

A View Direction-Driven Approach for Automatic Room Mapping in Mixed Reality

Dong Jun, Kim
University of South Florida, USA

Wanwan, Li
University of South Florida, USA



Figure 1: Before and after scene of the demo application made with the proposed method. The figure on the left shows an indoor space before using our proposed method. The middle and right figure shows the different applications of what users can do in the MR environment created using our proposed method.

ABSTRACT

Virtual Reality and Augmented Reality technologies have greatly improved recently, and developers are trying to make the experience as realistic as possible and close the gap between the physical world and the virtual world. In this paper, we propose an efficient and intuitive method to create an immersive Mixed Reality environment by automatically mapping your room. Our method is view direction driven, which allows the users to simply “look at” any indoor space to create a 3-dimensional model of the area the user is located in. This approach is easier and more intuitive for the users to use and reduces the time and effort compared to other MR environment generating methods. We use the Meta Quest 2’s cameras and gyroscope sensor and the Unity engine for the ray casting and the passthrough API. We will present the mathematical details of our method and show that the proposed method achieves better results than previous methods through the user study results.

CCS CONCEPTS

• **Human-Computer Interaction**; • **Room Mapping**; • **Mixed Reality**;

KEYWORDS

View Detection, Ray Casting, Corner Selection

ACM Reference Format:

Dong Jun, Kim and Wanwan, Li. 2023. A View Direction-Driven Approach for Automatic Room Mapping in Mixed Reality. In *2023 5th International Conference on Image Processing and Machine Vision (IPMV) (IPMV 2023)*, January 13–15, 2023, Macau, China. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3582177.3582183>

1 INTRODUCTION

An immersive virtual environment has been the goal of a lot of VR (Virtual Reality) researchers for many years. The immersive virtual environment’s goal is to let the user experience a computer-generated world as if it were real [1]. As VR technologies including 3D visual, auditory, haptic displays, and position tracking systems improved, VR researchers have been successful in producing a sense of presence in the user’s mind [1]. Since then, immersive experience has been a huge part of virtual reality technology. It has been used in many different fields such as military training [1], tourism [2], education [3], and the entertainment/gaming industry [4].

Even though VR researchers were able to provide immersive experiences using visual and auditory technologies, VR has its limitations as the environment created in VR is fully virtual [5] and therefore the user cannot touch or feel the object. There is no interaction between the VR environment and the physical world. Some experts even criticize that VR lacks social interactions and features an isolated user [5]. Also, many VR users get injured trying to sit down on a virtual chair that doesn’t exist or walking into a wall.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IPMV 2023, January 13–15, 2023, Macau, China

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9792-6/23/01...\$15.00
<https://doi.org/10.1145/3582177.3582183>

AR (Augmented Reality) is another technology aiming to achieve an immersive virtual environment that took a very different approach to VR. AR creates the illusion that virtual images are seamlessly blended with the real world [6]. The mobile game Pokémon GO is one example as the game displays Pokémon, a virtual object, in your backyard, a physical environment. AR mixes the virtual environment into the physical world, but its limitation is similar to the VR's as the user cannot interact freely with the physical world [5].

MR or Mixed Reality can be viewed as one in which the real world and the virtual world are presented together within a single display [7]. This idea has been getting more popular among the people trying to create an immersive virtual environment, especially a fully immersive environment. A fully immersive virtual environment is an environment where the user cannot distinguish whether they are in a virtual or a physical environment. A fully immersive virtual environment cannot be achieved using VR or AR as VR is fully virtual and AR is displaying a virtual object on top of a physical world. Mixed reality, however, is a combination of VR and AR, and theoretically, in a carefully designed MR environment, the user can experience full immersion.

One of the main challenges of creating an MR environment is mapping the room. Room mapping is a process of recreating the entire room so that you can create a 3D model of the room. This model can be displayed so that it overlaps the physical room. By changing the color of this model or allowing the model to collide with other virtual objects, it allows the user to experience the color of their room changing or a virtual object interacting with their room.

In this paper, we propose an easy, efficient view direction-driven method to map the room so that users can quickly and intuitively map their room, creating an MR environment wherever they are located. This is done using one of the most popular and the cheapest VR headset, the Meta Quest 2. Although the headset was made for VR, its cameras and gyroscope sensor allow MR features. This method has been implemented using Unity version 2021 and the passthrough API provided by meta. With the Unity engine, we can use ray casting to get the location of the corners. The passthrough feature is a technology that allows the users to use the cameras on their VR headset to see and interact with the real world while still wearing the headset [8].

The process we propose is automatic, meaning that the users just have to “look at” the corner rather than having to go through complicated setups or without even having to use their controllers. This view direction-driven approach can significantly reduce the time and effort of the users when they are creating an MR environment. With the MR environment created, the users can interact with virtual objects and decorate their rooms. Our method will provide a base step for users and developers as they can add endless features to this MR environment. With these features, users can play racquetball in their living room or change the color of their bedroom wall instead of imagining what it would look like. As you can see from figure 1, adding a painting to the wall or changing the color of the room creates realistic results. Contributions of our work include:

- Creating a more immersive MR environment compared to VR and AR environments.
- Mapping the user's room so that the user can interact with their room.
- Automating and simplifying the process so the user intuitively learns how to map a room just by looking at the corners
- Providing a better, easier solution than previous efforts to map rooms. For more details, please refer to this video: <https://www.youtube.com/watch?v=u6Lzeq8VRI>

2 RELATED WORKS

Most of the current studies combining the physical world and the virtual world are based on AR. Many of the studies are also based on object detection. Companies such as IKEA, Nike, and Amazon have already implemented AR to provide a realistic experience for their products [9] and a location-based AR game Pokémon GO became the most successful mobile game in both popularity and revenue generation at the time of its launch in July 2016 [10].

Reference [11] addresses the issues with marker-based AR tracking such as unstable and mismatch problems and proposes a better deep learning-based mobile AR tracking to conduct 3D spatial mapping without AR markers from single snap-shot-based RGB-D data. The study identifies object outlines by detecting objects and segments their regions in the RGB image. Then it matches the virtual 3D model onto the real object using an iterative closest point (ICP) method. Reference [12] proposes a method to inpaint a depth image using depth sensors on head-mounted displays (HMDs). The method reads data from an RGB-D sensor and then reduces the spike noise, which is usually included in a raw depth image, using the median filter. Then with the filtered image, the depth shape is reconstructed by extracting and filtering the edges of the object. Finally, smoothing is done on the image based on non-local means. This study successfully detects objects in a scene and can distinguish the border of different objects. Reference [13] presents a method that instantly recovers the lighting of objects in an MR environment using a Kinect camera and an inexpensive RGB-D camera. This method has successfully scanned objects instantly without any special setups such as using a green screen. The method scans the environment and then segments the object based on materials. Then by using inverse rendering it generates appropriate lighting on the scanned objects. Reference [14] proposes methods for providing an accurate real-time occlusion for MR. To achieve this, the authors scan an object using an RGB-D camera, then reconstruct a model based on the depth image captured. Reference [15] shows how real-time spatial mapping can be used to map human activity. The experiment maps human activity in real-time using RGB-D sensors and using the data, makes a robot mimic human motion in real-time.

Augmented reality has its limitations, however, as it can only position static information in space [16] or animated sequences [10] and thus lacks the physical interaction between the virtual information displayed and the physical world. Mixed Reality on the other hand is a fusion between virtual and real objects that allows interaction between the virtual and the real world [17]. In 2020, reference [18] showed how a robot can utilize a LiDAR sensor to

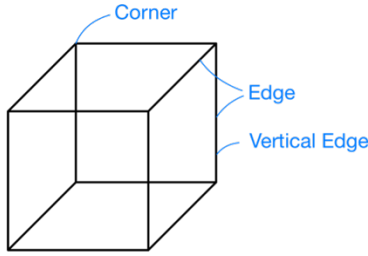


Figure 2: Edge and Corner Terminology

map the room in 2D. The authors showed that the technology can be useful in automatic navigation or object searching. However, this study was only focused on 2-dimensional room mapping and wasn't suitable for creating an MR environment. Also, the Meta Quest 2 doesn't have a LiDAR sensor, and proposing a method that doesn't require a specific sensor would allow our method to be used in any device with a camera and a gyroscope. Then in 2021, reference [19] proposed a method for indoor 3D mapping and modeling using LiDAR data. They used existing public datasets to achieve an accurate mapping of the environment with a single LiDAR scan. This method could be very useful in some cases, but our goal was to create an accurate mapping of an environment just by using the data we can obtain ourselves using the camera and the gyroscope.

In 2021, reference [20] showed that the creation of an MR environment is possible using Meta's passthrough API. This was the first high-quality room mapping application that created an MR environment, but the setup was complicated, and it required the user to manually enter the location of the edges and corners of the room. This application also involves creating unnecessary objects and requiring additional steps.

Using our view direction-driven approach can provide a method that is much simpler and doesn't use additional datasets or sensors. The details of the method are discussed in the next section.

3 TECHNICAL APPROACH

Figure 3 shows the overview of our technical approach for automatically mapping the room in Mixed Reality. We tried to find the most efficient and simple method to map the room. To make it intuitive for people to use this method, just "looking at" the corner for a few seconds automatically finds the corner without having to click the buttons or use the controllers at all. To avoid confusion regarding the terminology between edge and corner, we will use the terms from figure 2.

There are 2 conditions needed for this method to work. First, there must be at least 1 vertical edge of the room where the bottom corner and the top corner are both visible. Second, there must be at least 1 visible corner for every vertical edge of the wall. For example, for a basic cube-shaped room with 4 edges, one of the possible layouts could be 3 top corners and 2 bottom corners including the corners of all its 4 edges. Another possibility for this room could be 1 top corner and 4 bottom corners. With these 2 conditions, we can obtain the exact shape of the room. The method for getting the exact dimensions of the room is the following.

First, we aim to find the height of the room. To do this, we start with the edge where both corners are visible. We'll let g be the ground plane on which the user is standing above. Then the user "looks at" the bottom corner for a couple of seconds. The application counts down 3 seconds until it locks the bottom corner. If the user moves their head the counter restarts.

The software gets the position of that corner by using ray casting. By shooting a ray from the camera and finding the intersection of the ray with plane g , we can lock the position of the bottom corner and save the location to the memory until the program is terminated. By obtaining this position, we can get the distance between the corner and the user's feet. We will call that distance d . The ray casting is done through the equation below.

$$d = \mathbf{n} \cdot (\mathbf{x} - \mathbf{c})$$

Then, the user now "looks at" the top corner of the same edge. Using the Quest 2's gyroscope, we get the angle the user is viewing. Let's call this angle θ . We will also get the height of the user or more accurately, the height of the Quest 2. We'll call that h . Now with all the values we've obtained so far, we can calculate the height of the room, L . The equation is shown below.

$$L = d \tan(\theta) + h$$

$d \tan(\theta)$ would be the distance from the camera to the ceiling. Therefore, adding the camera's height to that value would be the height of the room, L .

Finally, now that we got the height of the room, we know the location of the first edge and the equation of the ground g and the ceiling plane c , each being $y = 0$ and $y = L$. From this point, all we need to do is find the location of the remaining vertical edges of the room. Because all the remaining corners are on top of the plane g or c , as long as there is a visible corner in a vertical edge, we can get the position of the entire edge. To do this, like in the first step, the user "looks at" an empty corner for a few seconds. After obtaining the position of all the corners, we can create a 3d model for this room. Displaying this 3d model so that it overlaps with the physical room allows us to create our immersive MR environment.

In this MR environment, endless features can be added such as adding shaders to the wall or allowing virtual objects to interact with the room. A simple example of real-life use of this environment could be placing posters on your wall or changing the color of your wall to imagine what your room would look like. As you can see from figure 1, you can add photo frames or change the wall color and be standing in your room to see what your room would look like. This is far better than imagining what your room would look like or using AR applications as our MR environment allows you to walk around your living room and experience it. You could also be playing racquetball or other indoor activities right in your living room. With a slight modification and by adding table/desk mapping, playing table tennis with your friends in your kitchen table is also possible.

4 NUMERICAL EXPERIMENT RESULTS

Figure 3 demonstrates the steps the user needs to take to create an MR environment using the demo application. First, the user aligns the cross to a bottom corner and stays still for 3 seconds (1). This allows the software to remember the location of the first

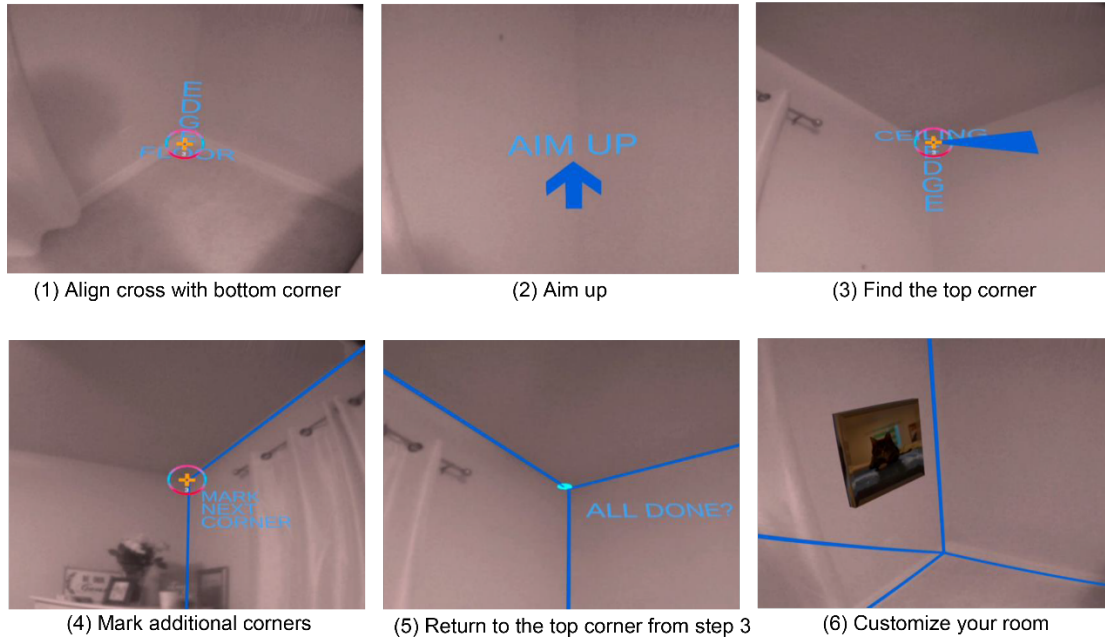


Figure 3: Instructions when using the demo application

bottom corner. Next, the user Moving the head more than 1 degree (measured with the gyroscope) would restart the timer.

Next, the user aims their head up towards the top corner of the same edge (2). The program guides the direction the user needs to move their head with an arrow. Then, the timer reappears allowing the user to find the top corner by waiting 3 seconds again (3).

Then, the user finds all the other corners of the room by repeating similar processes as steps 1 and 3 (4). The process ends when the user connects the last corner to the first corner (5). This will create an immersive MR environment.

After creating the environment there are endless possibilities of what you could do in this environment. In this demo, we tried putting photo frames on the wall and changing the color of the room (6).

5 USER STUDY DESIGN

We will conduct a user study to test the users' experience and the accuracy of the method. We recruited 15 people to test the room mapping application we made using our proposed method. These 15 people had no previous experience with any kind of VR or MR environment.

The user study will be comparing our application to Bob Berkebille's work [20] mentioned in the Related Works section as his application is the closest work to our application. We asked the 15 users to create an MR environment using our application and Berkebille's application. We recorded the time it took for the users to complete the room mapping process.

Then we surveyed how they felt and how convenient the proposed method was compared the Berkebille's work. The questions in the questionnaire include: 1. Do you agree that the proposed

method accurately maps the room? 2. Do you agree that the method was intuitive and easy to use? 3. Do you agree that the proposed method maps the room more accurately than Berkebille's work? 4. Do you agree that the method was more intuitive and easier to use compared to Berkebille's work? 5. Do you agree that the method creates an immersive, realistic environment? These questions asked the users to select a number between 1 and 7 where 1 stand for disagreeing and 7 stands for agreeing.

Additionally, to verify that the dimension of our room model accurately represents the actual room, we will compare the height of room L and the distance between the user and the vertical edge d calculated by our code to the actual distance that we measure. We will also get the difference in distances between the actual corner and the generated corner. This is done by modifying the program so that we can enter the actual location of the corner. This will allow us to know the average error when it comes to our MR environment's size and location.

6 USER STUDY RESULTS

As seen in figure 4, the results from the user study were as follows: (1): $M=6.2$ $SD=0.77$ (2): $M=6.6$, $SD=0.63$ (3): $M=3.93$, $SD=1.33$ (4) $M=6.2$, $SD=0.86$ (5): $M=5.73$, $SD=1.16$ From these results, we concluded that the users most agree that the proposed method accurately maps the room ($M=6.2$), is intuitive and easy to use ($M=6.6$), more intuitive and easier to use compared to Berkebille's work ($M=6.2$), and that the method creates an immersive, realistic environment ($M=5.73$). However, the user most disagreed that the proposed methods map the room more accurately than Berkebille's work ($M=3.93$).

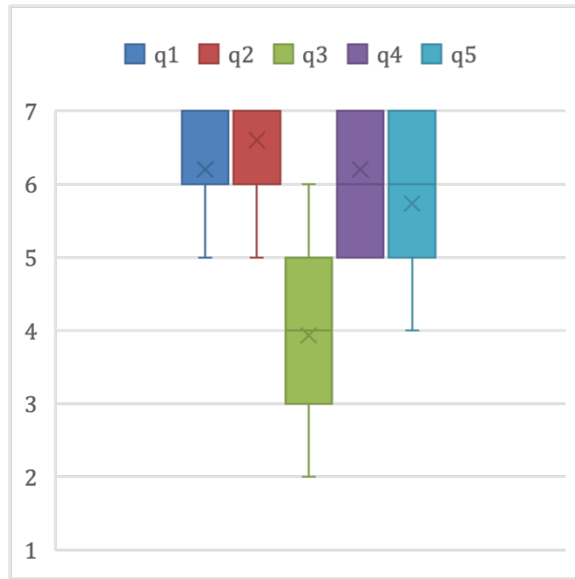


Figure 4: User survey results

In addition to the survey, we recorded the time that the user took to complete the room mapping using both methods. Our method took 54.73 seconds on average while Berkebile's method took 59.27 seconds which is an improvement of 8.3%. We also measured the distance between the actual corner and the generated corner. The mean distance was 0.14 meters.

From these results, we can conclude that our view direction-driven approach produces an MR environment faster and more accurately compared to other methods.

7 CONCLUSION

In this paper, we proposed a view direction-driven approach for automatic room mapping in Mixed Reality to create an immersive MR environment. The user study results show that the proposed method can get better results than previous methods.

In future work, we will simplify the process even more by using image processing and deep learning to support the current view direction-driven approach. Detecting the corner of the room using image processing could simplify the step where the user has to look at the corner. In the end, our goal is to be able to automatically map any indoor space in real-time.

REFERENCES

- [1] Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough?. *Computer*, 40(7), 36-43.
- [2] Guttentag, D. A. (2010). Virtual reality: Applications and implications for tourism. *Tourism management*, 31(5), 637-651.
- [3] Freina, L., & Ott, M. (2015, April). A literature review on immersive virtual reality in education: state of the art and perspectives. In *The international scientific conference elearning and software for education* (Vol. 1, No. 133, pp. 10-1007).
- [4] Cruz-Neira, C., Fernández, M., & Portalés, C. (2018). Virtual reality and games. *Multimodal Technologies and Interaction*, 2(1), 8.
- [5] Speicher, M., Hall, B. D., & Nebeling, M. (2019, May). What is mixed reality?. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).
- [6] Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality.
- [7] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE Trans. Information and Systems* 77, 12 (1994), 1321-1329.
- [8] Chaurasia, G., Nieuwoudt, A., Ichim, A. E., Szeliski, R., & Sorkine-Hornung, A. (2020). Passthrough+ real-time stereoscopic view synthesis for mobile mixed reality. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 3(1), 1-17.
- [9] McLean, G., & Wilson, A. (2019). Shopping in the digital world: Examining customer engagement through augmented reality mobile applications. *Computers in Human Behavior*, 101, 210-224.
- [10] Paavilainen, J., Korhonen, H., Alha, K., Stenros, J., Koskinen, E., & Mayra, F. (2017, May). The Pokémon GO experience: A location-based augmented reality mobile game goes mainstream. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 2493-2498).
- [11] Park, K. B., Choi, S. H., Kim, M., & Lee, J. Y. (2020). Deep learning-based mobile augmented reality for task assistance using 3D spatial mapping and snapshot-based RGB-D data. *Computers & Industrial Engineering*, 146, 106585.
- [12] Awano, N. (2020). Approximate Depth Shape Reconstruction for RGB-D Images Captured from HMDs for Mixed Reality Applications. *Journal of Imaging*, 6(3), 11.
- [13] Richter-Trummer, T., Kalkofen, D., Park, J., & Schmalstieg, D. (2016, September). Instant mixed reality lighting from casual scanning. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 27-36). IEEE.
- [14] Walton, D. R., & Steed, A. (2017, November). Accurate real-time occlusion for mixed reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (pp. 1-10).
- [15] Howard, A. M., Roberts, L., Garcia, S., & Quarells, R. (2012, November). Using mixed reality to map human exercise demonstrations to a robot exercise coach. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 291-292). IEEE.
- [16] Fedorov, R., Frajberg, D., & Fraternali, P. (2016, June). A framework for outdoor mobile augmented reality and its application to mountain peak detection. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (pp. 281-301). Springer, Cham.
- [17] Badías, A., González, D., Alfaro, I., Chinesta, F., & Cueto, E. (2020). Real-time interaction of virtual and physical objects in mixed reality applications. *International Journal for Numerical Methods in Engineering*, 121(17), 3849-3868.
- [18] Rivai, M., Hutabarat, D., & Nafis, Z. M. J. (2020). 2D mapping using omnidirectional mobile robot equipped with LiDAR. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(3), 1467-1474.
- [19] Wen, C., Tan, J., Li, F., Wu, C., Lin, Y., Wang, Z., & Wang, C. (2021). Cooperative indoor 3D mapping and modeling using LiDAR data. *Information Sciences*, 574, 192-209.
- [20] Hamilton, I. (2021, October 24). RoomMapper for Quest 2 maps surroundings for mixed reality. *RoomMapper For Quest 2 Maps Your Surroundings For Mixed Reality*. Retrieved March 1, 2022, from <https://uploadvr.com/quest-2-room-mapper/>