 2) or markerless approaches. Accordingly, the AR framework makes timely responses and offers user feedback (visual, audio, and haptics), e.g., 3D images popping

up on the user's screen. Although research prototypes attempt to manage the user feedback by leveraging the user's geographical information [258], many existing applications are limited by a moderate level of tracking precision due to GPS.

When AR frameworks are being deployed on a large scale, we foresee the number of candidate AR feedback in one specific location (no less than $5 \times 5 \text{ m}^2$) will surge exponentially. In such a way, AR frameworks may encounter a bottleneck in delivering optimised information flow (i.e., concise and context-aware visuals) on AR devices. In other words, such geographical location-based tracking requires auxiliary channels (e.g., camera-based detection) to get a more precise and detailed understanding of user's environments that guarantee the performance of adaptive user interfaces [129]. In return for improved detection and tracking precision, the framework will take additional overheads. However, the camera-based detection and tracking generate offloading tasks that demand high network traffic, and hence produces significant latency from the distant cloud-based servers. As AR users will be impacted significantly by degraded QoE, we have to restructure the client-server architectures in AR frameworks and SDKs.

6.6 Privacy preserving MAR frameworks

Multi-user collaboration has been rarely considered by existing frameworks. One key reason is that such remote collaborations may introduce privacy risks. Sharing user information in effective and just-in-need manners will be a grand challenge. We summarise several challenges to address potential privacy leakage in multi-user AR scenarios, as follows. First, the next generation MAR frameworks will need a content-sharing module to expose any sensitive information among multiple users across various applications. One remarkable example is the sharing of user locations in physical space, in which multiple users work on the same AR spatial environment. Some users may own their preferred privacy policies and hence will threaten such location-based user collaborations. Second, privacy policies from multiple users can conflict with each other. In other words, a user does not realise that one's action could deteriorate the privacy policy of another user in AR, primarily subject to the spatio-temporal relationship. Third, user interaction with AR objects could pose privacy leakage due to its historical record, namely user interaction trace, especially when we consider AR objects as some public properties in this emerging spatial environments. For example, a user may establish some virtual stickers and share to target users in an AR spatial environment. However, different users may own asynchronous views and hence unparalleled information due to their privacy preference, and hence they work differently on the virtual stickers due to different beliefs. Thus, the effectiveness of preventing privacy leakage highly depends on not only the user-centric and user-interdependent semantics on the application level, but also a multi-user framework level. It is important to strengthen the information-theoretic dimension in AR frameworks.

6.7 Federated learning for MAR

The traditional cloud-based MAR frameworks distribute the centralised trained models to AR clients, which may require prohibitively large amounts of time when models are often updated. Federated learning is an emerging ML paradigm to train local models in a decentralised manner and can cooperate with any ML method, which supports personalised models for users and significantly reduces the communication latency. Instead of centralising the data to train a global model for improved local predictions, federated learning downloads the global model to capable devices, where models are locally updated. Updates are sent to a central server, which then aggregates the local updates into an updated global model [45].

For example, to address rigorous network constraints of MAR systems, most existing MAR solutions focus on viewport adaptive schemes, which segments a video into multiple tiles. The system assesses the client's viewport position in the video and predicts the most probable future

viewport positions. The server then only transmits the tiles located in the field of view of the user and the tiles at the predicted viewport locations. A federated learning approach can train a local viewport prediction model and only transmits the model updates to the server, and such approaches that perform viewport prediction on-device and requests the corresponding tiles from the server. Therefore, federated learning-based approach allows training the global model in a distributed fashion while enabling fine-grained personalisation through the personal model.

7 CONCLUSIONS

This survey employs a top-down approach to review MAR frameworks supporting user interfaces and applications. We begin with the discussion of MAR applications, and the content management and generation that support AR visualisation adaptive to the user's environments of high dynamics and mobility. The article then meticulously selects 37 MAR frameworks either available in industry or proposed by researchers in academia, and further reviews and compares their functions and features. Additionally, the article covers the MML methods in the domain of MAR and provides a trendy analysis of how the methods support MAR becoming sensitive to user contexts and physical environments. We discussed the major challenges and unexplored topics of designing seamless yet user-centric MAR frameworks, potentially empowered by ML methods. Finally, this article hopes to provide a broader discussion within the community, and intends to invite researchers and practitioners, especially in the fields relevant to MAR, to shape the future of AR.

REFERENCES

- [1] A-Frame. 2020. A-Frame. <https://aframe.io/>
- [2] N. Ab Aziz, K. Ab Aziz, A. Paul, A. Yusof, et al. 2012. Providing augmented reality based education for students with attention deficit hyperactive disorder via cloud computing: Its advantages. In *ICACT*. IEEE, 577–581.
- [3] O. Abiodun, A. Jantan, A. Omolara, K. Dada, et al. 2018. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 4, 11 (2018), e00938.
- [4] A. Ahmadyan, T. Hou, J. Wei, L. Zhang, et al. 2020. Instant 3D object tracking with applications in augmented reality. *arXiv preprint arXiv:2006.13194* (2020).
- [5] K. Ahuja, S. Pareddy, R. Xiao, M. Goel, et al. 2019. Lightanchors: Appropriating point lights for spatially-anchored augmented reality interfaces. In *Proc. UIST*. 189–196.
- [6] A. Ajanki, M. Billinghurst, H. Gamper, T. Järvenpää, et al. 2011. An augmented reality interface to contextual information. *Virtual reality* 15, 2-3 (2011), 161–173.
- [7] O. Akgul, H. Penekli, and Y. Genc. 2016. Applying deep learning in augmented reality tracking. In *SITIS*. IEEE, 47–54.
- [8] S. Albawi, T. Mohammed, and S. Al-Zawi. 2017. Understanding of a convolutional neural network. In *ICET*. Ieee, 1–6.
- [9] Amazon Web Services, Inc. 2020. Amazon Sumerian Overview. <https://aws.amazon.com/sumerian/>
- [10] P. Amirian and A. Basiri. 2016. Landmark-based pedestrian navigation using augmented reality and machine learning. In *Progress in Cartography*. Springer, 451–465.
- [11] I. Ang and K. Lim. 2019. Enhancing STEM education using augmented reality and machine learning. In *ICSCC*. IEEE, 1–5.
- [12] K. Apicharttrisorin, X. Ran, J. Chen, S. Krishnamurthy, et al. 2019. Frugal following: Power thrifty object detection and tracking for mobile augmented reality. In *Proc. SenSys*. 96–109.
- [13] Aplicaciones de la Visión Artificial. 2020. ARUco. <https://www.uco.es/investiga/grupos/ava/node/26>
- [14] Apple Inc. 2020. ARKit. <https://developer.apple.com/documentation/arkit>
- [15] Apple Inc. 2020. Augmented Reality. <https://developer.apple.com/augmented-reality/>
- [16] Apple Inc. 2020. iOS 14 - Apple. <https://www.apple.com/ios/ios-14/>
- [17] R. Aramaki and M. Murakami. 2013. Investigating Appropriate Spatial Relationship between User and Ar Character Agent for Communication Using AR WoZ System. In *Proc. ICMI*. ACM, 397–404.
- [18] M. Arango, L. Bahler, P. Bates, M. Cochinwala, et al. 1993. The Touring Machine System. *Commun. ACM* 36, 1 (January 1993), 69–77.
- [19] Y. Arifin, T. Galih Sastria, and E. Barlian. 2018. User Experience Metric for Augmented Reality Application: A Review. *Proc. ICCSCI* 135 (2018), 648 – 656.
- [20] Augment. 2020. 3D and augmented reality product visualization platform | Augment. <https://www.augment.com/>
- [21] Augmented Pixels. 2020. Augmented Pixels. <http://augmentedpixels.com/>

- [22] AugmentedPro. 2020. AugmentedPro. <https://www.augmentedpro.com/>
- [23] R. Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 4 (1997), 355–385.
- [24] R. Azuma and C. Furmanski. 2003. Evaluating Label Placement for Augmented Reality View Management. In *Proc. ISMAR*. IEEE, 66.
- [25] A. Badiás, S. Curtit, D. González, I. Alfaro, et al. 2019. An augmented reality platform for interactive aerodynamic design and analysis. *Internat. J. Numer. Methods Engrg.* 120, 1 (2019), 125–138.
- [26] T. Bailey and H. Durrant-Whyte. 2006. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics & Automation Magazine* 13, 3 (2006), 108–117.
- [27] Banuba. 2020. Banuba Augmented Reality SDK. <https://www.banuba.com/>
- [28] R. Baruah. 2021. AR and VR Using the WebXR API. (2021).
- [29] B. Bell, S. Feiner, and T. Höllerer. 2001. View management for virtual and augmented reality. In *Proc. UIST*. ACM, 101–110.
- [30] F. Bergamasco, A. Albarelli, L. Cosmo, E. Rodola, et al. 2016. An accurate and robust artificial marker based on cyclic codes. *IEEE TPAMI* 38, 12 (2016), 2359–2373.
- [31] P. Bergmann, T. Meinhardt, and L. Leal-Taixe. 2019. Tracking without bells and whistles. In *Proc. ICCV*. 941–951.
- [32] S. Bernhardt, S. Nicolau, L. Soler, and C. Doignon. 2017. The status of augmented reality in laparoscopic surgery as of 2016. *Medical image analysis* 37 (2017), 66–90.
- [33] Blippar. 2020. Augmented Reality (AR) & Computer Vision Company | Blippar. <https://www.blippar.com/>
- [34] J. Bobadilla, S. Alonso, and A. Hern. 2020. Deep Learning Architecture for Collaborative Filtering Recommender Systems. *Applied Sciences* 10, 7 (2020), 2441.
- [35] B. Boom, S. Orts-Escolano, X. Ning, S. McDonagh, et al. 2013. Point Light Source Estimation based on Scenes Recorded by a RGB-D camera.. In *BMVC*.
- [36] B. Boom, S. Orts-Escolano, X. Ning, S. McDonagh, et al. 2017. Interactive light source position estimation for augmented reality with an RGB-D camera. *Computer Animation and Virtual Worlds* 28, 1 (2017), e1686.
- [37] A. Bornik, R. er, B. Reitingner, G. Gotschuli, et al. 2003. Computer-aided liver surgery planning: an augmented reality approach. In *Medical Imaging 2003*, Vol. 5029. International Society for Optics and Photonics, 395–406.
- [38] C. Bowles and M. Wu. 2010. Transparent Collaboration: Letting Users Simulate Another User’s World. In *Proc. CHIMIT*. ACM, Article 7, 4 pages.
- [39] JN. Brunet, A. Mendizabal, A. Petit, N Golse, et al. 2019. In *International Conference on Medical Image computing and computer-assisted intervention*. Springer, 137–145.
- [40] J. Carmigniani, M. Furht, B. fand Anisetti, P. Ceravolo, et al. 2011. Augmented reality technologies, systems and applications. *Multimedia tools and applications* 51, 1 (2011), 341–377.
- [41] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. 2017. Mobile augmented reality survey: From where we are to where we go. *IEEE Access* 5 (2017), 6917–6950.
- [42] D. Chatzopoulos and P. Hui. 2016. Readme: A real-time recommendation system for mobile augmented reality ecosystems. In *Proc. MM*. 312–316.
- [43] C. Chen, W. Chen, J. Peng, B. Cheng, et al. 2017. A real-time markerless augmented reality framework based on SLAM technique. In *ISPAN-FCST-ISCC*. IEEE, 127–132.
- [44] D. Chen, C. Tector, A. Br, H. Chen, et al. 1994. Case study: Observing a volume rendered fetus within a pregnant patient. In *Proc. VIS*. IEEE Computer Society Press, 364–368.
- [45] D. Chen, L. Xie, B. Kim, L. Wang, et al. 2020. Federated learning based mobile edge computing for augmented reality applications. In *ICNC*. IEEE, 767–773.
- [46] Q. Cheng, S. Zhang, S. Bo, D. Chen, et al. 2020. Augmented Reality Dynamic Image Recognition Technology Based on Deep Learning Algorithm. *IEEE Access* 8 (2020), 137370–137384.
- [47] M. Cidota, S. Lukosch, D. Datcu, and H. Lukosch. 2016. Workspace Awareness in Collaborative AR Using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proc. AHS*. ACM, Article 3, 8 pages.
- [48] CraftAR. 2020. Highest-Precision Image Recognition Solution. <https://catchoom.com/image-recognition-craftar/>
- [49] CraftAR. 2020. Marker-Based AR Solution. <https://catchoom.com/augmented-reality-craftar/>
- [50] E. Cruz, S. Orts-Escolano, F. Gomez-Donoso, C. Rizo, et al. 2019. An augmented reality application for improving shopping experience in large retail stores. *Virtual Reality* 23, 3 (2019), 281–291.
- [51] C. Cuevas, D. Berjón, F. Morán, and N. García. 2012. Moving object detection for real-time augmented reality applications in a GPGPU. *IEEE Transactions on Consumer Electronics* 58, 1 (2012), 117–125.
- [52] DAQRI. 2017. Read me for ARToolkit for Android. [https://github.com/artoolkit/artoolkit5/blob/master/README%20\(ARToolkit%20for%20Android\).md](https://github.com/artoolkit/artoolkit5/blob/master/README%20(ARToolkit%20for%20Android).md)
- [53] DAQRI. 2020. ARToolkit. <https://github.com/artoolkit>
- [54] M. de Sa, E. Churchill, and K. Isbister. 2011. Mobile Augmented Reality: Design Issues and Opportunities. In *Proc. MobileHCI*. ACM, 749–752.

- [55] P. Debevec. 2008. Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-Based Graphics with Global Illumination and High Dynamic Range Photography. In *SIGGRAPH*. ACM, Article 32, 10 pages.
- [56] Digi-Capital. 2019. For AR/VR 2.0 to live, AR/VR 1.0 must die. <https://www.digi-capital.com/news/2019/01/for-ar-vr-2-0-to-live-ar-vr-1-0-must-die/>
- [57] S. Drouin, D. Collins, and M. Kersten-Oertel. 2018. 7 Interaction in Augmented Reality Image-Guided Surgery. In *Mixed and Augmented Reality in Medicine*. CRC Press, 99–114.
- [58] R. Du, E. Turner, M. Dzitsiuk, L. Prasso, et al. 2020. DepthLab: Real-Time 3D Interaction With Depth Maps for Mobile Augmented Reality. In *Proc. UIST*. 829–843.
- [59] D. Dwyer, C. Ringstaff, and J. S. 1991. Changes in teachers' beliefs and practices in technology-rich classrooms. *Educational leadership* 48, 8 (1991), 45–52.
- [60] T. Ebesu and Y. Fang. 2017. Neural citation network for context-aware citation recommendation. In *Proc. SIGIR*. 1093–1096.
- [61] C. Elvezio, M. Sukan, O. Oda, S. Feiner, et al. 2017. Remote Collaboration in AR and VR Using Virtual Replicas. In *SIGGRAPH*. ACM, Article 13, 2 pages.
- [62] J. Etienne. 2020. AR.js. <https://github.com/jeromeetienne/AR.js/blob/master/README.md>
- [63] V. Ferrari, T. Tuytelaars, and L. Van Gool. 2001. Markerless augmented reality with a real-time affine region tracker. In *Proc. ISAR*. IEEE, 87–96.
- [64] M. Fiala. 2005. ARTag, a fiducial marker system using digital techniques. In *CVPR*, Vol. 2. IEEE, 590–596.
- [65] V. Fragoso, S. Gauglitz, S. Zamora, J. Kleban, et al. 2011. TranslatAR: A mobile augmented reality translator. In *WACV*. IEEE, 497–502.
- [66] H. Fuchs, M. Livingston, R. Raskar, K. Keller, et al. 1998. Augmented reality visualization for laparoscopic surgery. In *MICCAI*. Springer, 934–943.
- [67] M. Funk, T. Kosch, S. Greenwald, and A. Schmidt. 2015. A Benchmark for Interactive Augmented Reality Instructions for Assembly Tasks. In *Proc. MUM*. ACM, 253–257.
- [68] G. Kim. 2010. Tuscany+: Augmented Reality Tourism App. <https://mobilemarketingandtechnology.com/2010/05/07/tuscany-augmented-reality-tourism-app/>
- [69] G. Kim. 2011. Augmented Reality for Basel. <https://www.perey.com/AugmentedRealityForBasel/>
- [70] J. Gao, T. Zhang, X. Yang, and C. Xu. 2018. P2t: Part-to-target tracking via deep regression learning. *IEEE Transactions on Image Processing* 27, 6 (2018), 3074–3086.
- [71] M. Gardner, K. Sunkavalli, E. Yumer, X. Shen, et al. 2017. Learning to predict indoor illumination from a single image. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 1–14.
- [72] C. Gartenberg. 2021. TikTok debuts first augmented reality filter that uses iPhone 12 Pro's LIDAR camera. <https://www.theverge.com/2021/1/7/22217465/tiktok-augmented-reality-filter-iphone-12-pros-lidar-camera>
- [73] Y. Ghazwani and S. Smith. 2020. Interaction in Augmented Reality: Challenges to Enhance User Experience. In *Proc. ICVARS*. ACM, 39–44.
- [74] N. Ghouaiel, J. Cieutat, and J. Jessel. 2014. Adaptive Augmented Reality: Plasticity of Augmentations. In *Proc. VRIC*. Article 10, 4 pages.
- [75] J. Gibby, S. Swenson, S. Cvetko, R. Rao, et al. 2019. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. *IJCARS* 14, 3 (2019), 525–535.
- [76] R. Girshick, J. Donahue, T. Darrell, and J. Malik. 2015. Region-based convolutional networks for accurate object detection and segmentation. *IEEE TPAMI* 38, 1 (2015), 142–158.
- [77] Google Developers. 2020. ARCore. <https://developers.google.com/ar>
- [78] Google Developers. 2020. ARCore supported devices. https://developers.google.com/ar/discover/supported-devices#android_play
- [79] Google Developers. 2020. Get Android 11 | Android Developers. https://developer.android.com/about/versions/11/get#on_pixel
- [80] L. Gruber, T. Richter-Trummer, and D. Schmalstieg. 2012. Real-time photometric registration from arbitrary geometry. In *ISMAR*. IEEE, 119–128.
- [81] L. Gruber, J. Ventura, and D. Schmalstieg. 2015. Image-space illumination for augmented reality in dynamic environments. In *VR*. IEEE, 127–134.
- [82] D. Han, T. Jung, and A. Gibson. 2013. Dublin AR: implementing augmented reality in tourism. In *IFITT*. Springer, 511–523.
- [83] N. Hashimoto and D. Kobayashi. 2016. Dynamic spatial augmented reality with a single IR camera. In *SIGGRAPH*. 1–1.
- [84] Y. Hbali, M. Sadgal, and A. El Fazziki. 2013. Object detection based on HOG features: Faces and dual-eyes augmented reality. In *WCCIT*. IEEE, 1–4.
- [85] K. He, G. Gkioxari, P. Dollár, and R. Girshick. 2017. Mask r-cnn. In *Proc. ICCV*. 2961–2969.

- [86] K. He, X. Zhang, S. Ren, and J. Sun. 2015. Deep residual learning for image recognition. arXiv 2015. *arXiv preprint arXiv:1512.03385* (2015).
- [87] Y. He, J. Ren, G. Yu, and Y. ing. 2020. Optimizing the learning accuracy in mobile augmented reality systems with CNN. In *ICC. IEEE*, 1–6.
- [88] S. Hegde, J. Maurya, A. Kalkar, and R. Hebbalaguppe. 2020. SmartOverlays: A Visual Saliency Driven Label Placement for Intelligent Human-Computer Interfaces. In *Proc. WACV*. 1121–1130.
- [89] D. Held, S. Thrun, and S. Savarese. 2016. Learning to track at 100 fps with deep regression networks. In *ECCV*. Springer, 749–765.
- [90] HERE. 2020. Build apps with HERE Maps API and SDK Platform Access | HERE Developer. <https://developer.here.com/>
- [91] F. Herpich, R. Guarese, and L. Tarouco. 2017. A comparative analysis of augmented reality frameworks aimed at the development of educational applications. *Creative Education* 8, 09 (2017), 1433.
- [92] D. Holden, B. Duong, S. Datta, and D. Nowrouzezahrai. 2019. Subspace Neural Physics: Fast Data-Driven Interactive Simulation. In *Proc. SCA. ACM*, Article 6, 12 pages.
- [93] K. Holstein, B. McLaren, and V. Aleven. 2019. Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. In *AIED*. Springer, 157–171.
- [94] D. Hong, H. Schmidtke, and W. Woo. 2007. Linking context modelling and contextual reasoning. In *MRC*. Roskilde University, 37–48.
- [95] F. Hoyle, H. Dimakuta, K. Cabonita, V. dela Peña, et al. [n.d.]. Nervo: Augmented Reality Mobile Application for the Science Education of Central and Peripheral Nervous Systems. In *HNICEM. IEEE*, 1–7.
- [96] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger. 2017. Densely connected convolutional networks. In *Proc. CVPR*. 4700–4708.
- [97] S. Huang, Y. Yang, and C. Chu. 2012. Human-centric design personalization of 3D glasses frame in markerless augmented reality. *Advanced Engineering Informatics* 26, 1 (2012), 35–45.
- [98] I Love IceCream Ltd. 2020. Augmented Reality SDK. <https://www.deepar.ai/augmented-reality-sdk>
- [99] Apple Inc. 2020. Apple unveils new iPad Pro with LiDAR Scanner and trackpad support in iPadOS. <https://www.apple.com/newsroom/2020/03/apple-unveils-new-ipad-pro-with-lidar-scanner-and-trackpad-support-in-ipados/>
- [100] Inglobe Technologies. 2020. AR-media™ Developer Portal. <https://dev.inglobetechnologies.com/features/>
- [101] Mansoor Iqbal. 2019. Pokémon GO Revenue and Usage Statistics (2019). <https://www.businessofapps.com/data/pokemon-go-statistics/>
- [102] S. Irshad and D. Awang. 2018. A UX Oriented Evaluation Approach for Mobile Augmented Reality Applications. In *Proc. MoMM. ACM*, 108–112.
- [103] G. Jacucci, S. Andolina, D. Kalkhofen, D. Schmalstieg, et al. 2018. Combining intelligent recommendation and mixed reality in itineraries for urban exploration. *CreMedia* 2017/2 (2018), 18–23.
- [104] J. Jia, S. Elezovikj, H. Fan, S. Yang, et al. 2020. Semantic-aware label placement for augmented reality in street view. *The Visual Computer* (2020), 1–15.
- [105] S. Jiddi, P. Robert, and E. March. 2016. Reflectance and illumination estimation for realistic augmentations of real scenes. In *ISMAR. IEEE*, 244–249.
- [106] R. Jikadra, H. Shah, and P. Kulkarni. 2019. Video Calling with Augmented Reality Using WebRTC API. In *INDIACom. IEEE*, 363–367.
- [107] S. Julier, M. Lanzagorta, Y. Baillot, L. Rosenblum, et al. 2000. Information filtering for mobile augmented reality. In *Proc. ISAR. IEEE*, 3–11.
- [108] S. Julier, M. Livingston, I. Swan, and D. Brown. 2003. Adaptive user interfaces in augmented reality. In *STARS*.
- [109] A. Kamarainen, S. Metcalf, T. Grotzer, A. Browne, et al. 2013. EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers & Education* 68 (2013), 545–556.
- [110] P. Kán and H. Kafumann. 2019. Deeplight: light source estimation for augmented reality using deep learning. *The Visual Computer* 35, 6 (2019), 873–883.
- [111] P. Kán and H. Kaufmann. 2013. Differential irradiance caching for fast high-quality light transport between virtual and real worlds. In *ISMAR. IEEE*, 133–141.
- [112] T. Kan, C. Teng, and W. Chou. 2009. Applying QR code in augmented reality applications. In *Proc. VRCAI*. 253–257.
- [113] K. Kanno, E. Lamounier, A. Cardoso, E. re, et al. 2018. Augmented Reality System for Aiding Mild Alzheimer Patients and Caregivers. In *VR&3DUI. IEEE*, 593–594.
- [114] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, et al. 2014. Automatic scene inference for 3d object compositing. *ACM Transactions on Graphics (TOG)* 33, 3 (2014), 1–15.
- [115] M. Kasper, N. Keivan, G. Sibley, and C. Heckman. 2017. Light source estimation with analytical path-tracing. *arXiv preprint arXiv:1701.04101* (2017).
- [116] H. Kato and M. Billinghurst. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proc. IWAR. IEEE*, 85–94.

- [117] H. Kaufmann and D. Schmalstieg. 2002. Mathematics and geometry education with collaborative augmented reality. In *SIGGRAPH*. ACM, 37–41.
- [118] M. Kersten-Oertel, P. Jannin, and D. Collins. 2013. The state of the art of visualization in mixed reality image guided surgery. *Computerized Medical Imaging and Graphics* 37, 2 (2013), 98–112.
- [119] A. Khan and S. Khusro. 2015. The rise of augmented reality browsers: Trends, challenges and opportunities. *Pakistan Journal of Science* 67, 3 (2015).
- [120] H. Kim and T. Nam. 2015. Augmented Miniature Prototyping Toolkit for UX in Interactive Space. In *Proc. CHI EA*. ACM, 2229–2234.
- [121] S. Kim, G. Lee, W. Huang, H. Kim, et al. 2019. Evaluating the Combination of Visual Communication Cues for HMD-Based Mixed Reality Remote Collaboration. In *Proc. CHI*. ACM, 1–13.
- [122] S. Kim, H. Suk, J. Kang, J. Jung, et al. 2014. Using Unity 3D to facilitate mobile augmented reality game development. In *WF-IoT*. IEEE, 21–26.
- [123] K. Kiyokawa, M. Billingham, B. Campbell, and E. Woods. 2003. An Occlusion-Capable Optical See-through Head Mount Display for Supporting Co-Located Collaboration. In *Proc. ISMAR*. IEEE, 133.
- [124] C. Kounavis, A. Kasimati, and E. Zamani. 2012. Enhancing the tourism experience through mobile augmented reality: Challenges and prospects. *IJEBM* 4 (2012), 10.
- [125] S. Krebs, B. Duraisamy, and F. Flohr. 2017. A survey on leveraging deep neural networks for object tracking. In *ITSC*. IEEE, 411–418.
- [126] W. Lages and D. Bowman. 2019. Walking with adaptive augmented reality workspaces: design and usage patterns.. In *IUI*. 356–366.
- [127] Z. Lai, W. Tao, M. Leu, and Z. Yin. 2020. Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *J. of Manufacturing Systems* 55 (2020), 69–81.
- [128] J. Lalonde. 2018. Deep learning for augmented reality. In *WIO*. IEEE, 1–3.
- [129] K. Lam, L. Lee, T. Braud, and P. Hui. 2019. M2A: A Framework for Visualizing Information from Mobile Web to Mobile Augmented Reality. In *PerCom*. IEEE, 1–10.
- [130] Lamb, Philip. 2004. ARToolkit. <http://www.hitl.washington.edu/artoolkit/>
- [131] Lamb, Philip. 2007. ARToolkit. <https://sourceforge.net/projects/artoolkit/files/>
- [132] Layan. 2020. Augmented Reality | Interactive Print | Layan. <https://www.layan.com/>
- [133] L. Lee and P. Hui. 2018. Interaction methods for smart glasses: A survey. *IEEE Access* 6 (2018), 28712–28732.
- [134] L. Lee, K. Lam, Y. Yau, T. Braud, et al. 2019. HIBEY: Hide the Keyboard in Augmented Reality. In *PerCom*. IEEE.
- [135] X. Li, Y. Tian, F. Zhang, S. Quan, et al. 2020. Object Detection in the Context of Mobile Augmented Reality. In *ISMAR*. IEEE, 156–163.
- [136] H. Liao, T. Inomata, I. Sakuma, and T. Dohi. 2010. 3-D augmented reality for MRI-guided surgery using integral videography autostereoscopic image overlay. *IEEE transactions on biomedical engineering* 57, 6 (2010), 1476–1486.
- [137] C. Lin, Y. Chung, B. Chou, H. Chen, et al. 2018. A novel campus navigation APP with augmented reality and deep learning. In *ICASI*. IEEE, 1075–1077.
- [138] T. Lin, H. Duh, N. Li, H. Wang, et al. 2013. An investigation of learners’ collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. *Computers & Education* 68 (2013), 314–321.
- [139] T. Lin, P. Goyal, R. Girshick, K. He, et al. 2017. Focal loss for dense object detection. In *Proc. ICCV*. 2980–2988.
- [140] D. Lindlbauer, A. Feit, and O. Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proc. UIST*. ACM, 147–160.
- [141] L. Liu, W. Ouyang, X. Wang, P. Fieguth, et al. 2020. Deep learning for generic object detection: A survey. *IJCV* 128, 2 (2020), 261–318.
- [142] Shanhong Liu. 2019. Augmented reality (AR) market size worldwide in 2017, 2018 and 2025 (in billion U.S. dollars). <https://www.statista.com/statistics/897587/world-augmented-reality-market-value/>
- [143] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, et al. 2016. Ssd: Single shot multibox detector. In *ECCV*. Springer, 21–37.
- [144] Y. Liu, Y. Zhang, S. Zuo, and W. Fu. 2018. BoatAR: A Multi-User Augmented-Reality Platform for Boat. In *Proc. VRST*. ACM, Article 74, 2 pages.
- [145] J. Lopez-Moreno, E. Garces, S. Hadap, E. Reinhard, et al. 2013. Multiple light source estimation in a single image. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 170–182.
- [146] S. Lu and Y. Liu. 2015. Integrating augmented reality technology to enhance children’s learning in marine education. *Environmental Education Research* 21, 4 (2015), 525–541.
- [147] A. Luusua, J. Ylipulli, M. Jurmu, H. Pihlajaniemi, et al. 2015. Evaluation Probes. In *Proc. CHI*. ACM, 85–94.
- [148] B. MacIntyre and E. Coelho. 2000. Adapting to dynamic registration errors using level of error (LOE) filtering. In *Proc. ISAR*. IEEE, 85–88.

- [149] B. MacIntyre, E. Coelho, and S. Julier. 2002. Estimating and adapting to registration errors in augmented reality systems. In *Proc. VR. IEEE*, 73–80.
- [150] Magic Leap, Inc. 2020. Developer Portal | Magic Leap. <https://developer.magicleap.com/en-us/learn/guides/developer-portal>
- [151] F. Mahmood, E. Mahmood, R. Dorfman, J. Mitchell, et al. 2018. Augmented reality and ultrasound education: initial experience. *Journal of cardiothoracic and vascular anesthesia* 32, 3 (2018), 1363–1367.
- [152] I. Marneanu, M. Ebner, and T. Roessler. 2014. Evaluation of Augmented Reality Frameworks for Android Development. *International Journal of Interactive Mobile Technologies (ijIM)* 8, 4 (2014), 37–44. <https://www.online-journals.org/index.php/i-jim/article/view/3974>
- [153] J. Martín-Gutiérrez, P. Fabiani, W. Benesova, M. a, et al. 2015. Augmented reality to promote collaborative and autonomous learning in higher education. *Computers in human behavior* 51 (2015), 752–761.
- [154] J. Martín-Gutiérrez, J. Saorín, M. Contero, M. Alcañiz, et al. 2010. Design and validation of an augmented book for spatial abilities development in engineering students. *Computers & Graphics* 34, 1 (2010), 77–91.
- [155] MAXST Co., Ltd. 2020. MAXST | The most convenient and the easiest AR software development kit. <http://maxst.com/#/en/arsdk>
- [156] A. McNamara, K. Boyd, J. George, W. Jones, et al. 2019. Information Placement in Virtual Reality. In *VR&3DUI. IEEE*, 1765–1769.
- [157] A. McNamara and C. Kabeerdoss. 2016. Mobile augmented reality: Placing labels based on gaze position. In *ISMAR. IEEE*, 36–37.
- [158] D. Michie, D. Spiegelhalter, and C. Taylor. 1994. Machine learning. *Neural and Statistical Classification* 13, 1994 (1994), 1–298.
- [159] Microsoft. 2019. Microsoft HoloLens | Mixed Reality Technology for Business. <https://www.microsoft.com/en-us/hololens>
- [160] Microsoft. 2020. microsoft/MixedRealityToolkit-Unity. <https://github.com/microsoft/MixedRealityToolkit-Unity>
- [161] S. Miraftabzadeh, F. Foidell, M. Longo, and M. Pasetti. 2019. A survey of machine learning applications for power system analytics. In *IEEEIC/ICPS Europe. IEEE*, 1–5.
- [162] L. Naimark and E. Foxlin. 2002. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *ISMAR. IEEE*, 27–36.
- [163] K. Nakano, D. Horita, N. Sakata, K. Kiyokawa, et al. 2019. DeepTaste: Augmented reality gustatory manipulation with gan-based real-time food-to-food translation. In *ISMAR. IEEE*, 212–223.
- [164] H. Nam and B. Han. 2016. Learning multi-domain convolutional neural networks for visual tracking. In *Proc. CVPR*, 4293–4302.
- [165] K. Nantz and T. Lundgren. 1998. Lecturing with technology. *College Teaching* 46, 2 (1998), 53–56.
- [166] S. Nathan and M. Beham. 2020. LightNet: deep learning based illumination estimation from virtual images. In *ECCV. Springer*, 568–580.
- [167] M. Nebeling, M. Speicher, X. Wang, S. Rajaram, et al. 2020. MRAT: The Mixed Reality Analytics Toolkit. In *Proc. CHI. ACM*, 1–12.
- [168] U. Neumann and S. You. 1999. Natural feature tracking for augmented reality. *IEEE TMM* 1, 1 (1999), 53–64.
- [169] G. Ning, Z. Zhang, C. Huang, X. Ren, et al. 2017. Spatially supervised recurrent convolutional neural networks for visual object tracking. In *ISCV. IEEE*, 1–4.
- [170] S. Oh, K. Park, S. Kwon, and H. So. 2016. Designing a Multi-User Interactive Simulation Using AR Glasses. In *Proc. TEL. ACM*, 539–544.
- [171] Onirix. 2019. SDKs - Onirix. <https://www.onirix.com/ar-sdks/>
- [172] opuscope. 2020. Minsar Studio - Overview. <https://www.minsar.app/>
- [173] K. Park, M. Kim, S. Choi, and J. Lee. 2020. Deep learning-based smart task assistance in wearable augmented reality. *Robotics and Computer-Integrated Manufacturing* 63 (2020), 101887.
- [174] Y. Park, Y. Yang, H. Ro, J. Byun, et al. 2018. Meet AR-Bot: Meeting Anywhere, Anytime with Movable Spatial AR Robot. In *Proc. MM. ACM*, 1242–1243.
- [175] O. Pauly, B. Diotte, P. Fallavollita, S. Weidert, et al. 2015. Machine learning-based augmented reality for improved surgical scene understanding. *Computerized Medical Imaging and Graphics* 41 (2015), 55–60.
- [176] Y. Pei, Y. Wu, S. Wang, F. Wang, et al. 2019. WA VIS: A Web-based Augmented Reality Text Data Visual Analysis Tool. In *ICVRV. IEEE*, 11–17.
- [177] R. Perla, G. Gupta, R. Hebbalaguppe, and E. Hassan. 2016. Inspector: An augmented reality inspection framework for industry. In *ISMAR. IEEE*, 355–356.
- [178] T. Peters, C. Linte, Z. Yaniv, and J. Williams. 2018. *Mixed and augmented reality in medicine*. CRC Press.
- [179] J. Piao and S. Kim. 2017. Adaptive monocular visual-inertial SLAM for real-time augmented reality applications in mobile devices. *Sensors* 17, 11 (2017), 2567.

- [180] W. Piekarski and B. Thomas. 2001. Augmented reality with wearable computers running linux. In *Australian Linux Conference*. 1–14.
- [181] Pikkart. 2019. Pikkart AR SDK & ecosystem. <https://developer.pikkart.com/augmented-reality/sdk/>
- [182] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst. 2017. CoVAR: A Collaborative Virtual and Augmented Reality System for Remote Collaboration. In *SIGGRAPH*. ACM, Article 3, 2 pages.
- [183] C. Plaza de Miguel. 2014. ARLodge: context-aware recommender system based on augmented reality to assist on the accommodation search process. (2014).
- [184] D. Plecher, C. Eichhorn, K. Seyam, and G. Klinker. 2020. ARsinoë-Learning Egyptian Hieroglyphs with Augmented Reality and Machine Learning. In *ISMAR*. IEEE, 326–332.
- [185] PlugXR, Inc. 2020. A Cloud based Augmented Reality platform for everyone. <https://www.plugxr.com/>
- [186] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, et al. 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–36.
- [187] P. Pratt, M. Ives, G. Lawton, J. Simmons, et al. 2018. Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. *European radiology experimental* 2, 1 (2018), 2.
- [188] PTC. 2020. Equip your workforce with augmented reality to enhance productivity and improve your bottom line. <https://www.ptc.com/en/products/augmented-reality>
- [189] PTC Inc. 2018. Vuforia Fusion Supported Devices. <https://library.vuforia.com/articles/Solution/vuforia-fusion-supported-devices.html>
- [190] R. Iizuka. 2016. Welcome to NyARTToolkit.EN | NyARTToolkit project. https://nyatla.jp/nyartoolkit/wp/?page_id=198
- [191] N. Rakholia, S. Hegde, and R. Hebbalaguppe. 2018. Where to place: A real-time visual saliency based label placement for augmented reality applications. In *ICIP*. IEEE, 604–608.
- [192] K. Ramasubramaniam and R. Bhat. 2018. LcAR—Low cost augmented reality for the automotive industry. In *ICCE*. IEEE, 1–3.
- [193] J. Rambach, A. Pagani, M. Schneider, O. Artemenko, et al. 2018. 6DoF object tracking based on 3D scans for augmented reality remote live support. *Computers* 7, 1 (2018), 6.
- [194] D. Rambli and S. Irshad. 2015. UX Design Evaluation of Mobile Augmented Reality Marketing Products and Services for Asia Pacific Region. In *Proc. APCHIUX*. ACM, 42–45.
- [195] J. Rao, Y. Qiao, F. Ren, J. Wang, et al. 2017. A mobile outdoor augmented reality method combining deep learning object detection and spatial relationships for geovisualization. *Sensors* 17, 9 (2017), 1951.
- [196] T. Rasmussen and W. Huang. 2019. SceneCam: Using AR to Improve Multi-Camera Remote Collaboration. In *SIGGRAPH*. ACM, 36–37.
- [197] P. Rauschnabel, A. Rossmann, and M. er. 2017. An adoption framework for mobile augmented reality games: The case of Pokémon Go. *Computers in Human Behavior* 76 (2017), 276–286.
- [198] Realmx Inc. 2020. artoolkitX. <http://www.artoolkitx.org/>
- [199] J. Redmon and A. Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).
- [200] A. Redžepagić, C. Löffler, T. Feigl, and C. Mutschler. 2020. A Sense of Quality for Augmented Reality Assisted Process Guidance. In *ISMAR*. IEEE, 129–134.
- [201] G. Reitmayr, T. Langlotz, D. Wagner, A. Mulloni, et al. 2010. Simultaneous localization and mapping for augmented reality. In *ISUVR*. IEEE, 5–8.
- [202] P. Ren, X. Qiao, Y. Huang, L. Liu, et al. 2020. Edge-Assisted Distributed DNN Collaborative Computing Approach for Mobile Web Augmented Reality in 5G Networks. *IEEE Network* 34, 2 (2020), 254–261.
- [203] T. Richter-Trummer, D. Kalkofen, J. Park, and D. Schmalstieg. 2016. Instant mixed reality lighting from casual scanning. In *ISMAR*. IEEE, 27–36.
- [204] C. Robertson and B. MacIntyre. 2004. Adapting to registration error in an intent-based augmentation system. In *Virtual and Augmented Reality Applications in Manufacturing*. Springer, 147–167.
- [205] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer. 2018. Speeded up detection of squared fiducial markers. *Image and Vision Computing* 76 (2018), 38 – 47.
- [206] O. Ronneberger, P. Fischer, and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*. Springer, 234–241.
- [207] C. Sahu, C. an K, and R. Rai. 2020. Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review. *International Journal of Production Research* (2020), 1–57.
- [208] S. M. Salman, T. A. Sitompul, A. V. Papadopoulos, and T. Nolte. 2020. Fog Computing for Augmented Reality: Trends, Challenges and Opportunities. In *ICFC*. 56–63.
- [209] M. Santos, T. Taketomi, C. S, J. Polvi, et al. 2014. A Usability Scale for Handheld Augmented Reality. In *Proc. VRST*. ACM, 167–176.

- [210] N. Sayed, H. Zayed, and M. Sharawy. 2011. ARSC: Augmented reality student card an augmented reality solution for the education field. *Computers & Education* 56, 4 (2011), 1045–1061.
- [211] J. Schulman, P. Moritz, S. Levine, M. I. Jordan, et al. 2016. High-Dimensional Continuous Control Using Generalized Advantage Estimation. *CoRR* abs/1506.02438 (2016).
- [212] P. Sernani, R. Angeloni, A. Dragoni, R. Quattrini, et al. 2019. Combining image targets and SLAM for AR-based cultural heritage fruition. In *AVR*. Springer, 199–207.
- [213] J. Shuhaiber. 2004. Augmented reality in surgery. *Archives of surgery* 139, 2 (2004), 170–174.
- [214] Six to Start & Naomi Alderman. 2021. *Zombies, Run!* <https://zombiesrungame.com/>
- [215] E. Smirnova and F. Vasile. 2017. Contextual sequence modeling for recommendation with recurrent neural networks. In *Proc. DLRS*. 2–9.
- [216] Snap Inc. 2019. Lens Studio by Snap Inc. <https://lensstudio.snapchat.com/lenses/>
- [217] Y. Song, C. Ma, X. Wu, L. Gong, et al. 2018. Vital: Visual tracking via adversarial learning. In *Proc. CVPR*. 8990–8999.
- [218] M. Speicher, B. Hall, A. Yu, B. Zhang, et al. 2018. XD-AR: Challenges and Opportunities in Cross-Device Augmented Reality Application Development. *Proc. ACM Hum.-Comput. Interact.* 2, EICS, Article 7 (June 2018), 24 pages.
- [219] Y. Su, J. Rambach, N. Minaskan, P. Lesur, et al. 2019. Deep multi-state object pose estimation for augmented reality assembly. In *ISMAR*. IEEE, 222–227.
- [220] H. Subakti and J. Jiang. 2018. Indoor augmented reality using deep learning for industry 4.0 smart factories. In *COMPSAC*, Vol. 2. IEEE, 63–68.
- [221] C. Sun, S. Xia, D. Sun, D. Chen, et al. 2006. Transparent Adaptation of Single-User Applications for Multi-User Real-Time Collaboration. *ACM Trans. Comput.-Hum. Interact.* 13, 4 (December 2006), 531–582.
- [222] Y. Sun, A. Armengol-Urpi, S. re, J. Siegel, et al. 2019. Magichand: Interact with iot devices in augmented reality environment. In *VR&3DUI*. IEEE, 1738–1743.
- [223] S. Sutton, P. Foulkes, D. Kirk, and S. Lawson. 2019. Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In *Proc. CHI*. ACM, 1–14.
- [224] SZTAKI. 2020. ApertusVR. <http://apertusvr.org/>
- [225] M. Tan and Q. Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*. PMLR, 6105–6114.
- [226] K. Tanaka, Y. Kishino, M. Miyamae, T. Terada, et al. 2008. An information layout method for an optical see-through head mounted display focusing on the viewability. In *ISMAR*. IEEE, 139–142.
- [227] X. Tang, X. Hu, C. Fu, and D. Cohen-Or. 2020. GrabAR: Occlusion-aware Grabbing Virtual Objects in AR. In *Proc. UIST*. 697–708.
- [228] Y. Tassa, T. Erez, and E. Todorov. 2012. Synthesis and stabilization of complex behaviors through online trajectory optimization. In *IROS*. 4906–4913.
- [229] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg. 2016. Adaptive information density for augmented reality displays. In *VR*. 83–92.
- [230] Google Brain Team. [n.d.]. TensorFlow Lite | ML for Mobile and Edge Devices. <https://www.tensorflow.org/lite>
- [231] three.js. 2019. three.js. <https://threejs.org/>
- [232] K. Toczé, J. Lindqvist, and S. Nadjm-Tehrani. 2019. Performance Study of Mixed Reality for Edge Computing. In *Proc. UCC*. ACM, 285–294.
- [233] L. Trestioreanu, P. Glauner, J. Meira, and M. Gindt. 2020. Using Augmented Reality and Machine Learning in Radiology. In *Innovative Technologies for Market Leadership*. Springer, 89–106.
- [234] J. Um, J. Popper, and M. Ruskowski. 2018. Modular augmented reality platform for smart operator in production environment. In *ICPS*. IEEE, 720–725.
- [235] Unity Technologies. 2020. About AR Foundation | AR Foundation | 4.0.9 . <https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.0/manual/index.html>
- [236] G. Upadhyay, D. Aggarwal, A. Bansal, and G. Bhola. 2020. Augmented Reality and Machine Learning based Product Identification in Retail using Vuforia and MobileNets. In *ICICT*. IEEE, 479–485.
- [237] N. Vaidyanathan. 2020. Augmented Reality in Retail-A Case Study: Technology Implications to Utilitarian, Aesthetic and Enjoyment Values. In *Proc. ICVARS*. ACM, 27–34.
- [238] Vectary Inc. 2020. Vectary - The Easiest Online 3D Design and 3D Modeling Software. <https://www.vectary.com/>
- [239] Vidinoti. 2020. V-SDK - Vidinoti. <https://www.vidinoti.com/v-sdk/>
- [240] ViewAR GmbH. 2020. Introduction - Documentation. <https://documentation.viewar.com/docs/sdk/introduction>
- [241] VisionStar Information Technology (Shanghai) Co., Ltd. 2019. EasyAR. <https://www.easyar.com/>
- [242] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, et al. 2019. Mots: Multi-object tracking and segmentation. In *Proc. CVPR*. 7942–7951.
- [243] M. von Atzigen, F. Liebmann, A. Hoch, D. o, et al. 2021. HoloYolo: A proof-of-concept study for marker-less surgical navigation of spinal rod implants with augmented reality and on-device machine learning. *The International Journal*

- of Medical Robotics and Computer Assisted Surgery* 17, 1 (2021), 1–10.
- [244] VTT Technical Research Centre of Finland Ltd. 2019. Augmented Reality / 3D Tracking | ALVAR. <http://virtual.vtt.fi/virtual/proj2/multimedia/alvar/index.html>
- [245] G. Waltner, M. Schwarz, S. Ladstätter, A. Weber, et al. 2015. MANGO-mobile augmented reality with functional eating guidance and food awareness. In *ICIAP*. Springer, 425–432.
- [246] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, et al. 2019. Fast online object tracking and segmentation: A unifying approach. In *Proc. CVPR*. 1328–1338.
- [247] Z. Wang, L. Zheng, Y. Liu, and S. Wang. 2019. Towards real-time multi-object tracking. *arXiv preprint arXiv:1909.12605* 2, 3 (2019), 4.
- [248] Warner Bros. Entertainment Inc. 2020. Home - Harry Potter: Wizards Unite. <https://www.harrypotterwizardsunite.com/>
- [249] S. Wei, G. Ren, and E. O’Neill. 2014. Haptic and audio displays for augmented reality tourism applications. In *HAPTICS*. IEEE, 485–488.
- [250] E. Weng, R. Khan, S. Adruce, and O. Bee. 2013. Objects tracking from natural features in mobile augmented reality. *Procedia-Social and Behavioral Sciences* 97 (2013), 753–760.
- [251] T. Whelan, R. Salas-Moreno, B. Glocker, A. Davison, et al. 2016. ElasticFusion: Real-time dense SLAM and light source estimation. *The International Journal of Robotics Research* 35, 14 (2016), 1697–1716.
- [252] M. Wiczorek, A. Aichert, O. Kutter, C. Bichlmeier, et al. 2010. GPU-accelerated Rendering for Medical Augmented Reality in Minimally-invasive Procedures. *Bildverarbeitung für die Medizin* 574 (2010), 102–106.
- [253] Wikitude GmbH. 2020. Wikitude Cross Platform Augmented Reality SDK. <https://www.wikitude.com/products/wikitude-sdk/>
- [254] J. Wither, C. Coffin, J. Ventura, and T. Hollerer. 2008. Fast annotation and modeling with a single-point laser range finder. In *Proc. ISMAR*. IEEE Computer Society, 65–68.
- [255] J. Wither, S. Diverdi, and T. Hollerer. 2006. Using aerial photographs for improved mobile AR annotation. In *ISMAR*. IEEE, 159–162.
- [256] J. Wither and T. Hollerer. 2005. Pictorial depth cues for outdoor augmented reality. In *ISWC*. IEEE, 92–99.
- [257] N. Wojke, A. Bewley, and D. Paulus. 2017. Simple online and realtime tracking with a deep association metric. In *ICIP*. IEEE, 3645–3649.
- [258] C. Woo and D. Lee. 2011. A Development of a Framework for Building Knowledge based Augmented Reality System. *Journal of the Korea society of computer and information* 16, 7 (2011), 49–58.
- [259] C. Wu, D. Brooks, K. Chen, D. Chen, et al. 2019. Machine learning at facebook: Understanding inference at the edge. In *HPCA*. IEEE, 331–344.
- [260] XLsoft Corporation. 2019. AR (Augmented Reality) for Mobile Applications. <https://www.xlsoft.com/en/products/kudan/index.html>
- [261] XZIMG Limited. 2019. Home - XZIMG. <https://www.xzimg.com/Home>
- [262] A. Yilmaz, O. Javed, and M. Shah. 2006. Object tracking: A survey. *Acm computing surveys (CSUR)* 38, 4 (2006), 13–es.
- [263] N. Yuill and Y. Rogers. 2012. Mechanisms for Collaboration: A Design and Evaluation Framework for Multi-User Interfaces. *ACM Trans. Comput.-Hum. Interact.* 19, 1, Article 1 (May 2012), 25 pages.
- [264] Zappar Ltd. 2020. Universal AR SDK. <https://zap.works/universal-ar/>
- [265] T. Zarraonandia, I. Aedo, P. Díaz, and A. Montero. 2013. An augmented lecture feedback system to support learner and teacher communication. *British Journal of Educational Technology* 44, 4 (2013), 616–628.
- [266] J. Zhang, W. Xing, M. Xing, and G. Sun. 2018. Terahertz image detection with the improved faster region-based convolutional neural network. *Sensors* 18, 7 (2018), 2327.
- [267] L. Zhang, A. Sun, R. Shea, J. Liu, et al. 2019. Rendering Multi-Party Mobile Augmented Reality from Edge. In *Proc. NOSSDAV*. ACM, 67–72.
- [268] Z. Zhang, S. Shang, S. Kulkarni, and P. Hui. 2013. Improving augmented reality using recommender systems. In *Proc. RecSys*. 173–176.
- [269] S. Zhao, X. Xiao, Q. Wang, X. Zhang, et al. 2020. An Intelligent Augmented Reality Training Framework for Neonatal Endotracheal Intubation. In *ISMAR*. IEEE, 672–681.
- [270] P. Zhou, W. Zhang, T. Braud, P. Hui, et al. 2019. Enhanced augmented reality applications in vehicle-to-edge networks. In *ICIN*. IEEE, 167–174.
- [271] S. Zollmann, R. Grasset, T. Langlotz, W. Lo, et al. 2020. Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns. *IEEE Transactions on Visualization and Computer Graphics* (2020).
- [272] B. Zoph and Q. Le. 2016. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578* (2016).