SimVis – A Portable Framework for Simulating Virtual Environments

Timothy Parsons†
Brown University

ABSTRACT

We introduce a portable, generalizable, and accessible open-source framework (SimVis) for performing a large variety of Mixed Reality (MR) research. Using the framework, we conduct a user study whose results explore the differences between Brown's CAVE and Brown's YURT in performing insight-based tasks on volume data. The results of our user study (and future experiments using our framework) aim to help answer the question 'is MR Simulation a valid alternative for experimentation?'

Keywords: Virtual Reality, Mixed Reality Simulation

1 INTRODUCTION

The term 'Mixed Reality Simulation' (MRS) doesn't really show up in academic literature until around 2013. MRS is the field in which lower-fidelity virtual environments are simulated using higher-fidelity virtual environments. Basically, a virtual environment has many different objective measures of immersion – as identified by Slater et al. 2003. These objective measures of immersion include (and are not limited to) Field of View, Field of Regard, resolution, stereo/no stereo, brightness, etc. [1] A specific virtual environment is considered 'simulated' when all of these objective measures of immersion have been reproduced. This implies that the simulated environment should have less fidelity than the environment in which it is simulated.

There are a few sources of motivation behind this field. The first is that the field of virtual reality is still pretty young, and so there is still much we do not understand. We don't know exactly which aspects (i.e. which objective levels of immersion) of virtual environments actually affect a user's performance while performing an experiment or task. It is useful to know this when selecting an environment in which to perform an experiment. For instance, there is data (such as in Brooks et. al's paper Walking < Walking-In-Place < Flying) that gives support to the idea that fully immersive proprioception (movement that mimics reality, i.e. actually walking around while in the virtual environment versus navigating the scene by flying using a joystick) increases a user's performance in navigational tasks [5]. This data suggests that a research team performing navigational experiments should choose an environment that allows for full range of motion (i.e. a head-mounted display) to gather the best results.

One way to test which variables improve performance in different kinds of tasks would be to reproduce experiments in different virtual environments around the world. there are a wide variety of such virtual environments. Here at Brown we have two Caves – one built in 1998, and one that is in the act of being finished in 2015. The older of the two Caves is an 8'x8'x8' three-walled cube that contains 7 projectors – two for each wall and one for the floor. The new Cave – dubbed the 'Yurt', has 69 projectors with increased brightness and resolution that is comparable to that of the human eye. It is also a much bigger space and has curved walls instead of flat walls to increase the level of immersion for the user. There is the Reality Deck 2 at Stony Brook University in Long Island, NY – a 1.5 gigapixel display of ~500 high-definition monitors surrounding the user. [2] There is the StarCAVE at UC San Diego that surrounds the user in a pentagon shape with resolution that approximates 20/40 vision for the user [3]. There are also virtual environments across the ocean – one notable such environment is the Cave system currently at Kaust in Saudi Arabia.

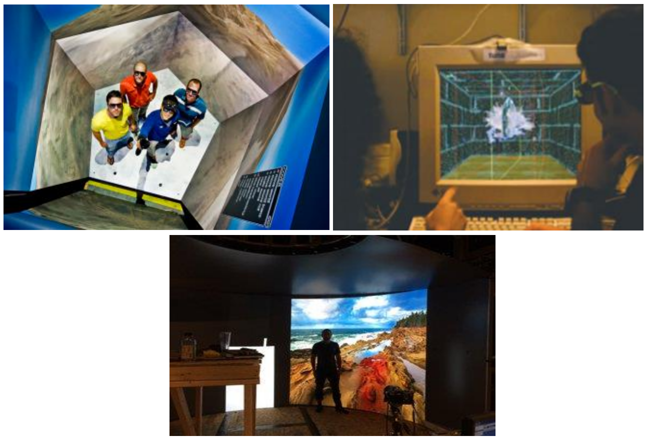


Figure 1: From left to right: StarCAVE at UC San Diego, Yurt at Brown University, Fish Tank apparatus

Reproducing experiments across great distances like this while keeping all the variables related to the experiment consistent is very difficult to achieve. The alternative to this would be to have many virtual environments with different objective measures of immersion in the same place. However, building virtual environments can be very expensive (state-of-the-art Caves cost millions at this point) and can take a lot of time to build, which is not financially feasible. Simulating virtual environments would allow for fine control of objective measures of immersion and the variables involved in the experiment, which is why we are interested in this field of study.

Currently, being a relatively young field, there is not enough evidence supporting Mixed Reality Simulation to even show whether it is a valid field of study and a valid alternative to physically using different virtual environments. Our framework's contributions are two-fold: 1) it gives an intuitive, portable, generalizable way to perform Mixed Reality Simulation, and 2) results from this framework will help give evidence toward the validity (or invalidity) of Mixed Reality Simulation.

2 DESIGN

The way in which we perform our Mixed Reality Simulation is to simulate all aspects of an existing Virtual Environment. For instance, if we want to simulate a CAVE$^{\text{TM}}$ with a front-facing wall, two-side facing walls, a clear line where the projectors meet, half of a floor, with a certain level of brightness and resolution, we can model all of this based on a configuration file and some 3D models created in any 3D modeling program. This is a description of the 1998 CAVE$^{\text{TM}}$ at Brown University. The explicit modeling of the Virtual Environment will implicitly and correctly model all of the objective measures of immersion as explained above.

We use VRG3D to implement our simulation framework. VRG3D is a portable framework for creating applications in any kind of virtual environment. Applications written in VRG3D can

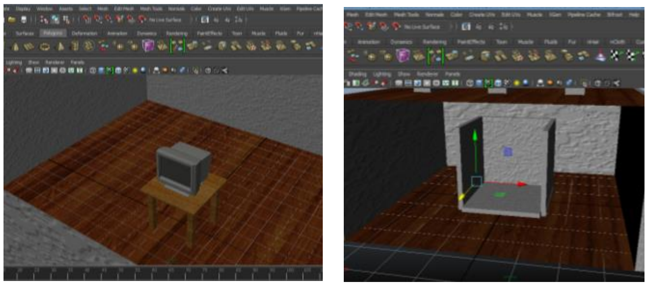


Figure 2: Examples of modelling different virtual environments.

theoretically be ported from one environment to another after some configuration is done. VRG3D has successfully been ported to both CAVE^{TM}s at Brown University. Our framework takes in a VRG3D program and runs that program within the simulation of the Virtual Environment. Because our simulation framework is built on top of VRG3D, it also is portable to any other virtual environment.

Our framework allows for the simulation of very specific Virtual Environments, which will allow for easier comparative testing between different environments around the world (i.e. StarCAVE at U.C. San Diego [2], etc.). Running experiments in these different places and keeping the variables consistent is tough, but it would be useful to know which existing environments are better suited for certain tasks/experiments. Our framework also theoretically allows for simulation of non-existent virtual environments, which allows for easier testing of these environments in that they don't have to be constructed beforehand.

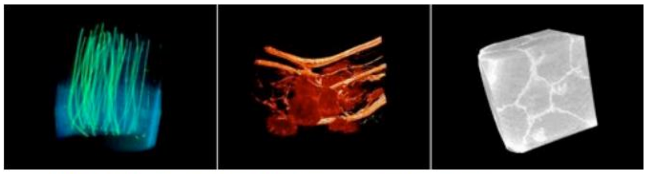
To simulate a specific virtual environment, a user would create 3D model of that virtual environment and export the model as a series of .obj files with textures associated with them. There are a few configuration files that will identify which .obj files are background objects (i.e. the wall of a room that you could see when turning around 180 degrees in a 3-sided CAVE^{TM}) that will always draw over whatever is being rendered in virtual reality, and the objects that contain whatever will be drawn in virtual reality (i.e. the screens of the CAVE^{TM}). Utilizing the Stencil Buffer in OpenGL, we can tell our program exactly where to render the 3D world in our simulated environment. Utilizing a frame buffer and performing some processing on a rendered 2-dimensional texture of the scene, we can control the brightness and resolution to match that of the simulated environment.

Our framework also allows for fine control over the objective levels of immersion we found to be most important – Field of View, Field of Regard, brightness, and resolution. The specific levels of these are indicated as fractional when compared to the environment where the simulation is taking place, and can be inputted via a configuration file. This makes for more easily-constructed experiments to test the effects of these specific measures of immersion.

To help verify whether this is a valid simulation, we have chosen an existing virtual environment (Brown's CAVE^{TM}), and we have chosen a higher-fidelity Virtual Environment (Brown's YURT) in which to simulate the Cave. We run a task-based experiment in these environments as well as in the simulation of the Cave and compare the results. We hypothesize that the higher-fidelity Yurt will have greater performance by users, and that performance in the Cave and the simulated Cave will be statistically similar.

2.1 Related Work

As mentioned previously in this paper, the field of MRS is relatively young, and the term Mixed Reality Simulation was not even used in this context until Cha Lee et al. 2013 [4]. In their work at UC Santa Barbara, Cha Lee et al. explore the effects of visual realism on navigational/search tasks in a simulated environment. The results of the study were mixed – no large effects were found regarding visual realism, but they do note that it is difficult to achieve realism close to the real world. They also explore the effect of differences in latency between an actual virtual environment and the simulated virtual environment and conclude that when the difference is small/negligible, there is no significant effect on performance and results.



(a) 3D Scaffold dataset (b) Mouse Limb dataset (c) Fossil dataset

Figure 3: datasets from Laha's PhD thesis experiment

Another MRS experiment was done in 2014 by Bireswar Laha in his PhD dissertation [7]. In this, Laha simulates a CAVE^{TM}-like system using a Head-Mounted display (HMD). The experiment he carried out involved qualitative (describing features) and quantitative (counting features) tasks on several different datasets. The results for this study were mixed – Laha notes that when the virtual environments are too dissimilar, simulation might not be feasible or helpful. In this case, the environments are rather dissimilar – a HMD engulfs the field of view and the user cannot see one's own limbs or other parts of his or her body. In a CAVE^{TM}, the user can see his or her own limbs. Laha notes this is likely the cause of inconsistencies in his data (See Figure 3).

(a) *Pterostichus* dataset (used for training) (b) *Platynus* dataset (used in the main study)

Figure 4: Laha et. al. 2014 isosurface renderings of beetle trachea.

Our experiment draws largely on an experiment done by Laha et al. at Virginia Tech [6]. In this experiment, Laha et al. simulated lower fidelity VR Environments in a CAVE^{TM} by reproducing different levels of Field of Regard (90, 180, and 270 degrees), Head Tracking (yes or no), and stereoscopy (yes or no). To compare these different simulations, they conducted a user study where each user performed insight-based tasks on isosurface renderings of beetle trachea and studied both qualitative and quantitative results. We do not reproduce their exact experiment, but we will be borrowing from them their insight-based tasks to compare several different VR Environments, including a simulation of one.

2.2 Procedure

Before the experiment begins, the user signs a consent form and fills out a questionnaire about prior experience in several different areas, including video games, interpreting volume datasets, and virtual reality. The user is then given a 3-minute timed spatial reasoning test where the user identifies whether two children's blocks could potentially be the same block or must be different given a view of three of the sides. If the user fails this test, we do not use their data. To pass the test, the user must score more than a 4 (grading scheme is number right minus number wrong). We had 3 users fail this test – one in each environment (just as a note, we also noticed no correlation between performance on this exam and performance in the tasks). This is all borrowed from Laha et al. 2014 [6].

At this point, we explain the experiment to them and load up a training dataset. The datasets are isosurface renderings of beetle trachea obtained using micro-CT scans. The user is given some time to become familiar with moving around and navigating the dataset – the controls for this are through a combination of joystick controls and button pushes on a wand that gives six degrees of freedom (three translational and three rotational). To help them get more familiar with navigating the data and with the sorts of tasks we will be having them complete, we give them five tasks on the training set – one from each category of task we will be testing them on. The categories of these tasks are as follows: search, pattern recognition, spatial judgment, quantitative estimation, and shape description.

Figure 5: A user explores the dataset in the Yurt

At this point, the test set is loaded up and the user is given fifteen randomly-permuted insight-based tasks. These tasks can be found in the appendix, and are taken directly from Laha et al. with their permission [6]. They are tasks that actual scientists in the field would perform, but they have been stripped of their technical terms to make them understandable to our non-expert users.Before each task, the beetle's position is reset and the user is read the task. For each training and testing task, we record the amount of time the user takes to complete the task, a quantitative score of their perceived difficulty of the task, a quantitative score of their perceived confidence in their answer, and a qualitative score of their result, to which we give a grade from 0 to 1.

We run the experiment in three different environments – Brown's 1998 CAVE™ (referenced to as 'the Cave' for the rest of this paper), Brown's new state-of-the-art CAVE™ (referenced to as 'the Yurt' for the rest of this paper), and in a simulation of the Cave, which takes place in the Yurt. The simulation is created using the framework discussed earlier in this paper.

3 DISCUSSION

Environment	Average Grade	Average Time	Average Difficulty	Average Confidence
Cave	.6837	95.017	3.9167	3.7842
Yurt	.6055	123.2	4.2167	3.7833
Cave Simulation	.5847	102.5	4.1333	4.0000

Table 1: Average Qualitative and Quantitative Scores

A total of 12 users were run in this experiment – four in the Cave, four in the Yurt, and 4 in the simulation of the Cave in the Yurt (for the rest of this paper, we will refer to this as 'the simulation of the Cave' or 'Cave Simulation'). The users were four female and seven male undergraduate students from Brown University and one male graduate student from Drexel University from various majors (ranging from Computer Science to Sociology to Public Health).

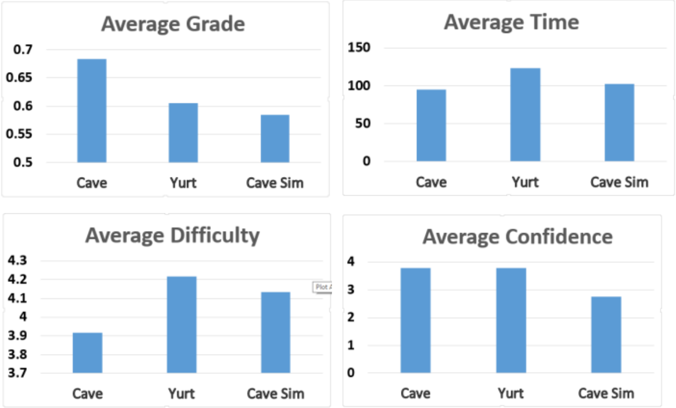


Figure 6: Averages in all the environments of grade, time per task, difficulty, and confidence.

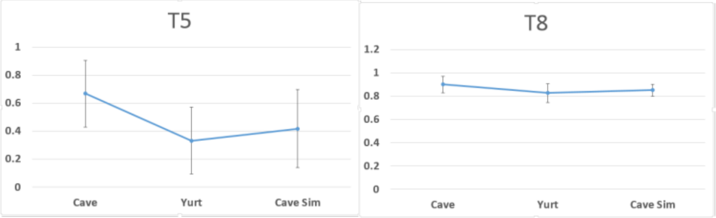


Figure 7: Task 5 and Task 6 Grades with Standard Deviation Bars. This shows the Cave outperforming the Yurt with no statistically significant measure of the Cave Simulation being similar to the Cave.

While we were unable to achieve statistical significance with the number of users run, we can talk about trends in the data that we do have. The results for Average Grade, Time, Confidence, and Difficulty are all somewhat surprising. The lower-fidelity Cave scores higher in all categories, with the Yurt and the Cave Simulation trading off between the last two performance spots. We will now comment on each of these.

The average grade for the Cave was .6837, the average grade for the Yurt was .6055, and the average grade for the Cave Simulation was .5847. There are two main observations we make from this – the users in the lower-fidelity Cave had higher performance than in the higher-fidelity Yurt, and the users in the Cave Simulation had significantly lower grades than in the actual Cave. One of the main reasons we think this occurred is because the floor is not currently functioning in the Yurt, and the floor does work in the Cave. This has two implications: 1) the Cave cannot be well simulated in the Yurt because it is missing the floor, and 2) the Yurt actually has lower fidelity in vertical field of regard than the Cave. As we ran the users, we noticed that the users in the Cave utilized the floor quite a lot, which could account for this unexpected result.

The average time taken in the Cave per task was 95.017 seconds, the average time taken in the Yurt per task was 123.2 seconds, and the average time taken in the Cave Simulation per task was 102.5. For the time taken, the data for the Cave Simulation is more similar to the Cave than the Yurt, and the Yurt clearly took more time. We have two observations we want to make on this: 1) we noticed that some people took a lot more time than others in general, and because of the wide variance between users we think a within-user study would be more effective in measuring how these different environments affect time taken to perform tasks, and 2) the users in the Yurt were able to walk around the data, which they tried to do often (we had one user that consistently crouched and moved from side to side of the Yurt, which is a range of mobility not available to the user in the Cave). The users in the Cave (and in the Cave Simulation) did not have as much room to move around, and so they were forced to manipulate the data much more than the users in the Yurt. We hypothesize that manipulating and moving the data around (and just in general having less ways to try and view the data) brought users to a conclusion faster.

The average difficulty for the Cave was 3.9167, the average difficulty for the Yurt was 4.2167, and the average difficulty for the Cave Simulation was 4.1333. Here once again we see unexpectedly that users felt the tasks were more difficult in the Yurt than in the Cave, and no significant similarity between the Cave Simulation and the Cave. We again hypothesize that this may be because of the lack of floor in the Yurt.

The average confidence for the Cave was 3.7842, the average confidence for the Yurt was 3.7833, and the average confidence for the Cave Simulation was 4.000. There was a decently large variance in confidence scores between users even within the same virtual environment, which makes us believe that testing these values within users in different environments might be more helpful.

There are a few conclusions we draw from all of this. The first is that we need more users to obtain statistical significance. The second is that perhaps the Yurt is not suitable for simulating the Cave because it is too different from the Cave without a functioning floor. We also would like to note that it is rather difficult to achieve perfect realism (as noted by Lee et. al [4]), and so the simulation of the Cave could have been improved. For instance, the textures could have been higher resolution, and the room that the Cave actually is in could have been modeled instead of an arbitrary (and imaginary) room that we did choose to model the Cave in. The third is that because there was no floor in the Yurt, there was actually higher fidelity in some respects in the Cave. The Yurt is still higher fidelity in brightness, contrast, and resolution, but because the data did not have much color and the resolution was adequate in the Cave for the tasks given (there was never anything that needed extremely close inspection at smaller scales), performance was not improved due to these higher fidelity levels as we expected. We do expect that in experiments where discerning different colors and/or inspection of very small objects/details are important, users the Yurt will outperform users in the Cave. The last is that because of high variance between users in many of these data categories, we feel that a study within users in different environments would be more helpful.

We had each user fill out a post-experiment questionnaire, and we will briefly summarize the results from them. As supported by our hypothesis, users felt that they moved the dataset around more in the Cave and in the Cave Simulation (average score of 6 out of 7) than in the Yurt (average score 5 out of 7), and users also felt that they moved themselves around the data more in the Yurt (6 out of 7) than in the Cave (5 out of 7) and in the Cave Simulation (4.75).

We also asked each user if they felt sickness and/or discomfort on a scale from 1 to 7. We found that users in the Yurt felt less sick (average 1 out of 7) than in the Cave or the Cave Simulation (2.25 out of 7). We hypothesize this is because walking around the dataset causes less sickness than actually moving the dataset, which was done more with users in the environments where they felt sicker.

4 CONCLUSION

We have created a portable, generalized framework for Mixed Reality Simulation. We have allowed for the simulation of very specific virtual environments (both existent and currently non-existent), which is an extension on past work in this field. We have allowed for fine-tuned control of specific measures of immersion, these being Field of View, Field of Regard, brightness, and resolution. The framework is publicly available on GitHub (taparson/SimVis). We have also run a user study using this framework, and the results are mixed. We do however hope that our observations and hypotheses help guide experiments in this field in the future.

ACKNOWLEDGEMENTS

We would like to that Professor David Laidlaw for overseeing and advising this research, John Huffman and Thomas Sgouros for helping us get this study running in the different virtual environments, and Bireswar Laha, John Socha, Wesley Miller, and Johannes Novotny for lending permission and resources to run this experiment.

APPENDIX

T1. Air sacs are parts of the tracheal system that are balloon-like in shape, and are distinguished from tracheal tubes, which are cylindrical. Does this specimen possess any air sacs? If yes, how many? (Search, counting)

T2. Look at this circular object near the head of the animal. Is this connected to the surrounding tracheal tubes? If yes, then show the connection point. (Spatial Judgment)

T3. Scan the entire body. Find the tracheal tubes of the largest and smallest diameters. How many times bigger is the biggest tube than the smallest one? When you are done, please let me know - I will show you five options to choose from. (Quantitative Estimation). Options: 5, 15, 30, 50, 60.

T4. How many legs are there? Please identify each one. (Search, counting)

T5. This is a leg. The leg connects to the body at the bend. How many tracheal tubes connect the body to this leg? (Search)

T6. Find the tracheal tubes in the abdomen. Are there any tracheal tubes in the top half of the abdomen that definitively connect the left and right portions of the system? To qualify, the tracheal tube reaching across the body must connect to the other side; it can't end blindly in the abdomen. If yes, are there multiple locations? (Spatial Judgment)

T7. Most tracheal tubes are circular in cross section, or nearly circular. Do any tracheal tubes exhibit a decidedly non-circular cross-section? If so, where in the body are they located? (Shape Description)

T8. The spiracles are the oval-shaped regions that act as valves between the tracheal system and the external air. This is an example of a spiracle inside this beetle. How many spiracles can you find in this entire sample? Search both the left and right sides of the beetle. (Search)

T9. Does the number of spiracles on the left side match the number of spiracles on the right side? If not, what is the difference? (Search)

T10. The manifold is the part just below the spiracle, where the tracheal tubes join. For this spiracle (third one on the left side), how many tracheal tubes connect to the manifold? (Spatial Judgment)

T11. Examine the number of tracheal tubes entering the manifold of the spiracle 5 on both the left and right sides. Are they equal? If no, by what number are they different? (Spatial Judgment)

T12. Is there a spiracle that is connected to only one tracheal tube? If yes, which one is it? (Pattern Recognition)

T13. This is the spiracle-1. Now trace this tracheal tube towards the head, and count the number of times it branches. At each branching point, always choose the larger branch. (Spatial Judgment)

T14. Look at this tracheal tube in the abdomen region. Please trace this tube to its closest spiracle. Which spiracle is it? (Spatial Judgment)

T15. What region of the body appears to have the highest density of tracheal tubes, in a one cubic foot space? These are the regions I want you to look at. I will ask you to arrange these regions in terms of decreasing density of tracheal tubes, from highest to lowest. (Quantitative Estimation)

REFERENCES

[1] Mel Slater. A Note on Presence Terminology. Presence Connect, 3(3): 1-5, 2003

[2] C. Papadopoulos, K. Petkov, A. E. Kaufman and K. Mueller. “The Reality Deck – Immersive Gigapixel Display”. IEEE Computer Graphics and Applications. 35(1), pp 33-45, 2015.

[3] DeFanti, T. A., Dawe, G., Sandin, D. J., Schulze, J. P., Otto, P., Girado, J., & Rao, R. The StarCAVE, a third-generation CAVE and virtual reality OptIPortal. Future Generation Computer Systems, 25(2), 169-178, 2009

[4] Lee, Cha, et al. "The effects of visual realism on search tasks in mixed reality simulation." Visualization and Computer Graphics, IEEE Transactions on Visualization and Computer Graphics, 19(4): 547-556, 2013.

[5] Usoh, Martin, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks Jr. "Walking> walking-in-place> flying, in virtual environments." Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 359-364,1999.

[6] B. Laha, D. A. Bowman, and J. J. Socha. Effects of VR system fidelity on analyzing isosurface visualization of volume datasets. IEEE Transactions on Visualization and Computer Graphics, 20(4):513–522, 2014.

[7] Papadopoulos, Charilaos, Bireswar Laha, and Arie E. Kaufman. "Interacting with Mixed Reality Systems", 2014.