Visual Coherence in Mixed Reality: A Systematic Enquiry

Jonny Collins
jonny.collins@otago.ac.nz

Holger Regenbrecht
holger.regenbrecht@otago.ac.nz

Tobias Langlotz
tobias.langlotz@otago.ac.nz

University of Otago
C/-Information Science Department
PO Box 56
Dunedin 9054
New Zealand

Abstract

Virtual and Augmented Reality, and other forms of Mixed Reality (MR), have become a focus of attention for companies and researchers. Before they can become successful in the market and in society, those MR systems must be able to deliver a convincing, novel experience for the users.

By definition, the experience of Mixed Reality relies on the perceptually successful blending of reality and virtuality. Any MR system has to provide a sensory, in particular visually coherent set of stimuli. Therefore, issues with visual coherence, i.e. a discontinued experience of a MR environment, must be avoided. While it is very easy for a user to detect issues with visual coherence it is very difficult to design and implement a system for coherence.

This paper presents a framework and exemplary implementation of a systematic enquiry into issues with visual coherence and possible solutions to address those issues. The focus is set on head-mounted display based systems, notwithstanding its applicability to other types of MR systems.

Our framework, together with a systematic discussion of tangible issues and solutions for visual coherence, aims at guiding developers of Mixed Reality systems for better and more effective user experiences.

1

1 Introduction

Mixed Reality by definition (Milgram & Kishino, 1994) combines Virtual Reality and Reality in a way that the two domains blend together. If possible (and often desirable) a user might be unable to distinguish between what is real and what is virtual in a given scene. Both domains form a (visually) coherent world, a mixed reality environment, in which users immerse. Nowadays, a head-mounted display (HMD) is often the target system when creating mixed-reality environments. Driven by commercial developments, HMDs are nowadays more readily available and come with specifications that are appealing when aiming for creating a mixed-reality experience. HMDs come in two predominant flavors: video see-through (VST) and optical see-through (OST). In video-see through the real-world is captured with cameras and mixed with the virtual world before being displayed within the integrated screens. Examples of such systems include the Oculus Rift (www.oculus.com/en-us/rift/) or HTC Vive (www.htcvive.com/) even though they have to be modified by adding camera support to provide a Mixed Reality experience (http://willsteptoe.com/). Optical-see through HMDs mix the real and the virtual using an optical combiner which is placed in the users' visual path. Recent examples for this technology are Microsoft's HoloLens (www.microsoft.com/microsoft-hololens/en-us/) or the Meta Glasses (www.metavision.com/). Despite the different technologies, both approaches for HMDs have been used to create mixed realities in which users can immerse. In today's MR systems, however, very often this mixed immersiveness is disrupted by issues in visual coherence - the elements of the two environments do not constitute a continuous, coherent space. This is not desirable as it breaks the illusion of the mixed-reality environment which can affect the users' perceived presence within this mixed-reality space as well. Applications like the

Augmented Mirror Box (Regenbrecht, Franz, McGregor, Dixon, & Hoermann, 2011) would not work without coherence and presence. Consequently, the ever present question for designers, developers, and researchers is how to identify and avoid those issues in coherence, in particular when aiming for HMD based mixed reality environments.

In this paper, we investigate in a very systematic manner the technological challenges in the provision of a mixed immersive environment and discuss current and potential solutions. We identified the basic, essential elements of a mixed immersive environment, namely interaction, lighting, objects, and the environment, notwithstanding the fact that actual applications need more than those four elements. However, it might be challenging to think of an application which does not contain each of those four. Based on those essential elements we formulate our EOIL (Environments, Objects, Interaction, Lighting) framework along the axis of the Reality-Virtuality Continuum (Milgram & Kishino, 1994). This EOIL framework allows us to systematically go through all combinations of virtuality and reality to explore the interactions between the elements in search for issues with visual coherence. This systematic enquiry is accompanied with the discussion of possible solutions and mitigating techniques.

To (a) be able to better communicate and illustrate our systematic approach and (b) practically investigate those issues with visual coherence, the EOIL framework was prototypically implemented as an exemplar research system. For each of the four EOIL elements we selected a representative implementation (user's hand, light, lamp, and desk) and exercised all possible interactions for a head-mounted-display-based MR system. We were able to collect content from each interaction which we used to support our analysis. The two predominant system types of MR-HMDs, video see-through and optical see-through, have been investigated leading to a total of 32 interaction combinations.

To the best of our knowledge this is the first systematic enquiry into issues with visual coherence in MR systems. Our EOIL framework and its exemplar implementation can be used to (a) evaluate on and (b) design MR systems for visual coherence. Without visual coherence there is no actual mixed immersive experience.

In the remainder of this paper we give a brief overview of the related work, specifically in the context of previous review work. We then describe the design of our EOIL framework followed by our EOIL implementation. Using the framework implementation we provide a systematic analysis of the issues in coherence that we established. Finally, a discussion of known or possible solutions is provided.

2 Related work

This section provides an outline of existing survey and review papers which focus on MR systems, and the known problem space. More detailed related work is provided throughout the remaining sections where appropriate.

Augmented Reality is a part of the mixed reality continuum defined by Milgram et al. (Milgram & Kishino, 1994). AR survey papers provide us with an accumulation of the concepts and technologies used, and also the issues that arise in trying to design such systems. There is a plethora of research attempting to solve some of the individual, inherent problems with mixed reality systems. In the late 1990’s, Ronald T. Azuma presented the first comprehensive survey on AR (1997). The main issues identified are focus and contrast, portability, scene generation, display device, registration, and sensing. Various solutions are presented which provide an AR system designer with choices for implementation. The solutions which are selected when building a system are usually picked due to their advantages with respect to the application. As an example, when building an outdoor system one might choose to use a video see-through

(VST) HMD approach as opposed to an optical display due to bright environmental light having a large impact on the contrast of an optical display. Most of these issues are mitigated using a specific approach so the decisions are still very dependent on environmental and human factors. In 2001, Azuma et al. presented an updated survey identifying new application spaces for the use of AR, namely mobile and collaborative applications (Azuma et al., 2001).

Van Krevelen et al. (2010) released a survey (Van Krevelen & Poelman, 2010) which provides a more recent take on AR availability. The presentation of recent surveys shows that complex system components such as tracking are still not mature enough to provide convincing augmentations to the real world. While there have been advances in some technologies (i.e. HMDs) a coherent MR system is still far from being available.

Besides pure research, industrial AR has been a topic of interest for the last two decades. An up-to-date survey of industrial AR applications was presented in 2011 (Fite-Georgel, 2011). This survey divided AR systems and applications into five categories: product design, manufacturing, commissioning, inspection and maintenance, and redesign and decommissioning. Due to the technological limitations of hardware and software required for AR systems, there was just one prototype which made it to the market, the so called Intelligent Welding Gun (Echtler et al., 2004). This was later removed from production. Regenbrecht et al (2005) also presented a paper on the use of AR systems for the automotive and aerospace industries (Regenbrecht, Baratoff, & Wilke, 2005) highlighting the challenges of implementing AR in that particular industry context.

Kruijff et al. presented a classification of perceptual issues in Augmented Reality using a conceptual visual processing pipeline (Kruijff, Swan, & Feiner, 2010). Issues are organized into ones of environment, capturing, augmentation, display, and individual user differences. This work provides us with a starting point for how to approach our analysis. Extending Kruijff et al.'s concept we subdivide a MR system into four components which we believe each MR system would probably have as a minimum set of components. We describe our derived framework in the following sections.

3 Framework

This section describes: 1) the design of our EOIL Framework, 2) our framework implementation and, 3) our prototypical implementation.

3.1 Design

For the basis of our framework we decompose an AR application into four abstract components: 1) The environment the user is perceptually surrounded by, 2) Objects within the environment, 3) Interaction, usually between one or more users and some other object(s) within the environment, and 4) Lighting within the environment. The environment component refers to the physical environment where one resides, and could include “environmental” objects such as furniture (if indoor) or trees (if outdoor). Objects can be anything that would reside within the environment such as books, cups, or a phone for example. The lighting component refers to any light within a scene, and the interaction component covers any method of interaction with the scene. This usually refers to some kind of user interaction. The components identified above exist as entities in our EOIL Framework and are represented in any given AR prototype, whether virtual or real. For this reason, those four components provide the foundation of the framework.

3.1.1 The EOIL Framework and Implementation

Considering the framework described above, we can take any AR prototype and break it down to each of the framework components and identify whether they are virtual, or real. Putting this in the context of the overall framework, we can create a visual representation of the possible permutations of the high level components in AR prototypes. To achieve this we created the EOIL Framework and Implementation table, shown in Figure 1.

Figure 1 here

The EOIL table helps us to visualize the strengths of the EOIL Framework. The table is comprised of the four scene components described earlier (found in the left most column). Each column labelled one ... sixteen represents a unique combination of virtual and/or real entities that we refer to as conditions. Additionally, we are able to see how the range of conditions spans the Reality-Virtuality Continuum (Milgram & Kishino, 1994) further demonstrating the comprehensive and systematic nature of the framework. The columns of the EOIL table are ordered according to Milgram’s continuum to facilitate understanding and to exemplify the flexible nature of the EOIL framework. The order in which conditions are analyzed during the enquiry is not important provided all conditions are covered. The method of analysis is described in the proceeding section.

Any AR systems engineer can use the given EOIL table as a high level guide for prototype implementations. We use this to implement a system which facilitates and exemplifies the enquiry contained in this work. Because our system is designed based on this framework, we can describe its components in the same form as the classification previously established: 1) an indoor office space with furniture such as a desk as an environment, 2) a desk lamp and a notebook as objects within that environment, 3) a user's hand is used to interact with the environment, and 4) light cast by the desk lamp, but also including other environmental light sources. This assembly of components forms an exemplary interactive indoor MR scenario. In the left most column in Figure 1, next to the framework components we discussed previously, we can see our prototype components (Desk, Lamp, Hand, Spot Light) labelled next to their respective representative components. This emphasizes the synonymy of a typical prototype with our EOIL Framework. With this we provide a comprehensive enquiry into the MR problem space.

Usually a system is designed considering a specific application context, however, the purpose of our system is to allow us to realize various MR scenarios that belong to different stages of the Reality-Virtuality Continuum (i.e. augmented reality, augmented virtuality). When a developer is given specific application requirements it means, with respect to MR, that some components may be virtual and some real. Because we have no specific context and because we wish to cover a large set of possible combinations of real and virtual components, our implementation must remain flexible.

3.2 Head-mounted Displays

Head-mounted displays are a medium used for enabling users to visualize and interact with MR environments, with the primary purpose being the visualization of combined real and virtual worlds. There are various approaches to implementing MR using HMD's however we focus on two primary ones: Video see-through (VST), and Optical see-through (OST) (Azuma et al., 2001). Less common techniques such as virtual retina displays (Pryor, Furness, & Viirre, 1998) or HMD projector systems (Rolland, Biocca, Hamza-Lup, Ha, & Martins, 2005) are not covered here.

HMD's provide an advantage over other visual mediums of having a hands-free approach which allows for two handed interaction. To take maintenance as an example, users are able to wear the HMD that provides instructions while they use their hands to perform a task (Feiner, Macintyre, & Seligmann, 1993). Furthermore, because the display is attached to the user's head, the augmented environment is displayed wherever the user looks in contrast to monitor setups that usually visualize only a specific region. These kinds of setups also have their own areas of application (Regenbrecht et al., 2012).

VST setups use either one (monoscopic) or two (stereoscopic) video cameras mounted on the HMD to retrieve a video stream of the real environment, and display that stream on an opaque monitor in the HMD (Figure 2 – left). Virtual content is then rendered to the monitor to achieve a combination of real and virtual worlds. The OST display projects the virtual content to a monitor inside a transparent glass display, and the user sees through the glass into the real world (Figure 2 – right).

Figure 2 here

Each approach described above has advantages and disadvantages in terms of display configuration, perceptual issues, and technical implementation of MR environments. In order to be comprehensive in our enquiry we need to consider both HMD approaches because of their differences.

4 Prototype

This section first discusses the standard scenario devised in order to create a process to be executed for each condition. We then briefly discuss the implementation details of the prototype.

4.1 The Scenario

For the purpose of our enquiry we present a basic scenario which conveys an everyday task that a user might perform. The scenario involves the components previously established within our EOIL Framework and Implementation table. The scene is based in an office space, and has a user sitting at a desk looking towards a desk lamp. The task is such that the user reaches towards the lamp, presses the button on the lamp which turns on the lamp’s light and subsequently shines light into the scenario space.

This scenario is executed for each of the 16 conditions in the EOIL table. As stated previously, to be comprehensive we consider both VST and OST HMD approaches and therefore the execution of the above scenario for each condition is performed once with a VST HMD, and once with an OST HMD resulting in 32 unique conditions. This provides us with a consistent set of content for each condition of the EOIL table (Table 1) which can be further analyzed.

4.2 Implementation

It would be possible to perform an enquiry into coherence issues on a theoretical level however we believe this would lack the comprehensibility that a physical implementation provides. A further benefit of a prototype is that it enables identification of “elements of surprise” which are issues that may arise in a counter-intuitive fashion; issues which a theoretical analysis may not detect. Additionally, the prototype allows us to collect supporting content which can be used as explanatory material. This helps to prevent issues in coherence being overlooked. This subsection discusses various implementation details including a brief description of the system hardware, modelling requirements, scene construction, and the combining of real and virtual content for our prototype.

4.2.1 The System

For the purpose of our investigation we used a Meta One HMD. This is naturally designed as an OST device however with slight modifications we are able to use it as a VST device. The Meta One is comprised of two relevant components: 1) a 960 X 540 display (qHD for each eye with an aspect ratio of 16:9), and 2) a DS325 Depthsense camera. The DS325 consists of a color sensor with a resolution of 1280 X 720 in addition to a Time of Flight (TOF) depth sensing camera with a resolution of 320 X 240.

4.2.2 Modelling

We can refer to the EOIL table to determine the application requirements. Our first consideration is the requirement for virtual content. One of the conditions of the EOIL table has each of the components in a virtual state, so we are therefore required to create virtual representations of each component. One of the defining characteristics of AR applications is such that virtual and real content is aligned (registered) in 3D space. Therefore when we create the virtual content, it must accurately represent the real objects at least with respect to physical measures. Three out of four of the components in our matrix are tangible, meaning they are physical objects, and one of the components (the light) is not. Within our sub-set of tangible objects, there are static and dynamic objects. Static objects are such that their form does not change (or is not intended to change). The Environment and Object components are static. The Interaction component (in our case the user's hand) is a dynamic object because the user will likely move and change the form of their hand while interacting with the environment.

In order to model the static tangible objects, we use a basic modelling program (Sketchup 8). We manually measure the objects as accurately as we can, and ensure we generate the models to scale of their real counterparts. Modelling the user's hand is more difficult, and needs to be done in real-time. We use a middleware SDK called IISU. This retrieves information from the DS325 camera (described above), and processes it so that we receive a point cloud that represents the user's hand in real time. These methods are used to virtualize our tangible components. The lighting component is modelled within Unity3D using built-in lighting models.

4.2.3 Scene Construction

Once the models are produced, they are imported into the primary environment generator which is Unity. Measurement of the real environment is required with respect to object's positions relative to each other. When this is known, the models can be positioned in the virtual environment corresponding to their real counterparts. The virtual environment should closely resemble the form of the real environment. Figure 3 illustrates the comparison.

Virtual environments always require a virtual camera to determine the point of view for the user. The consideration for real and virtual world alignment must also be considered when choosing the position for the virtual camera. There are several approaches to determining the correct position based on the user's pose in the real world. This is the problem of tracking in 3D space. This problem is given brief attention in a later section however it is worth noting for the purposes of this work that we chose to gather content using a statically positioned HMD. Given the focus of this work is visual coherence, tracking problems are not a priority. Errors in tracking implementation will impact visual aspects of MR systems, but they are indirect. Furthermore, by statically positioning the HMD we can use a much simpler method to align the virtual camera with the real HMD. Figure 3 provides a visual comparison of the virtual and real cameras.

Figure 3 here

The virtual camera possesses several properties that can alter the perceptual correctness of the system. The key property is the FOV and is relevant when discussing the combination of the real and virtual worlds. The next section will discuss the main considerations for combining virtual and real content.

4.2.4 Combining Real and Virtual Content

With the virtual environment constructed, the scenario described earlier can be executed for each end of the EOIL table (fully real and fully virtual). All other columns of the EOIL table are now considered. In order to combine real and virtual components, the real world needs to be integrated within the overall virtual environment in the scene editor. We are able to use the DS325 camera to acquire the view of the real world that we require for the addition of real content. The two most important considerations to make at this stage are 1) the aspect ratio of the video should be retained when imported into Unity, and the video should be positioned in space so it appears correctly to the virtual camera(s), and 2) the FOV of the virtual camera should match the FOV of the real camera. If there is a FOV disparity, there can be severe perceptual obscurities which can be very apparent to the user. In order to acquire the true FOV of the real camera, an intrinsic calibration is required. Various calibrations are discussed in a later section.

Once the cameras are aligned and the properties are set appropriately, the real world content should align correctly with the virtual world content. Figure 4 illustrates a scene where all components in the prototype are real, except the virtual lamp.

Figure 4 here

4.2.5 Supporting Content

The scenario described earlier replicates a typical action one might perform in an office. Once the system is implemented and ready for use, we proceed to execute our scenario for each of the conditions in the EOIL table. We briefly describe below the process of capturing the content used to support the enquiry for each of VST and OST approaches.

VST - When using a VST system, all of what the user sees is displayed on the monitors within the HMD. For the 16 conditions in which the user would be wearing a VST HMD we record two forms of content. The first is snapshots of two instances, before the lamp’s light is switched on in the scene, and after the lamp’s light is switched on. Taking a screen shot provides us with the same content the user would see on the monitors in the HMD. The second form of content was video recording of the conditions being executed. This way we could capture the whole scenario as it is performed. Figure 5 demonstrates the setup for the set of VST conditions.

Figure 5 here

OST - Recording the OST setup is more complex. In order to capture what the user sees through the HMD, an external camera is placed within the HMD where the user’s eye would be positioned. Two different perspectives are required for capturing content due to the lack of FOV of the OST implementation. To clarify, a screenshot function cannot be used for the OST setup because the real world components the user would see and interact with are actually real world, and not video mediated. When the HMD is placed in the same location as for the VST implementation, the limited FOV means that not enough of the environment is able to be virtualized. We increase the amount of the environment that is covered by the display's FOV by moving the HMD away from the scenario space. While this allows for the visualization of more virtual content in the scenario, the interaction aspect of the system is unable to be used. The range of the TOF depth sensor is between 15cm and one meter and in order to virtualize a good enough proportion of the scene, the HMD had to be moved back well out of this range. This means the alternative is to provide two different perspectives. The first is the distant view (shown in Figure 6 - left) which is given so the user is able to visualize all static components from one perspective, and the second is a close-up perspective (Figure 6 – right) so we are able to visualize component interactions with the hand component.

Figure 6 here

5 Issues in Visual Coherence

This section discusses 1) our categorization of problems we encountered throughout the systematic investigation, 2) common technological problems and 3) a detailed analysis of each problem category.

At this stage we have executed the scenario described in the previous section for each condition in our EOIL table. With the content we collected for each of the conditions we are able to analyze for issues in visual coherence. Table 1 depicts our categorization of the MR problem space. The space is further divided into software and hardware related problems. Related work is provided that aims to address the respective challenges. It should be noted that not all research is covered in this table but rather references to common approaches and outlines of techniques used in approaching the respective issues.

Table 1. Problem Categorization Table. The upper segment of the table provides an overview of software related issues and solutions, while the lower segment is focused on hardware related challenges.

	SOFTWARE					
Problems	Occlusion	Calibration (HMD – Eye)	Lighting + Shadows	Tracking + Registration	Finger Tracking	Stereoscopic Impl.
VST/OST HMD	(Shah et al., 2012); (Wloka et al., 1995); (Berger, 1997);	(Owen et al., 2004); (Itoh et al., 2014); (Tuceryan et al., 2002); (Gilson et al., 2008); (Hu, et al., 2014)	(Kanbara et al., 2004); (Agusanto et al., 2003); (Haller et al., 2003); (Gruber et al., 2014); (Knecht et al., 2010); (Kán et al., 2012);	(Park et al., 1999); (Naimark et al., 2005); (Comport et al., 2003); (Park et al., 2004)	(Dorfmuller-Ulhaas et al., 2001); (Oka,et al., 2002); (Regenbrecht et al., 2013);	(Kanbara et al., 2000); (Kanbara et al., 1999); (Urey et al., 2011);

	HARDWARE						
Problems	Field of View	Occlusion Capable	Focal Plane	Resolution	Portability	Color Blending	Contrast
VST/OST HMD	(Yamazaki et al., 1999); (Orlosky et al., 2014);	(Fischer et al., 2004); (Kiyokawa et al., 2003); (Gao et al., 2012);	(Liu et al., 2010); (Hu et al., 2014); (Dunn et al., 2017);	(Pryor et al., 1998); (Hu et al., 2014);	(Satoh et al., 2001); (Feiner et al., 1997); (Thomas et al., 1998);	(Hincapié-Ramos et al., 2014); (Fukiage et al., 2014); (Klein et al., 2008); (Langlotz et al., 2016);	(Pryor et al., 1998);

5.1 Common Problems

Within the large set of problems that are associated with HMD-based AR systems, there is a subset of problems (usually hardware related) that are present across the entire range of conditions which have an impact on a user’s visual coherence. This section discusses this subset of problems relative to the HMD approaches discussed earlier. Problems which reside outside of this subset are discussed in a later section.

Field-of-View The field-of-view (FOV) problem is related to how much of the user’s view is able to be virtualized. A smaller FOV means very little of the user’s view can be rendered on.

VST setups, using the Oculus Rift for example, have been developed to mitigate for this by providing upwards of a 100 degree horizontal display FOV. Current OST displays are unable to provide such coverage with recent solutions providing only up to 30-40 degree FOV (Yamazaki, Inoguchi, Saito, Morishima, & Taniguchi, 1999). Recently announced OST solutions are claiming to provide a 90 degree FOV (https://www.metavision.com).

Resolution This is a well-known problem and is applicable to both VST and OST HMD approaches. A lack of resolution means virtual content will appear rigid. Leading edge VST displays are able to provide resolutions up to 2560 x 1440 but the user views this at a very close distance. OST displays use projectors to display content, and current solutions are achieving resolutions of around 1268 x 720. Another OST display approach uses lasers pointed directly at the user's retina which provides very high resolution imaging (Pryor et al., 1998) however these approaches are far less common.

Focal Plane With regards to OST, the focal plane relates to where the display projection is in space. Current solutions use a static focal plane which presents difficulties for users. Current research focuses on adjustable focal planes often referred to as liquid lenses (Liu, Hua, & Cheng, 2010; Hu & Hua, 2014). The problem is slightly different for VST approaches and relates to the focus of the video-mediated real world. There is little research into this problem with respect to VST.

Contrast For VST approaches the user's view is isolated from the real world so contrast doesn't have much of an impact on the user's visual experience. However, OST displays have semi-transparent screens and the user sees the real world with their own eyes. This means that when a lot of bright or natural light is present in an environment, the display can easily be washed out. This problem has been addressed using laser projection (Pryor et al., 1998), although this is different display technology to conventional OST displays that use projection systems and is not widely used. An alternative method adopted in many solutions is to use a tinted lens however this is only a temporary solution.

Color Blending This is a known problem with OST displays and is due to the semi-transparent nature of the displays. Bright backgrounds (especially colorful) impact the way the user perceives the color of the projected image. Some research is available which attempts to address the problem (Fukiage, Oishi, & Ikeuchi, 2014; Hincapié-Ramos, Ivanchuk, Sridharan, & Irani, 2014), but the challenge goes largely unsolved. When considering color with respect to VST displays, the problem is more related to inaccurate color replication of the real environment. The colors that are rendered are limited to the color gamut of first the camera, and then the monitors. The advantage of VST is that the user will not necessarily perceive any disparity in color between real and virtual content because it is all displayed on the same gamut (Klein & Murray, 2008).

Occlusion Capability This refers to whether a display is physically capable of achieving occlusion. When developers implement systems using a VST display, they have full control over the rendering pipeline and hence the user's display, so it is possible to implement solutions achieving mutual occlusion (Fischer, Bartz, & Straßer, 2004). In contrast, OST displays have the semi-transparent monitor so it's naturally impossible with such solutions for virtual content to fully occlude real world content. Some research aims to address this problem (Kiyokawa, Billinghurst, Campbell, & Woods, 2003) Gao, Lin, & Hua, 2012), but currently none have been integrated in a portable manner.

Tracking and Registration As previously mentioned, virtual and real world alignment is one of the primary characteristics of Mixed Reality systems. This is the problem of registration in 3D space and the most common method for achieving registration is using tracking algorithms. Implementing tracking for our prototype was beyond our implementation scope for this work, as we are focused mainly on visualization aspects. That does not mean however that if tracking and registration goes unhandled, that it won't have a negative impact on a user's coherence. Instead we decided to statically position and orientate our HMD in space resulting in a less complicated implementation. Tracking has been a significant field of research for some time resulting in multiple solutions. These range from sensor based techniques (gyroscope, accelerometer) to vision based methods (marker-tracking, feature-tracking). Table 1 provides various approaches.

Each of the problems described above exists in current HMD solutions (VST, OST, or both) to the extent that, if unhandled, will affect a user's visual coherence. While it is important that these challenges are overcome, it is worth noting that these are mainly technological limitations. For example, an OST device's FOV will undoubtedly increase as technology evolves. Recently released HMDs such as the HTC Vive, Oculus Rift, and Microsoft HoloLens also provide an increase in resolution and tracking fidelity than the previous generation of HMDs. Therefore, we wish to emphasize the importance for research and development to focus on problems which may not simply be solved by the inevitable advance of technology.

5.2 Challenges

We will now proceed to discuss the main problems identified in our systematic analysis. The problems are divided into hardware and software related challenges. In addition, although not all of the challenges are applicable to each of VST and OST HMD setups, all problems are considered with respect to both.

5.2.1 Occlusion

Occlusion is a prevalent depth cue used to spatially determine object's positions relative to each other. In the context of Mixed Reality, this is otherwise known as mutual occlusion which refers to two base occlusion cases. There is the case where real objects should occlude virtual objects, and then the inverse. Additionally, there are cases where both are required. Occlusion is one of the most prominent depth cues we use to determine an objects' spatial placement in the real world, namely in depth. For example, if a mug is placed in front of a cardboard box then from the perspective of the viewer, some or all of the box will be blocked by the mug. For this case we would say "the mug is occluding the glass." When the problem of occlusion in AR systems goes unhandled, it is very apparent to the user that something is wrong with their environment, even if they can't identify the issue which inevitably causes a severe break in the visual coherence of the scene. Figure 7 illustrates an example of unhandled occlusion. It appears as though the lamp should be sitting on the desk, but once the user tries to move their hand in front of the lamp, it is disorienting and very clear that something is wrong with the scene. This in turn results in a scene which is difficult to interact with.

Figure 7 here

Both VST and OST share certain problems when it comes to occlusion however we will discuss occlusion with respect to each HMD approach remembering the idea of mutual occlusion described above.

We will first consider the cases on the extremes of the RV-Continuum. When all components are real it is obvious that occlusion is not a problem for either VST or OST approaches as all occlusion cases are handled naturally. VR requires a different methodology. Handling occlusion in VR is done automatically using the depth buffer (otherwise known as the Z-buffer). Objects are given a position in a 3D scene which means each has a 'z' coordinate (depth coordinate). Once the camera has been positioned and the view frustum established, the rasterization phase begins. For this process there exists a color buffer, and a depth buffer which are 2D arrays of the same size. For each pixel of each object in the scene the depth value of the pixel is compared to the corresponding value in the depth buffer, and if the new depth value is smaller (closer to the camera than the existing depth value), it will replace the existing one and the color buffer will also be updated with the new color. This process is used for fully virtual environments in both VST and OST systems.

One of the main advantages associated with VST systems is that all content the user sees is video mediated. That means developers have full control over what is rendered to the monitor. When considering this with respect to the problem of occlusion, the rendering of mutual occlusions is possible using software solutions. One of the more simple approaches for occlusion handling is Phantom Model Rendering which requires prior knowledge of the scene (Fischer et al., 2004). The phantom model approach utilizes the 3D models that are already generated as described earlier. To reiterate, it is especially important that the generated 3D models are modelled accurately based on their real world counter-part. Figure 8 demonstrates the obvious perceptual flaws when inaccurate 3D models are generated. The real lamp is now misaligned with the virtual scene.

Figure 8 here

Shaders are applied to the material that defines how the 3D model looks in terms of color. It calculates the resulting color of each pixel of the model based on the material configuration and lighting in the scene. The Phantom Model Rendering approach uses a custom shader to render each of the model’s pixels invisible so that it can’t be seen. Although the pixels are being rendered invisible, they still hold their position in the Z-buffer which creates what is referred to as a phantom model. Figure 9 (A) demonstrates the virtual cube occluding the real lamp. This is normal behavior however Figure 9 (B) shows the real lamp occluding the virtual cube once it is pushed behind. Figure 9 (C) and (D) are images from an arbitrary viewing point in the virtual space, taken for the same scene as (A) and (B) respectively. It is clear that with the lamp component in the scenario being made a phantom model, it still occludes the virtual desk and cube. The video texture can be seen in the 3D virtual space in the background though it does not align for (C) and (D) due to the arbitrary positioning of the view point. The blue book present in (C) and (D) is the real book sitting on the real desk which is being visualized on the video texture in the background. During development of our system we decided to make the virtual representation of the book green as opposed to blue so as to make distinguishing the virtual and real object representations easier during analysis.

Figure 9 here

Optical see-through setups can also implement such approaches, however they have an additional occlusion problem which is hardware related. Handling occlusions between virtual and real objects is discussed in the previous section however this only solves mutual occlusions for VST (on a software basis), but the problem of occlusion on OST displays persists through the software based solutions. If, for example, the user's hand is virtual and the user reaches into the scene, the virtual hand mesh will appear on top of the user's hand as one would expect, however, because the display is semi-transparent, it is not capable of rendering fully opaque content. This means that the user will still be able to see their real hand no matter what. The virtual mesh will appear somewhat ghostly as compared to VST implementation where the virtual hand mesh is able to fully occlude the real content in the background. This is one of the fundamental challenges faced with OST displays. No production HMD's have yet managed to implement a solution for this. There is research available that addresses the problem with novel solutions such as (Kiyokawa et al., 2003) and (Kiyokawa, Kurata, & Ohno, 2000).

The problem of occlusion was found to be prominent throughout the process of our systematic analyses. It is handled automatically in terms of fully virtual environments however when tangible real and virtual components are added to a scene together, occlusion must always be handled for both VST and OST systems. The software solutions provided here are enough to mitigate the problem of occlusion in VST systems, however OST presents an additional problem due the semi-transparent nature of the device. This is unsolved in application ready HMDs.

5.2.2 Lighting and Shadows

In any application that mixes real and virtual objects, both real and virtual lighting must be considered. In the context of our prototype the desk lamp is the object that casts either a real or a virtual spotlight into the environment. Before the spotlight is switched on, existing illumination must be considered. This is often referred to as ambient lighting in an environment. For example, Figure 10 demonstrates the scene before the lamp is turned on (left) and after (right). Before the light is switched on there is already illumination in the environment as is illustrated by the existing shadows falling on the desk behind the lamp.

Figure 10 here

Given these two types of illumination within an environment, there are two possible interactions between lighting and any other component. That is, 1) the application of real light to virtual content, or 2) the application of virtual light to real content. The case of virtual light illuminating virtual objects in a scene is solved, usually using ray-casting approaches. The challenge is in matching the virtual illumination model to the real world illumination model as to provide visual continuity. A general overview of illumination techniques is presented by Jacobs et al. (2006). Three main categories of techniques are given as: common illumination, relighting, and methods based on inverse illumination. The more information an illumination model attains from an environment the more effective the result will be. However, slower processing times are required rendering such techniques less effective for real-time application. A developer must find a balance appropriate for their application context.

In any system with a combination of virtual and real components, the real world illumination model must be considered. One of the more researched lighting-specific challenges is replication of environmental light. This is an important consideration: if virtual objects in the scene are illuminated with lighting (virtual) inconsistent with real world illumination, visual discontinuity will result. We describe here the desire to illuminate virtual objects with real illumination. Applying real lighting models to virtual content is desirable and achievable in both VST and OST HMDs. Kanbara et al., and Agusanto et al. provide various solutions for this problem (Agusanto, Li, Chuangui, & Sing, 2003; Kanbara, Yokoya, & Takemura 2004). In reversing the application of light to apply virtual light to real components we encounter an entirely different problem. This is far more of a challenge, namely for OST setups. In VST setups the developer is able to manipulate the entire visual medium, specifically, the color sensor data. In our implementation, we are able to apply a virtual spotlight to the real environment resulting in the illumination of the desk and notepad (shown in Figure 11.) These approaches are considered as image-based methods (Raskar, Welch, Low, & Bandyopadhyay, 2001).

Figure 11 here

It is possible for one to implement the above technical solutions for an OST HMD setup, however, the result will provide less satisfying visual coherence than on a VST setup. Using Figure 11 as an example, the illuminated desk and notepad is the result of operating on the data representing the real world. In an OST setup the user looks through the display to the actual world. It is possible to virtually illuminate the region where the illumination spot would be although this will only partially solve the problem. This is mostly due to the occlusion issues of OST devices discussed earlier.

Shadows are a direct result of illumination in an environment and they are more than a cosmetic element; they are used as visual cues for spatial relationships between objects. Computing shadows has been a topic of research for over three decades. Crow presents a comparison of earlier methods of producing shadows in synthesized images (1977). Similar approaches are used today for shadow generation in 3D scenes. Haller et al. present an approach utilizing shadow volumes which was deemed by Crow as the preferable technique (Haller, Drab, & Hartmann, 2003). Jacobs et al. present an alternative real-time shadow rendering solution for a MR system (2005).

More complex shadow rendering issues arise when multiple shadows are present. If there are, for example, two light sources in a room causing objects to cast two shadows, these should overlap in a realistic manner. This increases in complexity when we consider a mix between real and virtual shadows. Implementation of a solution to this problem is complex and is out of the scope of our prototype implementation, though it should be a serious consideration for any productive system.

5.2.3 Interaction

The introduction of the interactive component into MR environments once again emphasizes the need for real and virtual world alignment. While visual disparity is bad, the absence of an interaction component places a severe limitation on system development.

There are a lot of challenges that remain unsolved when it comes to interacting with virtual content. One of the biggest challenges is applying haptic and tactile feedback when interacting with virtual objects. Some research has achieved dynamic tactile feedback when a user touches a physical object (Bau & Poupyrev, 2012). This still requires a user to touch a real object which means once more, virtual and real world alignment is essential. Additionally, this requires instrumentation of one's environment. A lot of research looks to apply haptic feedback in VR and AR surgical training applications (Yudkowsky et al., 2013).

An additional challenge of real time interaction with virtual content is the ability to track any user's hand and fingers in space. The most reliable of current methods require the instrumentation of a user's hand with a glove (Dorfmuller-Ulhaas & Schmalstieg, 2001). A hybrid AR finger tracking method was implemented using 2D image processing to augment the user's hand into a virtual environment and tracking in 2D space, and a Leap Motion for the depth tracking of fingers (Regenbrecht, Collins, & Hoermann, 2013). Research targeting finger tracking is limited however various technologies for this purpose are being released in the form of TOF cameras and infrared sensors which can facilitate progressive research. It is worth noting here that most finger tracking solutions can be applied to either of OST or VST HMD setups.

There are two main approaches to hand and finger tracking in 3D space: 1) image based and, 2) depth sensor based. Image based approaches use a color sensor to capture the environment in front of a user and when the user moves their hand through the image, an algorithm processes the image data and works to identify the hands and fingers. These approaches are still infantile and have limitations (Dorfmuller-Ulhaas & Schmalstieg, 2001; Oka, Sato, & Koike, 2002; Regenbrecht et al., 2013).

3D sensor based approaches provide a more reliable result than image-based approaches however, limitations remain. A common limitation of all finger tracking approaches is the form of the hand. Usually only a single camera is used (color or depth) to acquire data for tracking, and given the user can move or rotate their hand in such a way that it cannot entirely be seen makes it difficult to track.

Within our implementation, we used the TOF camera. A problem case we found was each time the user would reach close to any other object such as the lamp or the surface of the desk, the finger tracking algorithm would lose the tracking of the hand. Figure 12 (left) demonstrates the image from the depth camera when the hand is in open space and it is clear which pixels belong to the hand. Figure 12 (right) shows the hand held up with a book being held behind it. The colors of the hand against the surface behind it are almost identical. This makes it difficult to identify the hand and fingers.

Figure 12 here

Due to the difficulties of hand and finger tracking in real time and the lack of available solutions, companies that have released the most recent VR systems use hardware components as interaction mediums for users. The HTC Vive system developed by HTC and Valve, and the Oculus Rift are examples of such systems. While handheld controllers can be used to achieve various forms of interaction, they do not allow for the user to use their hands and fingers directly which is the ideal case as this facilitates more complex interactions within a system.

5.2.4 Stereoscopic Implementation

The system developed for this work is a monoscopic implementation which means for one eye. A stereoscopic system is one which is implemented for both eyes of a user. OST devices have the advantage here that the user already sees the real world with a stereoscopic view which is their natural sight. The only thing that needs to be handled after that is the virtual content. The means for producing such a system is by placing two virtual cameras in the 3D scene instead of just one. The virtual cameras should then be placed the same distance apart as the user’s eyes. This means that another user specific calibration is required to determine the user’s interpupillary distance (IPD). This defines the exact distance, usually in mm, between a user's eyes. Once this is known, the virtual cameras (for each eye) can be placed the appropriate distance apart. That way the virtual camera on the left renders content to the left monitor and the right virtual camera renders content to the right monitor. The user will then see stereoscopically rendered virtual material added to their naturally stereoscopic view of the real world. It is also important that not only the virtual cameras are set according to the user's IPD, but that the HMD display monitors are also adjusted. If this is unhandled it can result in discomfort for the user (Sharples, Cobb, Moody, & Wilson, 2008).

Achieving such an implementation for a VST approach is more difficult. Due to the fact that the real world is mediated using a camera, the real world is only seen through one lens. This could be thought of as looking into the world with one eye. It can still be implemented however if implementation of virtual content is stereoscopic, the user will see a disparity between the monoscopic real world and the stereoscopic virtual content. To correctly implement stereoscopy, two color sensors are required. In addition to positioning the two virtual cameras that represent the user's eyes based on the users IPD (as specified previously), the physical cameras also need to be positioned based on the same value. Incorrect implementation of this can cause adverse effects in users ranging from general discomfort to headaches and nausea (Hakkinen, Vuori, & Paakka, 2002). If correctly implemented, the user will see a video-mediated stereoscopic version of the real world and virtual content will also be rendered stereoscopically.

The viewing angle is an additional consideration for developers. Ideally, the camera would view the world from where the user's eye is, but due to the HMD being worn, it is not possible. Usually the camera is mounted on top of the HMD however this means the video supplied to the display is from a different perspective than the user would expect to be looking from. As previously identified, this is known to cause possible discomfort in user's perception. Some implementations attempt to mitigate the problem by placing the camera (or two cameras for stereoscopic) on the front of the HMD as shown in Figure 13. This means that at least the viewing angle will be similar to the user's view however there is still the distance from the user's eyes to where the cameras are mounted to consider. Kanbara et al. present a stereoscopic VST implementation (M. Kanbara, Okuma, Takemura, & Yokoya, 2000). Kanbara et al. also discuss the composition of stereoscopic images (M. Kanbara, Okuma, Takemura, & Yokoya, 1999).

Figure 13 here

5.2.5 Calibration

There are different forms of calibration with regards to MR systems which vary between VST and OST approaches. The main forms of calibration include: 1) camera calibration, 2) HMD – eye calibration, and 3) hardware – hardware calibration.

Camera calibration is dependent on whether the hardware configuration of a system requires a camera. If the system is using a vision-based tracking algorithm which requires a color sensor, then generally a camera is required to be mounted on the HMD. Furthermore if the system implements vision-based finger tracking, whether image-based or depth-based, a camera will be required. In any of these cases where a camera is present, the sensor will need to be intrinsically calibrated as to attain the exact parameters which define the sensor. Different sensors hold different properties. Common properties considered in research systems include the focal length, principle point, skew coefficient, distortion coefficient, and the true FOV is determined from these values. The importance of the FOV property is emphasized in a previous section (4.2.4). Intrinsic calibration of color sensors is a well-known problem, and has been solved for some time. We used an OpenCV implementation for the calibration of our color sensor.

Depth sensor calibration is more difficult, however there is research which aims to solve this challenge. Because the typical checker-board approach is unreliable using infrared sensing technology, some researchers have combined a color and depth calibration resulting in more accurate results than the standard manufacturer calibration (Zhang & Zhang, 2014; Herrera, Kannala, & Heikkilä, 2012). There are cases however where a color camera may not be present. Linder et al. presents an approach for standalone calibration of time-of-flight sensors (Lindner, Schiller, Kolb, & Koch, 2010).

The second type of calibration is HMD – eye calibration, otherwise known as hand-eye calibration. The first consideration to make is common across both VST and OST HMDs and that is a user's physical characteristics (mainly with regards to their eyes). The most prominent of characteristics to consider is the user's IPD which is the distance between the centers of each eye. If these characteristics go unhandled, a range of effects are known to present themselves in users varying from dizziness and headaches to nausea or even vomiting (Hakkinen et al., 2002). One possible way to find the user's IPD is by using eye tracking solutions built into an HMD (Hu & Hua, 2014). When a user is wearing an OST HMD and visualizing virtual content, if the HMD moves on the user's head the HMD requires recalibration because the virtual content will no longer be rendered in the correct location. Eye tracking solutions can also help with this issue. There is a lack of solutions (especially consumer ready) that succeed in solving this problem efficiently. Other solutions for calibration of HMD - eye include (Itoh & Klinker, 2014; Owen, Zhou, Tang, & Xiao, 2004; Tuceryan, Genc, & Navab, 2002).

Hardware-hardware calibrations are generally simpler to solve. When a HMD setup is constructed it's important that the pose difference between the various components is known. If we consider the Meta One HMD used for this work. The DS325 camera mounted on the HMD has both a color sensor and a TOF sensor. If the distance between the two sensors is not accounted for, the hand detected by the depth sensor will not align with the image taken from the color sensor. There will be an offset which will need to be built into the system. Disparities of this form should always be considered or it will lead to a (possibly severe) break in the user's visual coherence.

6 Addressing Coherence Issues

This chapter provides a summary of the findings from the systematic enquiry of both VST and OST systems. Challenges that are encountered throughout the analysis are identified and then categorized by frequency, and the impact of each on the final system fidelity.

The categories are divided into hardware and software challenges. The primary hardware issues surrounding MR systems include: HMD FOV, occlusion capability, addressable focal plane, resolution, portability, color fidelity, and contrast. The software challenges consist of: occlusion, lighting and shadows, calibration, tracking and registration, finger tracking, and stereoscopic implementation.

Each category is weighted in two separate tables based on two significant classifiers:

- the availability of existing research, or commercial solutions that aim to address the problem
- the importance of addressing the problem, in order to enable the productive development of applications

For each weighting table, there are two main HMD approaches for the implementation of MR, namely VST and OST. The enquiry provides a systematic investigation. Given the comprehensive nature of the enquiry, one can establish weighting values to the identified categories for each of the said classifications.

With the categories defined, the first weight table can be discussed. Weight is given on a scale where:

- Little Coverage – there is very little research or available solutions addressing the problem
- Moderate Coverage – there is some research or available solutions addressing the problem
- Extensive Coverage – there is an abundance of research or available solutions addressing the problems

The first weighting table (shown in Table 2) is based on research and solution availability. The enquiry clarifies the difference in available solutions between VST and OST systems. In contrast to OST setups, VST implementations do not possess the same hardware limitations, i.e. field of view, contrast, or occlusion capabilities.

The concept of FOV is covered at multiple points in this work and covers the limitations of OST displays with regards to the FOV. Common VST HMD's such as the Oculus Rift provide a much larger FOV and while some research focuses on this issue in the context of OST setups, at this stage, only small improvements have been made on existing optical technologies.

The same can be said for occlusion capable hardware. VST systems provide full control over the rendering pipeline, so occlusion-oriented research is able to take advantage of this characteristic and provide multiple solutions for occlusion. Semi-transparent OST monitors do not provide the same flexibility which means the same software-based-occlusion solutions used for VST setups are unable to solve all occlusion cases for OST systems.

Research on the topic of focal planes is limited however a number of early prototypical solutions for variable focal plane implementation have been produced (Hu & Hua, 2014; Liu et al., 2010). This research is still infantile with ample room for improvement and reliable integration with wearable OST HMD's.

Color fidelity is well known as an OST specific problem however it is still relevant in VST setups. The Research and Solutions table (Table 2) specifies this as "Color Cohesion". Color cohesion refers to the integration of the colors from the real and virtual scenes and how much visual disparity exists.

Portability (in terms of MR systems) refers to the free physical movement of the HMD in space without diminishing the user experience. There are two common approaches to achieving portability: (1) wireless communication technology and (2) on-board computation. Wireless communication technology is still under development though solutions have been introduced for immersive VR systems such as the HTC Vive. This allows for the user to walk around freely (within the tracking space) without being tethered to a PC. On-board computation is an alternative which is exemplified with systems such as Microsoft's HoloLens, the Meta Company's Meta One, or the Samsung GearVR. These kinds of systems usually have a drawback that is a lack of computation power. There are some rendering workarounds that can be implemented by developers, though for obvious reasons these systems don't come close to solutions running on desktop computers.

Moreover, specific only to OST displays is the contrast problem. Retina display prototypes project the virtual display directly to the user's eye. While this provides an alternative display method that doesn't suffer contrast issues, users are often uncomfortable with the concept and the community has not embraced the technology. More focus can be given to this challenge.

Table 2. The Research & Solutions table provides weighting values providing a summary of the availability of research with respect to specific categorical problems.

RESEARCH & SOLUTIONS							
	Hardware						
Category	FOV	Occlusion Capable	Focal Plane	Resolution	Portability	Colour Cohesion	Contrast
VST	Extensive Coverage	Extensive Coverage	Little Coverage	Extensive Coverage	Moderate Coverage	Little Coverage	N/A
OST	Moderate Coverage	Moderate Coverage	Little Coverage	Moderate Coverage	Moderate Coverage	Little Coverage	Moderate Coverage

	Software					
Category	Occlusion	Lighting + Shadows	HMD-User Calibration	Tracking + Registration	Finger Tracking	Stereo Impl.
VST	Extensive Coverage	Extensive Coverage	Moderate Coverage	Extensive Coverage	Little Coverage	Extensive Coverage
OST	Little Coverage	Moderate Coverage	Little Coverage	Little Coverage	Little Coverage	Extensive Coverage

Due to the aforementioned additional hardware limitations of OST devices, research tends toward more frequently providing software solutions in the context of VST HMD's. As specified earlier, software based solutions which address the problem of occlusion can solve the problem with reasonable reliability. Because developers have access to the full rendering pipeline, there is more flexibility for controlling lighting conditions and various techniques of virtually illuminating real world content. This remains a complex problem and therefore there is still room for additional research, particularly for OST approaches.

Another large problem with OST setups is the requirement to calibrate the display to each individual user. Additionally, each single movement of the HMD after the calibration process means that the display will be misaligned with the user’s eye and another calibration must be carried out. This problem is under investigation and techniques have been presented, but once again, these are either not easily implemented, used, or do not provide a complete solution.

Tracking the HMD in space is an ongoing area of research which is applicable to both VST and OST. This problem has multiple solutions, each of which has advantages and disadvantages. Precise knowledge of the HMD’s position and orientation (pose) in space is required for accurate real and virtual world alignment which in most cases is essential for a productive system application. It is worth noting that these solutions may be utilized but are by no means 100 percent reliable. For this reason, continued research into tracking algorithms and technologies is essential.

Finger tracking has the potential to revolutionize the interaction component of MR systems. This technology is still infantile in its development however there has been an increased commercial interest in technologies aiming to provide solutions in the context of augmented and virtual realities such as Leap Motion. There have been some novel solutions presented in order to provide more precise interaction with virtual content but these are still prototypical systems that require more research and development to become useful for productive systems.

The identified gaps in research provide additional context for determining problematic areas (in the MR problem space) that require research and development. The next weighting table is now discussed where:

- Low Priority – This set of problems is solved to a reasonable standard with reliable, robust, and readily available solutions.

- Medium Priority – There exist some solutions to this set of problems, however the remaining issues are still severe enough that more research and development is justified.
- High Priority – This set of problems requires immediate attention in the form of research and development of prototypical systems in order for productive systems to be feasible.

The Priorities weighting table (Table 3) has an inverse relationship with the Research and Solutions table. Although VST systems have received more attention, or have at least had more successful prototypical research solutions provided, there are still some categories that require attention.

Table 3. The Priorities for Research & Development table provides values respective of how important it is for each categorical problem to be solved.

PRIORITIES FOR DEVELOPMENT							
	Hardware						
Category	FOV	Occlusion Capable	Focal Plane	Resolution	Portability	Colour Cohesion	Contrast
VST	Low Priority	Low Priority	Medium Priority	Low Priority	Medium Priority	Low Priority	N/A
OST	High Priority	High Priority	Medium Priority	Medium Priority	Low Priority	Medium Priority	High Priority
	Software						
Category	Occlusion	Lighting + Shadows	HMD-User Calibration	Tracking + Registration	Finger Tracking	Stereo Impl.	
VST	Low Priority	Medium Priority	Medium Priority	High Priority	High Priority	Medium Priority	
OST	High Priority	Medium Priority	High Priority	High Priority	High Priority	Medium Priority	

Tracking and registration is a very high priority challenge for both VST and OST approaches. This is because one of the defining properties of mixed reality is the combination and alignment of real and virtual worlds. As earlier specified, research exists aiming to solve the problem of which each solution is tailored for a somewhat specific environment. Within vision based techniques there are marker-based and marker-less tracking algorithms. Each of these different approaches is useful for a given environment. Marker tracking is usually only used for indoor MR systems because they usually have to be relatively close range, and positioned relative to simple geometry. This would not be useful for outdoor applications. Microsoft released their latest OST HMD, HoloLens, which reliably tracks the user in space with marker-less tracking techniques (SLAM) by reconstructing the surrounding geometry. This exemplifies the progress being made in the commercial sector though there is still room for improvement. While a lot of work has gone into tracking and registration, especially for VST approaches, more work needs to be done in order to accurately determine a user's pose in space and in a diverse set of environments.

Finger tracking is also given a very high priority for both HMD approaches. This is due to the high level interaction requirement that so many applications demand. Robust and readily available finger tracking software is not yet provided although the commercial sector appears to be putting work in to developing technologies. While this is promising and important, researchers should still give weight to prototypical development of such technologies to provide application context and provide more unique solutions for investors. Recent hardware solutions provide valid hardware-based interaction mediums, though finger tracking will provide preferred interaction techniques.

The two highest priority hardware challenges are both for OST displays. The FOV problem needs to be addressed because when an end user straps one of the newly hyped OST HMD’s on and realizes that only a very small portion of the room is able to be virtualized at any one time, they will feel disappointed, because it isn’t like the demonstration system in the promotional video. If this happens to enough end-users the technology will lose demand and investors will also potentially lose interest.

The other hardware challenge with a high priority is providing OST displays with occlusion capabilities. As discussed earlier, VST systems provide full control over the rendering pipeline, so occlusion is manageable. With typical OST setups, occlusion cannot be solved due to the semi-transparent nature of the monitors. A new hardware solution is required in a compact, robust, and reasonably affordable manner if occlusion is to be solvable for the OST HMD.

The focal plane problem is more obvious within OST setups, however it should be considered within the context of VST systems. VST systems should account for the difference between the focal plane of the real world mediation (from the camera/cameras) and the focus within the virtual environment. The problem is not as noticeable in VST setups, which is why it receives a moderate rating. Focal planes in OST are given a relatively high priority because static focal plane distance is more apparent to a user and can cause discomfort in their experience.

Lower weighted problems to be solved include resolution, color fidelity, and portability. While these are important problems, if their solutions are under-developed it will not necessarily break the system making it less likely productive implementations will be held back.

VST systems qualify for a higher weight for portability because they are generally heavier and require more instrumentation such as one or multiple cameras for real world mediation however a productive system may not be held back because of a slightly larger display setup. OST displays are naturally more lightweight in design, and some portable solutions exist.

VST systems aren't generally affected by lighting conditions because the user and the monitors are visually isolated from the environment. In OST systems, bright lighting has a significant effect on the contrast of the virtual display, and therefore virtual content can be washed out so the user cannot see. This problem needs to be addressed, especially in the context of outdoor mixed reality. Outside lighting is significantly brighter than indoor light. If the user cannot see the content, the system cannot effectively be used.

Calibration for OST systems receives a high priority weighting because this process needs drastic improvement before the everyday use of OST applications is possible. There are few solutions for this problem today. The calibration process is essential for correct perceptual alignment of virtual and real worlds, and therefore, for MR application using OST technology. As specified earlier, the difficulties lie in the uniqueness of each individual user, and also in the inevitable small and large movements of the HMD on the user's head. Calibration is still necessary for VST systems however there are less confounding factors. The HMD moving on the user's head does not have a large impact because the content is all rendered through one monitor.

Stereoscopic implementation is given moderate-high priority for each of VST and OST. It is more difficult for true stereoscopic implementation in VST approaches because the real world needs to be mediated which means two cameras need to be mounted where the user's eyes are. Because users look straight through OST displays they already have a real world stereoscopic view. When virtual content is rendered monoscopically through OST devices, there are perceptual disparities due to the difference between the user's natural vision and the configuration of the virtual environment. This can cause discomfort in a user resulting in various effects from headaches to nausea, depending on the user. If long term use applications are desired then stereoscopic concepts should be resolved.

Occlusion issues in VST systems are solved to a level where it will provide a robust enough solution for production level systems. A low priority weight is therefore given in contrast to OST that is given high priority. It is likely that a lot of research and development focus on hardware solutions will be required to solve systemic form-based problems OST displays possess.

Handling lighting and shadows in VST systems are more manageable than with OST systems. Lighting and shadows provide users with depth cues and visual acceptance of a scene. Replicating real lighting environments requires pre-processing methods and knowledge of the environment however if these methods are not implemented, the visual acceptance of the user may be reduced though it will not necessarily hold back a MR system.

Based on these summaries we are able to see where there is both space and need for research and development on various topics and which problem spaces are not critical to the proliferation of MR. Several companies are developing HMD solutions, mainly OST, such as Microsoft HoloLens, Meta Company, and Epson. Microsoft has been able to develop a reliable SLAM tracking system which enables them to achieve, to a reasonable level, certain occlusions. Additionally they have solved, to a degree, the HMD-eye calibration using cameras which track the user's eyes. Meta Company's solution specifies a 90 degree FOV claiming to be the widest FOV available in such HMDs. While these development efforts are pushing these technologies forward, it remains necessary that we progress and refine these solutions insofar as they become widely available.

7 Conclusion

This work provides two primary contributions. The first is a Visual Components Framework outlining the high level core components of any Augmented/Mixed Reality prototype. In order to exemplify the framework, we designed and developed an accompanying prototype that represents a very basic office scenario. The second contribution is a comprehensive enquiry which utilizes the framework and prototype. The enquiry identifies the challenges faced in the design and implementation of AR/MR systems, and provides existing works that aim to solve such problems. In addition, the enquiry distinguishes problems that, to this day, are yet to be solved; problems which continue to impede the productive development of MR systems. The principle scope of the enquiry is based on user's visual coherence, and the elements of a system (hardware or software related) which contribute to disturbances in the said visual coherence. In the context of the field of presence we provide a systematic analysis of MR systems, primarily the visualization and interaction aspects. Visualization is a significant factor contributing to a user's sense of presence and therefore the analysis provides insight into which components of a system may have an impact on presence. The interactive component of MR systems is also covered including not only interaction between user and world, but that between various high level components that may be present. Once more, the systematic nature of this enquiry provides comprehensive insight into the significance of various system components. This provides relevant material for the field of HCI.

8 Acknowledgements

We wish to thank the Information Science department for their support during this work. Further thanks go to Chris Edwards, Mohammed Alghamdi, and Abdulaziz Alshaer for assisting with the prototype development. Thanks also to Robin Alden of iMonitor for his support. This work was funded by Callaghan Innovation New Zealand.

9 References

Agusanto, K., Li, L., Chuangui, Z., & Sing, N. W. (2003). Photorealistic rendering for augmented reality using environment illumination. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings* (pp. 208–216). https://doi.org/10.1109/ISMAR.2003.1240704

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE, 21*(6), 34–47.

Azuma, R. T. (1997). A survey of augmented reality. *Presence*, *6*(4), 355–385.

Bau, O., & Poupyrev, I. (2012). REVEL: Tactile Feedback Technology for Augmented Reality. *ACM Transactions on Graphics, 31*(4), 89:1–89:11. https://doi.org/10.1145/2185520.2185585

Berger, M. O. (1997). Resolving occlusion in augmented reality: a contour based approach without 3D reconstruction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 91–96). https://doi.org/10.1109/CVPR.1997.609304

Comport, A., Marchand, É., & Chaumette, F. (2003). Robust and real-time image-based tracking for markerless augmented reality (report). INRIA. Retrieved from https://hal.inria.fr/inria-00071736/document

Crow, F. C. (1977). Shadow Algorithms for Computer Graphics. In *Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 242–248). New York, NY, USA: ACM. https://doi.org/10.1145/563858.563901

Dorfmuller-Ulhaas, K., & Schmalstieg, D. (2001). Finger tracking for interaction in augmented environments. In *IEEE and ACM International Symposium on Augmented Reality, 2001. Proceedings* (pp. 55–64). https://doi.org/10.1109/ISAR.2001.970515

Dunn, D., Tippets, C., Torell, K., Kellnhofer, P., Aksit, K., Didyk, P., … Fuchs, H. (2017). Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Transactions on Visualization and Computer Graphics*. Retrieved from http://ieeexplore.ieee.org/abstract/document/7829412/

Echtler, F., Sturm, F., Kindermann, K., Klinker, G., Stilla, J., Trilk, J., & Najafi, H. (2004). The Intelligent Welding Gun: Augmented Reality for Experimental Vehicle Construction. In S. K. Ong & A. Y. C. N. DEng (Eds.), *Virtual and Augmented Reality Applications in Manufacturing* (pp. 333–360). Springer London. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4471-3873-0_17

Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997). A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In *Digest of Papers. First International Symposium on Wearable Computers* (pp. 74–81). https://doi.org/10.1109/ISWC.1997.629922

Feiner, S., Macintyre, B., & Seligmann, D. (1993). Knowledge-based Augmented Reality. *Communications of the ACM*, *36*(7), 53–62. https://doi.org/10.1145/159544.159587

Fischer, J., Bartz, D., & Straßer, W. (2004). Occlusion Handling for Medical Augmented Reality Using a Volumetric Phantom Model. In *Proceedings of the ACM Symposium on Virtual*

Reality Software and Technology (pp. 174–177). New York, NY, USA: ACM. https://doi.org/10.1145/1077534.1077570

Fite-Georgel, P. (2011). Is there a reality in Industrial Augmented Reality? In *2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 201–210). https://doi.org/10.1109/ISMAR.2011.6092387

Fukiage, T., Oishi, T., & Ikeuchi, K. (2014). Visibility-based blending for real-time applications. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 63–72). IEEE. Retrieved from http://ieeexplore.ieee.org.ezproxy.otago.ac.nz/xpls/abs_all.jsp?arnumber=6948410

Gao, C., Lin, Y., & Hua, H. (2012). Occlusion capable optical see-through head-mounted display using freeform optics. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 281–282). https://doi.org/10.1109/ISMAR.2012.6402574

Gilson, S. J., Fitzgibbon, A. W., & Glennerster, A. (2008). Spatial calibration of an optical see-through head-mounted display. *Journal of Neuroscience Methods*, *173*(1), 140–146. https://doi.org/10.1016/j.jneumeth.2008.05.015

Gruber, L., Langlotz, T., Sen, P., Hoherer, T., & Schmalstieg, D. (2014). Efficient and robust radiance transfer for probeless photorealistic augmented reality. In *2014 IEEE Virtual Reality (VR)* (pp. 15–20). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/6802044/

Hakkinen, J., Vuori, T., & Paakka, M. (2002). Postural stability and sickness symptoms after HMD use. In *IEEE International Conference on Systems, Man and Cybernetics* (Vol. 1, pp. 147–152). Retrieved from http://s3.amazonaws.com/publicationslist.org/data/jukka.hakkinen/ref-22/Hakkinen%202002.pdf

Haller, M., Drab, S., & Hartmann, W. (2003). A Real-time Shadow Approach for an Augmented Reality Application Using Shadow Volumes. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (pp. 56–65). New York, NY, USA: ACM. https://doi.org/10.1145/1008653.1008665

Herrera C., D., Kannala, J., & Heikkilä, J. (2012). Joint Depth and Color Camera Calibration with Distortion Correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(10), 2058–2064. https://doi.org/10.1109/TPAMI.2012.125

Hincapié-Ramos, J. D., Ivanchuk, L., Sridharan, S. K., & Irani, P. (2014). SmartColor: Real-time color correction and contrast for optical see-through head-mounted displays. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 187–194). IEEE. Retrieved from http://ieeexplore.ieee.org.ezproxy.otago.ac.nz/xpls/abs_all.jsp?arnumber=6948426

Hu, X., & Hua, H. (2014). High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Optics Express*, *22*(11), 13896. https://doi.org/10.1364/OE.22.013896

Itoh, Y., & Klinker, G. (2014). Interaction-free calibration for optical see-through head-mounted displays based on 3D Eye localization. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 75–82). https://doi.org/10.1109/3DUI.2014.6798846

Jacobs, K., & Loscos, C. (2006). Classification of Illumination Methods for Mixed Reality. *Computer Graphics Forum*, *25*(1), 29–51. https://doi.org/10.1111/j.1467-8659.2006.00816.x

Jacobs, K., Nahmias, J.-D., Angus, C., Reche, A., Loscos, C., & Steed, A. (2005). Automatic Generation of Consistent Shadows for Augmented Reality. In *Proceedings of Graphics Interface 2005* (pp. 113–120). School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada: Canadian Human-Computer Communications Society. Retrieved from http://dl.acm.org/citation.cfm?id=1089508.1089527

Kán, P., & Kaufmann, H. (2012). High-quality reflections, refractions, and caustics in augmented reality and their contribution to visual coherence. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 99–108). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/6402546/

Kanbara, M., Okuma, T., Takemura, H., & Yokoya, N. (1999). Real-time composition of stereo images for video see-through augmented reality. In *IEEE International Conference on Multimedia Computing and Systems, 1999* (Vol. 1, pp. 213–219 vol.1). https://doi.org/10.1109/MMCS.1999.779195

Kanbara, M., Okuma, T., Takemura, H., & Yokoya, N. (2000). A stereoscopic video see-through augmented reality system based on real-time vision-based registration. In *IEEE Virtual Reality, 2000. Proceedings* (pp. 255–262). https://doi.org/10.1109/VR.2000.840506

Kanbara, Masayuki, & Yokoya, N. (2004). Real-time Estimation of Light Source Environment for Photorealistic Augmented Reality. In *International Conference on Pattern Recognition (2)* (pp. 911–914). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.4153&rep=rep1&type=pdf

Kiyokawa, K., Billinghurst, M., Campbell, B., & Woods, E. (2003). An Occlusion-Capable Optical See-through Head Mount Display for Supporting Co-located Collaboration. In *Proceedings of the 2nd IEEE International Symposium on Mixed and Augmented Reality*

(p. 133–). Washington, DC, USA: IEEE Computer Society. Retrieved from http://dl.acm.org/citation.cfm?id=946248.946788

Kiyokawa, K., Kurata, Y., & Ohno, H. (2000). An optical see-through display for mutual occlusion of real and virtual environments. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality* (ISAR 2000) (pp. 60–67). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/880924/

Klein, G., & Murray, D. (2008). Compositing for Small Cameras. In *Proceedings of the 7th IEEE International Symposium on Mixed and Augmented Reality* (pp. 57–60). Washington, DC, USA: IEEE Computer Society. https://doi.org/10.1109/ISMAR.2008.4637324

Knecht, M., Traxler, C., Mattausch, O., Purgathofer, W., & Wimmer, M. (2010). Differential instant radiosity for mixed reality. In *2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 99–107). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/5643556/

Kruijff, E., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 3–12). https://doi.org/10.1109/ISMAR.2010.5643530

Langlotz, T., Cook, M., & Regenbrecht, H. (2016). Real-Time Radiometric Compensation for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics, 22*(11), 2385–2394.

Lindner, M., Schiller, I., Kolb, A., & Koch, R. (2010). Time-of-Flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding, 114*(12), 1318–1328. https://doi.org/10.1016/j.cviu.2009.11.002

Liu, S., Hua, H., & Cheng, D. (2010). A Novel Prototype for an Optical See-Through Head-Mounted Display with Addressable Focus Cues. *IEEE Transactions on Visualization and Computer Graphics*, *16*(3), 381–393. https://doi.org/10.1109/TVCG.2009.95

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *77*(12), 1321–1329.

Naimark, L., & Foxlin, E. (2005). Encoded LED system for optical trackers. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)* (pp. 150–153). https://doi.org/10.1109/ISMAR.2005.28

Oka, K., Sato, Y., & Koike, H. (2002). Real-time fingertip tracking and gesture recognition. *IEEE Computer Graphics and Applications*, *22*(6), 64–71. https://doi.org/10.1109/MCG.2002.1046630

Orlosky, J., Wu, Q., Kiyokawa, K., Takemura, H., & Nitschke, C. (2014). Fisheye Vision: Peripheral Spatial Compression for Improved Field of View in Head Mounted Displays. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (pp. 54–61). New York, NY, USA: ACM. https://doi.org/10.1145/2659766.2659771

Owen, C. B., Zhou, J., Tang, A., & Xiao, F. (2004). Display-relative calibration for optical see-through head-mounted displays. In *Third IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 70–78). https://doi.org/10.1109/ISMAR.2004.28

Park, H., & Park, J.-I. (2004). Invisible Marker Tracking for AR. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 272–273). Washington, DC, USA: IEEE Computer Society. https://doi.org/10.1109/ISMAR.2004.37

Park, J., Jiang, B., & Neumann, U. (1999). Vision-based pose computation: robust and accurate augmented reality tracking. In *2nd IEEE and ACM International Workshop on Augmented Reality, 1999. (IWAR '99) Proceedings* (pp. 3–12). https://doi.org/10.1109/IWAR.1999.803801

Pryor, H. L., Furness, T. A., & Viirre, E. (1998). The Virtual Retinal Display: A new Display Technology using Scanned Laser Light. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(22), 1570–1574. https://doi.org/10.1177/154193129804202208

Raskar, R., Welch, G., Low, K.-L., & Bandyopadhyay, D. (2001). Shader Lamps: Animating Real Objects With Image-Based Illumination. In P. D. S. J. Gortler & P. D. K. Myszkowski (Eds.), *Rendering Techniques 2001* (pp. 89–102). Springer Vienna. Retrieved from http://link.springer.com/chapter/10.1007/978-3-7091-6242-2_9

Regenbrecht, H., Baratoff, G., & Wilke, W. (2005). Augmented reality projects in the automotive and aerospace industries. *IEEE Computer Graphics and Applications*, *25*(6), 48–56. https://doi.org/10.1109/MCG.2005.124

Regenbrecht, H., Collins, J., & Hoermann, S. (2013). A Leap-supported, Hybrid AR Interface Approach. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration* (pp. 281–284). New York, NY, USA: ACM. https://doi.org/10.1145/2541016.2541053

Regenbrecht, H., Franz, E. A., McGregor, G., Dixon, B. G., & Hoermann, S. (2011). Beyond the looking glass: Fooling the brain with the augmented mirror box. *Presence: Teleoperators & Virtual Environments*, *20*(6), 559–576.

Regenbrecht, H., Hoermann, S., McGregor, G., Dixon, B., Franz, E., Ott, C., … Hoermann, J. (2012). Visual manipulations for motor rehabilitation. *Computers & Graphics*, *36*(7), 819–834. https://doi.org/10.1016/j.cag.2012.04.012

Rolland, J. P., Biocca, F., Hamza-Lup, F., Ha, Y., & Martins, R. (2005). Development of Head-Mounted Projection Displays for Distributed, Collaborative, Augmented Reality Applications. *Presence: Teleoperators and Virtual Environments*, *14*(5), 528–549. https://doi.org/10.1162/105474605774918741

Satoh, N., & Hirai, K. (2001, August 14). Vibration motor holding apparatus and portable electronic equipment having the same. Retrieved from http://www.google.com/patents/US6274955

Shah, M. M., Arshad, H., & Sulaiman, R. (2012). Occlusion in augmented reality. In *2012 8th International Conference on Information Science and Digital Content Technology* (ICIDT2012) (Vol. 2, pp. 372–378).

Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, *29*(2), 58–69. https://doi.org/10.1016/j.displa.2007.09.005

Thomas, B., Demczuk, V., Piekarski, W., Hepworth, D., & Gunther, B. (1998). A wearable computer system with augmented reality to support terrestrial navigation. In *Digest of Papers. Second International Symposium on Wearable* Computers *(Cat. No.98EX215)* (pp. 168–171). https://doi.org/10.1109/ISWC.1998.729549

Tuceryan, M., Genc, Y., & Navab, N. (2002). Single-Point Active Alignment Method (SPAAM) for Optical See-Through HMD Calibration for Augmented Reality. *Presence:*

Teleoperators & Virtual Environments, *11*(3), 259–276. https://doi.org/10.1162/105474602317473213

Urey, H., Chellappan, K. V., Erden, E., & Surman, P. (2011). State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE*, *99*(4), 540–555.

Van Krevelen, D. W. F., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, *9*(2), 1.

Wloka, M. M., & Anderson, B. G. (1995). Resolving Occlusion in Augmented Reality. In *Proceedings of the 1995 Symposium on Interactive 3D Graphics* (pp. 5–12). New York, NY, USA: ACM. https://doi.org/10.1145/199404.199405

Yamazaki, S., Inoguchi, K., Saito, Y., Morishima, H., & Taniguchi, N. (1999). Thin wide-field-of-view HMD with free-form-surface prism and applications. Proc. SPIE 3639, Stereoscopic Displays and Virtual Reality Systems VI, 453 (May 24, 1999); doi:10.1117/12.349411.

Yudkowsky, R., Luciano, C., Banerjee, P., Schwartz, A., Alaraj, A., Lemole, G. M., … Frim, D. (2013). Practice on an Augmented Reality/Haptic Simulator and Library of Virtual Brains Improves Residents' Ability to Perform a Ventriculostomy: *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *8*(1), 25–31. https://doi.org/10.1097/SIH.0b013e3182662c69

Zhang, C., & Zhang, Z. (2014). Calibration Between Depth and Color Sensors for Commodity Depth Cameras. In L. Shao, J. Han, P. Kohli, & Z. Zhang (Eds.), *Computer Vision and Machine Learning with RGB-D Sensors* (pp. 47–64). Springer International Publishing. https://doi.org/10.1007/978-3-319-08651-4_3

10 Figure Captions

Figure 1. The EOIL table provides a visual representation of the EOIL Framework. The left most column outlines first, the abstract components, and then their counterparts from our prototype.

Figure 2. Two different types of see-through HMD: Video see-through (left) and Optical see-through (right).

Figure 3. The virtual camera (right) is positioned according to the real HMD camera (left)

Figure 4. The system from the user's perspective. Demonstrates a combination of virtual and real components (condition 4 from the EOIL table - real desk, virtual lamp, real hand, real lighting)

Figure 5. Shows the scenario space with the HMD setup in a static position. This is the perspective the VST content is taken from.

Figure 6. OST HMD positioned further from the scene so it is able to virtualize the entire scene (left), and then the HMD positioned close enough to the scene so the interaction component is able to be visualized (right).

Figure 7. An example of unhandled occlusion with a virtual lamp placed in an otherwise entirely real scene. The user would expect to see their hand above (or "in front of") the lamp.

Figure 8. If the 3D model of the real object is not of accurate form, it will become visually apparent to the user, breaking the visual continuity of the scene.

Figure 9. A) Virtual cube occluding real lamp, B) Real lamp occluding virtual cube, C) and D) An arbitrary view of the scene demonstrates the effect of the phantom model shader in both cases (A and B respectively)

Figure 10. The scene before the lamp is switched on (left), and after the light is on (right). This is condition one of the EOIL table (all components are real)

Figure 11. Demonstrating our approach to applying virtual lighting to real world components

Figure 12. TOF camera images demonstrating a clear hand contour when the hand is in open space (left) but once the hand is close to another surface, the hand becomes much less distinguishable (right)

Figure 13. VR immersion device (Oculus Rift DK2) modified for use as a stereoscopic video see-through HMD.

Table 1. Problem Categorization Table. The upper segment of the table provides an overview of software related issues and solutions, while the lower segment is focused on hardware related challenges.

Table 2. The Research & Solutions table provides weighting values providing a summary of the availability of research with respect to specific categorical problems.

Table 3. The Priorities for Research & Development table provides values respective of how important it is for each categorical problem to be solved.

EOIL Framework and Implementation																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Environment / Desk	R	R	R	R	R	R	R	R	V	V	V	V	V	V	V	V
Objects / Lamp	R	R	R	V	R	V	V	V	R	R	R	V	R	V	V	V
Interaction / Hand	R	R	V	R	V	V	R	V	R	V	R	R	V	R	V	V
Lighting / Spot Light	R	V	R	R	V	R	V	V	R	R	V	R	V	V	R	V

Reality — Augmented Reality — Augmented Virtuality — Virtual Reality

Figure 1. The EOIL table provides a visual representation of the EOIL Framework. The left most column first outlines the abstract components and then their counterparts from our prototype.

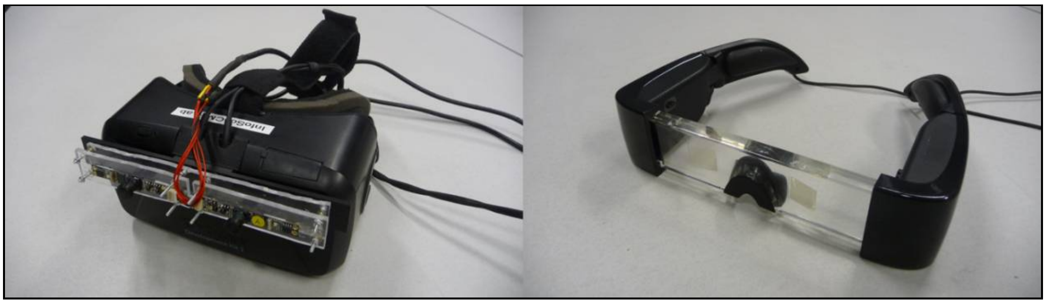

Figure 2. Two different types of see-through HMD: Video see-through (left) and Optical see-through (right).

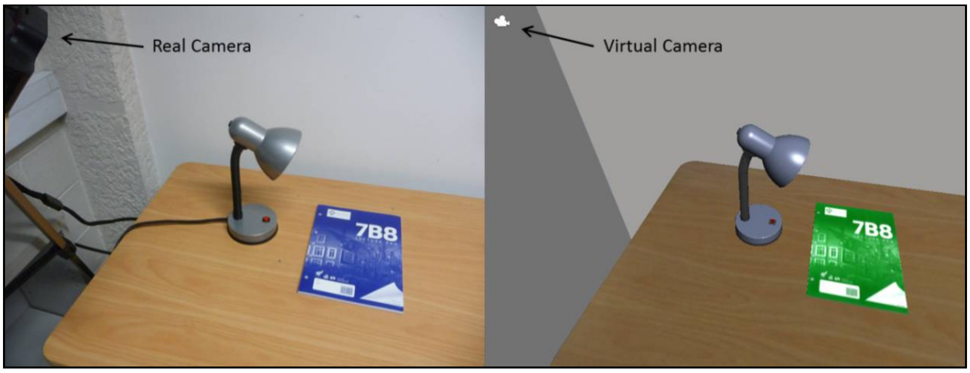

Figure 3. The virtual camera (right) is positioned according to the real HMD camera (left).

Figure 4. The system from the user's perspective. Demonstrates a combination of virtual and real components (condition 4 from the EOIL table - real desk, virtual lamp, real hand, real lighting).

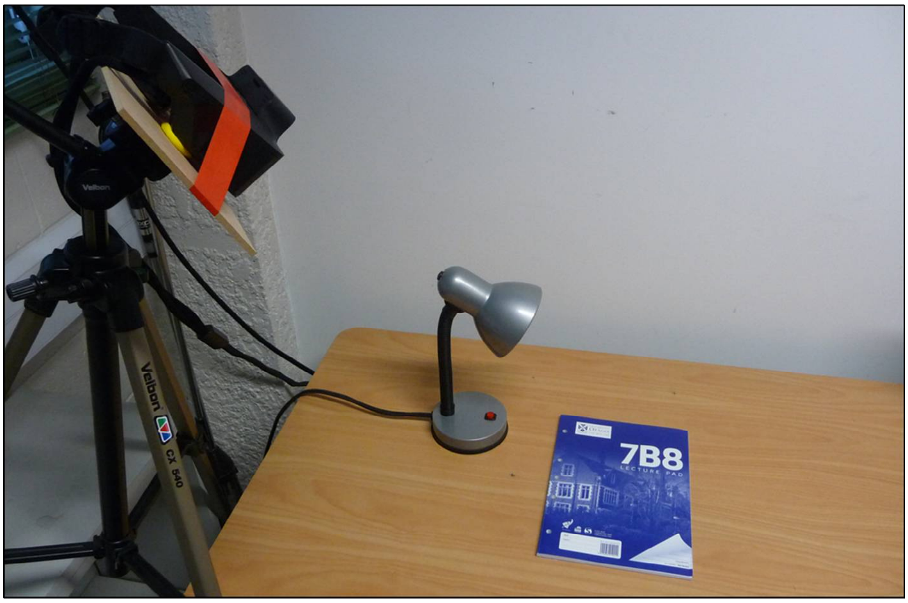

Figure 5. Shows the scenario space with the HMD setup in a static position. This is the perspective the VST content is taken from.

Figure 6. OST HMD positioned further from the scene so it is able to virtualize the entire scene (left), and then the HMD positioned close enough to the scene so the interaction component is able to be visualized (right).

Figure 7. An example of unhandled occlusion with a virtual lamp placed in an otherwise entirely real scene. The user would expect to see their hand above (or "in front of") the lamp.

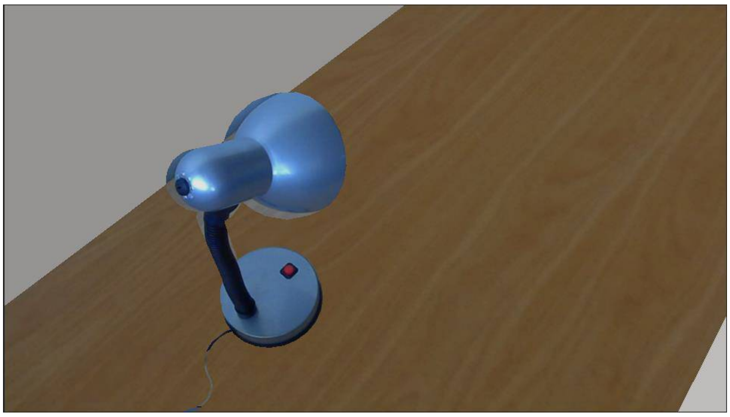

Figure 8. If the 3D model of the real object is not of accurate form, it will become visually apparent to the user, breaking the visual continuity of the scene.

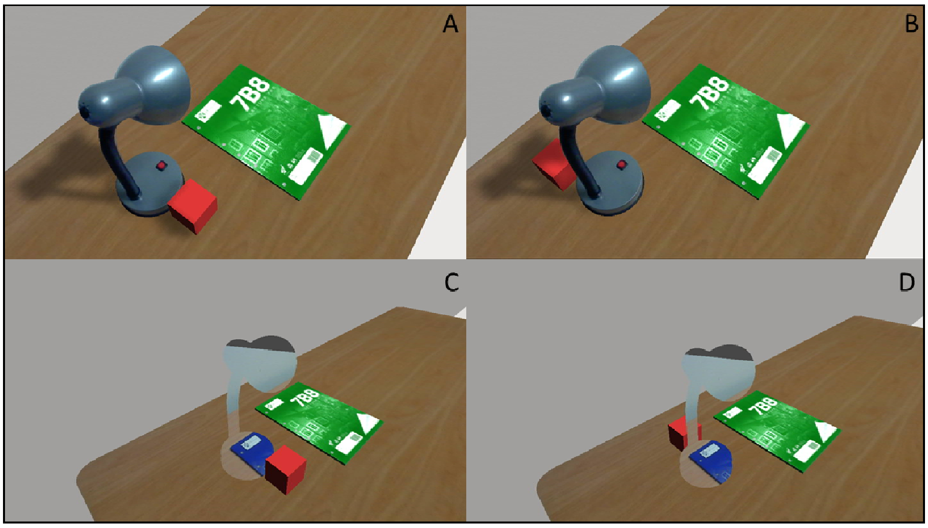

Figure 9. A) Virtual cube occluding real lamp, B) Real lamp occluding virtual cube, C) and D) An arbitrary view of the scene demonstrates the effect of the phantom model shader in both cases (A and B respectively).

Figure 10. The scene before the lamp is switched on (left), and after the light is on (right). This is condition one of the EOIL table (all components are real).

Figure 11. Demonstrating our approach to applying virtual lighting to real world components.

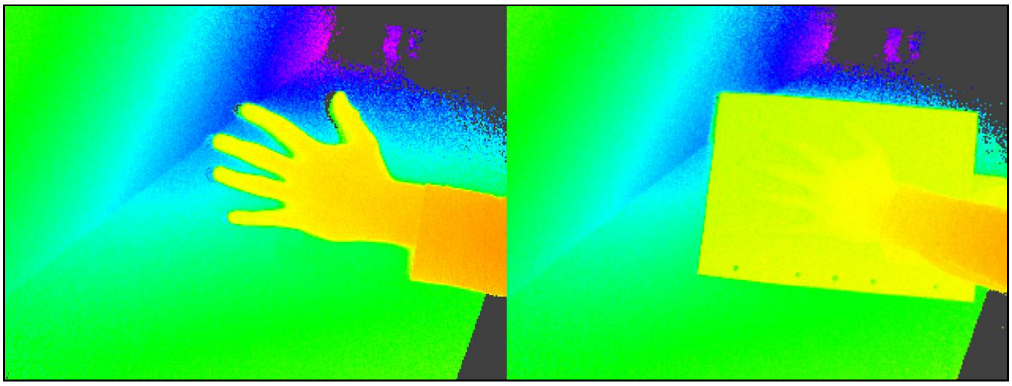

Figure 12. TOF camera images demonstrating a clear hand contour when the hand is in open space (left) but once the hand is close to another surface, the hand becomes much less distinguishable (right).

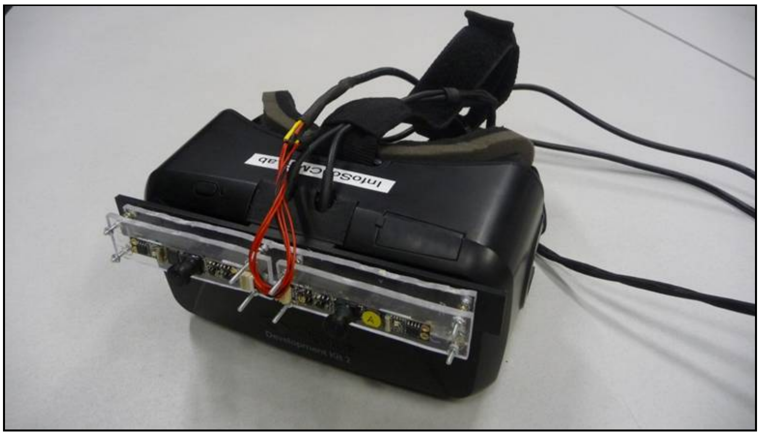

Figure 13. VR immersion device (Oculus Rift DK2) modified for use as a stereoscopic video see-through HMD.