Abstract—Augmented Reality applications have the capability of merging virtual objects into physical setting, or alternatively they can wrap physical objects within a virtual scene. Augmented reality applications are similar to virtual reality applications in that aspects of the visualizations are computer generated, but augmented reality apps also must contain a view of the physical world. Augmented reality applications are being utilized in service, manufacturing, product areas, as well as gaming. Mobile devices are becoming common runtime environments for augmented reality applications and the mobile device proliferation is enabling a wave of AR applications. Due to the combined nature of digital and physical objects, as well as the environmental and contextual constraints, a traditional test plan is not sufficient. A new quality model is proposed that takes these issues into account, and an example of how machine learning can assist with aspects of the model is discussed.

Keywords—Augmented Reality, Software Quality, Software Testing, Machine Learning, Test Automation, Quality Model

I. INTRODUCTION

Augmented reality is accomplished by creating an experience in an application’s user interface that merges both physical surroundings as well virtual objects, which may be models or scenes [1]. The term augmented reality was first used by Caudell and Mizell in 1992 [2]. The development of augmented reality applications has very much been democratized within the last decade, due to the rise of libraries that make building the applications easier, and the increase in compute power of devices such as mobile phones. Mobile phones have emerged into ubiquitous platforms as they fit into our lives, rather than requiring human change [3]. Similarly, augmented reality has allowed us to keep a physical plane as-is, and yet embellish upon it new digital objects. AR is becoming popular because it allows change to what we see and at the same time it allows the physical world to stay the same. According to one Gartner survey conducted in 2018 (Figure 1), “According to a Gartner Research Circle study, 27% and 17% of respondents were either using or evaluating AR and MR technologies “[4].

An additional Gartner survey focused on multi-experience application development. The analysts stated, “It is, however, surprising to see VR apps identified as the second most impactful type of multiexperience app (20%), as AR has more potential use cases and device support. But only 14% of the respondents thought that AR apps would be the most impactful, despite AR app development tools being more widely available. “ [5] (figure 2).

There are several development tools to assist developers in the creation of AR applications. The ARToolkit was
developed by Hirokazu Kato in 1999 which is now available for Windows, Mac, Linux, and Android platforms. ARCore, an SDK for Android and iOS phone, was developed by Google and released in 2018 [6]. Apple has their own native tooling, ARKit, which released with iOS 11 in 2017 [7]. Facebook released Camera Effects Platform in 2017. In October of 2018, Facebook rebranded the platform to Spark AR [8]. Qualcomm released the QCAR SDK in 2011 [9], it then sold to PTC in 2015 for $65 million [10]. Maxst started as a company in 2010, released an AR mobile game in 2011, 5 years before the release of Pokemon GO, and released their AR SDK in 2012 [11]. Wikitude’s Javascript SDK released in 2012 [12].

One driving factor of the popularity of mobile augmented reality applications, are the many SDKs available and the ease of use. The ease of development, however, has also led to applications having poor user experiences [13]. There are several choices of development tools, both open source, and commercial, which span across runtime environments. Table I lists several common examples of these types of augmented reality application development tools.

<table>
<thead>
<tr>
<th>Tool name</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vuforia Studio [14]</td>
<td>Android, iOS, UWP, Unity</td>
</tr>
<tr>
<td>Spark AR [15]</td>
<td>Android, iOS</td>
</tr>
<tr>
<td>ARKit [16]</td>
<td>iOS</td>
</tr>
<tr>
<td>ARCore [17]</td>
<td>Android, iOS</td>
</tr>
<tr>
<td>Wikitude [18]</td>
<td>Android, iOS, Microsoft Tablet, smart glasses</td>
</tr>
<tr>
<td>Maxst AR [19]</td>
<td>Android, iOS, Windows, Mac, Unity</td>
</tr>
</tbody>
</table>

Industry usage of augmented reality technology is now becoming widespread. AR applications have been utilized in manufacturing, service industry, and product design. AR has aided in the layout and design of factories [20]. Military machinery service and maintenance has also been enabled thru AR technology [21]. AR has also been utilized to assist in training new assembly skills [22]. In the automotive industry, augmented reality was utilized by Mercedes Benz to create a consumer application allowing potential buyers to customize a new car as it would look in their own personal driveway. In the first three months of this application being available to iOS users, there were over thirty-four thousand downloads and an average application rating of 4.3 out of 5 stars [23].

While the development tools and the application of augmented reality are becoming more abundant, AR specific testing tools and practices are lagging behind. One famous virtual reality headset, the “Sword of Damocles”, was invented in 1968 by Ivan Sutherland [24]. Yet, while the VR technology has existed for a longer period of time than AR, there are no common test automation practices for application interactions in VR [25]. Figure 3 gives the span of degree of similarity and relation between AR and VR [26].

Figure 3. The Real-Virtual Continuum

While previous research has identified a potential void in testing practices for novel interfaces, that is not to say that industry users won’t search them out, or rely upon them. Test automation solutions for web and mobile are available from both open source communities and commercial vendors. Test automation tools specifically for AR apps are not as easily found. Figure 4 identifies Gartner survey respondents answers to a question about technology adoption specifically in support enterprise multiexperience app development. Gartner analysts in their analysis of the survey have stated, “The respondents had fairly equal adoption rates across the technology categories we asked about, but Gartner advises application leaders to emphasize adoption of test automation tools and DevOps practices.” [5].

Figure 4. Gartner Survey Analysis: Insights to Kick-Start an Enterprise Multiexperience Development Strategy Findings on Adoption of Technologies in Support of Enterprise Multiexperience App Development

Assessing the quality of an AR application is more complex than the quality of a typical traditional software product. Early adopters of mobile AR applications have experienced bugs in production versions. Reports of quality
Evaluating AR systems will be problematic because the technology is young and we do not yet understand the expectations of end-users of AR apps [29]. In addition to those issues, traditional usability evaluation methods are not comprehensively adequate for detecting defects in applications having novel interfaces of which the evaluation methods were designed for and matured over time [30]. Nielsen defined usability as software attributes of learnability, efficiency, memorability, satisfaction, and error [31]. AR usability requires both the physical and virtual components to work together. Thus if either is degraded, such as a low-light physical scenario, the AR application quickly loses usability. There is also the concern of color sameness and collision between models and physical environments. Presence and collaboration in an augmented environment are also new quality concerns compared to traditional testing procedures [32]. The application must be intelligent in these cases to instruct the user to improve the environmental settings or move to a more traditional application interface to solve their needs. This implication of an augmented reality application being intelligent to understand a degraded physical environment’s impact to its own usability means quality concerns must be addressed at design.

Here an approach is proposed that produces a quality model to meet the growing expectations from the AR products. This model uses features from the ISO 25010 model.

II. SOFTWARE QUALITY MODEL OF AUGMENTED REALITY APPLICATIONS

A well-designed augmented reality application blends virtual and physical worlds into an experience that facilitates perspective, presence, interaction, and immersion [33]. Perspective is necessary for engaging virtual objects realistically within dynamic physical environments. Presence is the user’s ability to be within their physical environment, and engage in the virtual model or scene. This engagement is the interaction that occurs, as we manipulate the model, or as it changes in relation to the perspective of our environment. This all works to create immersion, our ability to be in both physical and virtual presence at the same time. However, to scale AR, the application must also have persistence. This allows the augmented models and scenes to be experienced by other application users, scaling the augmented reality experience. These characteristics of augmented reality applications require the developers to go beyond traditional testing concerns.

A. Environmental Context as it Pertains to Functionality

An augmented reality application’s functionality must be tested for two distinct requirements, occlusion and collision [34], as discussed below. Users of AR applications need realistic interactions between the physical and virtual objects. This means objects having proper shadows, depth and they should not occupy the same space at the same time.

Occlusion requires that if a virtual object is distanced away from physical objects, during rendering the physical objects should hide or occlude portions or the virtual object entirely in a realistic manner. Occlusion is difficult because AR engines must properly represent the physical world, even as our relation to it changes based upon physical movement. The AR engine must reconstruct the 3D model according to the properties of the physical world, and render the dynamic user interface appropriately. Test cases must be created that mimic the possible multitudes of physical realms the application will run in, and then load and engage the virtual objects to test propensity to fail occlusion or create an experience that does not mimic reality. Figure 8 represents failed occlusion and the existence of collision.

A collision occurs when a virtual object approaches a physical object and attempts to occupy the same 3D space at a time. Many AR applications will allow the virtual object to be manipulated as it appears in the physical plane. However, the virtual object should not occupy the same space as the physical objects, despite the manipulation. Collision detection needs to be tested in an implementation. Test cases should also include how color and light change the ability to detect and prevent collision.

B. Environmental Context as it Pertains to Visual Distinction

AR applications can lose visual distinction based upon the physical world they are operated within. Physical environments that make gradient distinction difficult, will reduce the usability of the applications. Engaging virtual objects is difficult and prone to fail in areas of intense light, where the color white is dominant, when the model and the physical scene seem chameleon in color, or when surfaces are smooth and reflective in general. Distinction testing must have multiple test cases that alter the physical environment to these types of extremes, to understand the application’s behavior, and to potentially present the user with options outside of the AR user interface, such as a traditional menu based interface, in the case failure to render virtual objects such that they are distinct.

C. Environmental Context as it Pertains to Portability

Another environmental concern to AR applications are the variations among the platforms (runtime OS and the
computational and display/interaction hardware). There are many physical form factors, and versions of device operating systems, thus an AR application will need to be tested for correct execution across devices and platforms. This is where utilizing a device farm, as we would for typical web and mobile development, will help assure the breadth of device compatibility needed. For those utilizing AR applications internally to an organization, and having a policy on acceptable hardware, the concern for variations in the mobile runtime platform may be diminished. There are many mobile device farm vendors to select from, some having both on premise and cloud options. Example vendors having both on premise and cloud offerings include Experitest, Mobile Labs and SmartBear, as well as others.

Crowd testing could also be utilized to establish a large test bed of diverse devices. Crowd testing can be facilitated internal or external to an organization, through crowd testing platforms. Crowd testing can accomplish quality goals such as user experience, functional or regression testing, and it offers diversity beyond devices, but also in geography and in the skills and experience of end users. Crowd testing platforms, as examples, could include RainforestQA, Applause or testIO.

D. Performance Engineering

Three-dimensional rendering while operating camera hardware and merging the image into a user interface is computationally expensive. Early releases of Google’s ARCore came with release notes that suggests that the anchors, (a type of feature point that can represents the virtual objects presence in the physical world), should be detached or deleted to prevent expensive CPU costs [35].

Augmented reality applications are often implemented using a distributed architecture, using both local client and some cloud or centralized services. In these cases, we must do performance analysis on the centralized resources, with the understanding that the computational capacities of those assets may be stretched when many clients engage the services. The client devices themselves may also have performance concerns, including battery depletion over long runtimes. Traditional performance testing will solve much of this analysis, and those utilizing cloud resources will need to add validation of the complexity and cost of automatically scaled resources to test plans. Cloud performance can be degraded when there is high network traffic to the cloud’s data center or even temperature within the data center [36]. The traffic could be heavy in a distributed AR application implementation.

E. Scaling Usability with Persistence

Scaling augmented reality experiences across multiple users requires centralizing physical maps and virtual object anchors so that multiple users can share immersion. ARKit and ARCore both have these capabilities [37]. Persistence plays into quality for single user scenarios, you place the virtual object and days later expect to see it embedded into the same physical location. Persistence is also a requirement of scaling users’ ability to share augmented experience. Test cases will require an amount of complexity in the numbers of testers that engage in one shared scenario and experience, as well as the duration of time the test case covers.

F. Creating the Quality Model

Test planning for an augmented reality application release has many environmentally contextual factors. The virtual object needs to behave within acceptable limits of the physical world. The device runtime environment and form factor are variables as well. Depending upon the nature of the application, cloud technologies such as shared anchor points may be required, which can also lead to having more complex functional and non-functional quality requirements.

Table II attempts to modernize aspects of the ISO 25010 model to reflect the user experience expectations, broken down into the quality characteristics of perspective, presence, interaction, immersion, persistence, and performance. User experience itself is a very influential quality attribute [38]. This model attributes eight characteristics to the user experience of an augmented reality application. One previous research has indicated three characteristics of user involvement, interaction between user, product and other agents, and what can be observed or measured [39]. While another work found five characteristics of aesthetics, appeal, joy, usability, and utility [40]. The model is meant to reflect the ISO 25010 model and relate to the original eight quality characteristics as possible.

<table>
<thead>
<tr>
<th>User Experience</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionality / Presence</td>
<td>Occlusion and Collision characteristics will make the AR application more realistic giving more functional uses. This characteristic allows users to be present in both physical and virtual worlds.</td>
</tr>
<tr>
<td>Visual distinction / Perspective</td>
<td>The environment, such as highly reflective and bright lighting may affect visual distinction of the application’s ability to inject virtual into physical consistently. Perspective is assured by testing for realistic rendering of virtual objects in relation to their physical counterparts so that height, width, and depth are accurately maintained.</td>
</tr>
</tbody>
</table>
Performance / Interaction

Multiple anchors achieved by having many virtual objects in one physical layout can impact device performance. This can limit the user’s ability to interact if interaction requires frequent or multiple virtual object creations, instances or mutations. Performance also depends upon cloud anchor mechanisms so that multiple users can share augmented reality experiences.

Portability / Immersion

The many mobile form factors and operating systems create a complex runtime environment to be tested cross-device. For many users of an augmented reality application to experience immersion, the application must be compatible across many device types and versions.

Usability / Persistence

Scaling to achieve shared augmented reality experience across multiple simultaneous users requires persistence. Persistence utilizes cloud technologies to enable distributed users to share anchor points and other object and scene metadata so that they can become immersed in each other’s virtual and physical worlds.

G. Applying Automated Testing to AR Applications using Machine Learning

Automation helps accelerate testing so that we can consider production release sooner and also so that we may find more bugs in a shorter amount of time. Similar to how traditional test practices do not consider novel user interfaces such as AR, testing tools are also lacking in this area. One case study by SmartBear is available to discuss their approach [41]. While they describe the physical mobile lab and physical robotics that reposition the devices allowing their cameras to adequately zoom and reset for incorporating the 3D model into the physical world, the case study does not portray how they tested the functionality and user experience as described in our quality model, or specifically how they tested for aspects such as collision or occlusion.

We present a framework that includes test automation, common tools and practices, and the use of machine learning to build an approach to testing that provides probability of accurate model presentation in the physical world, to accelerate manual testers and developers so that they may quickly hone in on problematic areas.

This approach begins with the model files being utilized within the AR development. Figure 5 is taken from an integrated development environment, Android Studio, and illustrates one example of how model files can be stored and organized during AR application development. These become a source of comparison to be utilized in the proposed automation framework.

![Figure 5 Android Studio Folders and sfb Model File](image)

What is important about these models is the return value when they are sent into common ML image recognition engines. Our framework uses a common open source test automation tool, Appium, to take screenshots during the application under test runtime.

This suggested framework has access to the model within the development environment, as well as the screenshot of the model when utilized within the mobile app. Figure 6 is an image of the model within the development environment. Figure 8 is a screenshot of an AR app running where the model is now occluding. The two images are sent to an ML driven image recognition engine. When the two resultants are within an acceptable value over time, we have accelerated through the testing effort by assuring the model is being detected with confidence when being displayed in the physical world. This is one aspect in how machine learning and test automation can assist in the entire software development lifecycle of augmented reality applications. The same type of harness can be created and run autonomously while the application is in production, again assuring that within the changing environments, the models are being adequately presented in the novel user interface of AR.

As a direct example of this framework, we can submit the model of the ketchup bottle as pictured in figure 6 to the AWS Rekognition service. As indicated in figure 7, the service returns a probability of 97.6% certainty that it is both
Food as classification and it is identifiable as a Ketchup bottle.

However, when we utilize a screenshot of the same model file utilized in an AR app, we see occlusion failing and the ML service returns concerning results. This would indicate potential defects and manual testing could then hone in on this particular test case.

Figure 8 indicates what failed occlusion looks like, and figure 9 gives the Rekognition return types that would help indicate a defect. The bottle is anchored to the back corner of the room, yet is also floating and occluding the guitar that is nearest the mobile device. The probabilities from the same AWS recognition engine have changed dramatically. We no longer see a bottle or food. Rather we have detected Furniture, Leisure Activities, and a Guitar.
III. Conclusions

Augmented reality applications merge virtual models and the physical world. These applications are becoming more popular in many verticals, and several SDKs are available to assist in creation. However further development of testing methodologies and tooling is needed. The novel features of an AR user interface is not adequately addressed by traditional testing methods, or of the ISO 25010 model. This research suggests creating test cases that focus on characteristics of perspective, presence, interaction, immersion, persistence, and performance. Automation and machine learning of image detection features also need to be leveraged to assist in the detection of potential defects in the AR applications. The revised quality model and the ML enabled automation framework seek to expand current capabilities and methods to enhance defect detection in AR applications.

REFERENCES


