Sharing Perspective in Remote Collaborative AR

Peter Buk
University of Calgary
pkpbuk@ucalgary.ca

ABSTRACT
Augmented reality provides an intuitive way to work with 3D models because of the natural mapping to our physical movement and the benefits of AR technology. However, it does not easily support remote collaboration due to difficulties in communicating spatial and contextual information. We investigate this problem by introducing the Perspective Block, a novel handheld interaction tool. This block has an embedded camera and sensors that can be used to capture important perspective information such as contextual location and video feeds. By using this block, collaborators will be able to share their perspective with each other using a variety of interaction techniques, allowing for more natural communication. These interaction techniques were evaluated through an informal study in order to gather feedback on their effectiveness and potential of use in a real world context.

Keywords
Augmented reality, Remote collaboration, View perspective, Contextual communication, Shared 3D models

1. INTRODUCTION
3D models are used in many domains as a representative model of our physical world. The problem is that current methods of navigating and working with these models often feel lacking. Navigating around digital 3D models through traditional 2D displays often feels clunky, and physical scale models are expensive to construct and difficult to manipulate afterwards. This problem is compounded when we have to do remote collaborative work involving these models. We often have to rely on screencasting software or video conferencing tools in order to collaborate around a shared 3D model because few collaborative tools currently exist for this type of task.

Previous research has shown that augmented reality (AR) provides a more intuitive way to navigate around 3D models [3, 7]. This is primarily because AR provides a natural grounding for 3D models since there is a physical mapping between our movements and the resulting observations. However, these tools are also not very well suited for use in remote collaborative situations. Important perspective information such as spatial and contextual data can be difficult to convey to remote collaborators. We often end up relying on giving deictic instructions such as "look over here", to each other which are hard to interpret without context. For example, in order to tell your partner to look at a particular target on a building model, you would have to instruct them to move to the right location and then orient their view to the correct angle to match the same perspective view that you see. This problem is illustrated in Figure 1.

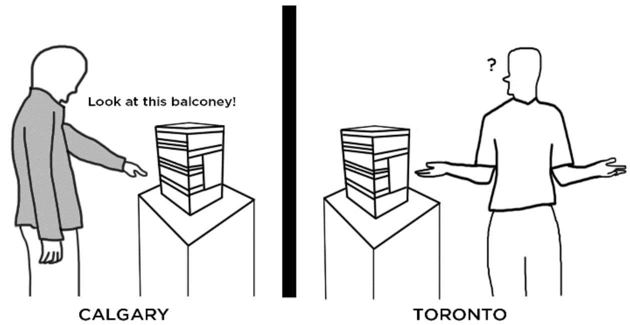

FIGURE 1: Communicating perspective information with a remote partner can be difficult without context or being able to see each other.

Our approach to this problem is to explore the concept of perspective sharing through the use of novel interaction techniques. In order to facilitate this, we designed a tool called the Perspective Block. This block will have an embedded camera that can capture a live video feed from a target perspective. It will also be tracked in 3D space so it can be used as a pointer and signifier to convey important contextual information. By using the Perspective Block, collaborators will be able to communicate more effectively because they no longer have to communicate the same perspective information vocally through ambiguous instructions. Some of the interaction techniques explored include: direct video feeds captured from your partner's block, a picture-in-picture view in order to keep constant track of your partner, and various visual representations to indicate your partner's perspective in your own 3D space. By using these techniques, a collaborator will be able to focus on the collaboration task at hand, rather than trying to orient and maintain awareness of their partner's perspective.

First we will take a look at the previous work done in this area in regards to remote collaboration and augmented reality tools previously explored. Then we discuss in detail the concept of the Perspective Block, our current prototype and the interaction techniques implemented. We conducted an informal study to test these techniques through a scenario where a pair of collaborators have to perform a critique on various architectural models. A post-study interview then gather qualitative feedback on the effectiveness of each interaction technique, and the potential of this early prototype. Finally, we conclude with a discussion on the strengths and weaknesses of the concept and the direction future work can take.

2. RELATED WORKS

In this section, we look at previous work done in the areas of augmented reality, view perspective information, and remote collaboration.

2.1 3D Models in Augmented Reality

The Boom Chameleon [7] is a movable display mounted on a mechanical boom which acted as a window into a 3D virtual environment. Since this display could maneuver around in 3D space, there was a natural mapping between the physical space and the virtual space. They explored how this device can be used effectively for 3D design reviews, and found it to be a quick and easy to learn method to navigate around 3D objects. The approach here is similar to what we are doing in that we are viewing a virtual 3D model through a physical display, except now we are using a modern head-mounted display instead of a display panel. This allows us to manipulate the view perspective more freely by naturally moving around instead of being anchored to a central device. They also did not explore how collaboration work might be done with this system which is the key focus of our approach. Nonetheless, they proved that AR can be an effective way to explore around a virtual 3D model.

2.2 View Perspective Information

An important focus of our interaction techniques are ways to communicate spatial information about a given perspective. ParaFrustum [4] looked at visualization techniques to guide a user to a certain viewing position and orientation. They defined a *ParaFrustum* as a geometric construct to represent the look-from and look-at volumes, which serve as the range of acceptable perspectives to view the target location from. One visualizations imposed a 3D shape into the physical world in which the user uses to orient themselves to the proper perspective. They found that it provided immediate information and was rather intuitive to use. The work done in this paper is important for us because it shows that a visual representation of the target perspective is an effective way to communicate spatial information. In our work, we are looking at ways to communicate spatial and contextual information between two remote partners. Parafrustum might be an effective way to convey that information in a useful manner.

Dyck and Gutwin [2] created GroupSpace, a 3D collaborative workspace. They looked at ways in which users can be aware of the locations, perspectives, and proximity of each other's avatar in the 3D space. They used embodiment enhancements such as a nose ray, a view cone, and the head light which is emitted from a user's avatar. User trials [1] have shown that these techniques are effective in supporting awareness in 3D environments. In particular, we were inspired by some of their visual representation when looking at our own system. We implemented a version of their view cone and head light, although in a slightly different manner because it is no longer tethered to a human head. By using these techniques in conjunction with augmented reality, we can create ways to aid in communicating perspective information through visual cues.

One advantage of AR is the ability to easily manipulate your viewpoint in ways that you cannot in the physical realm. Sekan [5] introduced a new interaction technique to rapidly switch between viewpoints using snapshots of an AR scene. Users can revisit the scene of a snapshot virtually, using a quick fluid technique to transition between them. They found that moving virtually between viewpoints provides a better way to navigate in the AR domain. This is relevant in our work because we are looking at perspective view points and how we can navigate between these. This paper showed that jumping between the grounded augmented view and a virtual one (the one seen by your partner for example) can be an effective way of easily exploring an augmented model.

2.3 View Independence

Tait and Billinghurst [6] described view independence as the ability for a remote user to view a scene independently of the local user. Other attempts at exploring remote collaborative AR often had the remote user fixed to the local user's view, akin to video conferencing. By using depth scanners, they were able to create an AR system that supported view independence where the worker was able to look around the scene freely. They varied the amount of view dependence and found that having more led to faster task completion time and greater confidence in users. View independence is important because it allows collaborators to be able to work independently on subtasks, as well as providing greater situation awareness. This aspect is something we want to keep as one of the main focuses of our design. A difference between their study and our work is that they looked at worker-expert collaboration where one is sitting at a desktop interface. In our system, both collaborators will have their own full (but identical) AR systems, so they naturally have full view independence. We will instead look at how collaborators can communicate information when they actually need to share their view with each other.

3. APPROACH

In this section, we describe the details of our prototype. We will discuss the design and goals of the Perspective Block, the prototype implementation, and the various interaction techniques.

3.1 Perspective Block Design

The Perspective Block is a handheld tool that is used when collaborators are working remotely around a shared 3D object. While our context here is using Augmented Reality, the core concept of this tool can be modified to be used with physical scale model as well as entirely virtual ones. The design of the block is inspired by handheld laser pointers that can be held with one hand and extends our natural tendency to point and direct at objects as we refer to them.

The goal of the block is to be able to easily share perspective information between collaborators. Perspective information is defined here as any data that provides a hint on where a collaborator is located, what they are looking at, or what they are referring to as they conduct collaborative work. This information is captured through sensors on the block such as cameras and gyroscopes. Important perspective and contextual information will be captured with respect to the 3D model anchored as the world center point.

This block will also be used with a head-mounted display that will display the actual augmented model to the collaborator. While we could have captured the perspective information straight from the HMD as it is a representation of what the collaborator sees directly, we instead chose to capture it from a separate block. Our intention was to separate the perspective of the block from the display view, so that the collaborator does not feel tethered to the constrained motions of their head as they explore and share a perspective with their partner.

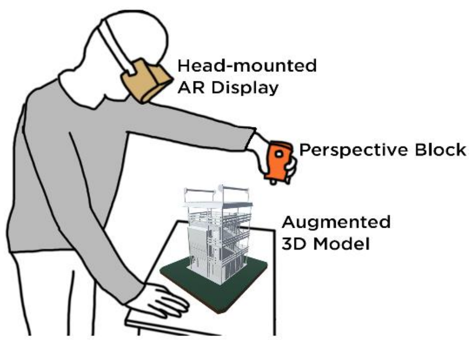

Figure 2. The Perspective Block concept.

3.2 Prototype Implementation

For the current prototype, we are using a Nexus 5 smartphone as the Perspective Block. This was chosen because of its compatibility with the AR framework we are using and is equipped with all the necessary sensors requires to capture spatial information and record a live camera feed. While this phone is usable in one hand, it does not match the ideal design of the block exactly. The phone has to be held in landscape mode in order to capture video properly, which takes away the affordance of using it as a natural pointer block as originally intended. Nonetheless, it provides a good proof of concept for evaluating how a handheld tool can be used when working with an augmented 3D model.

We display the augmented models through the Google Cardboard head-mounted display (HMD). Another Nexus 5 smartphone is inserted as the display device for the Cardboard. We chose this hardware combination because it provides an inexpensive and simple way to work with AR models. A HMD was chosen over a handheld display because it would be troublesome for the user to have to hold another device along with the Perspective Block. By using the smartphone camera pass-through, the user is able to maneuver normally in the real world while still being able to see augmented objects embedded on screen.

The implementation was created using the Unity3D game engine because of its strong support for manipulating 3D objects. We are also using the Qualcomm Vuforia AR platform [8] to render the augmented models and handle interactions involving AR. This platform was chosen because it has a very robust API for generating and manipulating augmented objects, as well as object tracking that is used to sync each collaborator's 3D scene to the same world coordinates. The 3D models are anchored to a physical target model that gets tracked using the camera of the AR device, which in this case is both the block and the HMD. Vuforia then renders the augmented model on top of the target and the user can then maneuver around the target as if it was a physical model as shown in Figure 3. This affords us a natural mapping to our movements and allows for physical exploration around the augmented model. Each collaborator will be viewing the same 3D object, but through their own remote image target which allows them full view independence as the objects themselves are not synced so the collaborators are not tethered to each other in any way.

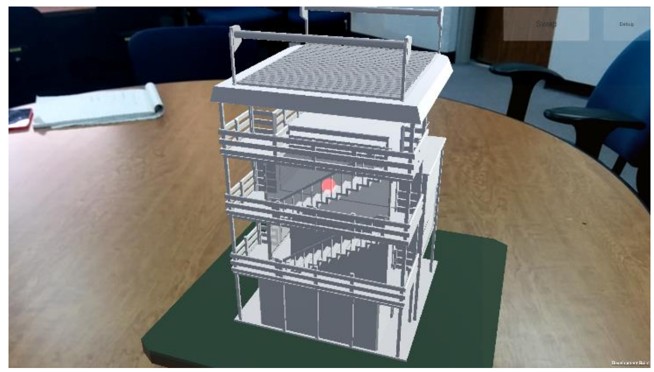

Figure 3. The AR model as seen through Vuforia. Normally this image is split into the stereoscopic view seen through the Google Cardboard to create the illusion of 3D depth.

3.3 Interaction Techniques

The block is able to capture the coordinates of its own position and rotation in 3D space with respect to the augmented model as the origin point. It is also capable of capturing a direct video feed of what it sees through its camera view. This information is sent through the network connecting the block and the view, forming the backend of our interaction techniques. These techniques are broken down into 2 categories, view pointers and video feeds.

3.3.1 View Pointers

We created various methods to embed a visual representation of the remote partner's view perspective into the 3D space of the collaborator. These techniques aim to subtlety cue a collaborator on their partner's perspective without being too distracting and preventing them from their own exploration task. By embedding these pointers into the same 3D space as the object, we can create the illusion of a shared workspace even though the collaborators are working independently in their own 3D space. Each method can be toggled through by touching a button through a cut-out on the Google Cardboard display.

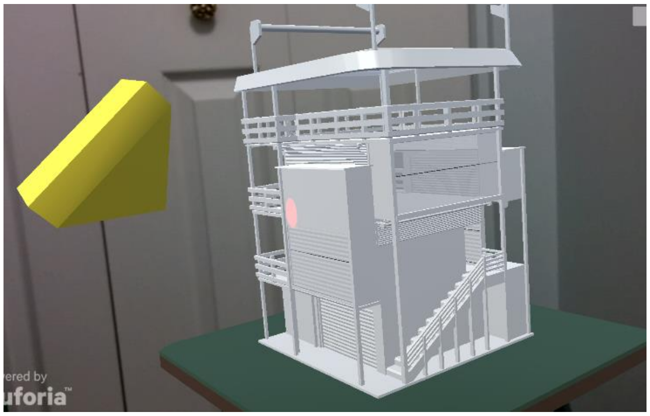

Figure 4. The yellow block is an embedded representation of the partner's block in 3D space.

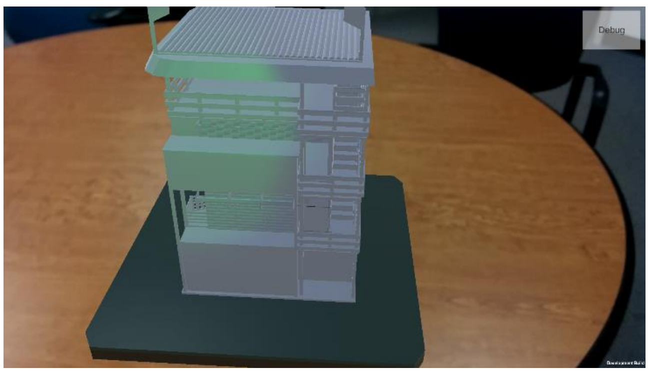

Figure 5. A green spotlight which subtlety shows the direction the partner is looking at.

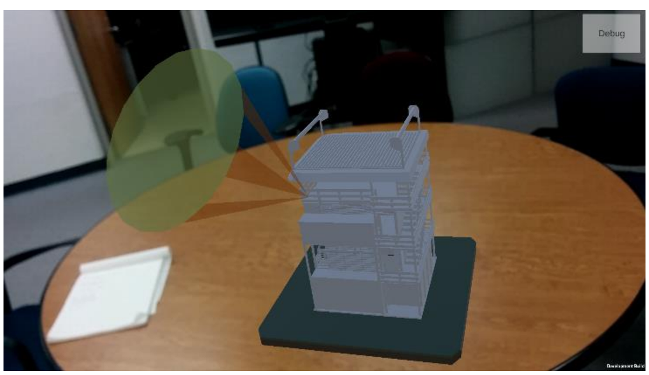

Figure 6. A view frustum inspired by the Parafrustum. This geometric construct represents the exact angle and direction of the partner's perspective.

Our first technique, shown in Figure 4, is a yellow floating pointer that directly represents the partner's block in 3D space as they move around in relation to the image target. By using the position and rotation data captured by the block, we can recreate it on the other side and embed it as an augmented object using the image target as the reference point. The angle of the block shows the direction that their block is currently facing. This pointer also shines a red laser point as a means to point at certain details of the model and give contextual information.

An alternative to the floating pointer is a green spotlight that is controlled by the partner in a similar manner. Instead of explicitly showing the partner's block in space, we show a green spotlight shone in the direction that the block is facing. This is shown in Figure 5. With this technique, we get a subtle representation of the partner's current perspective in 3D space, but it doesn't get in the way of the 3D model at all.

As a contrast to the subtle cueing of the spotlight, Figure 6 shows the last view pointer technique. Inspired by the Parafrustum [4], we created a geometric construct that represents the exact viewpoint and direction of a partner's perspective. This can be used to help directly line yourself up with your partner's perspective allowing you to see from their exact point of view.

3.3.2 Video Feed

By using the video feed captured from the partner's block, we can provide a live feed of what they are currently seeing.

In Figure 7, we see a video feed from our partner's block rather than our regular AR view with the camera-passthrough. We call this technique the Forced Perspective Shift as we are forced into seeing the model from another perspective as dictated by your partner. Through this technique, we end up voluntarily giving up our view independence in order to gain understanding of our partner's perspective. This is similar to classic video conferencing techniques where an expert assists a remote worker through a captured video feed.

If we want to continue exploring the model on our own, we can toggle the video feed to a smaller picture-in-picture frame instead as shown in Figure 8. This allows us to get the benefits of both worlds, giving us full view independence which still being able to keep track of your partner's perspective and the current view they are working with.

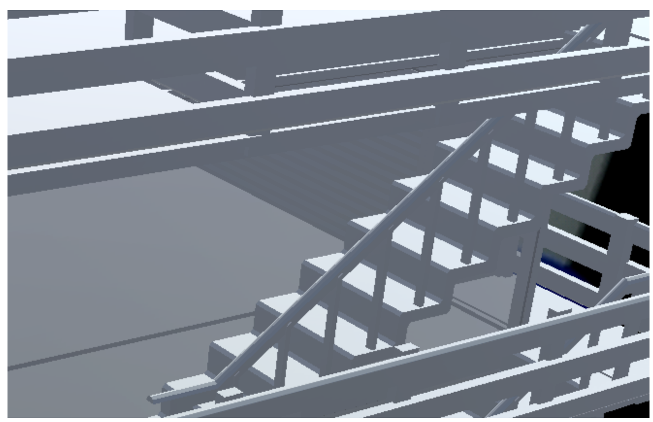

Figure 7. Forced Perspective Shift. We jump from our regular AR view to a live video feed from the partner's perspective.

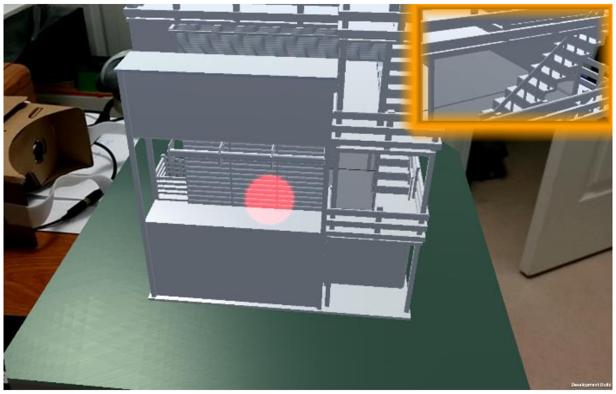

Figure 8. The video feed is shown in a picture-in-picture frame instead. This combined view allows for exploration while still being able to keep track of the partner's perspective.

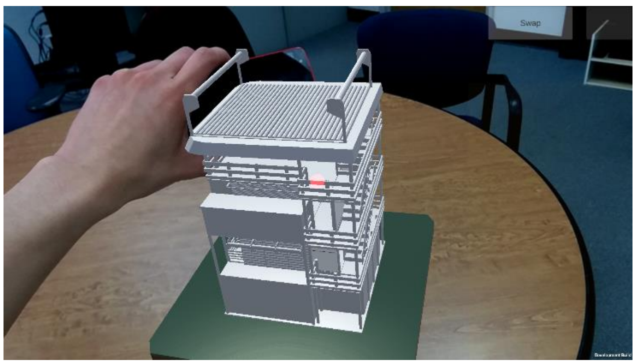

Figure 9. Depth occlusion problem. The Perspective Block cannot be used easily with the augmented model because it breaks the 3D depth illusion of the model.

3.4 Limitations

We faced numerous challenges as we were implementing this prototype. Many of these were related to limitations with the current state of the hardware, and unfortunately we were only able to compromise and work around them.

3.4.1 Depth Occlusion

While Vuforia's AR Camera works very well to detect and render an augmented object onto an image target, one major flaw is that the camera hardware doesn't support depth detection. This means that the rendered object is just overlaid in the correct position on the screen once the target is tracked. When you approach the augmented object in 3D space with a physical object, say the Perspective Block, the physical object can end up covered by the rendered object in an unrealistic manner since the camera pass-through does not account for the depth of the physical object relative to the image target. This problem is shown in Figure 9. It may seem like the phone is behind the image target, but in reality, it is actually pretty far in front of it, which the camera fails to capture. Attempts to overlay an augmented model on top of the physical block resulted in a clunky experience that broke immersion even more. To compromise for this problem, we tried making our 3D model bigger and skinner in an attempt to make it easier to position the block around the model without getting blocked by it, but this does not do much to address the core problem here. Since being able to position the block anywhere around the augmented object is a key concept of the design, we were not able to fully realize the potential of the Perspective Block due to this technological limitation.

3.4.2 Target Tracking and latency

Our choice of technology was quick and inexpensive to deploy and test, but it came with the downside of providing not the best of user experiences. The image targets frequently lost tracking, due to the low resolution of the smartphones. This severely limited the range of motion a user can have around the 3D model as the AR camera always has to be able to see the image target in some way. The camera-passthrough and video capture also suffered from latency issues that may have hampered the effectiveness of any techniques involving them. These issues will have to be taken into heavy consideration with the feedback of the study in order to decide if the problems exists with the technology used for the prototype, or if there are more fundamental problems with the core concept of the Perspective Block instead.

4. EVALUATION

An informal study session was run in order to gather feedback regarding the prototype. The goal of this session was not to necessary test the techniques against each other, but to gather feedback on the effectiveness and explore the potential of each idea. The feedback will be useful to determine if we are heading in the right direction with the Perspective Block and how it can potentially be used to solve problems in a real world context.

4.1 Participants

Two pairs of participants were recruited for this study session. One pair had architectural background and they both said they had experience collaborating remotely with 3D models before, although through tradition 2D software and conferencing tools. The other pair of participants had a Computer Science and Math background respectively. Two participants have worked with augmented reality before, but not very extensively.

4.2 Method

The study session consisted of an architectural critique task, where both participants were asked to critically evaluate various aspects of two architectural 3D house models. This task is representative of a real world scenario where architect collaborators may have to work remotely in different branch offices. Participants were given an introduction to the prototype system and then briefed on the various interaction techniques that are possible. They were then separated by a divider so that they could not see each other but could still communicate vocally. One participant wore the HMD and the other one held the block. Due to hardware constraints, we were not able to provide a full set of HMD and block to both participants. They were given 7 minutes to freely explore and critique the house with each other before switching roles to try out both devices. After the task was over, a post-study interview was conducted to gather qualitative feedback on the task and the prototype. The interview questions focused on both the overall potential of the concept, as well as the effectiveness of each interaction technique.

4.3 Feedback

We gathered a considerable amount of feedback regarding the prototype and core concept of the Perspective Block. The following is a synthesis on both the positive and negative aspects of the study.

The participants generally found the prototype to be novel and present interesting ideas, but also had some difficulties in using it. Perhaps the briefing at the beginning of the sessions was not through enough or the UI was not clear, but the participants had to be reminded frequently of what techniques were possible and how to switch between them. We also experienced a lot of issues regarding image tracking and participants often felt lost when they lost tracking entirely, or when the model ended up moving in unexpected ways because of faulty tracking.

It was unclear whether the tool provided much benefit over a traditional 2D display. It was brought up that if the alternative was for both architects to have a 3D software such as SketchUp open and communicating through Skype, this prototype does not offer much benefit over that due to the increased complexity introduced by the tool. Architects generally know their software well, so it feels much more natural to maneuver a 3D object on a traditional screen, so the benefits of AR is lost here. The affordances gained by moving from a static flat image to a manipulatable 3D model is huge, but moving to an actual AR 3D model is not so significant. One suggestion was that while this tool may not be very useful for architects working together, it may prove helpful when having to demonstrate to end-users. These end-users often have no experience working with 3D software and may benefit from the natural mapping provided by AR as a way to navigate around the 3D object.

One weakness of the 3D models used was that it only provided a single birds-eye view of the model. For an architectural model to be useful, architects would have to be able to evaluate not only the exterior of the model, but also the interior and surrounding environment. The fidelity on the models were also not very detailed compared to traditional software which may offer multiple views and cross-sections. A potentially more useful case where AR would excel in is if the model was seen through a one to one scale perspective view that the user can "explore". 3D Software tries to emulate this first-person perspective, but it lacks precision and immersion due to limited field of vision and control. With AR and VR we have to potential to create models that you can physically explore all around you.

Regarding the individual techniques, the picture-in-picture frame was found to be the most useful technique to use to share perspective. It provided a very easy way for one collaborator to refer to something and talk about it, knowing that their partner can see exactly what they are seeing with no additional effort. It allowed them to continue exploring independently while still maintaining awareness of what is going on in the other side. One participant watched the full Forced Perspective Shift video for a while, saying they wanted to just watch the person explore and guide them around the model. In fact, in both pairs of participants, there was usually someone who was leading the exploration. This may just be the nature of the task or the fact that the participants knew each other beforehand, but it would be interesting to see if the asymmetrical scenario setup played a role as well.

The view pointers were not found to be not very intuitive to make sense of. Both pairs expressed confusion at first for what they were and if it was related to the actual 3D model or not. The red laser light from the floating pointer made sense, as it was "like using a laser pointer" as expressed by one participant. The yellow floating was okay in terms of effectiveness after it was properly explained, but the green spotlight was found to be too subtle to be very helpful. The Parafrustum viewcone was potentially a good idea because of how big and visible it is, but the prototype did not anchor it very well to the model, making it difficult to actually use in practice. One remark was that even after deciphering the information from these view pointers, it would still take effort to manipulate to that spot and see what they are talking about, whereas the video frame would provide an instant contextual clue without having to move anywhere.

5. DISCUSSION

5.1 Comparison to current 3D collaborative solutions

Although it was found that AR did not add very much value to the exploration of 3D objects, it was revealed that current solutions do not fare much better. One participant mentioned that they had used Skype before and tried to sync their own programs through verbal directions with their partner. This was a very time consuming task as they had to describe a view and how to navigate to it. They also tried using screen sharing software to collaborate, but this resulted in having a shared workspace that only one collaborator can control at a time which doesn't allow for independent exploration. Since collaboration is common during design tasks, this discussion revealed a big gap in the market for solutions that offer strong collaboration with 3D software. Even though the Perspective Block may not be the right solution for this problem yet, it provides insight into a potential range of tools that can help.

5.2 One way communication

The biggest weakness of our study session was the asymmetric setup. The true concept of the Perspective Block has both participants wearing a HMD and using their own block to capture perspective. By limiting one participant to each, we effectively create a one way path of communication where information only flows from the block user to the HMD user. The block user has no idea where the HMD user is looking and deferred back to verbal deictic instructions to try and figure it out. This major oversight reinforced that this problem does exists but we did not attempt to solve it very well. Future work should definitely test the system with two Perspective Blocks before we can properly determine if the concept is a viable solution to the problem or not.

5.3 New interaction techniques

A number of new interaction techniques were suggested that can help enhance the current ones or replace them entirely.

One suggestion by a participant was to have a button that can be used to automatically "jump" to the view of the other collaborator. Since the models are rendered virtually, we can easily control and manipulate the orientation of it without requiring physical movement of the user. With a smooth transition, we can essentially guide the user from one view perspective to another while maintaining contextual awareness, as seen in the viewpoint switching done by Sekan [5]. An alternative to this idea would be a way to record a guided tour around the object that can be replayed to highlight the path taken in exploration.

An idea we did not fully explore was the concept of pointing. Despite it being one of the original sources of inspiration for the Perspective Block, our smartphone prototype did not allow us to point very naturally due to the way we have to hold and orient the phone in order to use the camera capture. Combined with the occlusion problem, we end up losing the natural tendency to point at objects with this prototype. In a later iteration with more refined hardware, the concept of pointing with regards to augmented objects should be explored more in depth as a way to provide contextual hints.

6. CONTRIBUTION

AR models have a lot of potential to be a cost-effective and intuitive alternative to both digital and physical scale 3D models. In order for this to happen, we require more intuitive tools such as the Perspective Block, which will allow for remote collaborative work to be conducted more easily. For example, instead of collaborators trying to describe vocally what they are seeing on an on-screen CAD model, they can instead be constantly aware of each other's perspective and orientation in their own 3D space with no additional effort, almost as if they were working co-located in the same room. The interaction techniques investigated here will help explore ways in how AR can be used in these types of remote collaboration situations, as well as pave the way towards future technologies involving full 3D projections models. Future work can take the interaction techniques explored here and apply it to more context specific situations.

7. CONCLUSION

In this paper, we introduced the Perspective Block, a novel interaction tool to help aid in remote collaboration. By using AR, we provided multiple interaction techniques that helped collaborators share their perspective with each other. These techniques include view pointers and video feeds that can provide important perspective and contextual information for collaborators as they work around a shared 3D object. We conducted an informal study session in order to gather feedback and rate the potential of this concept. Findings have shown that while there are numerous issues with the current prototype, the core ideas are on the right path and can be further explored given that the technological limitations can be overcome. The Perspective Block demonstrated the value of perspective sharing in remote collaborative work and provides a basis for future work to follow.

REFERENCES

1. Jeff Dyck, Carl Gutwin. Awareness in 3D Collaborative Workspaces, available from http://hci.usask.ca
2. Jeff Dyck, Cart Gutwin. 2002. Groupspace: A 3D Workspace Supporting User Awareness. In *Proceedings of Extended Abstracts on Human Factors in Computing Systems* (CHI 2002), 502-503. http://dx.doi.org/10.1145/506443.506450
3. Shen Fangyang, Jia Jun, Lu Xuejun, and Qi Yue. 2010. 3D modeling and augmented reality. In *Proceedings of the Universal Communication Symposium* (IUCS 2010), 185-192. http://dx.doi.org/10.1109/IUCS.2010.5666644
4. Mengu Sukan, Carmine Elvezio, Ohan Oda, Steven Feiner, and Barbara Tversky. 2014. Parafrustum: Visualization techniques for guiding a user to a constrained set of viewing positions and orientations. In *Proceedings of the ACM symposium on User interface software and technology* (UIST '14), 331-340. http://dx.doi.org/10.1145/2642918.2647417
5. Mengu Sekan, Steven Feiner, Barbara Tversky, Semih Energin. 2012. Quick viewpoint switching for manipulating virtual objects in hand-held augmented reality using stored snapshots. In *Proceedings of the IEEE symposium on Mixed and Augmented Reality* (ISMAR '12), 217-226. http://dx.doi.org/10.1109/ISMAR.2012.6402560
6. M. Tait, M. Billinghurst. 2015. The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work*, 1-27.
7. Michael Tsang, George Fitzmaurice, Gord Kurtenbach, Azam Khan, and Bill Buxton. 2002. Boom Chameleon: Simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. In *Proceedings of the ACM symposium on User Interface Software & Technology* (UIST '02), 111-120. http://dx.doi.org/10.1145/571985.572001
8. Qualcomm Vuforia. Retrieved January 22, 2016 from https://www.qualcomm.com/products/vuforia