Computers & Graphics 35 (2011) 810–822

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Mobile Augmented Reality

Indirect augmented reality

Jason Wither *, Yun-Ta Tsai, Ronald Azuma

Nokia Research Center-Hollywood, United States

ARTICLE INFO

Article history:
Received 11 February 2011
Received in revised form
15 April 2011
Accepted 28 April 2011
Available online 11 May 2011

Keywords:
Mixed Reality
Information Presentation
User Interface
Evaluation

ABSTRACT

Developing augmented reality (AR) applications for mobile devices and outdoor environments has historically required a number of technical trade-offs related to tracking. One approach is to rely on computer vision which provides very accurate tracking, but can be brittle, and limits the generality of the application. Another approach is to rely on sensor-based tracking which enables widespread use, but at the cost of generally poor tracking performance. In this paper we present and evaluate a new approach, which we call Indirect AR, that enables perfect alignment of virtual content in a much greater number of application scenarios.

To achieve this improved performance we replace the live camera view used in video see through AR with a previously captured panoramic image. By doing this we improve the perceived quality of the tracking while still maintaining a similar overall experience. There are some limitations of this technique, however, related to the use of panoramas. We evaluate these boundaries conditions on both a performance and experiential basis through two user studies. The result of these studies indicates that users preferred Indirect AR over traditional AR in most conditions, and when conditions do degrade to the point the experience changes, Indirect AR can still be a very useful tool in many outdoor application scenarios.

1. Introduction

Outdoor AR has become very popular in a number of application spaces in recent years with the growth in popularity of smartphones and other portable hand-held devices. There are several commercially available AR browsers that display both point-of-interest (POI) information and, increasingly, basic 3D content. There are also an increasing number of AR games available for these platforms. Academically there has also been an increase in the number of experience and game focused projects that make use of outdoor AR in some way. However, many of these projects suffer from poor registration because they rely primarily on built-in sensors (GPS, compass, and sometimes gyroscopes) for tracking. These sensors, especially those used in commodity products, do not have nearly the accuracy required for convincing tracking in AR. This limits both the scope of the applications that are possible and decreases the quality of the user experience for applications that do exist. On the plus side, smartphone platforms allow AR applications to have a much broader audience. It is possible to have well registered AR on a smartphone [20], but this generally requires vision based tracking which frequently relies on previously known textures, and often can be brittle in unknown outdoor environments.

The goal of this paper is to enable experiences in these outdoor unprepared environments that are traditionally nearly impossible to have compelling AR experiences in. The approach we are taking for this is not to perfect computer vision based tracking, but instead to minimize the visual disturbances in the experience when using a given, existing, tracking approach. We have named our technique to do this Indirect AR because we are making the entire scene inside of the device virtual, using panoramic images instead of a live camera view (see example in Fig. 1). The user is no longer looking directly at the scene through a live camera view, as in video see through AR, but is instead looking at the scene indirectly, by looking at a previously captured image of it. This moves visible registration error from being essentially inside of the device (between the real world and virtual content), to being on the border of the device itself, between the view on the device screen, and the real world around the device. This means that within the device, between the virtual content and the panoramic representation of the real world, there is no registration error. In Fig. 2 a somewhat abstracted example is shown illustrating the difference in how both AR and Indirect AR look with the same amount of registration error. Moving registration error to the edge of the screen is better in part because it is a more difficult place to detect error due to both the bezel around the screen, and the altered field-of-view parameters of the on-screen

* Corresponding author. Tel.: +1 310 266 8903.
E-mail addresses: jwither@cs.ucsb.edu, jason.wither@nokia.com (J. Wither), yun-ta.tsai@nokia.com (Y.-T. Tsai), ronald.azuma@nokia.com (R. Azuma).

0097-8493/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2011.04.010

image. In many ways, people are also already trained to believe that when they see a view of the real world on the screen of the mobile device it lines up with the world behind it. This is largely due to the proliferation of digital cameras already using the screen as a viewfinder. The viewfinder image people are used to seeing has already been manipulated in many ways though. It does not show the same view as a piece of glass in its place would. Instead the view is modified by the lens system to have a different field of view, depth of focus, etc. Even with these changes though people understand they are looking "through" the camera. We are taking this one step further by adding tracking error, essentially as another image modifier. Using built-in sensors, the portion of the panorama that is shown might be 5° or 10° off from what is directly behind the device, but people are not as likely to notice this because of all the other image modifiers.

While we focus on using panoramas as the representation of the real world in this paper, as the complexity and detail of the virtual representation of the real world increases, Indirect AR could become an even more powerful approach. The extreme conclusion of this increase in complexity might be the case of a real environment filled with a large array of video cameras and other sensors that would capture, in real time, the real environment and permit a perfect reconstruction of that environment, in real time, as seen from any arbitrary viewpoint. If this were possible, then an Indirect AR experience would become virtually indistinguishable from an ideal traditional AR experience. This vision may not be practical today; however, it is useful to keep in mind this ultimate expression of Indirect AR, much as Sutherlands Ultimate Display provided a vision of the ultimate expression of Virtual Reality.

In this paper, we focus on an implementation of Indirect AR that uses a more practical model of the real world that still delivers a very high quality, albeit temporally static, experience: panoramic images. Several companies, including Navteq, are collecting detailed models of urban environments by driving special vehicles that collect panoramic imagery and other data such as 3D point clouds. These are captured every few meters as the car drives down a road. Our approach assumes that a user stands at one location outdoors, downloads the nearest panoramic image, and then rotates around in place to examine the real world and the added augmentations. While this implementation is constrained compared to the ultimate vision, it still provides a very similar experience to traditional AR, and can potentially scale well today due to the wide availability of panoramic imagery. Since such imagery and related data are rapidly becoming available for most major urban areas across the world, Indirect AR could become a practical method for generating high quality AR experiences in urban environments. Using panoramas as a representation of the real world is somewhat limited in that panoramas are captured infrequently, meaning dynamic elements of the scene, such as traffic, weather, and lighting may not be represented correctly. As we will see throughout the rest of this paper, though, a high quality experience can still be had even if some of these limitations are not perfectly met.

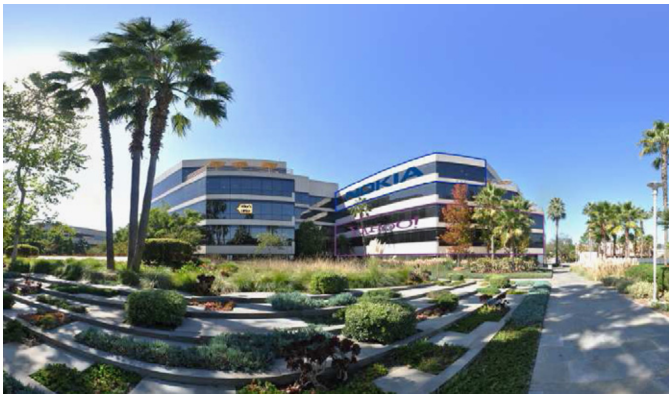


Fig. 1. A portion of an annotated panorama that could be used for indirect AR. This particular panorama was used in the study described in Section 6.

In Section 3 we will first define our research questions in more detail. We will then discuss some of the pure performance comparisons between AR and Indirect AR in Section 4, showing how orientation error can have a huge negative effect on traditional AR, while having a much more minor effect on Indirect AR. Next, in Section 5, we will examine a component of the common scenario in Indirect AR when the panorama is not colocated with the user. We will examine how well users can align their view of the real world with the view presented on the device by studying user performance at pointing out real world objects that are highlighted in an image taken from a different location. We will show that even in difficult conditions users are very good at correlating their real world view with the alternate view displayed on the device screen. This does not help determine how similar the actual experience is to traditional AR though. We will explore that question with a second more in depth study, presented in Section 6, examining user perception of the Indirect AR experience. This study looks at many of the boundary cases of Indirect AR, and how these boundary cases change the type of experience for the user. It also examines the user experience in "good" conditions comparing Indirect AR to traditional AR with several different types of content. In these good conditions we found Indirect AR was superior to traditional AR regardless of the type of content, and gave users the same type of experience. When the panorama was not colocated with the user the Indirect AR experience eventually degraded, but this was not instantaneous; there was a region around the user where the same type of experience was preserved.

2. Related work

Our work builds on many magic lens techniques in both VR and AR. Magic lenses were first used in AR as part of HMD-based systems and affected the way virtual content was viewed [10].

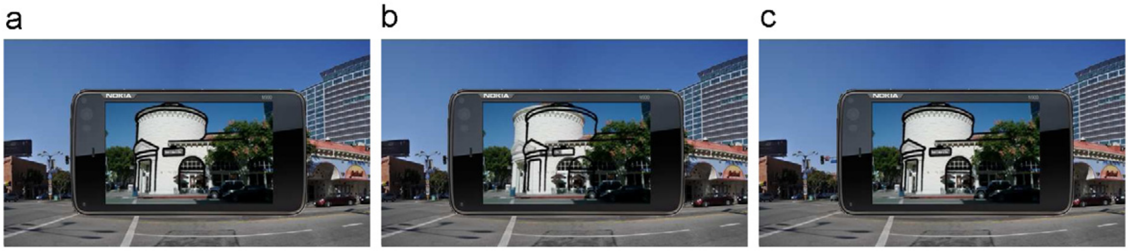


Fig. 2. Image (a) shows a mocked up representation of a perfectly aligned AR scene. The outline of the building (the virtual component) is directly lined up with the physical building. The image on the screen is also lined up with the world behind it. In (b) the traditional AR tracking problem is illustrated. In this case the building outline has been moved 20 pixels to the right. In (c), which represents Indirect AR, the entire onscreen image (both building and outline) are moved twenty pixels to the right (compared to (a)). As can be seen, the result of this is much less visually jarring than that in (b).

Some examples of magic lens use closer to our use are Brown and Hua's [2] work in VR, and Quarles et al.'s [13,14] work in AR. Both of these techniques allow the user to see a virtual version of the world (real or virtual) around them on their hand-held screen that is roughly registered to the world. Quarles et al.'s work is more similar to ours; however, they show an animated graphical representation of the machine that is being worked on rather than a complete annotated panoramic image of the world around them.

Liestoel et al. [7,8] have developed a technique they call Situated Simulations which is quite similar to Indirect AR. Their approach uses a hand built virtual world that is associated with a real place. As the user moves about in the real world, their avatar is moved through the virtual world displayed on a smartphone. Unlike Indirect AR though the virtual world they use is not as high fidelity of a representation of the real world as panoramas provide. The ultimate expression of Situated Simulations and Indirect AR are likely quite similar though. Ragen et al. [15] have also looked at an interesting new place on the MR spectrum. Rather than trying to mimic AR with a primarily virtual interface as we are, they simulate AR in VR in order to more easily test AR systems. Uyttendaele et al. [18] have used dense panoramas to explore a space, but their use case is for a remote viewer to see the panoramas, rather than an on-site viewer. Hill et al. [5] have used panoramas in a very similar way to what we are doing as part of their larger AR browser KHARMA. This work is also inspired by our previous work in The Westwood Experience [22] where we used an illustrated panorama and similar interface to connect the physical environment to the fictional one. It is also important to note that while we use pre-captured panoramas in this work, Wagner et al. [19] have built a system for the capture of panoramas in real time. These panoramas could also be used in place of previously captured ones.

An important part of this work is related to the question of presence in AR. In our case the broad question we are trying to answer is if the Indirect AR experience enables the same connection between the virtual content and physical world that AR does. There have been several people [4,12,21] previously looking at similar questions in AR and how things like presence and immersions translate from VR to AR. We feel though, at least for annotation focused AR, presence is not quite the most important question, and is superseded by the questions of believability and connectedness regarding the AR scene.

3. Using panoramas as an AR substitute

One thing we feel is important for a compelling AR application is for the virtual content to be convincingly placed in the AR scene. By this we mean that the technology for how the virtual and physical components are combined should be as hidden as possible. The user should not have to make additional assumptions about what the AR scene "should" look like if something worked better; the AR scene should be convincing as it is. There are many ways to do this, from having pixel accurate tracking, to having virtual content designed in a way to hide any existing tracking error, like having virtual characters ride in a balloon [11]. We hope that Indirect AR enables convincing AR-like-experiences more generally with sensor-based tracking, by lessening the impact of tracking error, and enabling enhanced content blending. Because the panorama is pre-captured, possibly along with depth information it is possible to greatly enhance the blending of the virtual and physical, inserting virtual content with matching lighting, correct occlusions, etc.

Throughout this paper we focus on the comparison between Indirect AR and traditional AR, but it is also interesting to think about what lies beyond Indirect AR. We have focused on interacting with the world by holding up the device and looking through it, as you would for traditional AR. Because the panoramic imagery exists independently of the real world though, it is also possible to interact with the panoramas directly using touch input. This interface would be essentially like the mobile version of Google's Street View application. The question then is: what is gained by holding the phone up and looking through it in the traditional AR manner instead of interacting with the panoramas using touch for panning? We feel that the difference between these two methods of interacting with the panoramas is in the type of experience that is generated. By holding up the device and having an AR-like interaction we think there is likely to be a much higher level of immersion in the AR scene, and along with that a tighter link between the virtual content and the real world. This does not imply though that a panning style interface does not have any use. For many search style applications where the user is trying to find simple information like directions or points of interest, a higher level of immersion is likely not as important. However, in many other application domains, particularly entertainment and games, having a close tie between the virtual and physical parts of the environment is very important because the experience itself is focused on building a rich Augmented Reality environment.

4. Registration

One of the biggest benefits of using Indirect AR in place of traditional AR is the greater accuracy in registration. In traditional AR any error in registration is visible directly between the physical object and virtual annotation. In Indirect AR the same registration error is only visible between the device and surroundings. That means that the registration between virtual annotations and the panorama that represents the real world is *always perfect*, even if the registration between the device and the real world is not. Because both the panorama and virtual content are moved together there is never any offset between them. This is a very important difference, since any registration error, including jitter and lag can have a huge negative impact on the overall experience. Those problems will not be present in Indirect AR.

It is illustrative to quantify exactly how much orientation tracking error influences the AR experience. Because annotated objects in outdoor AR are generally quite far away, even small amounts of orientation error can result in large amounts of registration error. For instance, at 20 m, only 5° of orientation error will generate 1.75 m of registration error on a perpendicular plane. Ten degrees, a not uncommon amount of error with standard sensors, will generate 3.53 m of registration error, while even a much better system with only 0.5° of error will still generate 0.17 m of registration error. All of this error can have a detrimental effect on the overall AR experience. Livingston and Ai [9] previously conducted a more in depth look at registration error, showing how different types of registration error affect user tracking of distant objects. While they found that latency, noise, and overall orientation error all negatively impacted user performance, latency seemed to have the greatest impact.

There are naturally many ways to try to ameliorate the problem of inaccurate registration in AR through thoughtful interface design, but there is in the end a certain amount of error that is present when orientation tracking is not accurate. From these numbers though it is clear that even a small amount of rotational error can have highly detrimental consequences for AR applications. From a pure functionality point of view these errors do not occur in Indirect AR because the virtual and representation of the physical move together. This also reduces the visual impact of other types of tracking error including jitter and lag.

5. Localization from disparate viewpoints

One of the biggest drawbacks of our current implementation of Indirect AR is the reliance on pre-captured panoramas. For the naïve, ideal experience it would be necessary to have a panorama located exactly where the user was standing. In the case of deploying Indirect AR generally this would mean having panoramas everywhere. While we do not have panoramas everywhere, we can approximate this by using panoramas collected by Navteq as they drive down most streets. There are two possible problems with this though: would users be able to look at the panorama and find the nearby point of interest in their own view of the real world, and if so would using those panoramas be the same type of experience as traditional AR, or would the difference in location change the experience? This section looks to answer the first of those questions, while Section 6 explores the second.

To study the first question we conducted a small study to look at how well people can localize points of interest between two disparate views, which in this case would be the view on-screen, and the user's direct view of the environment.

5.1. Study description

The goal of this study was simply to determine if users could even determine how panoramas corresponded to the real world if they were taken from a different viewpoint. This is a similar question to that asked in Iachini and Logie's previous work [6]. They were determining if a person could find their location on a map when taken to a novel viewing point of a building. Our question differs because we do not necessarily care if users have built a mental map around their location; we are interested in simply seeing if they can match their real world view to the view they are shown on-screen. To this end the questions we hoped to answer with this study were:

- How successful are people at locating buildings when shown them from a different perspective?
- How does the location the panorama was taken from (relative to the user) affect this?
- How does the angle the picture was taken at (relative to the building of interest) affect this?
- How does the perspective difference between the image view and the user view affect this?

The second through fourth questions above are essentially sub-questions directly looking at the factors we think could negatively affect a user's ability to correlate views. Our hypotheses are that people will do well at recognizing buildings, but will have a more difficult time as conditions in these three sub-factors become worse.

To conduct this study we actually simplified the question even further and did not use tracked panoramic imagery, instead just using static images of the target displayed on the screen of a Nokia N900. There are advantages and disadvantages of this simplification, but we think there were sufficient disadvantages in using a tracked panorama to justify the simplification. Some of the advantages of using a tracked panorama that were not present in our study are that when using a panorama more of the scene could be viewable since users could turn to examine more of the panorama. Similarly, once users locate the target building in the panorama they know the direction from the panorama location to the building, although they still do not know their physical relationship to either the panorama location or the target building. Knowledge of the panorama to building target orientation might also be a disadvantage and confuse the study though. If the panorama location is quite distant from the user, knowing the orientation to the target might prompt the user to look in the same direction, which could be incorrect. An even larger problem (in terms of conducting a well controlled study) is that the target building might not always be visible on the panorama initially. Because the orientation of the panorama is tied to user orientation, they would first have to turn to find the target before trying to find the same target in the real world. This extra task is not what we were interested in exploring in the study, and so would likely confuse results.

What we did instead was simply to show users images of storefronts, like those seen in Fig. 3 on a pedestrian only street, from different perspectives with the store of interest highlighted. The user then had to look around in the physical world, find the same building (which could be seen from their location), and physically point it out to the study administrator. A timer was started when the user was first shown the view of the next target image and stopped when the user pointed out the building. The administrator also kept track of incorrect guesses.

More specifically, in the study users stopped at five viewpoints along the 3rd Street Promenade, a pedestrian only shopping street in Santa Monica. At each viewpoint users stood on the southwest side of the street, approximately three meters from the building façade. Users were then shown images from five nearby locations, all of which were in a line along the center of the street. These image locations were spaced approximately 15 m apart, and were picked to simulate a sparse set of panoramas that might be collected by Navteq or Google. At each viewpoint users were shown a total of nine images taken from the five image locations. Three of these images were looking directly at the side of the street, perpendicular to the building facade, three were taken at a medium angle (around 50° from perpendicular), and three at a high angle (around 75° from perpendicular). Exemplars of each of these image types can be seen in Fig. 3. The order of the views the user was shown for each viewpoint was counter balanced using a Latin square technique, and included three pictures each from

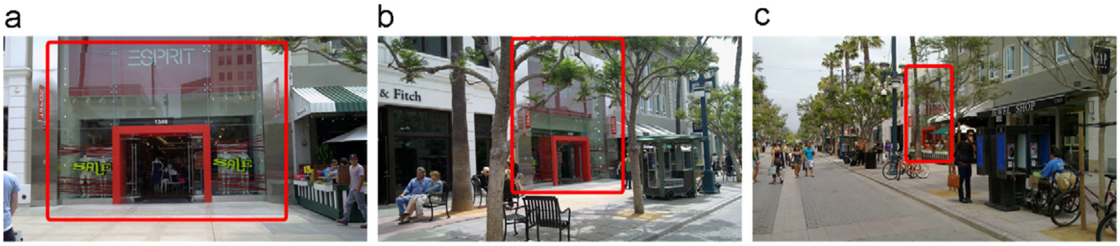


Fig. 3. Three example images from the study. Users would be shown these images on the N900, and asked to find the annotated building in their physical surroundings. Straight, medium, and high viewing angles of the same target (used for illustrative reasons) are shown in (a), (b), and (c), respectively.

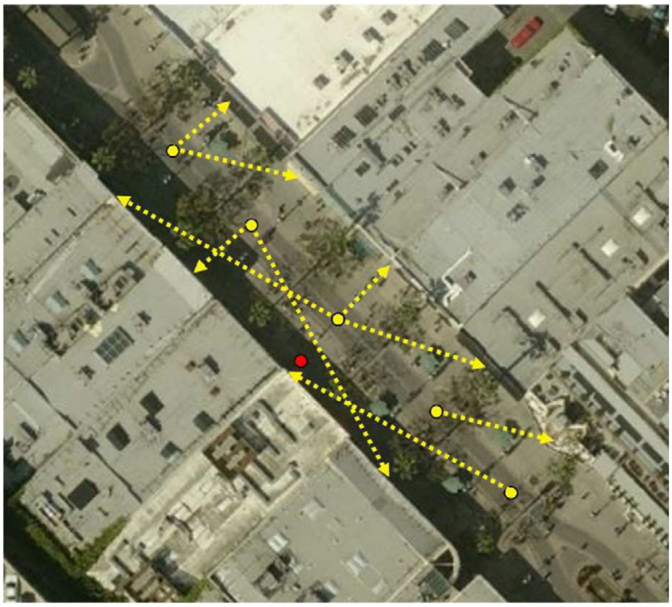


Fig. 4. A aerial view of one of the locations used in the study. The user stood at the location indicated by the red dot. The five yellow dots in the center of the street are locations images were captured from, and the yellow arrows point from the capture location to the target location. Target buildings were on both sides of the street, and the choice of image location/target was randomized amongst the five locations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

locations near the user, and at medium and far distances. Target buildings were not reused, and were also balanced to have the same number of targets on both sides of the street. An example of the image and target building locations can be seen in Fig. 4. Before the study proper began there was a brief explanation and training session with an example image that was not related to the rest of the study locations. Locations were presented in the same order for every user, but we saw no learning effects because of this, likely because the task was quite simple. Nearly all results were tabulated across location as well, so the order of location was not important. Results of the study were evaluated between subjects.

We had nine users participate in the study, all of whom were Nokia Research Center employees, and all of whom were male and between the ages of 25 and 48. The study took approximately 30 min per user, and in that time the user completed 45 matching tasks, nine each at the five locations. Some users were somewhat familiar with the environment, but casual familiarity seemed to have little impact on user performance. There were very few times when users knew the location of a certain store from previously being in the environment. User viewpoint locations were spaced far enough apart that there was no overlap between locations, except locations 1 and 2 which had a very small amount of overlap.

5.2. Results and discussion

Overall, we found people were very good at locating the indicated buildings around them. In Fig. 5 a summary of these results can be seen, for the nine possible location/orientation image combinations. While in some difficult scenarios users were much slower at locating the building shown in the image, there were very few errors. In total we had a 4% error rate, and half of those errors were from the most difficult condition. Task completion took users between 1.7 and 78.2 s; times which were influenced by individual differences in strategy. There were also learning effects that occurred per location as users became more familiar with their surroundings in a particular spot. Because of the counter balancing this should not have affected overall average scores, but did affect the variance for each score.

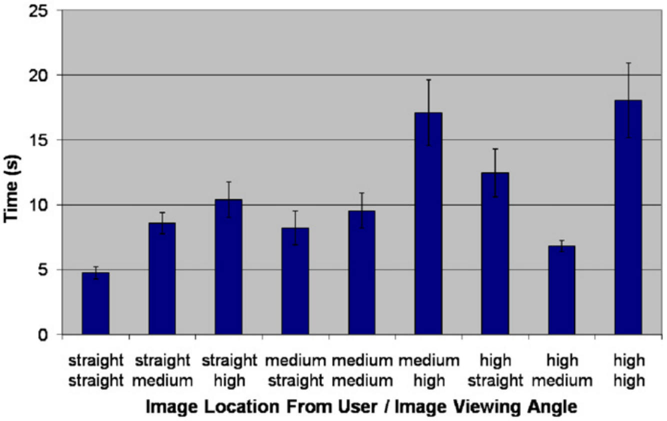


Fig. 5. Results showing how long it took users to find the indicated building in different conditions. The label for each column indicates how far the image location was from the user's location, and the angle between the target and image location.

Two of the primary questions we had for this study were: does the location of the image matter for finding the target, and does the orientation from the image location to the target matter? We found both of these things did matter, in a two way ANOVA, location was significant ($p<0.01$) as was the orientation of the image ($p \ll 0.01$). There was also a significant interaction between the location and orientation ($p \ll 0.01$).

To look at the effect the location the image was taken from had on user response we chose to do further analysis on a subset of images from each location. We chose all images from each location with a straight viewing direction, providing a full frontal view of the target building. Looking just at this subset of images we found significant differences (ANOVA $p \ll 0.01$) between locations, with a fairly linear progression of time to completion based on the distance from the user to the image location. Images taken from the location straight in front of the user took on average 4.7 s for the user to recognize and locate, images from a medium distance took 8.2 s, and images from a far distance took 12.2 s. This result is not particularly surprising because storefronts further from the user are not only more distant, but often harder to see because of both minor occlusions and the more acute viewing angle. The appearance of the building will also be different because the user's direct view will be the most significantly different from the view presented on-screen.

We were also interested in how the angle between the location the image was taken from and the target building would impact results (independent of image location). We evaluated our three sets of image angles, those directed straight at the building, at a medium angle, and a high angle, and again found significant differences (ANOVA $p \ll 0.01$). In this case the straight and medium angled pictures were very similar in time to completion, while the high angle pictures took users nearly twice as long to localize (15 s on average vs. 8.4 for both straight and medium). In addition, 14 of the 18 errors that users made in identification also came from high angle images. The fact that the high angle images did so poorly was not a particularly surprising result. Many of those images were exceedingly difficult, because of distance and angle, as well as foreground occlusions. What was somewhat surprising, and reassuring for Indirect AR, was that medium angled pictures did as well as pictures with a completely straight view. We think this is likely because in both of those views people had a fairly clear view of the target building. That clear view made it much easier to notice all the relevant details to look for in their physical surroundings.

Variance between the angle to the target from the user and from the image capture location also seemed to be an important factor. Nine of the 18 errors in the study were from the set of pictures taken with nearly opposite image and user views. This was also a significant percentage (20%) of the total number of images shown with those parameters. These errors were largely due to people using landmarks near the target as an aid, but because the views were so different these landmarks were unreliable.

What we can conclude from this study is that in general people are quite good at matching an image on a screen to their surroundings. In the easiest case (straight straight) when the target was directly in front of the user it took users on average 4.7 s to recognize the target. For most users the majority of this time was spent on the mechanics of the study (looking at the picture on the phone, confirming the choice, etc.) rather than on actively searching for the target. If we subtract this time then from the other conditions to get a rough idea of how long people were actively searching we can see in most cases that people needed to search for less than 5 s to find the desired target. In addition to this, even in cases where the majority of the target of interest was occluded by other foreground objects, users were generally able to locate the same object correctly even if it was a more time consuming search. This study provides several guidelines for building an Indirect AR system as well when using street captured panoramas. It seems that picking a panorama close to the user should be first priority, followed by picking one with a good view of the object of interest. These two constraints combined would result in picking a panorama between the user and target (if available) to provide the best user experience. Clearly though, not all of the scenarios examined in this study would result in an AR like experience with Indirect AR. When users were taking more than 15 s to even determine what was being annotated, the overall experience would be quite different than looking directly through a camera as you do in AR. In the next section we will look at this question further to try to determine if using panoramas from different locations can provide the same type of experience in addition to being a useful tool.

6. User experience vs. AR

Now that we have established both that the link between augmentations and a representation of the real world is tighter in Indirect AR than in traditional AR, and panoramas from a variety of locations can at least be recognized by a user, we can ask questions about how the actual experience compares to AR. We are particularly interested in seeing if two related experiential components of traditional AR are also present in Indirect AR. In well done traditional AR there is a very strong visual tie between the virtual content and the physical world. This is a general goal of AR in fact, to have the virtual content look like it is actually present in the real world. This is important both for purely functional reasons, and because it introduces a higher level of immersion in the Augmented Reality world. In many cases the same information could be presented in another format, but it would not be the same type of experience. Part of our goals for this study were to determine if Indirect AR can also provide the same level of immersion in the AR environment that traditional AR does. This type of immersion is much more important for some applications than it is for others. In a AR browser style application immersion is not as important because users are primarily interested in task completion, not exploring the AR environment. In many entertainment or game applications though the experience is primarily about exploring the environment around the user, making immersion in that environment very important.

In this study the primary question we were interested in was: does the Indirect AR experience feel the same (offer the same level of immersion, etc.) as a traditional AR experience? We looked at this question in a number of conditions, first comparing the two under good conditions, and then introducing things that we thought were likely to adversely affect Indirect AR to see if the experience was still similar. These boundary conditions included testing with a variety of styles of virtual content, making the physical environment dynamic, and changing the location where the panorama was taken from to not be at the user's location. There are other boundary conditions that we did not test in detail, including mismatches of lighting and seasons between the panorama and real world. A large summer/winter season mismatch could have detrimental effects on the overall experience, but we were unable to test that in Los Angeles. We did include differences in weather conditions (cloudy vs. sunny) and lighting in a pilot run of the study, but did not find significant differences in user response to these conditions. In order to expedite the final study we did not include varying weather conditions, although there were quite different lighting conditions throughout the day that were particularly noticeable due to shadowing at the end of the day. Even in this case though very few users noticed a difference between the panorama and real world.

The following key points summarize our study findings:

- In good conditions the vast majority of users will not notice a difference between Indirect AR and traditional AR without previous knowledge.
- In good conditions Indirect AR was *preferred* to traditional AR across all conditions tested.
- When the physical environment was modified to include dynamic elements users found that the Indirect AR experience degraded *less* than the comparable traditional AR experience.
- Having the panorama used in the Indirect AR experience captured from a different location than the user's viewing position does reduce immersion and change the experience. This change is only significant though when the two locations are sufficiently spread apart.

Through the rest of this section we will explain our study design, present our hypotheses and results, and discuss the importance of these results.

6.1. Study design

The primary questions we had when designing this study were experiential rather than task based. We had already established that Indirect AR could be successful for completing many AR tasks (as can be seen in Sections 4 and 5), so the primary questions we still needed to evaluate involved the type of experience. It seemed counter to our purposes to design another task based study since in many cases subjects who are focused on completing a task do not pay attention to the overall experience. Because of this our evaluation metrics for the study were questionnaire and interview based rather than strictly task performance based. To get qualitative data from users we had several points within the study where we would interview subjects about things they had just completed. This interview occurred after any related questionnaire responses, and was open ended in order to probe further on questions individual to each user. The quantitative data used in the evaluation below is from analyzing questionnaire responses between subjects.

When designing our questionnaire we looked at previous work in a number of fields for inspiration. In one sense the questions we are interested in relate to presence, as defined in the VR community. However, presence in this sense does not translate very well to AR since users are always present in the physical world. Because of this, VR presence questionnaires like those presented by Slater and Steed [16], or Witmer and Singer [23] do not translate particularly well. Gandy et al. [3] recently presented work trying to explore the question of presence in AR further. Our questionnaire was developed after close examination of all of these questionnaires, particularly Gandy's. However, we felt the environment we were working in, and questions we wanted answered, were different enough from Gandy's that we were not able to adopt their questionnaire directly. Their situation largely involved users interacting with a VR environment that was placed within the real world. What we were interested in on the other hand was the interaction between the real and virtual parts of the environment.

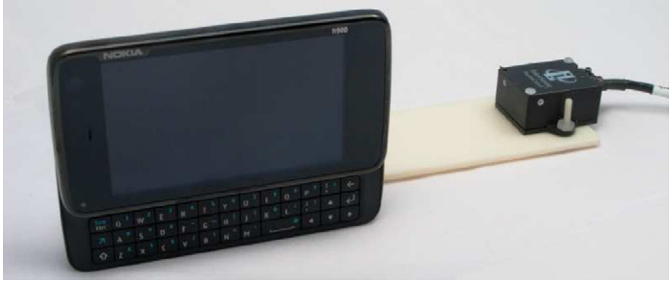


Fig. 6. The Nokia N900 used in our study with InterSense InertiaCube3 attached. The InteriaCube was attached at a distance to avoid magnetic disturbances from the device.

In constructing our study we first ran a pilot with nine users who worked at Nokia and were familiar with AR. The final study was run with 18 recruited users with ages 18–55. Fourteen of these users had not previously heard of AR, while the rest were somewhat familiar with the technology, primarily from AR browsers. The study took place in an outdoor environment in a large office complex. Users stood on a sidewalk and primarily focused on one building in the complex which was annotated with the same content in both AR and Indirect AR conditions. In both AR and Indirect AR conditions users used a Nokia N900 to view the Augmented Reality scene. Because we were interested in testing how both conditions performed in real world conditions, we did not use vision based tracking, since it can often fail in unprepared environments, instead falling back on sensor-based orientation tracking, motivated by the fact that nearly all smartphones now include orientation sensors. Because we wanted this study to be forward looking, we used a high quality sensor box, the InterSense InertiaCube3, for orientation tracking, rather than built-in accelerometers. Since tracking error is more noticeable in AR than in Indirect AR. We also calibrated the system to a known orientation before each trial. This provided a fairly stable, although far from pixel accurate, tracking solution, similar to what we imagine most smartphones will include in the near future once gyroscopes become a common feature. Using higher quality sensors also made the comparison between AR and Indirect AR more even since AR degrades more quickly with poor tracking than Indirect AR. The N900 with the InertiaCube attached can be seen in Fig. 6.

The results of our preliminary study motivated our final study design, which had four main parts. The first component of the study was to compare as directly as possible Indirect AR to traditional AR in good conditions. Good conditions in this case meant users stood at the same location the panoramas were captured from, and the real world environment was predominantly static. The choice of content

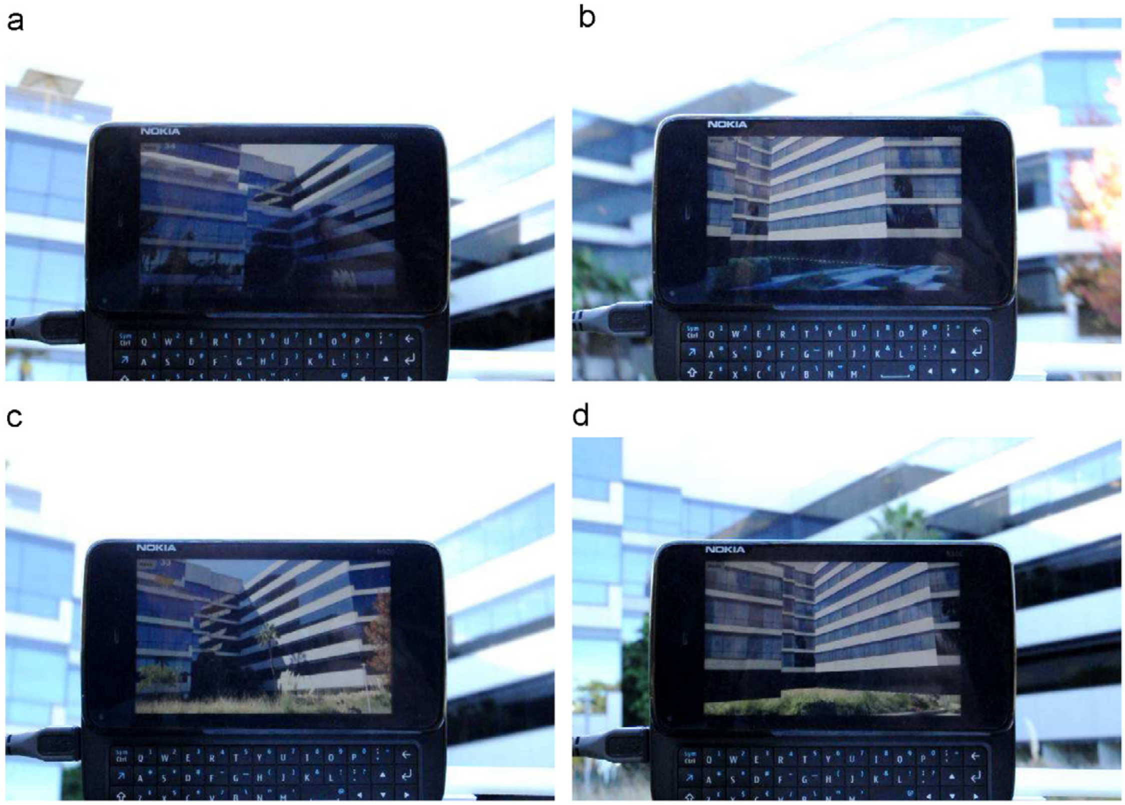


Fig. 7. The four methods of presenting the augmented building. In every case the five storey building has two additional floors added to the top of it. (a) 2D AR. (b) 3D AR. (c) 2D Indirect AR. (d) 3D indirect AR.

was important for this part of the study, and was motivated by what we see as the most important application domains for Indirect AR. Because we think Indirect AR should be used primarily in content rich AR environments we chose to display 3D photo-realistic virtual content. The content we chose, which can be seen in Fig. 7 under various conditions, visualizes a virtual addition to a building. This virtual addition gave us the chance to test several different conditions easily and did not introduce features that might confound the study like animated virtual content. It is also still in an application space where the visualization of the cohesive Augmented Reality scene was important. It allowed us to easily create content with both correct occlusions in AR and Indirect AR, as well as more traditional model based content where the virtual content occludes the real world. Because the content itself is not flashy it also allowed users to focus on the overall experience, and how the Augmented Reality world was presented without the content itself highly coloring their view.

We displayed the augmented building to users in four different ways (as seen in Fig. 7) which we felt covered standard techniques. In what we call the 3D case (because it was implemented using a 3D model) we have a virtual model of the entire building with two extra floors added. This model is displayed on top of the real world background in both AR and Indirect AR, meaning in both cases occlusions of the building by the real world are not correct. The second way of presenting the building tried to preserve correct real world occlusions, and is what we call the 2D case, because it was implemented as a 2D image alteration. This was also possible in AR because users were at a static location. For the Indirect AR case we simply altered the panoramic image to include the modified building. In AR we only added the extra floors virtually, preserving the direct view at the rest of the world. This meant that there were no incorrect occlusions in either case; however, the registration in AR was quite demanding. To evaluate these four techniques we presented them pairwise to users (using a Latin square approach to counter balance the order of the 6 possible pairs). Using this approach every technique (3D AR, 3D Indirect AR, 2D AR, and 2D Indirect AR) was compared directly to every other technique. Users were asked comparative questions after each pairing was shown in order to give us a direct view of how all combinations compared. After all comparisons were done users were shown all four techniques for presenting the scene again and asked to rate them all on an absolute scale, as well as provide written and verbal background about their choices.

Once the interview was finished, users completed the same type of comparison task once more but with different content. We wanted to check if our results also extended to current navigation style applications, so we presented users with the content seen in Fig. 8 and asked them to compare the AR and Indirect AR cases. We then, again, had them rate both AR and Indirect AR for this style of information on an absolute scale, and talked about motivations for their answers.

For the rest of the study we explored some of the things that we thought would degrade the Indirect AR experience. To this point in the study the scene was fairly static. There were occasionally people walking by, but they were far enough away that they were predominately ignored by users. We did want to test Indirect AR in more dynamic environments as well though, so for the next test we had a study administrator walk back and forth two meters in front of the user. We then asked users to compare the experience with or without the mobile occluder in both the AR and Indirect AR conditions, and to compare the two conditions when both were occluded. Once these three comparison tests were done we again had users evaluate each occluded condition on an absolute scale and asked for their opinions.

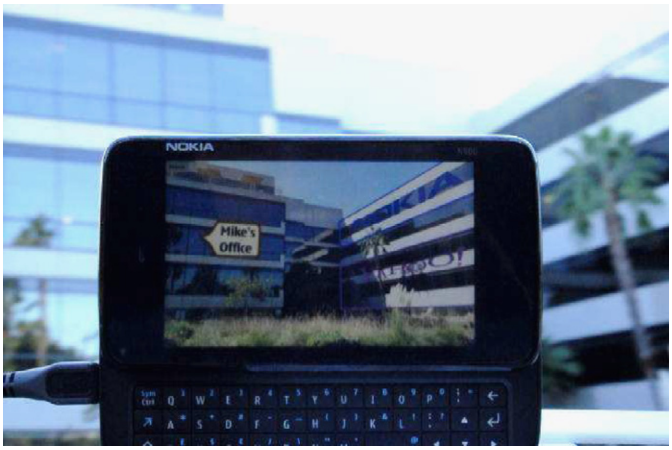


Fig. 8. The simple information based content presented to users. Both an individual office, and blocks of offices are labeled on the exterior of the building. This is the Indirect AR condition; the traditional AR one looks similar, but frequently has more tracking error.

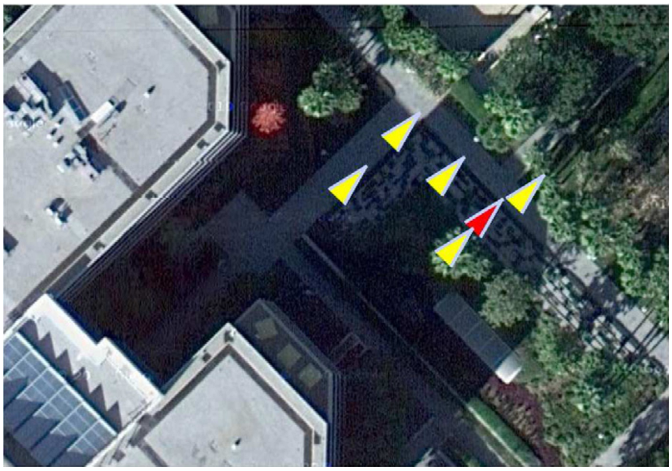
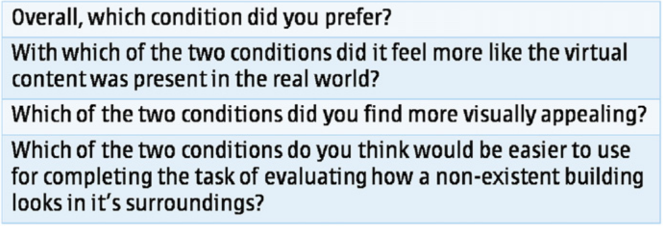


Fig. 9. An aerial photograph of the study location. The augmented building is at the lower left. The user's location (and location of the centered panoramas) is marked with the red triangle, and is around 30 m from the building. The user's location is at the point of the triangle used to mark the viewing frustum. The yellow triangles mark the locations of the other panoramas used which were between 5 and 15 m away from the user's location. From left to right the locations correspond to labels f2, s2, s1, f1, c, b in Figs. 12 and 13. The locations that were chosen also had unoccluded views to the building of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The last part of the study was designed to look at panorama location. In many cases it will likely not be possible to use panoramas that are located exactly where the user is for Indirect AR. It is clear that at some point using a panorama from a different location will change the experience. At some point it will no longer be analogous to the AR experience of looking through a camera, and will instead feel more like the user is looking at a pre-captured alternate view of the environment, similar to using Google's mobile Street View. The goal of this part of the study was to determine if the transitions between experiences was sudden or gradual, and how large an area around the user (if any) gave the same experience as when the panorama is colocated with the user. To test this, we first showed users the Indirect AR condition with the panorama taken from their central position (throughout this part of the test we used the simple navigation style content). We then showed them the same content from six other positions. These other positions were around the user in every direction, and included again the original position. Some positions were quite close to the user, while others were far away. In Fig. 9 the locations of the user and captured



Overall, which condition did you prefer?
With which of the two conditions did it feel more like the virtual content was present in the real world?
Which of the two conditions did you find more visually appealing?
Which of the two conditions do you think would be easier to use for completing the task of evaluating how a non-existent building looks in it's surroundings?

Fig. 10. Questions from our user questionnaire used to compare AR and Indirect AR techniques for displaying complex virtual content.

panoramas can be seen, along with the building they were primarily focused on in the lower left. Users were asked to compare each new view they were shown to the original centered view with a variety of questions (view order was counter balanced). After being presented with all the different views, users were interviewed about how the change in viewing location impacted the experience for them. This part of the study was carried out twice, the first time with everything below horizontal blocked so that users could only see objects that were relatively far away. The second time through (in a different order) users were able to see the entire panoramas including the ground. This repetition was done to see how much of an impact near-field objects had on results even when the objects of interest were still far away.

6.2. Hypotheses

When creating our study we formed several hypotheses relating to different parts of the study. They are presented here:

- Indirect AR will provide the same type of experience as traditional AR when using a panorama centered at the user's position.
- When comparing Indirect AR to traditional AR using complex content (modifying the existing building), Indirect AR will be preferred because of the perceived improvement in tracking.
- When using simple content (labels) the perceived benefits of Indirect AR over AR will be present but less strong.
- Having a dynamic physical environment will degrade both Indirect AR and traditional AR. When the dynamic portion of the environment occludes things of interest to the user the degradation will be similar in both.
- The Indirect AR experience will degrade quite quickly as the user moves further from the location of the panorama. When the distance between the panorama and the user is greater than 10% of the distance between the user and the object of interest the experience will no longer feel like traditional AR.

6.3. Results and discussion

We will present our study results in the same order as the study itself was conducted in, beginning with comparing Indirect AR and traditional AR with complex content (see Fig. 7) when the user was located at the panorama capture location.

Before we begin the specific discussion about each portion of the study there is one more general result that was very important. When starting the study we did not explain the differences between techniques to users; they were only told that they were going to be viewing information in four different ways. Half way through the study, before the portion of the study covering dynamic scenes, the techniques were explained in more detail. At that point we first asked users about the differences between techniques, and only 2 of the 18 users (11%) had noticed that they were not looking through the camera in the Indirect AR conditions. Nearly 90% of users did not realize they were not looking directly at the real world. This is a very strong result indicating that Indirect AR can provide the same type of experience as AR while also providing extra benefits.

6.3.1. AR/Indirect AR comparison with complex content

In this portion of the study we compared four cases (two in AR and two in Indirect AR) pairwise, resulting in six comparisons. With direct comparisons already done by users, analysis is fairly simple. For each question in the questionnaire users responded both with which condition they preferred, and how much they preferred it by. Preference was scored on a 7 point Likert scale, which when combined with the preferred technique gave us a score from -7 to 7. A score of 0 would mean the user thought there was no difference between techniques. To test for significant results we can see if 0 (the no difference score) is contained within ± 1.96 standard errors, a 95% confidence interval, of the mean. If it is not, than that score is statistically significant.

The results for each of the six comparisons can be seen in Fig. 11. The 2D AR technique was the least preferred of all techniques, followed by the 3D AR technique, 3D Indirect AR technique, and 2D Indirect AR technique in order. This ordering of results was also mostly confirmed with the results from the absolute scores users gave to the different techniques. When asked "Overall how convinced were you using [each] technique that the virtual content was actually present in the real world?" Users' mean scores on a 7 point Likert scale were 1.72, 3.22, 3.83, and 6.83 for the 2D AR, 3D AR, 3D Indirect AR, and 2D Indirect AR techniques in order. An ANOVA run on these results showed a strongly significant difference between results ($p \ll 0.01$), and a Tukey Post Hoc test showed significant differences between all techniques except the two 3D techniques ($p = 0.61$). We think this result in particular was less strong in the overall ratings both because the techniques were most similar, and because users viewed all four techniques before responding making it less likely that they would remember the subtle differences that they might have noticed in the direct comparison.

These results essentially agreed with our hypothesis. The 2D AR case was broadly thought of as the worst condition in large part because it made registration errors the most obvious. Although occlusions of trees in front of the building were correct, there was frequently a problem with the roof line where the virtual content was supposed to line up with the physical building. The actual tracking error was not any larger in this condition than in any other, but it was perceived to be much larger because there was an obvious seam where direct comparison between the physical

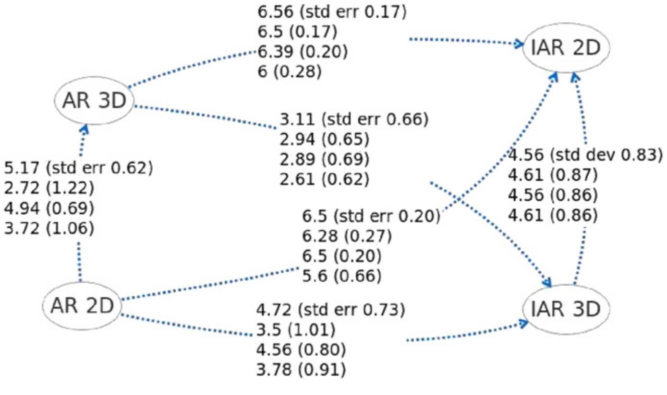


Fig. 11. Results comparing the four presentation techniques with complex content. The four numbers for each comparison are responses to the four questions in Fig. 10. The preferred technique is at the arrow head. All results are significant.

building and the location of the virtual addition was possible. To many users it appeared as if the additional floors were "floating" over the rest of the building rather than being in any way attached. One user said "On that particular one I was taken out of the experience because the top two floors were floating above the rest of the building". The gap between the virtual and physical environments may or may not have made it possible for this user to understand what was being annotated, but it definitely drastically decreased the quality of the experience for him.

The 3D AR technique obviously had the same registration problems as the 2D AR case, but it was more subtle in a way because the entire building was represented as a virtual model, making the seams between real and virtual much less obvious. Naturally, this also introduced huge occlusion problems as all of the landscaping around the building was no longer visible. Because the virtual building was on top though registration errors were much less obvious. There were some, of course, that were noticeable, the most common being when part of the physical building was visible, or when the virtual building covered something that was obviously incorrect, like the sidewalk next to the building. These registration errors were only noticeable on occasion though, rather than nearly all the time, making them much less visually intrusive. The occlusion problems that were introduced with this approach on the other hand were quite intrusive, making it difficult to place the virtual building in the real world because the ground did not line up correctly. One user commented "Where the building meets the ground there's a sharp edge, and obviously the bottom of the building shouldn't really be on top of the stuff that's in the foreground. On the street where there's just pavement and the building it wouldn't be as much of an issue, but here there's all this vegetation".

The 3D Indirect AR technique was visually very similar to the 3D AR technique. As was previously mentioned, very few users noticed a difference between the live camera view and the panorama background, so the primary difference that was noticeable to users was how AR and Indirect AR handled the tracking error differently. This was in some ways a more fair comparison between AR and Indirect AR than the 2D versions because the content was largely virtual (and visually very similar) for both cases. Unfortunately, this is frequently the best case for traditional AR (ignoring occlusion, but having a good representative model), and one of the worst cases for Indirect AR because the knowledge of the previously captured panorama is ignored. Most users could tell a difference between the two techniques (14 of 18 users preferred Indirect AR), especially when directly comparing them as there was a significant difference in that case. Some users did have a difficult time verbalizing what the difference was though. Because there was not a strong visual seam, like there was in the 2D AR case, registration error in 3D AR case was still noticeable to many users and just made the experience feel a little off. In Indirect AR on the other hand because the building and background were completely locked together many users found the overall experience more convincing because the building felt more solid. One user explained their preference for the Indirect AR condition by saying: "The building was more stable; it wasn't floating quite as much. I noticed when I moved the camera the building stayed solid against the trees, and that was a little more convincing to me. There was just a slight less floating effect with the 3D indirect AR".

Lastly, the 2D Indirect AR cases was far and away the preferred technique across all users. With an average score of 6.83 on a 7 point Likert scale users were very convinced that the virtual content was in the physical world. Many users said things like: "That's the one I actually had to do a double take and go that's pretty insane, b'cause it looked pretty seamless and flawless. There was no floating effect, nothing glitchy about it. Everything in the background stayed consistent with what was in the foreground. It was solid for lack of a better word".By matching the content to the pre-captured panorama we were able to overcome many of the problems that occur in traditional AR when less is known about the real world. We were able to add the extra content with pixel accuracy, including occlusions by foreground objects (this was done manually, but could also be done automatically if correct depth information was known). There was also no inaccuracy introduced on the screen by tracking error, and because the virtual content was added with image editing software it could be made to look properly lit to match the lighting conditions within the panorama. All of this enabled an essentially photo-realistic Augmented Reality scene that proved to be very convincing to users.

6.3.2. AR/Indirect AR comparison with simple content

As previously mentioned, we also compared AR and Indirect AR with simple informative annotation style content. We expected to see weaker results here than with more complex content because the merging of real and virtual is not as important. We still found a very strong preference for Indirect AR though. Responses to the first question "Overall which condition did you prefer?" gave a mean result of 4.28 (on our -7 to 7 scale) with a standard error of 0.57, a significant favoring of Indirect AR. We think this is in large part because the Indirect AR case was more visually pleasing, without any tracking error. One user said he felt the AR condition "looked kind of sloppy" because of lag and registration error. Users also favored the Indirect AR case because some of the annotations required a level of accuracy that made some users uncomfortable when using AR. One user commented that in the AR condition: "It didn't match up with the building so badly that I wouldn't trust it. It's saying your office is here, but I don't really know where that is". We feel these similarly strong results in an application area we did not think would show as strong of a difference strengthens the case for Indirect AR being an important approach to consider in a broad range of applications.

6.3.3. Dynamic environments

So far we have only tested Indirect AR in good conditions, but one of the largest potential limitations of Indirect AR is that these ideal conditions will not always be available. In this sub-section and the next we will see how Indirect AR performs in degraded conditions and if the performance still provides the same type of experience.

For the next component of the study we compared AR and Indirect AR in static and dynamic environments using the complex 3D model based content. We had three comparisons between the four items we were testing (the fourth comparison, static AR vs static Indirect AR, was completed previously), as well as an overall rating section, and a short interview.

We found that, somewhat contrary to our expectations, users much preferred Indirect AR to traditional AR in environments where there were dynamic real world occlusions. In the direct comparison between the two (on our -7 to 7 scale) Indirect AR was preferred significantly, with a score of 5.67 (standard error 0.40). There was also a significant difference when comparing the scores of how convinced users were that the virtual content was present in the real world overall (4.62 for Indirect AR vs. 3.06 for AR on a 7 point scale), using a Wilcoxon signed rank test for significance found $p < 0.05$. When comparing the static and dynamic Indirect AR scenarios there was no significant difference (static preferred at 0.38 with standard error of 0.84). On the other hand, the AR condition was significantly degraded in the dynamic environment (static preferred at 5.44 with standard error of 0.42).

From interviewing users we can explain this seemingly counter-intuitive result. When talking about the Indirect AR case many users said things like: ‘‘To me it’s all the same. You’re getting the same information [as you are in the unoccluded case] which is what you want to know’’. The fact that the on-screen environment did not match the real environment did not seem to bother users either. Many felt that they were looking at an idealized view of the environment. The fact that the idealized view did not include some portions of the physical environment did not bother users because, in this case, that was not the part of the environment they were interested in. One user talked about how the presence of the person on the screen was not important because the person was not important to the experience by comparing the person to ‘‘windshield wipers. If it’s not on the screen you just focus on the screen and it doesn’t make a big difference’’. Naturally if the application in some way involved the dynamic portions of the environment as well this result would likely be different. In the AR case on the other hand, there was a significant difference between the static and dynamic environment, largely because of the way occlusions break the illusion of the physical and virtual. One user did not see a strong match between the physical and virtual in this case, saying: ‘‘It looked like two different scenarios. It’s almost like the building was just slapped on top of reality’’. There was also a secondary problem in this case with the AR approach. Because the virtual content was a large model, it occluded a large part of the real world. When a person walked between the user and the building, that person would naturally be mostly occluded by the model, but their legs would generally still be visible, since their legs were below the horizon line. Being able to only see part of the dynamic person, instead of not being able to seem them at all, as in the Indirect AR case, disturbed many users: ‘‘Having half a person there and sort of cut off by the building was more of a distraction. It was obvious it wasn’t real’’.

While the Indirect AR scenario does not represent the world as accurately when the physical world is not static, for many use cases this does not seem to degrade the experience because many of those dynamic portions of the environment are not critical to the overall experience. In traditional AR on the other hand dynamic occluders can drastically alter the experience, breaking the link between physical and virtual content, and looking visually unappealing as they are partially occluded by virtual content that is always on top.

6.3.4. *Non-centered panoramic imagery for Indirect AR*

One of the largest potential failure points for Indirect AR (at least in its current form) is its reliance on pre-captured panoramic imagery. With this portion of the study we hoped to further explore how users reacted to panoramas showing the same virtual content from different perspectives around the user.

As previously described we had users compare images from six locations (including the center) to an original view from their centered location. We did this twice, the first time with everything below the horizon line blocked so that the ground was not visible. We had users fill out a questionnaire after viewing the scene from each different panorama location, comparing it to the original view they were shown which was the same content in a panorama from their viewing location. Of the questions we asked we found two to be the most important. The first: ‘‘Did you perceive a difference between the two conditions?’’ we asked to see if users would notice a difference between the centered panorama, and the more distant ones. Answers to this question were on a seven point Likert scale. The second question that we felt was most important was: ‘‘Did you feel you were looking directly at the physical world, or at a virtual world?’’ The goal of this question was to help determine if the experience still felt like true AR, or if it felt more like a locally relevant panoramic view, but not like AR.

When the ground was blocked there was significance based on location to the first answer (ANOVA $p \ll 0.01$), however, all of this significance was due to a single location, f2 (as seen in Fig. 9), which was the farthest from the users location. In a Tukey Post Hoc analysis the response users gave at that location was significantly different than all other location (p-values < 0.05). We think this result occurred for two reasons: first, and most importantly, the panorama from that location was the most visibly different. Also, users were not particularly successful at picking out the center location (giving it a score of 2.17 in the seven point scale), which on average was the lowest score but not significantly lower than any others except f2.

There was no significant difference per location in response to the second question (ANOVA $p=0.72$) which again indicates that users were not noticing many differences between the different views. Responses to this question and the previous can be seen in Fig. 12. While there were visual differences that some users noticed, many others either did not notice any difference, or did not feel they impacted the experience. One user said: ‘‘Sometimes the landscape changed a little, but that wasn’t part of my focus. For all intents and purposes it could have been reality, and I probably wouldn’t have noticed anyway’’. This result was also influenced somewhat by the effect of having the ground blocked. Many users commented that having the ground blocked made the whole experience feel more like a virtual world: ‘‘The fact that it was blocked already tells you that something was different about the picture. [The unblocked view] was closer to reality overall because it didn’t have the block’’. The fact that users felt the blocker itself was impacting the experience likely reduced the variance in scores between views since for some users just having the blocker made the experience feel more virtual.

In the non-blocked case there were several differences in the results, which can be seen in Fig. 13. When asked if they could tell a difference between the current view and a centered view there were more significant differences compared to the blocked case, which makes sense because users had more information (the ground) available to look at. In this case a Tukey Post Hoc analysis found significant differences between the centered view and every other view (p-values < 0.05), except the s1 view ($p=0.41$). In many ways these two views shared the most similar views of the building of interest, and absolutely shared the most common foreground view. Interestingly though, there were fewer differences in responses to the question about how the experience felt (real or virtual). In this case there were only significant differences between the centered view and f2 (Tukey Post Hoc $p \ll 0.01$) and s2 ($p < 0.01$) views. We found no significant differences between the center view and other nearby views. These two results are somewhat juxtaposed.

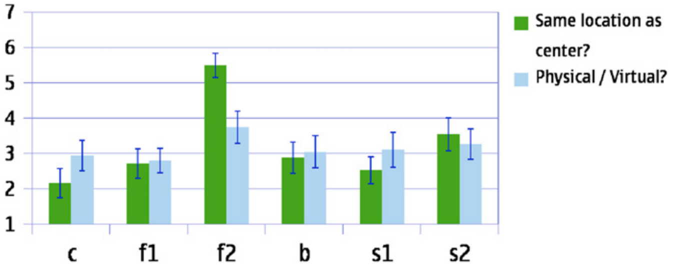


Fig. 12. Questionnaire results when everything below the horizon was blocked. For the first question ‘‘Did you perceive a difference between this condition and the center?’’ a lower score means less difference. For the second question ‘‘Did you feel like you were looking directly at the physical world, or at a virtual world?’’ a lower score means physical world.

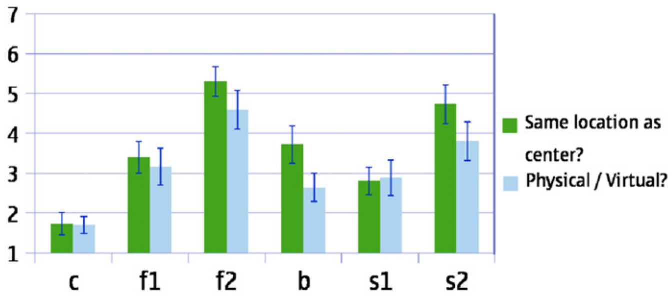


Fig. 13. Results when users could see the entire panorama.

One explanation for this is that although users could notice small differences between the onscreen view and an ideal view, when those differences were significantly minor (i.e. the panorama as taken sufficiently close by) users still felt that the overall experience was very similar, or the same.

The differences between the blocked and non-blocked results are also interesting. It is clear that blocking everything below the horizon line does have an effect, since people were much better at differentiating panoramas when they were not blocked. Also of interest were users' responses when asked if they were looking at a physical or virtual scene. When the ground was blocked these answers were all quite close together, while when the ground was not blocked they covered a broader range in more or less the expected order. The distraction of the blocker itself certainly was part of the reason for this result, but it is also quite likely that the effect of panorama location was lessened when only more distant objects were visible.

Defining what distance the location of the panorama can be from the user is still a very difficult, and ill-defined problem. What does seem likely though from our results is that there is a region near the user where the overall experience is very similar for most users to that of traditional AR, or Indirect AR when the panorama is centered. The size of this region is dependent on several factors though. Large differences in the near field scenery might play a role, and the distance to the objects users are interested in definitely plays a role. In our case the building of interest was approximately 30 m away, and the three nearby locations which still seemed to have a fairly similar overall experience to the centered panorama were each around 5 m away. If the distance to the objects of interest was also only 5 m this result would clearly not still hold. We could possibly speculate that the same ratio might be consistent, that the distance to the panorama location can be up to 16% of the distance to the target of interest, but we have insufficient data to confirm this claim. It is also not clear if the direction to the panorama's location is important. In our case, we did not notice a strong effect from the direction to the panorama location. In fact, users gave very conflicting results on direction. Some liked the closer view more, saying "when it was bigger and filled the frame more it felt just like it was zoomed in". Others felt the same way about the more distant views, preferring to be "at a distance so you could see the whole building". In general though it does seem that using panoramas from nearby locations can provide the same overall type of Indirect AR experience as those colocated with the user.

We were predominately interested in using existing panoramas in this study to see if Indirect AR could be used with panoramas collected automatically by companies like Navteq and Google. While it does seem that this could work well, in many cases it might be possible to improve the experience even further by using those existing panoramas to generate a new novel view (either panoramic, or 3D) closer to the user's actual position. There is a large body of existing work [1,17,24] in the image based rendering community to achieve this type of effect, and much of it would likely translate well to Indirect AR.

7. Conclusions

In this paper we have presented a new type of Mixed Reality experience that is similar to AR experientially, but can be used with lower quality tracking while still maintaining a good overall user experience. Using panoramas in place of the live camera view enables pixel accurate matching between the virtual content and representation of the real world. Through the work presented in this paper we also showed that even when conditions for Indirect AR degrade to the point that the experience changes, people can still use the application functionally since users are very good at matching their real world view with the on-screen view. The most important take away message though is that in good conditions Indirect AR provides a user preferred experience to traditional AR which can be used with lower quality tracking.

In presenting Indirect AR we think it is very important to think about both good and bad use cases. Our user study results suggest that Indirect AR does very well in outdoor applications where the user is more than a few meters away from the physical objects of interest. Because of the pixel accurate registration between the virtual and physical, Indirect AR also excels in use cases where the tie between the virtual content and real world is important to the overall experience. That is particularly true for entertainment based applications where the believability of the Augmented Reality environment is central to the overall experience. Indirect AR is just one tool in an MR toolbox, however, and there are certain times when other approaches will have a better end user experience. For instance, when the user is interacting with objects in the near field a vision based AR approach will likely be superior given its ability to better handle user motion parallax. While Indirect AR may not be perfect for every scenario, we do feel that it has great potential to enable immersive AR-like applications in many places where that was not previously possible.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version of 10.1016/j.cag.2011.04.010.

References

[1] Aoki T, Tanikawa T, Hirose M. Virtual 3d world construction by interconnecting photograph-based 3d models. In: Proceedings of the IEEE virtual reality conference; 2008. p. 243–4.

[2] Brown LD, Hua H. Magic lenses for augmented virtual environments. IEEE Computer Graphics and Applications 2006;26(July):64–73.

[3] Gandy M, Catrambone R, MacIntyre B, Alvarez C, Eiriksdottir E, Hilimire M, et al. Experiences with an ar evaluation test bed: presence, performance, and physiological measurement. In: ISMAR '10: proceedings of the international symposium on mixed and augmented reality; 2010. p. 127–36.

[4] Goldiez B, Dawson JW. Is presence present in augmented reality systems? In: Proceedings of presence 2004, international workshop on presence; 2004. p. 294–7.

[5] Hill A, MacIntyre B, Gandy M, Davidson B, Rouzati H. Kharma: An open kml/html architecture for mobile augmented reality applications. In: ISMAR '10: proceedings of the international symposium on mixed and augmented reality; 2010. p. 233–4.

[6] Iachini T, Logie RH. The role of perspective in locating position in a real-world, unfamiliar environment. Applied Cognitive Psychology 2003;17(6):715–32.

[7] Liestol G. Augmented reality and digital genre design—situated simulations on the iphone. In: ISMAR '09: proceedings of the international symposium on mixed and augmented reality—arts, media and humanities; 2009. p. 29–34.

[8] Liestol G, Rasmussen T. In the presence of the past. a field trial evaluation of a situated simulation design reconstructing a viking burial scene. In: Media inspirations for learning. Proceedings of EDEN 2010; 2010.

[9] Livingston M, Ai Z. The effect of registration error on tracking distant augmented objects. In: ISMAR '08: proceedings of the international symposium on mixed and augmented reality; 2008. p. 77–86.

[10] Looser J, Billinghurst M, Cockburn A. Through the looking glass: the use of lenses as an interface tool for augmented reality interfaces. In: GRAPHITE '04: proceedings of the second international conference on computer graphics and interactive techniques in Australasia and South East Asia; 2004. p. 204–11.

[11] Miyashita T, Meier P, Tachikawa T, Orlic S, Eble T, Scholz V, et al. An augmented reality museum guide. In: Proceedings of the international symposium on mixed and augmented reality; 2008. p. 103–6.

[12] Papagiannakis G, Kim H, Magnenat-Thalmann N. Believability and presence in mobile mixed reality environments. In: IEEE workshop on virtuality structures (VR2005); 2005.

[13] Quarles J, Fishwick P, Lampotang S, Fischler I, Lok B. A mixed reality approach for interactively blending dynamic models with corresponding physical phenomena. ACM Transactions on Modeling and Computer Simulation 2010;20(November):22:1–23.

[14] Quarles J, Lampotang S, Fischler I, Fishwick P, Lok B. A mixed reality approach for merging abstract and concrete knowledge. In: VR '08: proceedings of the IEEE virtual reality conference; 2008. p. 27–34.

[15] Ragan E, Wilkes C, Bowman D, Hollerer T. Simulation of augmented reality systems in purely virtual environments. In: VR '09: proceedings of the IEEE virtual reality conference; 2009. p. 287–8.

[16] Slater M, Steed A. A virtual presence counter. Presence: Teleoperators and Virtual Environments 2000;9(October):413–34.

[17] Takahashi T, Kawasaki H, Ikeuchi K, Sakauchi M. Arbitrary view position and direction rendering for large-scale scenes. In: CVPR '00: proceedings of the IEEE conference on computer vision and pattern recognition; 2000. p. 296–303.

[18] Uyttendaele M, Criminisi A, Kang S, Winder S, Szeliski R, Hartley R. Image-based interactive exploration of real-world environments. Computer Graphics and Applications, IEEE 2004;24(3):52–63.

[19] Wagner D, Mulloni A, Langlotz T, Schmalstieg D. Real-time panoramic mapping and tracking on mobile phones. In: VR 2010: proceedings of the IEEE virtual reality conference; 2010. p. 211–8.

[20] Wagner D, Reitmayr G, Mulloni A, Drummond T, Schmalstieg D. Real-time detection and tracking for augmented reality on mobile phones. IEEE Transactions on Visualization and Computer Graphics 2010;16:355–68.

[21] Wagner I, Broll W, Jacucci G, Kuutii K, McCall R, Morrison A, et al. On the role of presence in mixed reality. Presence: Teleoperators and Virtual Environments 2009;18(August):249–76.

[22] Wither J, Allen R, Samanta V, Hemanus J, Tsai Y-T, Azuma R, et al. The westwood experience: connecting story to locations via mixed reality. In: ISMAR '10: proceedings of the international symposium on mixed and augmented reality—arts media and humanities; 2010. p. 39–46.

[23] Witmer B, Singer M. Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoperators and Virtual Envrionments 1998;7(3):225–40.

[24] Zomet A, Feldman D, Peleg S, Weinshall D. Mosaicing new views: the crossed-slits projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 2003;25(6):741–54.