Advanced Visual Interfaces: Virtual Control in Augmented Reality

Brady Ledger
Brady.Ledger.6525@student.uu.se
Uppsala Universitet

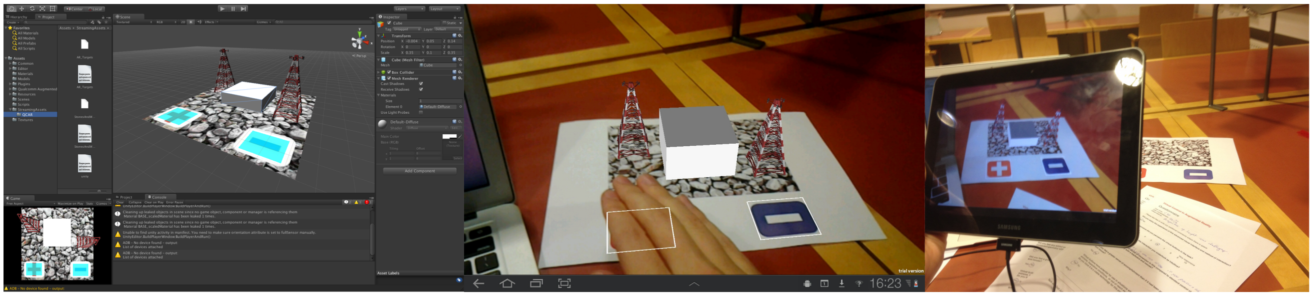

Figure 1: *AR scene editing inside Unity, Interaction with virtual buttons, Augmenting the image target.*

Abstract

Augmented reality (AR) is an upcoming area in interactive computer graphics. AR is a field where a view of the physical world has elements augmented with computer-generated features such as graphics or audiovisual media.

Virtual control places the user interaction within the rendered augmented scene. Physical interaction is made with *virtual buttons*. This method of interaction is uncommon within AR applications.

Virtual control is enabled by registering *occulusion* of image features within an image target. The virtual buttons are scripted to change the height property of a virtual mesh in the augmented scene.

A study is taken on the usability and experience of performing a task related manipulating the height properties using the virtual buttons.

Results show great interest in the interaction medium, provoking thought on the application of virtual control in multiple areas.

1 Introduction

1.1 Background

Augmented Reality is a field in computer graphics and computer vision where the live view of the physical world has been enhanced with virtual (computer-generated) features such as graphics or audiovisual media. These can be invoked by sensory input such as markers being hit by a camera, or GPS and direction data. [Wik] [Damgrave 2014]

Augmented reality is a form of mixed reality where the live view of a real-world environment is enhanced by virtual (interactive) overlay techniques.

Virtual objects can display information the user is unable to detect. This increases the user's perception of the real world and intereraction within it, provididing potential for assistance. [Azuma 1997]

An output device is required to display the virtual object, either optically through systems such as virtual retinal displays (the augmentation is projected directly on to the retina) and optical see-through (the augmentation is projected on to a half mirrored optical lens overlaying a direct view on reality), video see-through (the overlayed world is viewed through a camera's live feed), or spatially (the augmentation is projected). It is worth noting that spatial displays are limited for interaction. [van Krevelen and Poelman 2010]

Input devices are required to register interaction with AR virtual objects, this is most commonly part of the output hardware or a peripheral to the output hardware rendering the scene, but this is not always the case; common examples includ, a touch screen interface on a mobile device rending the image in video see-through or a gyroscope on a headmounted display. Interaction is performed in relation to what is seen on the output device.

Multimodal AR interaction is a recently emerging alternate method of interaction. The observer uses natural physical interaction with the virtual objects in the scene (as if the objects were real). This can be hand gestures, tracking and interaction with the scene itself. [Furht 2011]

1.2 Outline of project

This project serves to analyse virtual control in augmented reality as an advanced visual interface.

Virtual control is both in reference to multimodal augmented reality interaction, and usability of the interaction medium to directly influence the behaviour of the augmentation.

The virtual objects that give the means to interaction within the scope of the project are *virtual buttons*. Virtual buttons provide for a form of multimodal interaction; the button itself is part of the augmentation and interaction is performed through natural physical means. The virtual buttons are created to enable a user to manipulate another virtual object in the scene.

The notion of control is introduced by creating a manipulation task for the user to perform with the virtual buttons to achieve some measurable outcome.

A user study is established where the participants of the task are observed and asked to complete a questionnaire covering their experiences of using it, their opinions on the state of the technology and views on scenarios for its potential use.The resulting information is used in analysis to draw insight into the usability and sentiments on this advanced visual interface and its interation form.

Figure 2: *An example of an image target used with clear physical reference to the virtual button areas on which the virtual scene is rendered.*

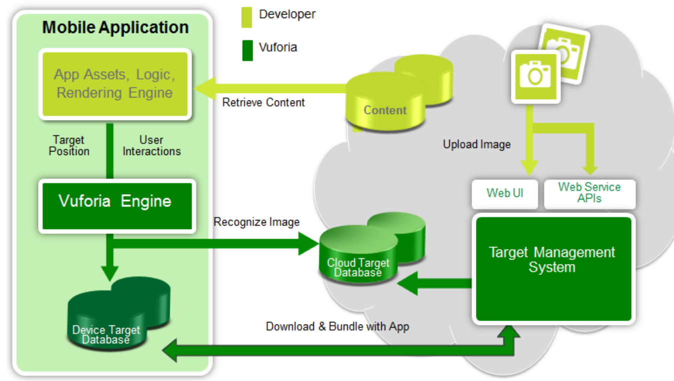

Figure 3: *Overview of the application development process with Vuforia; comprising of the Vuforia Engine (within the SDK), the Target Management System hosted by Vuforia, and its subscription option, the Cloud Target Database. [Vuf a]*

2 Implementation

The project requires both the development of an augmented reality application and construction of a usability study with associated questionnaire.

2.1 AR application

2.1.1 Tools

The AR rendering device used, is a video see-through Samsung Galaxy Tab 10.1 touchscreen tablet device. The device runs a version of the Android operating system.

Development of the augmented reality app is performed by the use of several different components:

Unity A game creation system which includes a game engine and integrated development environment (IDE). [Uni]

Qualcomm Vuforia SDK A software development kit that facilitates creation of augmented reality applications. [Vuf c]

Android SDK Tools A software development kit for the building and debugging of Android applications. [And]

2.1.2 Realising AR

The virtual objects are rendered on image targets. Image targets are used as opposed to markerless tracking, for clear spatial arrangement of the virtual buttons. The image that acts as the target for augmentation is adorned with clear separate areas signifying buttons (Fig 2.).

Images are uploaded to a server in the service provided by Qualcomm Vuforia. The server then uses image feature extraction to create a recognition mask. The images are then placed in a target database, which is downloaded to become part of the AR Android application (Fig 3.). [Vuf a]

The Vuforia SDK for interacting with the image targets is imported into Unity, where an `Image Target` is placed into the `Game Scene` as a plane for the virtual objects to be relatively spatially assigned.

The main camera used within Unity is replaced with an `AR Camera` from the assets imported through the Vuforia package, which is placed above the scene.

Within the area of the image target reserved for the visible virtual objects, three models are placed. Two are constant, namely two Radio Tower models available for free from Unity on the Asset Store; the third is a simple `3D cube` mesh of reasonable width and breadth, set at at a much lower scale transform value on the y-axis. The cube is the object under control by the user.

`Virtual Button` objects are dragged into the scene from the `prefabs` within the assets imported as part of the Vuforia SDK. They are positioned over the entirety of the button section within the `Image Target` plane.

Hierarchically, all virtual objects that are to be part of the augmentation, both visible and unseen (in the case of the virtual buttons), are placed under the `Image Target`. This is to signify that the image target needs to be present in order to render the virtual objects.

The interaction of the virtual button with the cube has been scripted in the MonoDevelop utility that is a part of the Unity suite. The scipts are written in C#. Vuforia has a number of scripts found within the SDK assets to get the virtual object values from each button object in the scene; from this, a number of method loops are written to alter the y-axis scale transform properties of the cube in response to an action on a certain virtual button object (Listing 1.).

The registered interaction with the virtual buttons provided in the SDK is achieved by monitoring occulusion of `Virtual Button` objects in reference to the image features located beneath them on the image target upon which they are placed.

Once the interaction has been scripted, the scene is complete. The application is then built through the Android settings within Unity which, with the Android SDK Tools, pushes the developed app to the tethered Samsung tablet device (set in developer debugging mode).

```
if(mCuboid.transform.localScale.y < 1){
  mCuboid.transform.localScale +=
                    new Vector3(0f,0.1f,0f);
  mCuboid.transform.localPosition +=
                    new Vector3(0f,0.05f,0f);
}
```

Listing 1: *Code snippet of function receiving the plus button* `switch` *case that increases the height of the virtual cube by increments of* 0.1 *to a max of* 1.0*.*

The application is launched and has the capability to produce multimodal interaction within the set augmented reality environment.

2.2 User study

2.2.1 Tools

The user study is made possible via the use of a number of tools:

Device with AR Application The device (in this case, a Samsung Galaxy Tab 10.1) with the built AR application developed at the previous stage (2.1) enables the user study.

Image Target A representation of the image used as the assigned target during development, printed, needs to be present for the participant to carry out the task else the AR application will not find any place to render the virtual scene.

Survey/Questionnaire A means to record user observations, time taken, accuracy of result. A series of open and closed questions for the participant to answer.

Stopwatch To measure the time taken by the user.

Pen To write on the survey/questionnaire.

2.2.2 Method

As multimodal interaction with AR is a recently emerging interaction medium, finding participants new to the concept is not a difficult task. [Furht 2011]

Participants of the study are given a printed image target (where possible) and a device with the AR application open. Most participants are informed of the method of interaction (a new concept), some are not. All are given the task: using the virtual buttons, change the height of the cube to match that of the radio towers[1].

Participants of the task are measured in time and overall accuracy of completing the task. Observations are made during the task and are written on the survey/questionnaire. No identifiable data is taken.

To measure usability and experience, a series of open and closed questions have been created. Closed questions (yes/no answers) ask the participant if:

- they enjoyed the experience,
- they considered the interaction as intuitive,
- they felt and added sense of immersion compared to standard control of computed geometry with a keyboard and mouse,
- they felt any greater sense of accomplishment than if they had performed the task with a keyboard and mouse,
- they think this technology and method of interaction is viable for control in hazardous environments or control of things out of reach.

There is one further closed question:

- how responsive did the participant perceive the virtual buttons in the task?

where answers are written on a one-to-ten scale; *one* labeled as unresponsive, and *ten* labeled as very responsive.

The questions help detail the usability of virtual buttons in this context, and the experience of the participants using virtual buttons for control. The participants that are not informed on how to interact provide an insight into the current understanding of virtual interaction.

[1]The radio towers are set at 0.4 and the cube is initially set at 0.1

Accompanying the closed question there is one optional open question asking the participants:

- why is this a viable technology for control, where can they see it being used?

In addition to this the participant is told that they made add additional comment to any of the questions or otherwise where there is space on the survey/questionnaire.

2.2.3 Data set

- 10 participants in the study
- 3 participants are not told how to interact

This is a small data set, primarily due to the time allocated towards the project. Nevertheless, a set size such as this provides some initial description of trends in user data.

3 Results[2]

3.1 Survey

Out of the ten participants in the study:

- The fastest participant took 13 seconds
- The slowest participant took 115 seconds

The mean time was 46.9 seconds, the median time 50.29 seconds. The range was 102 seconds. The slowest two participants were part of the 3 that were not informed how to interact. The second slowest taking 90 seconds, a full 30 seconds longer than the slowest of the participants that were shown how to interact. The quickest of the 3 not shown was only 5 seconds faster than the slowest of those that were.

Almost all participants completed the test successfully; 2 participants did not (Fig 4.).

3.2 Questionnaire and comments

Everybody enjoyed the experience of interacting this way (Fig 5.).

No participants assessed the responsiveness by giving a rating between 0 and 2. One participant gave a rating of 9. The majority of ratings for how responsive the task was fell between 5 and 8 (Fig 6.). The mean value was 6.2, the median was 6.5, the mode was 7 and the results had a range of 5.

8 of 10 participants perceived the interaction as intuitive (Fig 7.). One participant commented that keeping the buttons in view was tedious. Another comment stated that without the (tactile) feeling of a button, having the binary state buttons factor time in to iteratively increment or decrement the values as long as the button was occluded would serve better interaction. One further comment mentioned how it would be more intuitive if it was possible to grab and extend the box. One participant mentioned that their initial response would have been no, as it was not clear how to interact; this participant was part of the 3 that were not shown how to interact. 1 of the 2 users saying “no” was part of the 3 not shown how to interact. This user also commented saying not to let users try and figure it out and said during the task that they felt “stupid” not knowing what to do. Other users in that group appeared similarly confused, though less outwardly emotive.

[2]Actual survey/questionnaires with hand written observations form part of the submitted material.

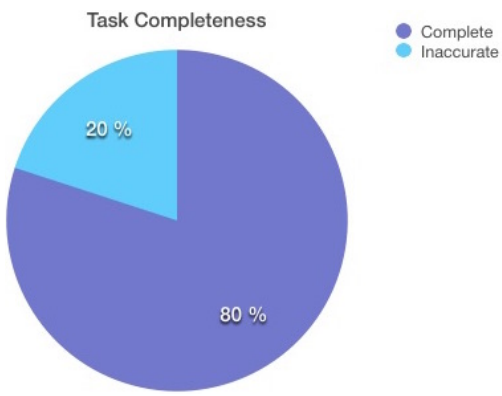

Figure 4: *Chart showing the percentage of participants accurately completing the task.*

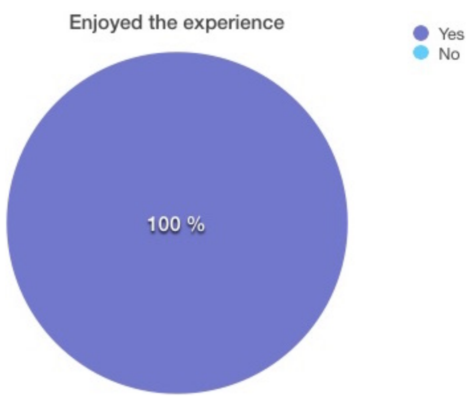

Figure 5: *Chart showing the percentage of participants finding enjoyment in the task.*

8 of 10 participants felt an added sense of immersion interacting with AR than standard control of computed geometry using a keyboard and mouse. One participant was undecided on the matter (Fig 8.).

Half of the participants felt a greater sense of accomplishment than performing the same task in comparison to performing it with a keyboard and mouse. Half did not (Fig 9.).

All participants agreed that this form of technology is viable for hazardous or out of reach environments (Fig 10.) For the open question extending this, one participant wrote that it was a good fit for wearable technology. Another user suggested a possible use in care or education for people with physical or mental difficulties, as they may be more aware of understanding the situation. The participant familiar with AR said it could be built into AR glasses enabling wearers to "push" surrounding buildings and stores to get information about the place. There was one participant that suggested use for it in gaming and security, and brought up a scenario for entering numbers on lockpads, with the advantage that it was unsusceptible to uncovering prints on the keys.

4 Analysis

4.1 Time

The disparity in the user times has a large amount of relation to the amount of intruige the participant had in the method of interaction.

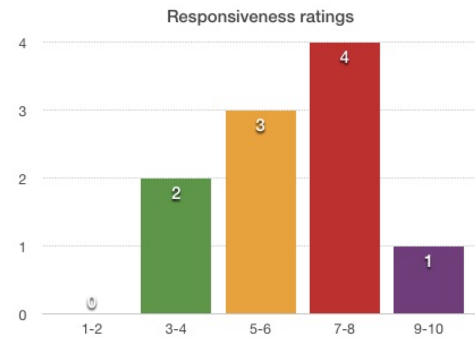

Figure 6: *Chart showing the spread of responsiveness ratings given by participants*

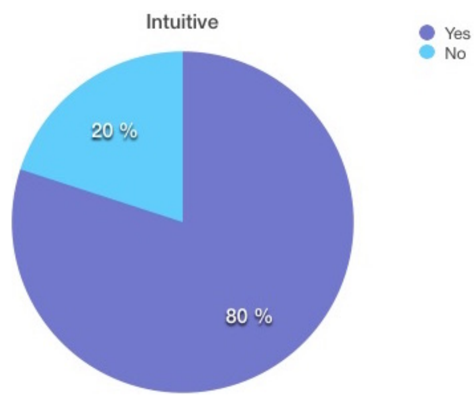

Figure 7: *Chart showing the percentage of participants that considered the interaction within the task as "intuitive".*

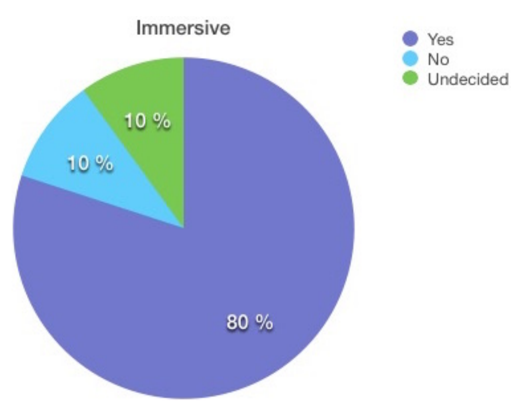

Figure 8: *Chart showing the percentage of participants that felt an added sense of immersion compared to using a keyboard and mouse.*

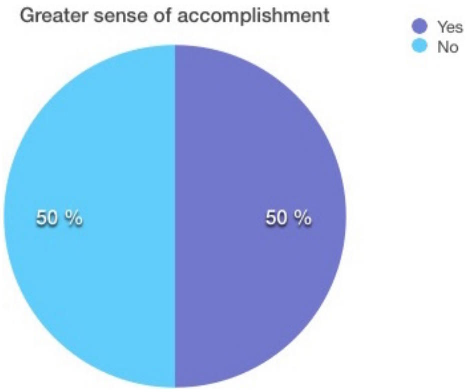

Figure 9: *Chart showing the percentage of participants that felt a greater sense of accomplishment compared to using a keyboard and mouse.*

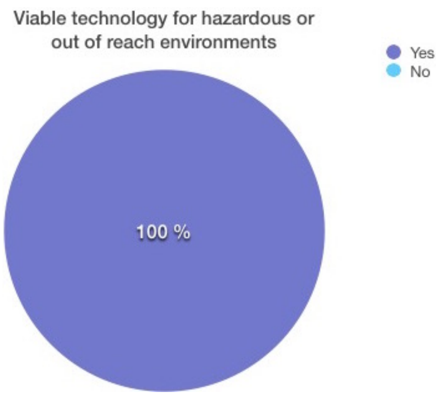

Figure 10: *Chart showing the percentage of participants that saw virtual buttons as viable for use in hazardous environments and out of reach scenarios.*

Some would sit testing the interaction, others would want to try both buttons even though the task completion would not require it.

4.2 Completeness

The test was designed to use different aspects of augmented reality. The two participants that did not complete the task (i.e. they said they had finished when the height levels were not the same) did not adjust perspective with the device. All participants started the task with the device angled down towards the image target. It is harder to perceive height from above, and the increments in height were subtle enough that a side view of the scene would give more information. This shows a limited understanding within those participants toward the tracking functionalities of AR.

4.3 Enjoyment

The fact that all participants enjoyed the task, which is fairly menial, highlights the "novel" nature of AR and virtual object interaction at this present state in time.

4.4 Responsiveness

The mean responsiveness value is 6.2, this is not particularly robust. The main reason being the interaction with the virtual buttons were not perceived as binary switches requiring repeated occulsions. The participants' general discourse approaching the buttons was that it of expecting it to iteratively increment or decrement while the button was being occuluded. Only one participant commented on this however, the other's assuming that it was just an issue with responsiveness instead of form factor.

4.5 Intuitiveness

A large majority of the participants agreed that the interaction was intuitive; the method of interaction was natural, easy to use and understand. This was also the impression given by 2 of the 3 that were not shown how to interact. Though one commented on saying it was only intuitive once they had figured it out - providing further discernment toward the footprint augmented reality and multimodal interaction disseminates.

The comment on replacing functionality to match a timed increment or decrement on prolonged occulsion is something that would contribute to the perceived intuition of using the interaction medium. The commenter was not alone in this perceived method of interaction. If the participants received something that they'd expect from a using multimodal interaction (nature is far from a binary concept), then this would increase the intuitiveness.

Further research into haptic devices would have to be done to fulfil the want to be able to grabbing the virtual object, which is beyond the scope of the project.

The participant in the small group that were not shown how to interact commented that "users should not have to figure this out" and felt "stupid". Again, this highlights the exposure surrounding this emerging form of interaction.

4.6 Immersion

A participant commented complaining on how keeping the scene in view the whole time was frustrating. The main component of this problem is the use of a hand-held video see-through device. If the study was performed again using an optical see-through device, or head-mounted-display video see-through device, the user can see the virtual objects without holding a device. Having to carry something around to view a scene you manipulate presents more reason for the interaction to be done on the device as opposed to within the virtual scene.

8 out of 10 shows the promise of this form of technology and interaction, it is more involving than that of sitting at a desk in the office.

4.7 Accomplishment

Half of the participants agreed with the statement saying they felt greater accomplishment in the task they undertook than had they performed it with keyboard and mouse. Half said they did not feel any extra accomplishment. A reason these results are so divided could be the importance a participant places on performing tasks with their hands; which raises the question of if the participants considered the interaction as a skill.

4.8 Thoughts on viability

The fact that all participants agreed on the viability of this technology for out of reach and hazardous environments indicates that every one of them understood the applicability of occlusion as an input facilitator for interaction with virtual objects and what that means in real-world scenarios.

As one participant had written, this form of interaction is "a good fit for wearable technology". There is growing research being done on such wearable systems for multimodal interaction in augmented reality. [Mistry et al. 2009] [Siu and Herskovic 2014] [Roggen et al. 2014]

The suggested use in care or education for persons with physical or mental difficulties is another area in research. Combining multimodal interaction with virtual objects using haptics would provide a new area in tangible interfaces and simulation possibilities, giving rise to a more accessible means of understanding for both health-care and persons with disabilities. This collaboration of virtual objects and haptics lends itself well to the previous proposal of being able to grab the cube. [Escobedo et al. 2014] [Harders et al. 2007]

Spatial awareness is a considerable part of using augmented reality with virtual buttons for hazardous and out-of-reach situations. The proposition by the participant familiar with AR to be able to "push" surrounding buildings and stores to ascertain related information lends itself well to the application of the technology.

The idea by a participant for its use in security is interesting. Proposing a virtual button system for opening locked doors traditionally locked with keypads to eradicate dusting for fingerprint residue is an intruiging concept. Systems akin to this will require throughly extensive analysis and testing. The other idea was for gaming, presenting a move towards bringing portability and collaboration as the user is only restricted by the surface to project upon - though there are systems for users to define their own image targets, in addition to markerless tracking. [Vuf b] [T-I]

5 Discussion

5.1 Limitations

As stated in 2.2.3, the data set was small. This was due to time contraints of development the project.

There are certain issues to consider with the features contained in the target images. Even the image in Fig 2. is not ideal, due to the lack of features present on the placeholders for the virtual buttons. This sometimes caused autonomous manipulation of the virtual objects due to the low amount of features to track as occluded - a discrepency in the received feed could be treated as an occlusion. There are ways to combat this, but the easiest would be to add more features to the target areas that are subject to occlusion for virtual button functionality.

The use of a hand-held video see-through device was not optimal for a task that required hands-on interaction. A head-mounted device (e.g. AR glasses) would have been a better fit for interaction with the virtual objects.

5.2 Concluding remarks

This project has detailed the user approach and opinion on virtual control in augmented reality through interaction.

The overall impression of the advanced visual interface from the participants was of genuine interest, though the binary state approach of the buttons was not optimal for interaction. However, the method of virtual control in augmented reality was widely received as having tremendous potential in many fields, as is in line with many authors, and was interesting to witness how thought provoking the technology made the participants, judging by the responses they gave. [Furht 2011] [Damgrave 2014] [Roggen et al. 2014] [Azuma 1997] [van Krevelen and Poelman 2010]

References

Android SDK Tools. `https://developer.android.com/tools/sdk/tools-notes.html`. Accessed: 2014-12-17.

AZUMA, R. T. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments 6*, 355–385.

DAMGRAVE, R. 2014. Augmented reality. In *CIRP Encyclopedia of Production Engineering*, L. Laperrière and G. Reinhart, Eds. Springer Berlin Heidelberg, 66–67.

ESCOBEDO, L., TENTORI, M., QUINTANA, E., FAVELA, J., AND GARCIA-ROSAS, D. 2014. Using augmented reality to help children with autism stay focused. *Pervasive Computing, IEEE 13*, 1 (Jan), 38–46.

FURHT, B., Ed. 2011. *Handbook of Augmented Reality*. Springer.

HARDERS, M., BIANCHI, G., AND KNOERLEIN, B. 2007. Multimodal augmented reality in medicine. In *Universal Access in Human-Computer Interaction. Ambient Interaction*, C. Stephanidis, Ed., vol. 4555 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 652–658.

MISTRY, P., MAES, P., AND CHANG, L. 2009. Wuw - wear ur world: A wearable gestural interface. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, USA, CHI EA '09, 4111–4116.

ROGGEN, D., PEREZ, D. G., FUKUMOTO, M., AND VAN LAERHOVEN, K. 2014. Iswc 2013–wearables are here to stay. *Pervasive Computing, IEEE 13*, 1 (Jan), 14–18.

SIU, T., AND HERSKOVIC, V. 2014. Mobile augmented reality and context-awareness for firefighters. *Latin America Transactions, IEEE (Revista IEEE America Latina) 12*, 1 (Jan), 42–47.

T-Immersion D'Fusion SDK. `http://www.t-immersion.com/products/dfusion-adobe-flash/dfusion-sdk-adobe-flash#`. Accessed: 2014-12-17.

What is Unity? `http://unity3d.com/pages/what-is-unity`. Accessed: 2014-12-17.

VAN KREVELEN, D. W. F., AND POELMAN, R. 2010. A Survey of Augmented Reality Technologies, Applications and Limitations. *The International Journal of Virtual Reality 9*, 2 (June), 1–20.

Developing with Vuforia. `https://developer.vuforia.com/resources/dev-guide/getting-started`. Accessed: 2014-12-17.

User Defined Targets — Vuforia. `https://developer.vuforia.com/resources/dev-guide/user-defined-targets`. Accessed: 2014-12-17.

Vuforia SDK. `https://developer.vuforia.com/resources/sdk/unity`. Accessed: 2014-12-17.

Wikipedia: Augmented Reality. `http://en.wikipedia.org/wiki/Augmented_reality`. Accessed: 2014-12-17.