Touch Screen vs Vision Based Augmented Reality: A Comparison Between Interfaces

Andrew Lau
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Ontario, Canada M3J 1P3
pellost@my.yorku.ca

ABSTRACT
To be added later.

Keywords
Augmented reality, mobile device, android, Vuforia SDK

INTRODUCTION

With the popularity of mobile devices now reaching a point where people cannot survive a day without using their smartphones, the importance and relevance of how people consume their information has become that much more present. As mobile devices continue to accumulate increasingly more functions and features, the amount of information that that users can consume has also increased. Since the introduction of smartphones, the main means of displaying information has been very similar. Majority of the information is delivered through the screen of the smartphone, where overflow can be accessed using a scroll bar or a next-page button.

Limitations of Touch Screen interfaces

Unlike non-mobile devices such as desktop computers, mobile devices suffer a major drawback. The smaller screen space restricts the amount of information that can be presented. This problem is generally managed through scrolling interfaces or split-screening, however doing so still has its limitations. Although these methods allow the user to navigate interfaces that are larger than the available screen size, the amount of information presented to the user at a time is always the same. When multitasking or working with large sets of items, users must resort to swapping between applications or items. The inability to display everything that the user requires to complete a task increases overhead time and in turn reduces productivity. This problem is apparent in applications such as the photo gallery app on most mobile devices. The problem with most photo gallery is that they suffer from this problem with limited screen space, where only a number of photos can fit on screen at a time. This restricts actions such as when users are searching for a photo in a large set of photos. Being able to look through a larger set of photos without having to make an action to see more affects how fast that photo can be found. One of the main cause of this is that the layout of the photo within the interface are static and do not change. Most mobile photo galleries do not allow the user to resize the images within the layout to allow more photos to fit on screen at a time such as Google's *Photos* app. This is where augmented reality might be able to address this problem.

Potential of Augmented Reality

Virtual reality (VR) is the technology of immersing the user within an artificially generated environment and augmented reality (AR) is a branch of this concept. Rather than surrounding the user with a virtual environment augmented reality superimposes artificial elements, generally generated by a computer, into the real world. These artificial elements can range from vision-based components to tactile feedback. According to a definition by Azuma[4], AR systems must combine real and virtual, be interactive in real time, and be registered in 3D.

There are many ways to achieve AR, such as the technology used in Intel's recently announced smart glasses, called *Vaunt* seen in Figure 1. These glasses provide the user with a head up display (HUD) in one eye. This is achieved through the use of monochrome projector that projects a low-power laser directly onto the retina, a method called retinal projection. A more common way of achieving AR is using the camera of mobile phones and projecting computer generated components into the view of the camera.

Figure 1. Intel's *Vaunt* smart glasses

In the context of mobile devices, AR has already made an impact in the mobile industry. One of the most well-known AR applications is a mobile game developed by Niantic called *Pokémon GO*. This mobile game that allows users to visually see and interact with elements of the game in the real world through the device's camera, as seen in Figure 2. In order to find these game elements, the user must physically move to certain locations in the real world and look around their environment through their phone's camera. This use of AR adds a new dimension to how information is perceived on mobile devices. By generating these game elements as AR components, the application makes a connection between the physical location of the person and the objective of the game.

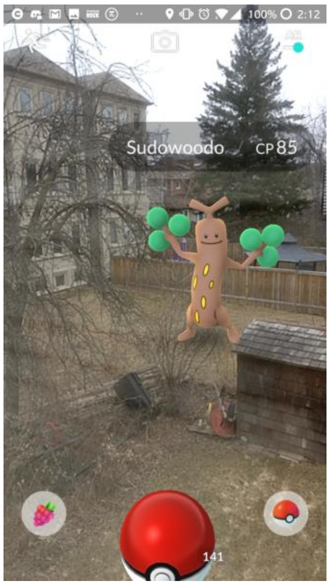

Figure 2. Screenshot of *Pokémon GO*.

Pokémon GO is an example of an AR browser that renders geo-bound content. An AR browser is an application that only renders content into the real world through the camera view without understanding the environment [2].

Image recognition AR applications do the opposite, it tries to understand what is in view and renders content accordingly. An example of this is *Snapchat*'s photo filters. These filters recognize and track human faces and render different effects onto the face as seen in Figure 3. The application searches for known patterns from what it receives from the camera. For a human face, the system looks for key points defined by darker areas around the eyes and mouth. Once this point of references is established in the camera frame, the system is free to render content according to the perspective of the camera on the tracker. These recognizable patterns can be generated any image with strong features, such as the general shape of a face or the specific shape of a quick response (QR) code.

Image recognition AR applications allows for a stronger connection between the physical world and the generated digital components compared to AR browsers. Using the camera and a combination of sensors built into most modern mobile devices the size of usable "screen space" effectively becomes the size of the user's environment.

Figure 3. Screenshot of a *Snapchat* filter being applied to the view of the camera

There are a several studies that try to explore the potentials of this technology. A study conducted at the Lehigh University were set out to find how students would react to a new method of teaching using visual AR educational games [1]. Results showed that the use of AR in education has minimized user frustration and has reduced cognitive overload during the user's experience with the technology.

METHOD

Participants

9 participants in the Richmond Hill area where asked to participate in the study. The participants fell into two major age ranges, 20 to 30 and 40 to 60. All participants have a general understanding of using an Android interface. Participants were given no incentive for volunteering their time.

Apparatus

For this study, an Android smartphone was used as the main test apparatus. The device was a OnePlus 3t running Android Oreo 8.0.0 as seen in Figure 4.

Figure 4. OnePlus *3t* Android phone.

The main sensors that were used on this device during the experiment was the camera, accelerometer, gyro sensor, and touch sensor. Two sizes of printed trackers were also used to assist the AR part of the experiment. This is to test whether the scale of AR applications has any affect on user performance. These will come in the form of QR codes. The smartphone will be hosting two implementations of a mobile photo gallery: one implemented using a 2D on-screen interfaces similar to the default Android photo gallery, and the other implemented as an image recognizing AR application. The on-screen interface is a modified version of a demo program developed by Professor Scott MacKenzie called Demo_GridView as seen in Figure 5. This application takes the photos in a directory on the phone and displays it in an Android grid view component.

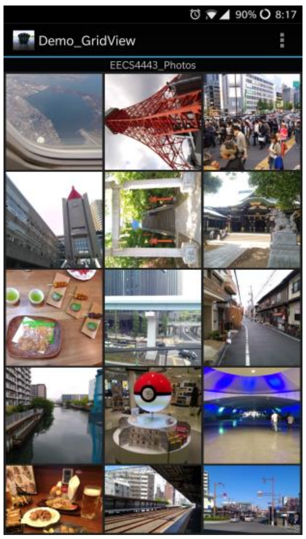

Figure 5. Modified version of Demo_GridView

The AR application was developed in Unity using the *Vuforia* software development kit (SDK). This SDK allows elements to be overlaid on-screen relative to the position of the recognized tracker within the view of the camera, as seen in Figure 6. The image to be used as a tracker for the application (for this study a QR code) is put through an algorithm provided in the SDK which calculates key features of the image and stores it in a dictionary for the application. Images with a lot of detail and high contrast are good candidates to be used as target tracker. The value of the QR code used for this study does not have any meaning to the AR application and is just a tracker that produces clean key features when passed through the preprocess algorithm. The AR applications also utilizes the extended tracking feature within *Vuforia.* This feature allows the AR application to render the tracked components even when the camera moves out of view of the tracker. During this state, the application attempts to alter the perspective of the rendered component using only the information obtained from internal sensors such as accelerometers, gyroscopes, and gravity sensors. Although this may be inaccurate, it is necessary to keep the photo gallery rendered when navigating the interface.

Figure 6. Augmented Reality application made with *Vuforia*

Procedure

Participants were informed about the subject of the study after they have agreed to participate. They were then given instructions on how to complete the user task for all types of interfaces. Participants then completed 5 trials for each of the interfaces. A participant performing the user task can be seen in Figure 7.

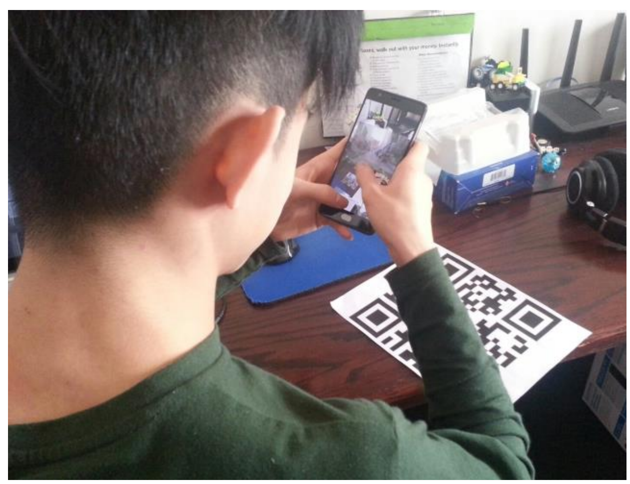

Figure 7. Participant doing the large target AR test

the task the participants were asked to perform was to locate a single image from an image gallery using each types of interfaces. The order of tests was balanced between participants to reduce order effect between the tests. During the trials, participants where shown a random image, and told to find that image by going through a list of photos. The completion time is then recorded. For the AR interface, the user was asked to complete the task using two different sized trackers, five times each. Participants are to point the camera at the tracker to generate the AR photo gallery interface. They can navigate this view by moving the phone around changing the perspective of the camera and select the image by tapping the image on screen. Participants are asked to find a randomly selected photo. Completion time is then recorded. Once all testing is complete, the user will be asked additional questions on which interface did they find more better to use, and their reasons for their choice.

Design

The study was a 3 x 5 within-subjects design with the following two independent variables and levels:

- Type of Interface: on-screen interface, small target AR interface, large target AR interface
- Trials: 1, 2, 3, 4, 5

Each participant completed 10 trails for each type of interface. Testing lasted approximately 10 minutes per participant. The total number of trials for the experiment was 9 participants x 2 types of interface x 5 trials = 90.

RESULTS AND DISCUSSION

The data collected from the user task were transcribed into a spreadsheet to be analyzed. Summary measures were calculated, and chart was generated.

Overall Average Time

The main dependent variable was time per trial. The overall mean time for all 90 trials was 11.5 seconds. The mean of the on-screen interface was 8.9 seconds, small target AR interface with a mean of 10.7 seconds, and large target AR interface with a mean of 14.8 seconds. The on-screen interface was 21% faster than the small tracker interface, and 49% faster than the large tracker interface. This relationship between the mean times of on-screen and AR interface is most likely due to lack of experience with AR interfaces. Only two (22%) participants of the participants noted that they had some experience with AR before the study. All participants also have experience using mobile devices, explaining the low average time when using the on-screen interface due to their familiarity with the controls. Comparing the two AR interfaces we find that the small target interface is 39% faster than using a large target. This can be explained by the size of the target tracker. When the size of the target tracker increases the scale of movement the user must make to navigate the augmented component also increases.

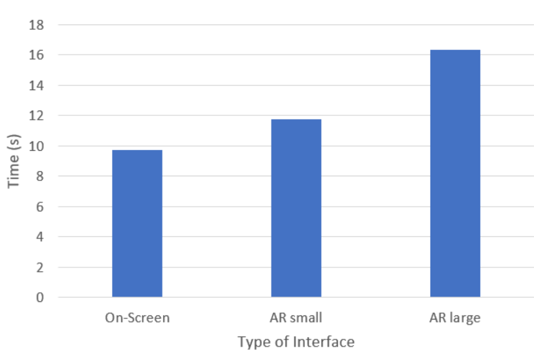

Figure 8. Mean trail completion time over five trails

Average Time per Trial

Throughout the study, 30 trials of each interface type were completed. The large target AR interface showed the most improvement of 5 trails with 34.7% improvement in completion time. Where as the other two interfaces only had an improvement of 4.6% for the on-screen interfaces and 6.4% for the small target AR interface.

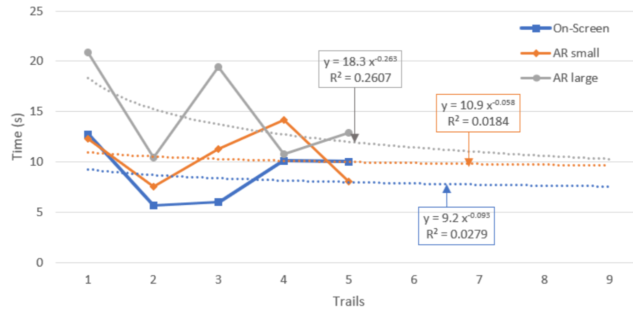

Figure 9. Average time per trail extrapolated 3 trials

Using trail completion times, models for each type of interface made using the power law of practice and extrapolated to the 9th trial. The explanations for learning that these models provide are weak due to how the test only spans 5 trials. These models can be seen in Figure 8. The predictor on-screen interface (R^2 = .0279) and the small target AR interface (R^2 = .0184) were both low, which suggest that there is little room for improvement. However, the model for the large target AR interface is interesting because the predictor for interface was much higher than expected. This suggested that the interface is quite easy to learn, makes sense when consider the effects of having a larger target tracker. The larger tracker requires the user to make larger movements to navigate the interface. Participants that were not used to working with AR interfaces had an easier time getting used to the actions required to use AR. While participants were using the small target interface it was observed that the participant would hold the phone in place since all the photos already fell in view of the camera, instead of moving the phone closer to ‘zoom in’ on the photos that were too small some to see. This would explain why there was so little improvement in time on the small target AR interface as the participant would only rely on their ability to identify small images through colour recognition and scanning to complete the task. It is also important to note that all times are affected by a randomness of the photo chosen for the participants to find which explains the large variance between max and min times within each interface.

Participant Preferences

When asked whether they would prefer to use an AR based photo gallery over an on-screen interface, all nine participants stated they would rather stay with using an on-screen interface. However, when asked if they had to choose between using a small target AR interfaces or a large target one, 4 (44%) participants favored the large target AR interface, and 5 (55%) favored the small target

AR interface. Those who favored the large target interface mostly fell in the age range of 31 to 60, where as those who favored the small target interface were those who fell in the 20 to 30 age range. This breakdown of age can also explain the improvement seen in the large target interface, as older generations many not have the same dexterity and quality of vision as younger generations required a longer practice time to reach the same level of performance. This explains why the large tracker interface generally took longer to complete. This is backed up by the comments made about the interfaces after testing. Many participants noted that when using the AR interfaces navigating to the edges of the interface was difficult as the camera view would occasionally leave the tracker.

Future Work

Future work would include a further investigation into large target AR interfaces as well as expanding the study to different applications that can potentially benefit from this technology.

CONCLUSION

In this study, the use of on-screen interfaces resulted in better task completion times overall. However, due to the large amount of experience participants had with the interface little improvement was observed over five trails. On the other hand, the large target interface showed 34.7% improvement over five trials, however remained the slowest method of the three. All participants preferred to use the on-screen interface. But when asked if they had to choose between the two AR interfaces, 44% of participants chose the large target interface and 55% chose to the small target interface.

REFERENCES

1. Bressler, D. M., and A. M. Bodzin., A mixed methods assessment of students' flow experiences during a mobile augmented reality science game, *Journal of Computer Assisted Learning,* 29.6, 2013, 505-517.
2. Olsson, T., and Salo, M. and Markus Salo. Online user survey on current mobile augmented reality applications. *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. IEEE, 2011, 75-84.
3. Ekengren, Björn. *Mobile augmented reality*. Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan, 2009.
4. Azuma, R. T., A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 1997, 355-385.