Augmented Reality (AR) Mirror on Mass Market Hardware

Considering Mass Market AR on Commercial Off The Shelf (COTS) Hardware like Laptops, Tablets and Phones

MÅRTEN NORMAN
Augmented Reality (AR) Mirror on Mass Market Hardware

Considering Mass Market AR on Commercial Off The Shelf (COTS) Hardware like Laptops, Tablets and Phones

MÄRTEN NORMAN

Bachelor’s Thesis at CSC
Supervisor: Björn Thuresson
Examiner: Katarina Gustavsson
Abstract

Augmented Reality (AR) is studied in this thesis from a desire to deploy AR products in large scale within a few years. This assumes a plain software solution, using Commercial Off The Shelf (COTS) hardware as existing laptops, tablets and phones. The thesis scope is a mirror metaphor, the device mimics a mirror and AR is presented in the mirror. The mirror is from looping video from the camera facing the user to the screen and real-time AR graphics is added to the video stream. A video mirror is slightly different from a physical mirror and the consequences in this context are considered.

The more obvious video mirror from operating a device in selfie-mode – recording a video of oneself – is recognized to be used for AR features in large scale without being marketed as an AR product. This is seen as sign of AR maturity in the discussion.

A less recognized video mirror, by looping video from a work area on a physical desk and apply AR on objects in the work area is studied. The focus is the potential of using a device’s face oriented camera to create an AR mirror from unmodified COTS hardware. Even if AR is a well researched area, the mirror metaphor by video loop has been almost neglected. This can be explained by a) physical mirrors are generally better for AR, especially in lab situations or unique equipment, b) the AR research field matured long before common COTS screen devices had cameras and c) AR in mirrors has less immersion than AR in a direct view.

The basic properties of an AR implementation in a laptop and a tablet are studied, obstacles are found and possible ways to handle them are presented. The general recommendation though is to add a USB camera to the laptop or a tiny mirror to the tablet. Literature research indicates that AR using the mirror metaphor offers a noticeable improvement compared to screen or paper based presentations when used in relevant situations.

It is concluded that AR is maturing. There will not be an AR revolution but we are in the beginning of an evolution where AR is included for its ability to improve user interaction rather than for being impressing technology.
Sammanfattning


En uppenbar video-spegel – där användaren ser sig själv – används redan i stor skala, utan att egentligen marknadsföras som AR. Detta betraktas i diskussionen som ett tecken på att AR mognat.

En mer ovanlig video-spegel – där en arbetsyta på ett fysiskt skrivbord speglas och AR kombineras med objekt på arbetsytan – studeras. Fokus ligger på möjligheten att använda en omodifierad standardenhets ansikts-kamera. AR är ett välstuderat område, men delområdet AR i en videospegel är nästan helt försummat. Detta kan förklaras med a) fysiska speglar är oftast bättre för AR, speciellt i laboratoriummiljö och unika installationer b) forskningsområdet AR mognade långt innan konsumentelektronik fick kamera i skärmarna c) AR i speglar ger sämre närvarokänsla än AR direkt i synfältet.


# Contents

1 Introduction ........................................... 1
   1.1 Purpose ........................................ 1
   1.2 Technical Innovation – Incremental and Breakthrough .... 2
   1.3 Augmented Reality (AR) ................................ 3
       1.3.1 What is AR? – Some Definitions ............... 4
       1.3.2 Tracking and Registration .................... 4
       1.3.3 Fiducial Markers ............................... 5
       1.3.4 Natural Feature Markers ...................... 5
       1.3.5 Simultaneous Location And Mapping (SLAM) ....... 5
   1.4 The Human Vision System and Image Technology as Cameras and Displays ........................................... 6
       1.4.1 The Human Vision system and Image Processing .... 6
       1.4.2 Mirror Physics and the Human Brain – the Psychology of mirrors ........................................... 7
       1.4.3 Digital Cameras and Video Processing ............ 8
       1.4.4 Display Technology ............................... 8
   1.5 Problem Statement .................................. 10
       1.5.1 Research Question ............................... 10
       1.5.2 Scope and Delimitations .......................... 10

2 Background and Related Work ........................................... 13
   2.1 Surveys of AR ...................................... 13
   2.2 Product: The Snapshat Filter Feature .................... 14
   2.3 Product: Osmo Accessories to iPad brings AR to the Desktop .... 16
   2.4 Product: Ericsson Piero AR ................................ 17
   2.5 Study: Miragetable .................................. 18
   2.6 Study: Virtual Try-on using Kinect and HD camera ........... 19
   2.7 Study: Monitor based AR for industry .................... 19
   2.8 Study: Augmented Reflection of Reality ................... 21
   2.9 Study: Through the Combining Glass ..................... 21
   2.10 Study: AR on Screen for Architects and Engineers in Civil Construction ................................. 24
   2.11 Study: AR as User Interface ............................ 24
2.12 Study: Intelligent Augmented Reality Training for Motherboard Assembly

2.13 Study: Computer-Augmented Mirror to Aid Makeup Application

2.14 Study: Metamorphosis Hand: Humans Can Accept Certain Interface Errors

3 Method – Mirror Workspace Feasibility Test

3.1 Test Background and Discussion

3.2 Test Setup

3.3 Device: ASUS 14” Laptop

3.4 Device: Samsung Galaxy Note

4 Results – Mirror Workspace Feasibility Test

4.1 Device: ASUS 14” Laptop

4.2 Device: Samsung Galaxy Note 10.1 2014 edition

5 Discussion

5.1 Mass Market Products Employing Augmented Reality

5.2 Mirror Metaphor Looking at the Workbench

5.3 Mirror Metaphor Looking at Yourself

5.4 AR Impact on Society and Ethical aspects

6 Future Work

6.1 Product Design Related

6.1.1 Evaluation of Cameras in Commercial Off The Shelf (COTS) Hardware

6.1.2 Study Potential Applications for AR mirrors

6.2 User Interaction Related

6.2.1 AR Mirror on COTS Laptop – user interaction

6.2.2 Human Interaction in the Mirror World

6.2.3 Potential of AR to Make Architects and Similar Professionals to Give up Paper

6.3 AR Technology Related

6.3.1 Studies to add mirror metaphor to AR mainstream

6.3.2 Activity recognition

6.3.3 AR Mirror on COTS Tablet – Occlusion

6.3.4 Hand and face tracking

6.3.5 Generic Tracking Inspired by Face Tracking Algorithms

6.3.6 Merge Hands and Arm Tracking and Human Interaction Algorithms from Robotics

Bibliography
Chapter 1

Introduction

1.1 Purpose

Augmented Reality (AR) is a technology with potential to be an obvious part of everyday human life within a few decades. Currently, the concept level of Augmented Reality is well researched. The implementation is far from mature though. There are some products in niches, but most activities are still experiments in lab environments. Hence Augmented Reality has an evolution progress right now. It offers incremental improvement in niches, and the number of niches is likely to increase as Augmented Reality technology improves.

Still though, there can be a sudden revolutionary progress of AR. There is a potential opportunity to instantly upgrade millions of devices with Augmented Reality by a simple app installation. Most of the current laptops, tablets and smartphones can run Augmented Reality applications. The main blocker is not technology, but an interesting application of the technology. Hence it is interesting to look for an intermediate level of Augmented Reality that fits within current limits of Commercial Off The Shelf (COTS) hardware.

This thesis will focus on exploring a mirror metaphor variant of Augmented Reality that has promising properties for use on a large base of Commercial Off the Shelf Hardware and software. The mirror metaphor exist in literature, but is not acknowledged in the rather complete mainstream taxonomy of Augmented Reality and is thus a less explored area of Augmented Reality.
CHAPTER 1. INTRODUCTION

1.2 Technical Innovation – Incremental and Breakthrough

An important key to sustainable innovation is risk management (figure 1.1, adapted from ). A technical breakthrough is very desirable, but hard to predict and it can take decades of hard work until it happens. Also during the breakthrough there is significant insecurity and risks. Most great innovations has failed several times before they reached recognition and acceptance. Hence a reliable and predictable progress is also a desire, suggesting a conservative choice of mature technology and domain. When adding the business risks, there may be further reason to reduce the technical risk. One popular example of successful innovation is the Apple iPhone in 2007. But already in 2005, it was an attempt to merge Apple’s success of iPod

---

1.3. AUGMENTED REALITY (AR)

and iTunes with Motorola’s successful phones. The result, Motorola Rokr, failed\(^4\)\(^,\)\(^5\) though. Later, in 2007 Apple took the risk of releasing their own phone. This was a significant risk, considering that Apple had no history of making phones. There was a lot of risk reduction for the iPhone project though. Apple leveraged both web browsing by their mature Safari browser and iTunes music by positioning the iPhone as the most advanced iPod product.\(^6\) And it worked. A significant number of loyal Apple customers trusted the brand enough to consider a technically average GSM telephone superior to competition – including 3G phones from Nokia and Sony Ericsson. By using mature technology, Apple reduced their technical risk and by using their existing customer base Apple reduced the market risk. Hence the iPhone utilized the middle zones of figure 1.1 to reduce the innovation risk. And the next iPhone included 3G technology, offering an incentive for upgrading.

1.3 Augmented Reality (AR)

Augment means “Make (something) greater by adding to it; increase.”\(^7\)

Ronald Azuma is one of the pioneers in recent AR. In a seminal paper\(^8\), he provides a thorough survey of the research in 1997 together with a good introduction to the AR field. During this period a lot happened in AR, resulting in a complementing update\(^9\) in 2001. AR research is still prospering, but focus is shifted into different research topics with their own surveys. A few of them are mentioned in section 2.1.

Oliver Bimber and Ramesh Raskar has performed thorough and seminal studies in the AR subtopic of spatial AR, that is presented further in section 1.4.4. Their research is documented in a book\(^10\) and an extensive siggraph paper\(^11\).

---

\(^4\)Tynan, Dan (2006). *The 25 Worst Tech Products of All Time*. “Motorola Rokr E1 (2005): The world’s most popular digital music player meets the world’s coolest looking phones; what could possibly go wrong? Well, plenty. The Rokr E1 held only about a hundred songs, file transfers were painfully slow, the iTunes interface was sluggish, and—duh—you couldn’t download tunes via a cell connection. This phone ain’t rockin’, so don’t bother knockin’ ”. [url](http://www.pcworld.com/article/125772/worst_products_ever.html?page=7) (visited on 03/26/2016).

\(^5\)West, Joel and Mace, Michael (2010). “Browsing as the killer app: Explaining the rapid success of Apple’s iPhone”. In: *Telecommunications Policy* 34.5, pp. 270–286.

\(^6\)Ibid.

\(^7\)Oxford Dictionaries.


1.3.1 What is AR? – Some Definitions

Azuma offers a technical definition; “1) Combines real and virtual; 2) Interactive in real time; 3) Registered in 3-D”\textsuperscript{12}. In this context, registered means alignment of the real and virtual views.

There can be other definitions though; Prochazka and Koubek describe it as: “The general principle of the augmented reality (AR) is embedding digital information into the real world scene. Thus it is a step between virtual reality and the real world. The embedded information is usually based on the content of the scene. A selected real object could be augmented by a virtual object or completely replaced. Well-known examples are the presentation of the fighter status report on the head-up display before a pilot or a navigation information projected on the wind shield of a car.”\textsuperscript{13}

1.3.2 Tracking and Registration

To add useful 3D graphics into a user’s view-field, it has to be scaled, rotated and rendered to be aligned with the real world from the user’s viewpoint. Hence registration is a central term in Augmented Reality – the alignment of the virtual world and the real one. Good registration depends on good tracking of position and rotation for all relevant objects in the real world scene.

Poor registration result in 3D graphics looking disconnected from the real world by appearing in the wrong position or not moving properly. This can be a major problem if the 3D object has an exact visual relation to a real world object, for example should appear to be standing on it. For other cases poor registration may be a lesser problem, for example a text label floating beside an object.

This thesis will focus on tracking methods using the existing 2D camera hardware of a device, but knowing alternative tracking methods helps understanding the potential and limitations of the chosen tracking methods. There is a broad survey of tracking methods\textsuperscript{14} that helps understanding most – maybe even all – tracking options that can be used for AR.

\textsuperscript{12}Azuma (1997).
1.3. AUGMENTED REALITY (AR)

1.3.3 Fiducial Markers

A fiducial is a familiar object in a picture\textsuperscript{15}. The presence of a coin, a human or a millimetre ruler in a photograph is a well known way to help the viewer understand the scale of the image\textsuperscript{16}. For a 2D camera, it is fairly easy to track a known pattern like an image of a QR-code with predefined size. The QR-code is associated to a surface in the room or on a movable object – a position. The size of the QR-code tells distance, while deformation allows calculation of a surface normal.

1.3.4 Natural Feature Markers

It is also possible to use ordinary objects as fiducial markers. The object must be known by the tracker though – a marker is supposed to be recognized – including shape and size. Recent commercial trackers can dynamically load markers from an online database\textsuperscript{17} and even recognize 3D objects\textsuperscript{18}. The texture has to be unique and easy to track in different light conditions though.

A human face is also possible to use as a marker\textsuperscript{19,20} by using appropriate algorithms. The same is valid for human hands\textsuperscript{21,22}. This implies that almost any object can be tracked, given sufficiently advanced and optimized algorithms are developed.

1.3.5 Simultaneous Location And Mapping (SLAM)

SLAM is a tracking algorithm with roots in robotics. An autonomous robot should be able to navigate by vision in unknown areas. SLAM allows tracking by finding markers, dynamically selecting visual details in the view-field. This allows for robust tracking and registration of camera position relative to the environment. There is often need for a known marker to get a starting position, but SLAM maintains

\textsuperscript{15}Azuma (1997).
\textsuperscript{16}Example: The centimetre scale in figure 4.1.
stable tracking if the known markers disappear or get occluded.

The most relevant SLAM algorithm, in the thesis context, is MonoSLAM\(^{23}\) that works for a single camera without depth-sense, like the cameras on current mass market COTS hardware. MonoSLAM or similar algorithms are included in recent commercial trackers\(^{24,25}\) for cell phones and tablets. Users want to move the camera freely and suspension of disbelief\(^{26}\) is lost if the AR image suddenly disappear when the marker is at the edge of the screen. SLAM maintains tracking by visual details and improves robustness significantly without the hassle of adding several physical marker objects.

### 1.4 The Human Vision System and Image Technology as Cameras and Displays

The human eye and a digital camera share some properties, but there are massive differences too. Common properties of an eye and a typical camera unit in a tablet or phone are: a) image detector b) lens c) aperture d) real-time image processing. The camera is a passive receiver though, while the eye is extremely active. It works in close connection to a conscious brain, an active body and our other humans senses to interact with the world. The human eye is part of a complicated system\(^{27}\). Understanding human vision includes considering the complete system including a conscious human mind.

#### 1.4.1 The Human Vision system and Image Processing

The human eye has several properties\(^{28}\) that differs significantly from a digital camera. The image sensor – the retina – has very good resolution for an area equal to a thumbnail\(^{29}\) at arm-length’s distance. Then the density of colour image sensors – cones – drops, but keeps reasonable colour vision within an area equal to a face at arm-lengths distance. The remaining view-field is dominated by black-and-white image sensors – rods – optimized for low light conditions and detection of moving objects.


\(^{26}\)Voluntary suspension of disbelief is a frequent term in media and culture, describing the user’s ability to voluntarily and temporarily accept an illusion or unbelievable story to take part of it.


1.4. THE HUMAN VISION SYSTEM AND IMAGE TECHNOLOGY AS CAMERAS AND DISPLAYS

We are generally not aware of the eye’s limitations though. The human brain tracks all relevant objects nearby and predict their ability to move. Our attention is focused on the most urgent or interesting objects and the eyes shifts quickly between them – a mostly unconscious activity handled by the spatial parts of the brain. The eyes are deeply involved, but supported by head and body movements while the ears, nose and touch senses support the tracking. One extreme but interesting example is when we see an unfamiliar object and feel an urge touching it to quicker understand its shape and properties.

Both the movie and comics industries depends on our visual system’s ability to reconstruct short samples of images into a meaningful series of events. We expect sound to be continuous, but the vision is optimized for quick samples of key moments. Hence it is possible to communicate a long walk home by a few clips of a person walking in different places – as long as the walking direction is consistent. The brain can easily imagine the long walk between the clips without observing every single step.

The most important aspects of human vision in the context of this thesis are the human abilities to a) accept and compensate for at least some visual defects and anomalies b) shift visual view frequently while maintaining continuity of events. The human visual system has superior tracking and registration compared to recent AR.

1.4.2 Mirror Physics and the Human Brain – the Psychology of mirrors

While a point on a 2D screen sends the same light in all directions, each point on the mirror reflects light from several angles. Moving the head or eyes will give a different view through the mirror – similar to a view through a window. This makes a physical mirror image much richer than a similar image presented on a 2D screen. Still, humans tend to recognize a mirror-like video stream on a 2D screen as a mirror and suspend disbelief. This can be due to our poor ability to predict how things should look in a mirror. Our mental models are situation specific and adapts to what we expect to see rather than following the actual physics of

---

32This is actually true for most animals with eyes, including lizards and insects. Vision is an extremely mature and proven system in nature.
CHAPTER 1. INTRODUCTION

Figure 1.2: Smartphones with built-in camera are able to replace some low end Digital Still-image Cameras (DSC) while the advanced cameras (with interchangeable lens) keeps a stable market position. Figure source

a mirror. This human property makes it relevant to assume that mirror-metaphor Augmented Reality may have less immersion than see-through implementations while still having reasonable potential to be accepted by a user.

1.4.3 Digital Cameras and Video Processing

The digital cameras of a tablet, phone or laptop are cost optimized and limited by size constraints. The image sensor and lens are much smaller compared to a good full-size camera – saving space and cost. They produce images of reasonable quality, but full size and especially professional cameras do make better images. The phone cameras are good enough to take market shares from basic digital cameras (figure 1.2) though. In reasonably good light, even\textsuperscript{36} the face oriented camera is able to deliver an HD\textsuperscript{37} video stream – at least on recent devices. The face oriented camera does often have a fixed focus lens resulting in infinite depth of field. The small camera size allows using a fixed focus lens without the usual penalty of noticeable unsharp image at close range. Examples in section 2.1 also indicates that the face camera quality is sufficient for AR. This is good since the scope of this thesis is to use the face oriented camera for AR.

1.4.4 Display Technology

There are several ways to add the computer generated image into the view. All methods have their advantages and disadvantages depending on situation.

\textsuperscript{35}Battiato, Sebastiano and Moltisanti, Marco (2015). The future of consumer cameras. URL: http://dx.doi.org/10.1117/12.2083853.

\textsuperscript{36}The face oriented camera is usually the device’s lowest performing camera. The limited case of taking snapshots within armlength distance allows for cost and size optimizations.

\textsuperscript{37}A typical face camera has an 1-2 Mpixel sensor and delivers 1280×720 or 1920×1080 pixels at 30 fps
1.4. THE HUMAN VISION SYSTEM AND IMAGE TECHNOLOGY AS CAMERAS AND DISPLAYS

Figure 1.3: Augmented reality can be divided into classes from how they project the image. It can be a display between eye and object or the image can be projected directly on the object. The projector or display may be either head-attached, hand-held or spatial. (Adapted from\textsuperscript{39})

Figure 1.4: Table of display technologies and their characteristics. Figure from\textsuperscript{40} with interesting properties for this thesis marked in red.

This thesis will focus on a variant of spatial screen-based video see-through display.\textsuperscript{38} (Figure 1.3) Spatial means it is not attached to the viewer. Screen-based is a normal monitor, rather than for example a projector. Video see-through means that reality is viewed through a camera. Creating AR in a mirror view of a work area – rather than a view of the person looking in the mirror – by looping video appear to be novel for this thesis though.

\textsuperscript{38}Bimber and Raskar (2006).
\textsuperscript{39}Ibid.
1.5 Problem Statement

The problem is related to understanding AR better and in particular why there still is more hype than actual products used by ordinary people. This is a bit surprising since there is widely available hardware that should be able to use AR. Right now, in 2016, most computers are laptops equipped with a face oriented camera. The same is true for tablets and phones which are equipped with both a face oriented and a front oriented camera.

AR is a large and active research area due to the potential of improving user interaction. Hence there is a general interest to understand how to turn this potential into useful products that improves life as much as possible for as many people as possible. This thesis studies the potential to make this happen in near time by implementing AR on already available COTS devices.

Since AR is a very large field, I decided to focus on a very narrow implementation that I am labelling AR mirror metaphor. It loops video from the face oriented camera device back on the screen and adds AR in the mirror view. There is a huge available hardware base that is able to interact with users by implementing this AR mirror metaphor. The user can, at least in theory, just install software to start interacting in this way.

1.5.1 Research Question

The problem can be captured more formally in research questions:

- Is it really a good idea to interact via AR using a mirror metaphor? There should be an advantage in performance or experience compared to existing interfaces or ways of working.

- Is it technically possible to create a useful product that can be launched in large scale in the near time? And if the answer is “no”, what are the main blocking problems?

1.5.2 Scope and Delimitations

Focus is to understand the consequences of the limitations listed below and discuss the potential to create a useful product within those limits.

Large scale (from the research question) gives some premises:

- All necessary hardware and software must be available to ordinary customers\(^{41}\). This limits the thesis scope to Commercial Off The Shelf (COTS) hardware that is either on the market already or likely to be on market soon.

\(^{41}\)Available through normal retail channels at a reasonable price level for the context and not limited to developers or short in supply.
1.5. PROBLEM STATEMENT

- The studied application has to address a mass market need or be generic for a large number of applications that can be considered as a mass market together.

- The entrance barrier should be low compared to the perceived value of starting to use the application. This is mandatory for reaching large scale in a short time-frame.

Hence the scope is narrowed to:

- Only study the less explored video mirror metaphor and ignore the more common “see-through” metaphor. Reason: To bring something new to the AR research field by focusing exploration into a less studied area.

- Use video from the built in camera facing the user, ignoring options to add cameras or optics like lenses or mirrors. Reasons: a) The premise of low entrance barrier, assuming that it is hard to reach a perceived value that justifies acquiring and installing new hardware. b) Focus on applications that can use COTS hardware according to the premise listed above.

- An extra focus on a mirror metaphor variant that mirrors a workspace on a physical desk rather than the user’s face. This variant of video mirror loop appear to be even less researched than a face video mirror.
Chapter 2

Background and Related Work

This chapter presents a wide number of related applications and papers. They are selected for one or more of a) representing AR research that is close to the video mirror metaphor – most of it is recent b) contains results that can be re-purposed for the context of discussion in this thesis. c) indicates potential areas where COTS based AR is relevant and can be deployed in quantity. It is pointed out in the description of each work why it is interesting in the scope of this thesis.

2.1 Surveys of AR

AR is a well researched field and a number of reviews presents a structure for the research areas. The Mirror metaphor concept of mimicking a mirror by looping video from a camera to a display is effectively neglected, while other AR areas are well covered. Hence this thesis relates to a less explored area of AR research.

Azuma\(^1\) presents a complete and structured overview of AR as known in 1997. There is a complementing update\(^2\) in 2001. Both surveys are frequently cited seminal papers for recent AR research, providing common definitions and taxonomy. Here display technology is grouped as; See-through Head Mounted Displays, projection based and hand-held displays i.e. today’s typical phones and tablets. An independent survey\(^3\) in 2010 by Van Krevelen and Poelman covers the development since 2001. During recent years, AR research is over-viewed in more narrow areas, like ISMAR papers\(^4\) for the recent 10 years and a survey of AR assembly research\(^5\).

\(^{1}\)Azuma (1997).
\(^{2}\)Azuma et al. (2001).
\(^{3}\)Van Krevelen and Poelman (2010).
CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.1: A collage of some Snapchat filters. The common property is that they add or modify something in the face and maintains the change in real-time allowing it to be an effect in a video clip.

2.2 Product: The Snapchat Filter Feature

Snapchat is a well known social media app that allows their users to send movie clips and photos to other users. The key feature is that received media disappears after viewing or a certain time. Hence there is no time-line or history, following their slogan “Life’s more fun when you live in the moment”.

The Filter Feature in Snapchat proves that it is possible to deploy an interactive video mirror with AR on a significant share of the phones and tablets owned by users in 2016. It uses human faces as AR markers and offers several AR effects (figure 2.1) related to the face like a) change eye colour b) make skin smoother c) turning into a cyclops d) add cat features like whiskers and ears e) swap faces with another person in the picture. The main theme is entertainment and practical-joke-style special effects like turning into a zombie when screaming.

The AR tracking and registration is fairly reliable but not perfect. It appear to be optimized for robustness, sometimes detecting random patterns as a face – offering the spooky option to swap face with someone behind you who really is not there.

Snapchat filters is also an example of successful technical and business innovation. The original feature was developed by Looksery inc and a crowd funding campaign in 2014 resulted in app releases for Android and IOS devices. The app

http://dx.doi.org/10.1007/s40436-015-0131-4.


8Most devices with either Android 4.3 or above IOS) iPhone 5, iPad 4 or above.

2.2. PRODUCT: THE SNAPSHAT FILTER FEATURE

was technically successful; both Looksery and the app received media attention and awards. This app never reached mass market scale on its own though. This changed in September 2015, when Snapchat inc. integrated the technology\(^\text{10}\) as a feature in their already widespread and successful app. Snapchat inc. also acquired Looksery. AR did not reach a mass market by itself, but made it when integrated with other technology and in a relevant context.

Since lenses is a commercial product rather than research, there is no direct information in published papers. There are patents though – filed by Looksery founders and key developers – presenting some important features in detail\(^\text{11}\):

- Patent\(^\text{12}\) related to a filter component for changing the colour of for example an eye.

- Patent\(^\text{13}\) related to the trigger component for effects, for example “… opening a mouth turns a person to a zombie, closing one eye turns a person to a cat, raising a nose with a finger turns a person to a pig, or showing horns with fingers turns a person to a deer.”

- Patent\(^\text{14}\) related to tracking and modifying objects, but the more detailed legal claims are focusing on a front image of a face and moderate computing power like for a portable device\(^\text{15}\). “… parameterised face mask specifically developed for model-based coding of human faces. Its low number of polygons (approximately 100) allows fast reconstruction with moderate computing power. … The local Action Units control the mimics of the face so that different expressions can be obtained.”

Snapchat filters are included as a proof that a) it is possible to implement AR on a significant number of iOS and Android devices by downloading software b) the face oriented camera on mentioned devices can be used for an AR mirror c) an AR mirror application can be launched in large scale, given suitable infrastructure and context d) AR tracking and registration on said devices is at least good enough for casual use.


\(^{11}\)Patent law demands patents to disclose all information needed to easily reproduce the invention.


\(^{14}\)Shaburova (2015).

\(^{15}\)Both in direct words and by choice of algorithms suitable for limited processing power and optimized for a front image of a face.
Figure 2.2: The Osmo system is adding a mirror to the face oriented camera of an ordinary iPad. The rest is software, in this case a physics education program for kids. The objects in front of the device are mirrored to the screen and interacts with bouncing balls on the screen that can hit the round targets, also on screen. (Osmo)

2.3 Product: Osmo Accessories to iPad brings AR to the Desktop

Osmo is a combination of an iPad application, a tiny mirror for the face oriented camera and a set of toys that works as AR markers (figure 2.2). The iPad hardware is slightly modified by putting it in a stand and attaching a piece of plastic containing a tiny mirror for the face oriented camera. This redirects the camera to view a work area on the desk top in front of the standing iPad.

Osmo\textsuperscript{16} is extremely close to the AR mirror metaphor studied in this thesis. The only apparent difference is the choice to include a kit of physical items as a mandatory part of the product. This is likely to be a deliberate choice: Osmo’s solution depends on a mirror for the camera and it is significantly easier to have full control over the AR markers. Packaging the product as a physical box of items is an obvious solution – but physical distribution in combination with a $100 level price tag, on top of already owning an iPad, gives a high entrance barrier.

Osmo is important for this thesis as a proof that a) it is possible to implement AR on an iPad by downloading software b) the face oriented camera on iPad can be used for an AR mirror.

\textsuperscript{16}Osmo Web Site (2016). URL: https://www.playosmo.com (visited on 05/07/2016).
2.4 Product: Ericsson Piero AR

Ericsson AB is mainly known for telephone systems, but a significant area is also Broadcasting and Media systems. In April 2016, Ericsson released\textsuperscript{17} the Piero AR system for live 3D AR overlays during sports broadcasting. (Figures 2.3 and 2.4) The AR features are presented as an extension to the established Ericsson Piero system that is used for television productions worldwide. Significant parts of tracking, registration and 3D graphics is already in the Piero system or available through interfaces from the studio or productions systems.

Piero AR is included as an example of a) AR used in a mass market context b) AR launched in large scale as an add-on to existing products and c) Incremental AR development. The 2001 survey\textsuperscript{18} by Azuma et al. presented a similar – but less graphically advanced – barrier used in American Football broadcasting.


\textsuperscript{18}Azuma et al. (2001).
CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.5: Full view of the video mirror, including the photographer taking the picture. This is not a trivial reflection, but an image processed by the video camera above the screen and projected on the screen together with the mirrored doll.

Figure 2.6: A) View when the projector is off. Kinect depth camera on top. Glasses for the 3D projector is on the left. B) The artefacts are mirrored in 3D. C) User holds a virtual 3D ball. D) Physics simulation of domino bricks.

2.5 Study: Miragetable

Hrvoje Benko, Ricardo Jota and Andrew Wilson did a frequently cited study at Microsoft Research, exploring a mirror-like user interface.19 (Figures 2.5 and 2.6) The 3D volume between the user and a mirror is explored together with virtual and tangible objects as artefacts. This part of Augmented Research is considered new and fundamental user perception is tested. Generally, users accepted distorted graphics and other defects well and interaction was effective. It was not obvious that the 3D stereo projection is necessary for a good experience though.

Miragetable is included for its user interaction involving AR in a mirror and tangible objects handled by the user. The researched interaction is relevant for any AR mirror that is tracking physical objects that are handled by the user.

2.6 **Study: Virtual Try-on using Kinect and HD camera**

Giovanni et al.\(^{20}\) presents a mirror that augments the viewer’s body with virtual clothes from a catalogue. The system is in daily use at a high profile fashion shop and is popular among customers. Their key difference from other virtual fitting rooms is that clothes are generated from 2D images rather than created by hand in 3D. It can bring in new clothes from their ordinary catalogue due to the ability to take 2D images.

Virtual try-on is included as an example of an AR mirror variant implemented as a real product. Important properties are a) interaction with ordinary customers as a part of normal shopping b) integration with relevant systems (update from ordinare catalogue).

2.7 **Study: Monitor based AR for industry**

Re et al.\(^{21}\) studies the potential of using a *spatial see-through video* with separated camera and screen (figure 2.8 and 2.9). They recognize that Head Mounted Displays are popular in Augmented Reality research, while industrial applications may benefit better from a separate monitor and a camera facing the work bench. Stated reasons are a) cost of equipment b) robustness and durability c) better tracking d) less intrusion of work space and e) more realistic to work extended periods if not having to wear or hold AR equipment. They find literature support that Augmented Reality is more effective than paper based instructions for complex tasks. For easier tasks, there was no difference though. There is also potential to reduce stress due to less mental load when performing a task with Augmented Reality Support.

---


\(^{21}\)Re, Guido Maria, Oliver, James, and Bordegoni, Monica (2016). “Impact of monitor-based augmented reality for on-site industrial manual operations”. In: *Cognition, Technology & Work*, pp. 1–14. ISSN: 1435-5566. URL: [http://dx.doi.org/10.1007/s10111-016-0365-3](http://dx.doi.org/10.1007/s10111-016-0365-3).
CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.8: Overview of monitor set-up. Left: A camera above the workbench captures real-time video of a board in production. Mid: A standard computer runs AR software. Right: A standard screen shows real-time augmented video.

Figure 2.9: Example of AR projection as shown on the screen. The fixed camera and screen setup allows for exact and robust tracking/registration.

There is a user test of a set-up with a down facing camera that streams video to a fixed screen and adds AR on that screen. AR gives a clear but not massive improvement in time for doing the task and reduced mental load compared to paper.

Monitor based AR is similar enough to an AR mirror to infer conclusions from. The specific research case is not mass-market but it may be generalized into a standardized media system or browser that can read a standardized media file. Then it should be able to load a specific media file for a certain object and provide relevant AR interaction for this object. This would be a mass market solution similar to a web browser or video player. The research indicates if and when the AR interaction improves work performance compared to printed instruction – in other words how urgent it is to implement AR. In the researched context, there is low urgency. AR can give noticeable improvement if used in the right context, but there is no revolutionary improvement to expect.
2.8 Study: Augmented Reflection of Reality

Li et al.\textsuperscript{22} presents a virtual drum set behind a see-through mirror. A user with drumsticks can see his own reflection together with the drums and play on them. (Figure 2.10)

The Drum Set is included for the user interaction with a virtual object in the mirror. This indicates that humans are able to easily coordinate tangible objects in the real world (drumsticks in the hands) with virtual objects (drum) in a mirror. The focus point of activity (drum stick hitting drum) is in the virtual world, but there is an apparently very little effort to adapt.

2.9 Study: Through the Combining Glass

Plasencia et al.\textsuperscript{23} explores the “magic” volume behind a combining mirror that allows the mirrored hands to interact with objects inside the volume. The objects and the glass box are real, like in a museum or jewellery store box, but the mirrored hand or its shadow can “touch” the objects inside the box. This virtual shadow-based


interaction (Figure 2.11) is used to request extended information, get a magnified image or shine light on delicate objects that normally resides in darkness.

One advantage of using a physical mirror is that 3D depth and parallax is maintained. Hence the reflected view works for more than one viewer which is good for applications in a museum. There is also a discussion around how to design interaction in this kind of volume and how to educate new users. A significant share of museum visitors are there for the first time and should be able to learn interaction with the new technology fast. One useful enabler was live video screens in the entrance hall showing visitors interacting with the glass boxes. It served both to build interest for the attraction and to incept the idea of touching tangible objects in a sealed box with a mirror copy or shadow of their own hand. The visitors were indeed able to interact, but they had to understand what they were supposed to do with the box.

Through the Combining Glass is included for two reasons. It is AR implemented as a real product and it is a large scale example of users interacting with a virtual focus point. Important properties are a) deployed in a public museum interacting with ordinary visitors b) AR is used as presentation media rather than an attraction in itself c) information how the museum educates user interaction to new visitors.

![Figure 2.11](image)

Figure 2.11: Ways of interacting through a mirror: (A) Video see-through augmented mirrors, (B and C) different implementations of augmented mirrors. (D) Situated Augmented Reality. (E) Shadow-based interaction that is used in the museum.
2.9. STUDY: THROUGH THE COMBINING GLASS

Figure 2.12: Comparing presentation methods between 2D paper drawing, 3D Building Information Model (BIM) and Augmented Reality in study.

Figure 2.13: Result of comparision according to figure. Note that the AR was the only computer format that the architects preferred over paper drawings.
CHAPTER 2. BACKGROUND AND RELATED WORK

2.10 Study: AR on Screen for Architects and Engineers in Civil Construction

Meža et al.\textsuperscript{24} compares paper drawings, traditional computer presented information and AR presentation. (Figure 2.12) They conclude that AR is definitely an improvement over current methods but it is only incremental. (Figure 2.13) AR is very likely to enter this market, but in an evolutionary manner since the improvement is not sufficient to motivate a rapid change of technology.

This study is included since it compares AR with state of art in both paper drawings and 3D computer drawings used in real life work situations. It is possible to infer from the results that AR is better in many situations but not all. Introducing AR can be useful but it requires care to make it effective. The study is also important since the construction industry performs systematic standardization, research and development to implement AR in their tools.

2.11 Study: AR as User Interface

Krichenbauer et al. presents work in progress to define an AR user interface for 3D media production\textsuperscript{25}. This indicates there is a desire, but not large scale momentum. Even in an area where the need is sufficient, they are just preparing for introducing AR in large scale by defining standards for it.

A survey\textsuperscript{26} by Fite-Georgel indicates that real-world applications in the industry tend to stay in the lab. They emphasize strong improvement of the technology, but the applications are still not deployed in scale.

This is included for improving understanding of the evolutionary nature of AR. There is development of AR, but implementation for everyday use in production environment is slow.

\textsuperscript{24}Meža, Sebastjan, Turk, Žiga, and Dolenc, Matevž (2015). “Measuring the potential of augmented reality in civil engineering”. In: Advances in Engineering Software 90, pp. 1–10. ISSN: 0965-9978. DOI: http://dx.doi.org/10.1016/j.advengsoft.2015.06.005.


2.12 STUDY: INTELLIGENT AUGMENTED REALITY TRAINING FOR MOTHERBOARD ASSEMBLY

Westerfield et al. compared learning performance for test subjects using AR with traditional training how to assemble a computer by inserting cards in a motherboard. The AR implementation used line-of-sight AR through glasses. They tested two kinds of AR guiding systems with different ability to adapt information to the user and in recognizing progress and failure (Figure 2.14). They did also study how much defects in registration impacted actual performance. The important results in the context of this thesis are that: a) Assembly of products is faster at least for untrained subjects if aided by AR. b) Even poor registration or information not aligned with the object is still better than no AR.

This is included since the study results makes it possible to infer a) Similar or worse performance for AR using the mirror metaphor compared to AR glasses. b) AR is likely to be useful even with some defects in the AR presentation.

---

2.13 Study: Computer-Augmented Mirror to Aid Makeup Application

Iwabuchi et al. studied a computer based make-up mirror. A camera captures the user’s face and loops it to the screen mimicking a mirror. Image processing and object detection is used to automatically zoom and pan eyes or lips for precision work (figure 2.15). This is a significant improvement from the awkward positions for getting a similar and often required view in a traditional mirror. The mirror also simulates different light conditions, left/right reversed view and ability to see snapshots from different angles.

The mirror is included due to its mass market potential and ability to inspire new ideas. It may not fit the formal scope as a plain AR application but there is real-time tracking and some elements of image manipulations in connection to markers. The interesting potential is the combination of a) technical feasibility for COTS hardware b) make-up is an everyday activity for a significant number of people c) the usability gain and user interest is reported as noticeable d) the application is connected to a habit and adds convenience.

---

2.14. STUDY: METAMORPHOSIS HAND: HUMANS CAN ACCEPT CERTAIN INTERFACE ERRORS

Ogava et al. presents an AR box\(^{29}\) where a user can insert hands and play on a virtual piano (figure 2.16). The top of the box visualizes the hands, but with a strange ability to transform shape and reach longer than normal hands. Around four hundred people tested the box during a public university event. Feelings of surprise were common. Still, the users felt a strong ownership of the hands, accepts the new properties and were – often to their own surprise – able to play the piano.

This is included as an indication how humans are able to fairly easy accept and compensate for significant differences between reality and a virtual world. The findings here are consistent with section 1.4.2 that describes how humans tend to accept and adapt to mirrors rather than really understanding how physics transformed the view. Hence it is possible for AR to work even through a mirror in an odd or remote position.

Chapter 3

Method – Mirror Workspace Feasibility Test

This chapter describes the actual tests performed for this thesis. It is worth noting the limited number of tests. This is due to the somewhat surprising large amount of information available from previous work – mainly in research, but also from products and patents. For the discussion, the only urgent remaining questions relate to the properties of face oriented cameras for the novel case considered by this thesis: using the face oriented camera to view a work space on a physical desk.

3.1 Test Background and Discussion

Can a Commercial Off the Shelf (COTS) device be oriented to show a mirror view of a workspace area on a physical desk? The purpose of this test is to find the minimum premises for this set-up to work at all on widely deployed COTS products.

The most important assumption for this thesis is that we can create a reasonably useful Augmented Reality mirror by just downloading software into COTS hardware equipped with a) a face oriented camera, b) a screen and c) sufficient processing capacity. This test should verify or dismiss that assumption. It may also give important information as limitations in usage and reveal characteristics that need further assessment.

The desired outcome is to give the user a decent mirror view through the unit’s face oriented camera while sitting in a convenient work position. This view should cover a work volume formed by an area roughly similar to a landscape A4 size paper and a height of 10 cm. The standard paper size A4 is selected out of convenience due to easy availability. The area, volume and aspect ratio is a good compromise between typical camera view field and the desire for a large workspace. A4 is roughly the size of two human hands flat beside each other and a) represents the convenient range of two-handed work with small hand-tools on b) suits the human eye’s high
CHAPTER 3. METHOD – MIRROR WORKSPACE FEASIBILITY TEST

Figure 3.1: Illustration of the camera view-field of a typical laptop with camera integrated into the screen. The normal use of this camera is to capture the face of a user looking at the screen. To capture a work area on the desk, the camera must be aimed down by tilting the screen down. This can result in problems viewing the screen.

resolution view field (section 1.4.2). The height 10 cm is also based on hand size and some room to handle small objects like for example a circuit board, smart phone or piece of artwork.

One obvious issue of aiming a camera into this volume is that the face oriented camera is optimized for capturing the users face. Typical use is snapshot photographs and video conversations. But in this case, we want to capture video from the workbench in front of the user while showing the looped video on the screen. The same screen as the camera is integrated into (figure 3.1). This can result in problems viewing the screen if it has to be tilted down too much in order to give a useful view field. There is potential for discovering problems, hence there can be need for experiments with workarounds during the tests.

Workarounds considered are:

- Tilting the screen to get better camera or screen view angle. Optimal camera and screen tilt is likely to be antagonistic and the selected angle will be a compromise that has to be negotiated during the test. Tilting is easy for most laptops while tablets may need special support.

- Turn the screen upside down or on the side to get a better camera position or screen angle. This is easy for most tablets, while laptops may need special support.

- Change distance and position between user, workspace and screen. This is indeed a natural part of calibrating the setup, but experiments with extreme changes should be considered if there is need for a workaround.

- Raising the screen or work area to get a height difference.

- Use a mirror to change camera or screen view angle independently.

- Separate camera and screen.
3.2. TEST SETUP

The list is not to be exhausted for each test, but is used to select methods from until an acceptable configuration is found with as little hardware modifications and extra equipment as possible. Further workarounds and changes can be made in later tests or future studies.

The test should also verify informally that resolution, camera field-of-depth and frame-rate is suitable for real-time work.

3.2 Test Setup

See figure 3.1 for an example setup with a laptop. A tablet or phone can be set up in a similar way but may need support to stand in an upright position. Also, refer to the list in section 3.1 for alternative set-ups.

The test is centered around an intended work volume that is represented by an area of a landscape oriented A4\(^1\) paper (210 mm \times 297 mm) and a height of about 10 cm. The set-up is using an ordinary table and a chair for the user to sit on and adjusted to allow the user to sit close to the table in a convenient work position with 90 degrees elbows and lower arms resting horizontally on the table. The A4 paper is located in landscape orientation and positioned by the user first resting both hands on it, then moving it to a centred location in front of the body while moving arms to 90 degrees elbows in the previously defined work position.

The device is configured to loop live video from the face oriented camera to the screen. Then the tablet is located to offer a reasonably useful mirror image to the user sitting in the previously defined work position. This may require one or several workarounds from the list in section 3.1. The set-up should be adjusted until it gives best possible mirror image for the user.

A number of wooden blocks of the size 75 \times 25 \times 15 mm are used as physical 3D markers of the desired volume. The blocks are initially positioned in each corner of the paper and stacked in at least one relevant\(^2\) position to a height of at least 10 cm.

The set-up will depend on the user’s body size. Hence the measured numbers are not expected to be exact, but to give a qualitative understanding of the properties of a working mirror configuration. A significant outcome of the test is the level of workarounds required to make the mirror configuration to work at all. It is by no means obvious that any COTS hardware can perform this task. The mirror mode targeting a workbench is definitely outside the intended mode of operation

\(^{1}\)The US letter paper size of 8.5 in \times 11 in (216 mm \times 279 mm) is close to A4 size. In this rough context, they can be considered equal.

\(^{2}\)Relevant means ensuring 10 cm height for the full A4 area by putting the marker at the spot with lowest height. Slightly lower height may be accepted in a few positions, like corners, if there is a significant gain on other parameters.
for this kind of devices. They are not designed for this use case. Even if the camera and screen angles may work, there can be issues with camera depth or autofocus.

Informal verification of sufficient resolution, camera field-of-depth and frame-rate is performed by holding and moving the hand or an object in different parts of the workspace and look for unsharp or slow mirroring.

3.3 Device: ASUS 14” Laptop

This is a typical classic laptop with 16:9 screen and a face oriented camera on top. The size is common and a compromise; big enough for nearly full sized keyboard and screen, while still small enough to be easily carried around.

3.4 Device: Samsung Galaxy Note

This is a typical large tablet, designed to be used in landscape mode. This is mainly reflected by location of the face oriented camera on the long side.
Chapter 4

Results – Mirror Workspace Feasibility Test

This chapter presents the actual results from the test. It is indeed possible to view a workbench area in front of the device, but the number of workarounds and limitations are noticeable.

4.1 Device: ASUS 14” Laptop

This is a typical laptop with face oriented camera in the top of the screen (figure 4.1). Video is looped by running the Vuforia multitarget\textsuperscript{1} tutorial in full screen mode. Aiming the camera to the work bench required tilting the lid forward into a position of about 50 degrees compared to the 90 degrees of an upright position. The screen was still reasonably viewable, making this a reasonable tradeoff since the front corners were slightly below 10 cm. This should be studied further to learn if the tilted lid can be accepted by users and if it presents problems.


Figure 4.1: ASUS 14” laptop in mirror mode adjusted to cover the proper area. The screen is tilted forward in a 50 degrees position but still viewable. Raising the screen will cut the view of the area close to the laptop. Lowering it further will first reduce the user’s view further, then cut the 10 cm height at the area closest to the user. The image on screen is not mirrored due to inflexible software.
Refer to figure 4.2. An upside down position allows full coverage of the volume, but there is need for a solution to mechanically support this position. This view reminds more of a physical mirror than the previous one, mainly due to the low camera position. It does rise questions about occlusion though. Will the tracker be able to follow objects located in the work area for a typical use case? This should be studied further to learn if occlusion is a problem.

A 90 degrees rotated view was also tested but proved to not improve anything. The computer did indeed stand by itself on the side, but the camera still had to be tilted down to get an useful image. Hence the first case proved to be more relevant.

Informal test of the looped image quality indicated some noise in indoor light and significant noise in shaded parts of the image. Increased light improved the image, but the general impression is that the camera needs good light conditions. Otherwise no apparent issues in speed or resolution – the noise dominates. Camera and light performance need further studies. Performance may be sufficient for AR tracking, but how is the noise perceived by users? This should be studied further as indicated in section 6.1.1.

4.2 Device: Samsung Galaxy Note 10.1 2014 edition

This is a typical large tablet, designed to be used in landscape mode. This is mainly reflected by location of the face oriented camera on the long side. Mirror loop is performed by activating the tablet’s camera function and selecting the face oriented camera. This gives a full size mirror loop. The face oriented camera allows up to $1920 \times 1080$ (2 Mpixel) resolution and this mode is used.

The position with camera on top works if the tablet is tilted forward. Support is required to maintain this position and it was hard to create a reliable stand from easily found objects like books and boxes.

Upside-down with the camera close to the table proved to work better (figure 4.3).
4.2. DEVICE: SAMSUNG GALAXY NOTE 10.1 2014 EDITION

Figure 4.3: Samsung Galaxy Note tablet upside down. It works well in upright position due to the low camera position. The tablet is supported by a box behind it. Almost any object of sufficient size and weight works as support due to the almost upright position.

Figure 4.4: Samsung Galaxy Note tablet on the side. Note the camera position on the side of the screen. The tablet has to be located rather far away to cover the work area. This results in a very small part of the screen to show the work area. The marked work area on the screen is barely visible.

The tablet still need support, but in an almost upright position leaning slightly backwards – making it easy to find a supporting object at will. Coverage of the work area is good, but there may be issues with occlusion. Even a low object located close to the camera will occlude a significant area behind it due to the low camera position. This should be studied further as indicated in section 6.3.3. Tablet on the side was also tested (figure 4.4). The camera position requires the tablet to be rather far away or tilted forward to show the works. The most useful position is upright, but it may be more relevant for AR including the user’s face and body.

Informal test of the looped image quality indicated no apparent issues in speed or resolution. There was indeed some noise in indoor light, but it was hard to notice and not distracting. With increased light, there was no noticeable noise.
Chapter 5

Discussion

This chapter is concluding the thesis by discussions around the topics a) AR found in mass market products. b) Mirror metaphor AR looking at an area on a workbench. c) Mirror metaphor AR when the user looks at himself or herself. d) Ethical and societal aspects.

5.1 Mass Market Products Employing Augmented Reality

Augmented Reality (AR) is already part of products with significant market presence, even if the uptake is in an early phase. Examples in this thesis are a) Snapchat Filters presenting AR effects on the user’s face (section 2.2) b) Osmo connecting AR to toys (section 2.3) and c) Ericsson Piero sport broadcasting AR (section 2.4). A common factor is that AR is used without any marketing hype or special treatment. The products are indeed technically advanced, but nothing is driven by technology. AR is selected for its abilities to improve user interaction rather than hype or coolness factors. There can be marketing, but it is the complete solution or feature that is marketed rather than the fact that AR is involved. This is a sign of maturity and it is possible that the real AR revolution has started, by slowly evolving into new products as the best available option. The sports broadcast and Snapchat examples were also well prepared for including AR in large scale. Both products were already well established and already related to video streams. Those examples also align well with the desire to reduce technical and business risk in innovation (section 1.2) since the platforms and markets are rather mature – compensating the risk of adding AR.

For Snapchat, AR is an incremental improvement. The product is a success even without AR, but the new filter features offers a set of well integrated fun options likely to help keeping Snapchat’s momentum of success.

For Osmo, AR is indeed essential but the focus is fun interaction. AR is not mentioned at all and it makes sense, since AR is just one piece of all the software and hardware details comprising Osmo. It is a well integrated product and there is no
need for understanding any technology to use it.

For broadcast, advanced graphics and effects are essential to attract viewers. Hence it is reasonable that AR is adopted soon in this context. The market is mature and the broadcasting industry has high technical maturity for video and image processing. Often, there is already a tracking and registration infrastructure in place, supporting a sports production. Player positions, balls, vehicles or other artefacts are frequently tracked to support everything from commentator’s game understanding over camera and production logistics into animations and video effects. AR is entering as just another feature in the system.

There are also signs of early developments in other established areas like a) Fashion retail (section 2.6) b) Industry (sections 2.7 and 2.12) c) Museums (section 2.9) d) Civil construction (section 2.10) and e) Media production (2.11). The AR progress is slow but this can be due to the incremental nature of AR (sections 2.7, section 2.10 and 2.12). When there is an advantage, it may be sufficient to motivate AR in a new design or upgrade, but not to justify the cost of replacing a working system.

5.2 Mirror Metaphor Looking at the Workbench

Using the mirror metaphor with unmodified Commercial Off The Shelf (COTS) hardware can be relevant, but there are significant restrictions. The Snapchat filter’s AR effects (section 2.2) proves that it is possible to deploy an interactive video mirror with AR, on a significant share of the phones and tablets owned by users in 2016. A significant number of devices fulfil the needs for resolution, frame-rate and real-time tracking that is needed for useful AR with the face oriented camera towards a face. Using a workbench area though, requires remedies like tilting the screen forward or turning the device upside-down to aim the face-oriented camera to the bench-top or desktop. The looped image still looks acceptable but further studies are needed to understand how AR and user interaction is affected. There are also questions about image noise due to light – at least on devices with low performing face oriented cameras – that also need further studies (section 6.1.1).

The general conclusion is that the workbench scenario requires too many workarounds to be useful. It may work for some limited case but adding a USB camera to a laptop (section 2.7) or a tiny mirror and a stand (section 2.3) to a tablet or phone removes all the restrictions at a rather low cost. The workarounds considered in this thesis are indeed free in terms of cost, but a user who can afford an advanced device is likely to prefer spending a bit more money rather than performing workarounds.
5.3 Mirror Metaphor Looking at Yourself

The implementation in Snapchat’s app (section 2.2) gives technical relevance to every solution based on the user looking at herself or himself via a video loop through the face oriented camera. Hence the question is about relevance for a particular implementation rather than technical feasibility. Technically, it works and it is robust enough for mass deployment – given a proper and careful implementation.

5.4 AR Impact on Society and Ethical aspects

Augmented reality (AR) in itself is just a technical component that has to be integrated into a technical system and a context. Hence it is hard to discuss impact without knowing the context. Still, the combination of a) the significant implementation scale assumed in this thesis and b) literature reporting advantages of AR in at least some areas, gives that AR may have significant impact on everyday life. In the long run, reality itself may change by AR, since physical objects will have new virtual dimensions added to them – at least for people using AR technology. The concrete effects on society and ethics, though, will depend on design and operation of the actual system or systems.

AR as a component cannot assume full responsibility, but it is important to educate decision makers about the potential impacts of AR. By AR’s ability to enable easy access to information, there may be information that becomes too available. Consider for example an application with real-time access to personal information utilizing AR: While AR is just a presentation component, it is also an active part of a system that may enable the user to, for example, access personal information about random people by just looking at them. Possible ethical issues for that example may be: a) If people using AR can see personal information floating beside everyone they look at, is that fair to privacy? b) AR may create a new digital divide, like in the previous example where AR users may get a significant information advantage over non-AR users.
Chapter 6

Future Work

This chapter presents a rather broad collection of potential topics for further studies, found during the writing of this thesis. The list is intended as brief inspiration for the reader who wants to come up with a subject for a thesis or paper in the area. Feel free to explore any of the topics if you are interested. I am indeed interested in the listed topics, but I know from experience that I am unlikely to find time exploring more than a tiny fraction of them.

To provide a better overview, the topics are grouped into areas a) product design b) user interaction research and c) AR technology.

6.1 Product Design Related

This section is development oriented. The scope ranges from increasing knowledge about making products to actually consider prototypes that may be turned into real products.

6.1.1 Evaluation of Cameras in Commercial Off The Shelf (COTS) Hardware

Investigate if the devices evaluated (section 4.1 and 4.2) are representative for the COTS products by looking at a larger number of devices. It is likely that a few camera and screen components are used by all products on the market. Then it is possible to make an almost complete market analysis by working on high level statistics for those components, like numbers manufactured and products they go into.

The differences in camera noise due to low light between the tablet and laptop is also interesting. Is there a general trend to have cheaper cameras in laptops? What is the background for the manufacturers to decide on a particular camera quality?
6.1.2 Study Potential Applications for AR mirrors

Start with the make-up mirror in section 2.13 and verify that it works for a reasonable number of tablets and laptops. Weak points are probably screen size and the face camera’s ability to provide sufficient resolution for a useful zoom. Also look into similar but new applications or features like a) zoom for aiding insert/remove of contact lenses and b) applications in tele-medicine, making it easy to zoom into different skin areas during a video conference with a doctor or nurse.

6.2 User Interaction Related

This section is research oriented and related to the user interaction aspects of AR.

6.2.1 AR Mirror on COTS Laptop – user interaction

Perform user studies to evaluate the potential of using a COTS laptop with tilted lid and a low quality built-in camera based on the results in section 4.1. Is it possible to find a user interaction solution where the users accept the limitations? The advantage of finding a working solution is the potential to create products based on the solution for a very large number of users who already owns the hardware.

6.2.2 Human Interaction in the Mirror World

Explore relevance of using interaction with objects and other humans using the mirror metaphor explored in this thesis. Consider the paper in section 2.5 as a starting point, then add a second person or even more people. Use the same screen or work remotely over several screens. This kind of interaction is suggested as advantageous when using Virtual Reality (VR), but using AR may be more convenient since AR is not shielding off vision like a VR helmet does.

6.2.3 Potential of AR to Make Architects and Similar Professionals to Give up Paper

Refer to the graph in figure 2.13 hinting that architects may prefer paper for some information rather than today’s computers, while AR was a concrete improvement. Can general conclusions be drawn about artistic users and what they need to switch from a traditional user interaction like paper to a new one, like computer? Consider looking for studies around Apple products and artistic users. A study of this kind may give useful insights about unexplored user needs.
6.3  AR Technology Related

This section is research oriented and focused on the engineering aspects of implementing AR and research of AR in itself.

6.3.1  Studies to add mirror metaphor to AR mainstream

Consider the mirror metaphor briefly studied in this thesis and discuss the relevance and need of recognizing it as an AR research niche. If there is a need and relevance, this is a new AR research niche available for exploration.

6.3.2  Activity recognition

Improve understanding of disassembly of objects and tools. Recognize pieces of an object and how e.g. a screwdriver merges or disjunct a screw from an object. This includes some level of Artificial Intelligence for the computer to understand how the pieces connect and how it affects their tracking features. A successful solution may be an enabler for interactive instructions “movies” where the computer guides the user based on the current situation and need.

6.3.3  AR Mirror on COTS Tablet – Occlusion

Study impact on occlusion issues when using a COTS tablet upside-down based on the results in section 4.2. This is interesting to find the limitations and abilities of an unmodified tablet to view a workbench area in front of it.

6.3.4  Hand and face tracking

Future cameras are likely to be 3D depth sensing. Recognizing the hands holding and moving the objects should improve tracking/occlusion performance but also enable gestures related to objects. Consider opportunities for improved or less computing intensive tracking.

6.3.5  Generic Tracking Inspired by Face Tracking Algorithms

The dedicated face tracking algorithms referred to by sources in section 2.2 indicates that it should be possible to create dedicated tracking algorithms for almost any object. Study the potential for a generic parametric model for the tracker. If this works, it should be possible to create a parameter library for every relevant object. Thus it should be possible for a generic AR tracker to run optimal tracking by just identifying an object, then downloading the optimal parameters for it.
6.3.6 Merge Hands and Arm Tracking and Human Interaction
Algorithms from Robotics

Robots generally have more advanced computing hardware than the consumer devices considered in this thesis, but future consumer devices may have capacity for their more advanced algorithms. Consider the blockers for moving algorithms and when they are likely to disappear. Note: During research, it can be efficient to use expensive equipment to get a result fast rather than spending time optimizing performance. Hence some blockers may be related to development time for optimization rather than a true need for advanced hardware.


BIBLIOGRAPHY


BIBLIOGRAPHY


Tynan, Dan (2006). The 25 Worst Tech Products of All Time. “Motorola Rokr E1 (2005): The world’s most popular digital music player meets the world’s coolest looking phones; what could possibly go wrong? Well, plenty. The Rokr E1 held only about a hundred songs, file transfers were painfully slow, the iTunes interface was sluggish, and–duh–you couldn’t download tunes via a cell connection. This phone ain’t rockin’, so don’t bother knockin’ ”. URL: http://www.pcworld.com/article/125772/worst_products_ever.html?page=7 (visited on 03/26/2016).


West, Joel and Michael Mace (2010). “Browsing as the killer app: Explaining the rapid success of Apple’s iPhone”. In: Telecommunications Policy 34.5, pp. 270–286.

