Measuring Occlusion Performance in Augmented Reality Training Systems

Michael Martin, PhD, Scott Johnson, John Baker,
Juan Castillo, Jaime Cisneros, Kiel Ewing
ML Horizons
Plantation, FL
mike.martin@mlhorizons.com,
scott.johnson@mlhorizons.com,
john.baker@mlhorizons.com,
juan.castillo@mlhorizons.com,
jaime.cisneros@mlhorizons.com,
kiel.ewing@mlhorizons.com

Patrick Garrity
U.S. Army Combat Capabilities Development
Command Soldier Center
Orlando, FL
patrick.j.garrity4.civ@mail.mil

ABSTRACT

Augmented Reality (AR) is rapidly maturing as a training solution and promises to provide the training community with unprecedented capabilities. However, there remain several key technical challenges that must be addressed to deliver credible and immersive experiences. At the Training & Simulation Industry Symposium (TSIS) 2019 and in subsequent public forums, the US Army's Simulation & Training Technology Center (STTC) identified dynamic occlusion as a key technological challenge critical to the successful implementation of AR applications and specifically critical to AR training.

Dynamic occlusion is the ability of an AR system to realistically integrate synthetic and real-world content. Successfully implemented dynamic occlusion allows moving real-world objects, such as people, to occlude virtual content in a credible and natural manner. For example, if a person walks in front of a digital/synthetic box, the person should occlude the box and the box should not be visible "through" the person. This concept is intuitive to observers, but there currently does not exist a commonly understood metric or industry standard for measuring dynamic occlusion performance.

In this paper, we propose a metric for assessing dynamic occlusion. We break dynamic occlusion down into constituent factors of false positive and false negative occlusion. Through this, we create a conceptual framework that can facilitate the establishment of requirements as well as objective comparisons of performance across various AR systems. We test this approach using a survey instrument to assess viewer acceptability of the false negative and false positive occlusion. Finally, we present our recommendation for a revised occlusion metric, based on the survey results, that allows occlusion solutions to be objectively compared in a manner that may better reflect their utility in AR training systems.

ABOUT THE AUTHORS

Michael Martin is a Principal and the Manager of ML Horizons. He is currently the Project Manager for ML Horizons' Augmented Reality Dismounted Soldier Training (ARDST) Phase III SBIR with the Combat Capability Development Center's Simulation & Training Technology Center. Mike is a retired US Army Lieutenant Colonel with a background in Armor, Cavalry, and Simulation Operations. He earned a B.S. in Computer Science from the United States Military Academy in 1997, an M.S. in Modeling, Virtual Environments, and Simulations from the Naval Postgraduate School in 2004, and a PhD in Modeling and Simulation from Old Dominion University in 2012.

Patrick Garrity is a Chief Engineer for the U.S. Army Combat Capabilities Development Command Soldier Center (CCDC SC). He currently works in Dismounted Soldier Simulation Technologies conducting research and development in the area of dismounted soldier training and simulation where he served as the Army's Science and Technology Manager for the Augmented Reality for Training Science and Technology Objective (STO). His current interests include Human-In-The-Loop (HITL) networked simulators, virtual and augmented reality, and immersive dismounted training applications. He earned his B.S. in Computer Engineering from the University of South Florida in 1985 and his M.S. in Simulation Systems from the University of Central Florida in 1994.

Scott Johnson is the Director of Software Development at ML Horizons. He is the lead author of an IITSEC paper that won Best Paper in Simulation in 2017. He shared the 2012 Modelling and Simulation Award from PEO STRI for his work as Lead Software Engineer on the Dismounted Soldier Training System (DSTS). Scott has 15 years of experience in defense and 10 years of experience in the video game industry where he worked primarily in the areas of Animation and Physics. He earned his Bachelor's degree in Electrical Engineering from Purdue University and his Masters in Computer Science and Engineering from the University of Michigan.

John Baker is a Senior Director of ML Horizons. John has 25 years of experience in defense, intelligence, and civil sectors. He has built a wide variety of systems - and, specifically, using virtual and augmented reality, ranging from small research projects to large programs of record. He pioneered the research and development behind the modern-day mixed reality-based wearable training systems – having honed in on high precision performance, low cost, and exceptional portability. He earned his B.S. in Industrial Engineering and Systems Management from the University of Central Florida in 1995 and his M.S. in Industrial Engineering, with a focus on artificial intelligence, from the Pennsylvania State University in 1998.

Juan Castillo is a multi-discipline Content Producer at ML Horizons. He has 10 years of experience in mixed reality, gaming, and graphics industry. In addition to that for the past 3 years, he has actively participated in the development and research of the ARDST system for the Army and the Navy. He earned a Bachelor's degree in Computer Animation from FullSail University.

Jaime Cisneros is a Principal Software Engineer at ML Horizons, a wholly owned subsidiary of Magic Leap, Inc. He has 27 years of experience in constructive, virtual and augmented reality simulation and training. Mr. Cisneros was the simulation lead for the Dismounted Soldier Training System (DSTS). As a researcher, he has designed and developed advance prototypes for a next generation virtual reality dismounted soldier trainer, as well as performed applied research in the areas of outdoor localization, human pose estimation, and their integration in an augmented reality dismounted soldier trainer. He earned a Bachelor's and Master's degrees in Computer Science from the University of Central Florida.

Kiel Ewing is a Software Engineer for ML Horizons. He currently works in research and development related to augmented reality and virtual reality systems applied to dismounted soldier training. He earned a Bachelor's of Science in Game Development from Full Sail University in 2007.

Measuring Occlusion Performance in Augmented Reality Training Systems

Michael Martin, PhD, Scott Johnson, John Baker, Juan Castillo, Jaime Cisneros, Kiel Ewing
ML Horizons
Plantation, FL
mike.martin@mlhorizons.com, scott.johnson@mlhorizons.com, john.baker@mlhorizons.com, juan.castillo@mlhorizons.com, jaime.cisneros@mlhorizons.com, kiel.ewing@mlhorizons.com

Patrick Garrity
U.S. Army Combat Capabilities Development Command Soldier Center
Orlando, FL
patrick.j.garrity4.civ@mail.mil

INTRODUCTION

For the past four years, ML Horizons has been working with the SFC Paul Ray Smith Simulation & Training Technology Center (STTC) on the Augmented Reality Dismounted Soldier Training (ARDST) Small Business Innovation Research (SBIR) topic area. The stated objective of the SBIR is to “Design and fabricate an integrated Augmented Reality system for use by Dismounted Soldiers that demonstrate high levels of immersion in live indoor and outdoor environments and demonstrate future interoperability in both single and multiplayer (collective) configurations with evolving Synthetic Training Environment (STE).” The SBIR is currently in Phase III, and continues to push the boundaries of the state-of-the-art. Because of this, there are many research topic areas in which there are no readily accepted industry-wide metrics available to objectively measure progress.

One area in which we have found this to be true is dynamic occlusion. When this research began, the focus was on the implementation of dynamic occlusion in the dismounted Soldier training use case that is central to our SBIR work. Over the past two years, however, we have progressed well beyond initial implementation, and are now dealing with the challenges of improving the performance of available solutions. As with any research, a key part of improving a system is having an objective metric to benchmark progress. Our initial assessment depended on a rote interpretation of what we termed “overlap error”. However, through implementation and experience, we developed an intuition that this initial attempt to quantify dynamic occlusion may have not accurately reflected the physiological and psychological nuances of human perception. As a result, we undertook this study to better understand the levels and types of dynamic occlusion that are acceptable to AR users. We present the results of our initial findings in the hopes that this work might inform future research and future requirements on systems that make use of dynamic occlusion.

BACKGROUND

Occlusion

Dr. Jennifer Esposito, Vice President of Health for Magic Leap, describes spatial computing as “…really the idea that the digital world and the physical world are fully interacting. They are aware of each other” (Parmar, 2019). This awareness, primarily of the digital world interacting realistically with the real world, begins with simple visual cues that help the user believe that the two are blended. Spatial computing is not just a display overlaid on the real world, but rather the creation of fully 3D digital content that exists within the user’s surroundings. When a digital object is “in front” of a real object, it should obscure or occlude the real object. And conversely, when a real object is in front of a digital object, the spatial computing system must be smart enough to render that digital object so that it appears as if the digital content is behind the real object. This is intuitive, and when done successfully, should not even be noticed by the user. However, creating this unnoticeable and intuitive result proves to be very technically challenging.

The nature of augmented reality is that the display surface is always physically closer to the viewer than the real-world content in their environment. This is illustrated in panel 'a' of Figure 1 below. The eye represents the viewer. When experiencing augmented reality (AR), the viewer will have some sort of display interposed between them and the real world. Synthetic objects, represented by the red and green shapes in the diagram, are then rendered to appear as if they are in the world. However, for these objects to appear correctly to the viewer, the synthetic objects must be properly obscured by real world objects. Note that it is trivial for the synthetic objects to occlude real world objects – that is, to have synthetic objects appear in front of real-world objects. After all, the display surface (e.g., a head mounted display) will always be closer to the viewer than the real objects, so simply by rendering the objects they should cover up the real world. The challenge of occlusion is having real world objects obscure the synthetic object rendered on the display surface. For this to happen, the synthetic objects must be clipped or modified so that the real-world objects show 'through' the synthetic objects. Obviously this will not change the location of the display surface with respect to the real world, but if done properly, it should provide the viewer with the illusion that the synthetic content is behind the real world content, rather than in front of it.

Panel 'b' of Figure 1 illustrates what happens when there is no occlusion, or when occlusion is not performed correctly. Due to the mechanics described in Panel 'a', the synthetic content will simply appear to the viewer on top of the real world. This effect can be jarring and break the suspension of disbelief and sense of immersion that is critical to successfully applying AR. Panel 'c' illustrates successful occlusion – the synthetic content is displayed to the viewer in a natural manner. Content that should be in front of the person in view, in this case Synthetic Object 1, will appear in front of the person and obscure them. But Synthetic Object 2, which is behind the person, will appear to be obscured by the person. Such a natural blending of the real and synthetic world is a key aspect of successful spatial computing.

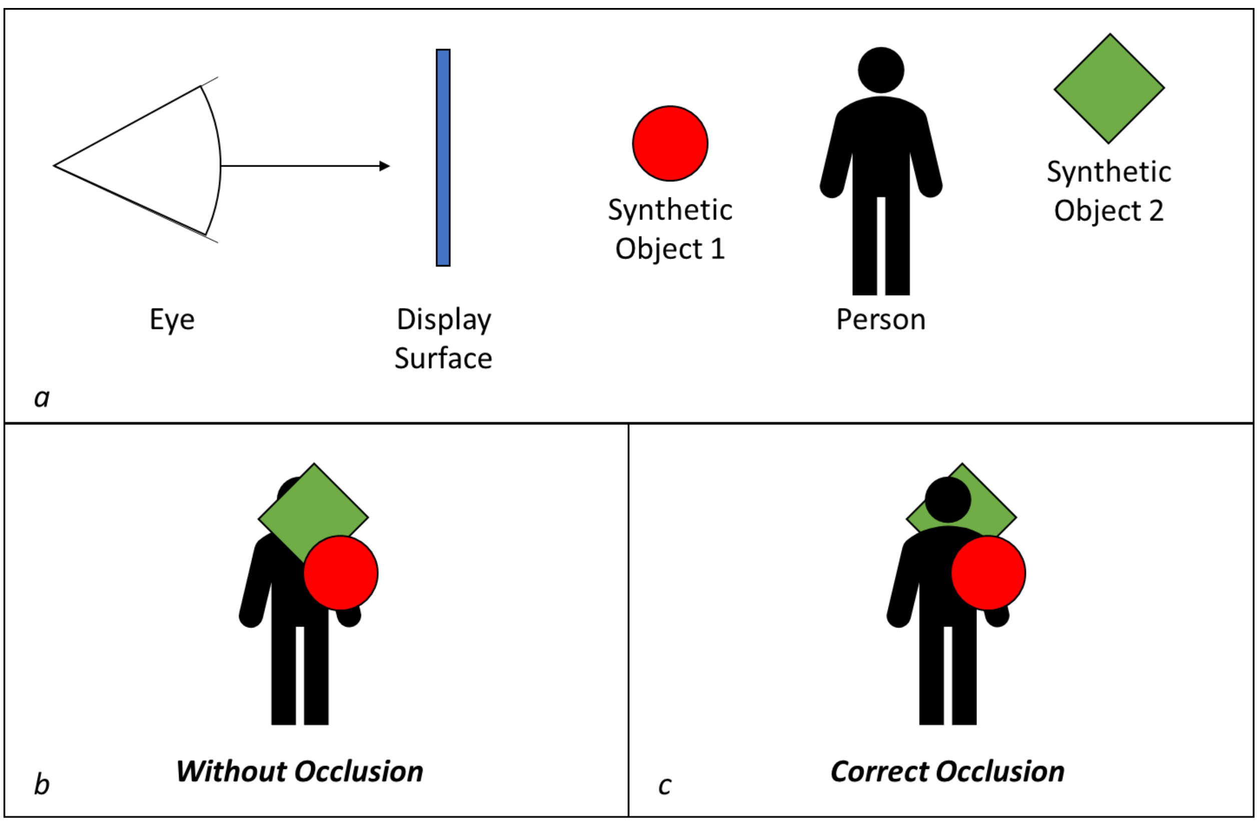

Figure 1. Occlusion

For an AR system to successfully implement occlusion, the system must track the given objects, understand how they relate spatially to the synthetic content, and properly obscure the rendering of their synthetic content to match the viewer's perspective of that content. As the display surface remains closer to the viewer than the real world, effectively the system must 'punch' out an accurate 'hole' in the shape of the real world objects in order for the synthetic content to appear as if it is behind the real world objects. Referring again to Panels ‘b’ and ‘c’ above, one can imagine that to render this image properly, the green diamond shape of Synthetic Object 2 would need to have a hole in the shape of the person’s head and upper shoulders occluded from it so that the person is visible in a natural fashion.

Notes on Generalizability

For our research, we differentiate static occlusion from dynamic occlusion. Static occlusion is the obscuration of synthetic content by real world static objects, such as walls, floors, and some types of furniture. These objects tend to be relatively permanent in an environment. Typically, AR systems will implement static occlusion by having an accurate mesh of the physical environment, and using that mesh to occlude digital objects, just as they would using standard computer graphics occlusion techniques. These meshes are typically created in near-real time, rather than real time, and may necessitate dedicated preparation to create. Static occlusion should be considered a baseline requirement for any AR system capable of spatial computing.

In ARDST, we have primarily concerned ourselves instead with dynamic occlusion. Dynamic occlusion poses unique challenges specific to our use case and has a very detrimental effect if not implemented properly. Dynamic occlusion is the obscuration of synthetic content by mobile real-world objects, such as people. It is more technically demanding than static occlusion, due largely to the extremely low latency demands for both sensors and spatial processing. This real-time capability is required for successful implementation of dynamic occlusion. This paper will not focus on the techniques we have developed or implemented for dynamic occlusion, but will instead focus on our understanding of the metrics necessary to provide a meaningful assessment of how successful any given occlusion implementation is. The techniques and metrics we describe could be applied to both static and dynamic occlusion, but again, dynamic occlusion is our key focus area. The examples and survey described in this study may reflect that bias.

Occlusion is not a universal concern for all AR applications, or even for all features within the same application. There are times when the use case requires occlusion and times when it would be better to omit it. For example, if you have a system in which you want to create an immersive tactical situation for Soldiers, like ARDST, then both static and dynamic occlusion will be needed. Trainees in this system must believe, for example, that a synthetic enemy is hiding behind a real-world tree. If the Soldier were to see the enemy through the tree, then the value of the training delivered would be diminished and may even create negative training transfer. Conversely, in the same training simulation, perhaps an exercise Observer Controller (O/C) has administrative digital content that helps them facilitate the execution of the exercise. In this case, the digital content may be better off not abiding by any occlusion, static or dynamic. This would allow the O/C access to key information regardless of who might walk in front of them or the size of the room they were in.

Additionally, there exist two notably different approaches to AR that have a significant impact on the implementation of occlusion: optical-see-through and video-pass-through augmented reality. Optical-see-through is a paradigm in which the viewer sees the world through a semi-transparent lens, and digital content is displayed onto that lens. Optical-see-though is the type of augmented reality implemented by Magic Leap and Microsoft headsets, in their respective ML1 and HoloLens products. Conversely, in video-pass-through augmented reality, the viewer sees a display captured on video with digital content added to it. The user is not seeing the real-world directly in this case, but rather is seeing a modified real-time video. This may be seen on any smartphone AR application, or in headset implementations where video cameras are added to VR headsets like those offered from the HTC Vive and Oculus Rift product lines.

Our research in ARDST is primarily concerned with optical-see-through implementations, and our use cases and study questions reflect that bias as well. We believe that the subject of this study and the conclusion we draw from it are applicable in either paradigm, but for transparency we note that our research is focused on the optical-see-through paradigm.

LITERATURE REVIEW

There are a large number of survey works that generally discuss occlusion and its challenges, but few works that describe metrics for assessing occlusion. Popular tech outlets like the website Hacker Noon describe the gamut of technical challenges associated with occlusion in AR at large. (Mathew, 2018). And though the article describes the challenges of mesh generated occlusion versus other forms of occlusion, the author does not use the same categorizations that we do, in terms of static and dynamic occlusion. Shah (2018) covers much of the same ground as Mathew and describes a variety of different approaches to implementing occlusion. Sandström (2018) provides a more in-depth survey of varying techniques, focusing mainly on ARKIT, Unity, and OpenCV. The focus in these works tends towards implementation, and they do not dig into developing metrics to assess the relative merits of each approach.

It is also worth noting that much of the previous work found in the area of occlusion is based in the video-see-through paradigm, and much of it is heavily based on handheld AR systems, like smartphones or tablets. In fact, while many of the approaches could be widely applicable, all of the references that described novel implementations of occlusion were based on using tablets or cameras connected to powerful computers. The results of these research papers were interesting to consider, but many depended on per-pixel analysis of imagery upon which the occlusion was performed. While these methods may be adapted for optical-see-through AR, the implementation is challenging, and any computational processing risks inducing noticeable lag that may be less detectable in video-pass-through. In video-pass-through, the entirety of the images may lag, but in optical-see-through, any lag in processing, whether associated with occlusion or not, will be more easily detected by simple contrast with the real world.

Fischet, et al. (2003) present an occlusion based on static backgrounds in the natural surroundings, which must be acquired beforehand. This would be unsuitable for moving AR systems. Maidi, et al. (2010) proposed a tracking algorithm based on the combination of tracking fiducials and optical flow to track visible points and maintain virtual graphics overlaying when targets are not identified. Dong, et al. (2013) make use of a time-of-flight (TOF) camera to create a depth map, and then uses OpenGL shading language (GLSL) and render to texture (RTT) techniques. While most modern AR systems have TOF capability, the reliance on per-pixel analysis via GLSL make adaptation to optical-see-through applications challenging. Pavan Kumar (2018) on Medium also presents an implementation to demonstrate capabilities of the Selerio SDK on smartphones, but he does not offer metrics or means of assessing performance. Jorge, et al. (2019) use raw depth information of the scene to create a rough foreground / background segmentation. They combine this data with color data to blend the virtual objects with real objects, but this approach is heavily dependent on RGB camera capture, and again very demanding and challenging to implement on an optical-see-through headset. Unfortunately, these papers offer little in the way of generalizable metrics and focus more on documenting their implementation techniques.

Walton, et al. (2017) implements their own approach to occlusion, though mostly focused on static objects. However, they do provide a metric for assessing performance. The authors do this by measuring the mean-square error between where pixels should be and where they were via a matte and green-screen process. This approach has its merits, but the focus on pixel-based mean-square error may hide the true impact of inaccurate clipping along object edges. This purely technical assessment may not accurately reflect how acceptable the final results are to viewers. Walton does bring up a valid point regarding temporal noise, and it is worth keeping in mind that occlusion takes place not only in 2D images as observed in 3D environments, but across time as well.

Yuan, et al. (2010) provide some of the most in depth analysis of metrics to assess performance. However, much of the metrics they offer seem uniquely suited to assessing the performance of their particular implementation. They provide three different measurements to assess performance: occlusion error, texture error, and temporal instability. Occlusion error penalizes depth edges for not being sharp. This error measures how crisp and well localized the edge in the depth map is for annotated occlusion pixels. It extracts a profile of ten depth samples perpendicular to the edge, five on either side, and measures the deviation from an ideal step edge, after removing the mean and standard deviation. This is the closest to “overlap error” but it does not try to quantify how much bleed-in or bleed-out occurs, and instead only provides an overall error score. Texture error penalizes depth at texture edges for not being smooth. Texture edges are color changes in regions that are not occlusion boundaries. The expectation is that the depth map would be smooth at these locations because false depth discontinuities would cause self-occlusion artifacts or cracks in objects. This appears to be something unique to their method and may not be generalizable to a wider variety of applications. However, it is something to consider when assessing the acceptability of differing occlusion techniques. Finally, temporal instability penalizes temporal jitter of static points. Abrupt depth changes in the video in their solution can cause flickering when rendering effects. These measures seem directly related to video-pass-through implementation and should be noted for anyone looking to apply that form of AR in their solutions. They also bring up the same interesting theme as Walton did; that there is a temporal aspect to dynamic occlusion that needs to be considered.

In our literature, we found descriptive coverage of the challenges of occlusion and especially the added challenges of dynamic occlusion, and several novel implementation approaches. Unfortunately, the majority of these studies focused on implementation approaches, and few discussed metrics needed to objectively assess performance. For those papers that did address metrics, we found their metrics to be too specific to their unique solutions, and unsuitable for our applications. Just as we acknowledge that our study may be biased because of our focus on optical-see-through AR solutions, we found that the majority of the research was biased the opposite way towards video-pass-through solutions.

PROBLEM DESCRIPTION AND HYPOTHESIS

During the course of our research, we tested various means of implementing dynamic occlusion and assessed the performance of each approach. Our objective was always to reduce the absolute error in occlusion. We calculated this as a strict percentage which we termed overlap error. Overlap error was objectively the size of the errant occlusion divided by the total size of the object being occluded (typically most easily measured in pixels).

$$Overlap\ Error = \frac{Total\ Area\ of\ Occlusion\ Error}{Total\ Area\ of\ Target\ Object\ Being\ Occluded} \tag{1}$$

As we considered the occlusion problem further, we subdivided the total area of occlusion error into two distinct components, false positive and false negative error. False negative error would be parts of the target object that were not occluded – in other words, the synthetic content covers up the real-world content, even though it should be behind it. Conversely, false positive occlusion is where the synthetic content is clipped to make a hole for the real-world content, but the real-world content is not actually there.

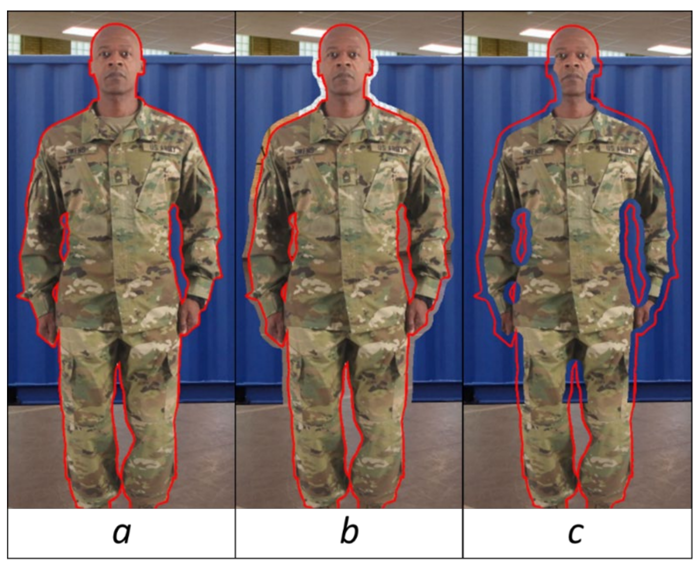

Figure 2. Examples of False Positive and False Negative Occlusion Error

Figure 2 demonstrates the difference in false positive and false negative occlusion error. Panel ‘a’ shows no error – the Soldier is standing in an auditorium, and a blue container is rendered as synthetic content behind the Soldier. The blue container appears as if it truly is behind the Soldier because the Soldier perfectly occludes the container. The red outline is added for reference to highlight the true occlusion perimeter of the Soldier. Panel ‘b’ in Figure 2 illustrates false positive occlusion error. In this case, the synthetic content is occluded outside of the Soldier’s true occlusion perimeter. The resulting effect is a halo around the Soldier through the container. The Soldier still appears in front of the container, but clearly the occlusion has error. Panel ‘c’ illustrates false negative occlusion error. In this case, the synthetic content, which should appear to the viewer to be behind the Soldier, incorrectly cuts off too much of the soldier, infringing upon his occlusion perimeter shown in red. As a result, the Soldier’s outline is distorted, and he appears unnaturally obscured at the edges.

As we saw more and more of these results from our dynamic occlusion tests, the ARDST research team developed an intuition about what was more pleasing or acceptable to see. Prior research has indicated that aesthetics can play a significant impact on learner outcomes (Martin, et al., 2014). The assumption is that the less acceptable an image is, the more it will break the suspension of disbelief and sense of immersion in training simulations, and thus negatively impact the training value. While we strive for perfect occlusion and the reduction of all error, we internally began to suspect that false positive occlusion was more acceptable to users than false negative occlusion. If this is the case, then a metric designed to reflect the effectiveness of the occlusion should also reflect that preference. In order to test this intuition, we developed our hypothesis, that a viewer will find false positive occlusion preferable to false negative occlusion.

STUDY METHODOLOGY

In order to study our hypothesized user preference for false positive versus false negative occlusion, we devised a simple survey. In each question, the respondents were presented with two images, one with false negative error, and one with false positive error. Sample images from the survey can be seen in Figure 3. We varied the amount of false positive and false negative error in each image by increments of 5%, from 5% to 25%. Figure 3 shows 20% false positive error on the right and 20% false negative error on the left.

Figure 3. Survey Image Samples

Our initial instinct was to simply test one image at a time, rather than having the respondent compare the two images. However, upon preliminary assessment of the images and further consideration, we became concerned that this would simply result in 'floor bounded' responses where all images were given the lowest score. Without a benchmark for comparison, we feared that respondents would simply mark down any occlusion error to the lowest value and thus confound any analytic power of the study. A flat response surface where every response was marked at the lowest value would not provide us with insightful analysis, and we would have no ability to determine the relative effects of the false positive and false negative occlusion. We therefore elected to pursue the double-image/contrast approach.

In our initial limited pilot study, we mirrored the error between false positive and false negative. For example, for each image, the survey participant would choose between 15% false positive and 15% false negative error, or 20% false positive and 20% false negative error. Though the sample size for this pilot was not statistically significant, our assessment was that this format failed to deliver sufficient assessment power for us to derive any valuable conclusions from.

We determined that in order to get a true assessment of preference, we should test the spectrum of combinations of the varying occlusion errors. However, this would have resulted in a long and ungainly survey instrument, and we suspected that it would negatively affect our ability to collect data. Our assessment was that, in order to gather statistically significant sample sizes and not skew the results from induced survey fatigue, we wanted to keep the survey as short as possible. We turned to Non-Orthogonal Latin Hypercubes to assist us in developing our Design of Experiments. Information on Design of Experiment best practices can be found at the Air Force Institute of Technology STAT Center of Excellence web site on 'Best Practices' (n.d.). Additionally, the Naval Postgraduate School SEED Center offers tools to easily implement Latin Hypercube designs (Naval Postgraduate School Simulation

Experiments and Efficient Designs (SEED) Center. n.d.). The Latin Hypercube design provides us with added assessment power while reducing the number of combinations which need to be tested. This approach gives us confidence that even by testing a reduced number of the possible combinations factors in the survey, we will still have sufficient coverage to derive beneficial analytical results. The result was a 17-question survey which, on average, took less than three minutes to complete. For each question, the respondent indicated their preference on a 5-point Likert scale, with '1' indicating a preference for false positive error (the halo effect), and a '5' indicating a preference for false negative error (the cut off effect).

Based on this instrument, we can restate our hypothesis as:

H_0: Survey Respondents will not prefer false positive over false negative occlusion error
H_1: Survey Respondents will prefer false positive over false negative occlusion error

RESULTS

We administered the survey via email requests and generated 62 responses. To understand the data better, we first examined the descriptive statistics of the response data. In this case, we are specifically examining the 62 averages of each respondents' answers. As can be seen from Table 2, the average response is below '3'. In the survey, a response of '3' indicates no preference and a response lower than '3' indicates a preference for the false positive image rather than the false negative image.

Table 1. Descriptive Statistics

Mean	*Std Error*	*Median*	*Mode*	*Std Dev*	*Samp Var*	*Kurtosis*	*Skewness*	*Range*	*Min*	*Max*	*Count*
2.2989	0.0878	2.2353	2.2353	0.6913	0.4779	-0.3799	0.2803	2.8824	1.0000	3.8824	62

We can further investigate the nature of the response data by plotting the average responses for each participant in a histogram. We are relying on the Latin Hypercube design to provide statistical coverage over the possible combinations, so it seems reasonable to compare each survey participant's average answer. The result is shown in Figure 4. The x-axis indicates the average response on the 5-point Likert scale, in bins of 0.25. The y-axis indicates the number of survey respondents whose average score fit in that respective bin.

With the histogram, we can see that indeed the preponderance of the responses are below 3, indicating that there is a greater preference for the false positive images. However, it is also notable that the histogram shows a potential bi-modal distribution, and that we will need more statistical evidence to refute or confirm our hypothesis.

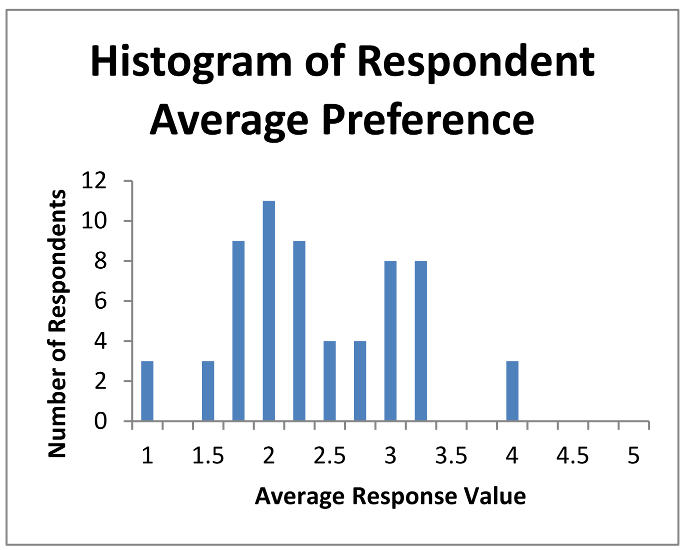

Figure 4. Histogram of Response Averages

With the response data, we conducted a linear regression to assess the effect size and statistical significance. We selected a P-value of 0.01 for our hypothesis test. We then compared the respective effect size and statistical significance of each factor. The results are shown in Table 2 below.

Table 2. Linear Regression Results

	Coefficients	*Std. Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3.6720	0.4324	8.4919	6.79156E-07	2.7445	4.5994
False Pos %	0.0081	0.0182	0.4430	0.6645	-0.0309	0.0470
False Neg %	-0.0961	0.0182	-5.2880	1.1464E-04	-0.1351	-0.0571

As we can see, false positive occlusion has a P-value of 0.664. This indicates that there is no statistical significance in our responses to false positive occlusion error. In other words, we cannot discern with any statistical confidence that the amount of false positive error, or halo effect, had an impact on the survey participant's responses.

In contrast, false negative occlusion error has a P-value of 0.00011, which indicates that false negative occlusion error does have a statistically significant impact on user preference. We also see that the coefficient for the false negative occlusion error is negative, indicating that as the amount of false negative occlusion error increases, respondents are more likely to prefer the false positive images.

For explorational purposes, we extended our statistical investigation to include a univariate regression upon the difference between the percentages of false negative and false positive occlusion error, with the hopes of understanding the relationship between these two factors a little better. The results of this analysis are shown in Table 3 below.

Table 3. Univariate Linear Regression on the Difference Between False Negative and False Positive Error

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.2989	0.1491	15.4204	1.3132E-10	1.9811	2.6166
Diff (Neg-Pos)	-0.0521	0.0159	-3.2826	0.0050	-0.0859	-0.0183

The independent variable is the difference between the size of the negative error and the size of the positive error in each image pair presented. As this number increases, we would find that the cut off false negative image has more error than the corresponding halo false positive image. The p-value is 0.005, indicating that it is statistically significant. However, the coefficient is nearly half of what the coefficient of the false negative error percentage alone was (-0.096 for false positive alone versus -0.052 for the difference between false negative and false positive error). This would seem to indicate that consideration of the relative difference between false positive and false negative error may actually confound the false negative error effect, and that false negative error alone is a better indicator of respondent preference.

These results suggest that we can reject the null hypothesis and accept our alternative hypothesis. We cannot conclude with statistical confidence that false positive error had an effect on user preference, but we can conclude that false negative occlusion error made images less preferable. To be strict, however, we cannot say that there is a general preference for false positive error. Rather it would be more accurate to conclude that when given the choice between false positive and false negative error, the respondents' dislike for false negative error was evident, while the false positive error did not seem to impact their response.

DISCUSSION

While we conclude that false negative occlusion error has a greater effect than false positive error, it is worth taking a look at how each functions and why these results may have come about. False positive error tends to create a halo effect, as described above and shown in Figures 2, 3, and 5. Clearly, this effect has a chance of breaking the sense of immersion for the user as they see 'through' the synthetic objects. However, for the object or person being occluded, the positive aspect of this error is that the outline of the person remains unperturbed.

In contrast, when there is false negative error, the outline of the target person being occluded becomes rapidly disturbed, especially as the amount of error increases. The effect is that the person becomes distorted and may become unrecognizable. Most significantly, the effect is pronounced when key features of a person are distorted. In the image in Figure 5, note how the false negative occlusion error significantly affects the arms and face of the target person subject to the occlusion error. The distortion makes the person look almost unrecognizable, as if they are someone else. While the false positive occlusion error creates a halo effect that detracts from the experience, at least the person remains recognizable. And perhaps more importantly for training use cases, the subject's actions and key features such as what equipment they may be wearing or using will be fully visible to the observer.

Figure 5. Example Images

It is worth noting that certain features unique to each person may provide a buffer that mitigates the impact of the error. For example, a person with more hair may be less 'distorted' by negative dynamic occlusion error than someone with less hair. The error may cut into their hair, but as hair is less rigid, we accept more deformation in how we perceive hair than in how we perceive someone's face. However, when examining the true occlusion shape of someone without hair, it is obvious that they lack the benefit of this visual buffer, and that false negative occlusion error may instead immediately begin affecting salient facial features. This is also true when examining the effect of the error in someone who is in profile versus someone who is seen head on. In future experiments, we will look to quantify the degree to which the image of a subject may provide buffer features that mitigate or confound the effect of the dynamic occlusion error.

Implications for Practice

The driving motivation for this research was to assess the validity of our initial simple formula for calculating Overlap Error:

$$Overlap\ Error = \frac{Total\ Area\ of\ Occlusion\ Error}{Total\ Area\ of\ Target\ Object\ Being\ Occluded} \quad (1)$$

Accounting for the distinction between false positive and false negative error, this equation may be rewritten as:

$$Overlap\ Error = \frac{False\ Positive\ Occlusion\ Error + False\ Negative\ Occlusion\ Error}{Total\ Area\ of\ Target\ Object\ Being\ Occluded} \quad (2)$$

The results of this study indicate that these two types of error should not be equally weighted, and that false negative occlusion error has a much more significant impact on user acceptance. We can surmise that it would likely have a more detrimental effect on user immersion and suspension of disbelief and may significantly detract from training. Despite the results of our analysis, from a practical standpoint, we would not recommend completely dismissing the effect of false positive occlusion error. While the regression above indicated that false positive occlusion error is not statistically significant in the survey responses, this is only true in contrast to the false negative error, which seems to have an overwhelming effect upon the viewer's preference. A metric that omits false positive error would not be able to effectively compare the true utility of dynamic occlusion techniques. This is a similar concern to the one we identified with prior metrics, such as that proposed by Walton, et al. (2017). The metric might be technically accurate but would not reflect practical merit. As a simple reductio ad absurdum, such a metric would give a perfect score of 0% overlap error to a system that shows no virtual content at all, because there would never be any false negative occlusion error, only false positive occlusion error over all content. Clearly such a result is not helpful in being able to compare optimal solutions for use in future implementations.

We feel that more research is needed to better quantify the relative effect of each type of error on user acceptance, and we intend to pursue that research in future efforts. In the meantime, we know that false negative error has a greater effect than false positive, yet we should not discount the effect of false positive error. Our efforts are confounded by the lack of statistical significance in the false positive error. In the spirit of Ockham's Razor, we propose an interim solution in which the false positive occlusion error very simply be weighted by a factor of 0.5, while false negative occlusion error maintain its full weight. We recognize that this proposed solution likely does not fully capture the nuanced dynamics at play, but we also cannot confidently assert that any more sophisticated equation would be more supported by the results of our experiment. We are continuing this line of research to better define the relationship between false positive and false negative error, but in the meantime, we believe it is worth making this simple adjustment to current efforts to assess the relative effectiveness of dynamic occlusion techniques.

Accordingly, we propose the following simple revised formula for calculating Overlap Error as a metric to quantify the performance of dynamic occlusion in augmented reality systems:

$$Overlap\ Error = \frac{(0.5\ \times Area\ of\ False\ Pos.Occ.Error) + Area\ of\ False\ Neg.Occ.Error}{Total\ Area\ of\ Target\ Object\ Being\ Occluded} \tag{3}$$

CONCLUSION

This study was designed to assess preference for false positive versus false negative occlusion error. Our initial supposition was that given the choice between false positive and false negative error, viewers would prefer false positive. Through the survey instrument we created, we were able to determine that false positive error did not have a statistical impact on image acceptability, but that false negative error did have a statistically significant impact on viewer preference. As more false negative error was shown, survey respondents liked the results less. Accordingly, we recommend that a generalized overlap error metric for occlusion count only half the false positive error and all of the false negative error when determining total percentage of error. This formulation will incentivize the use of occlusion techniques which are more acceptable to the end user and thus more likely to aid in the creation of compelling and immersive simulation training experiences via augmented reality and spatial computing.

ACKNOWLEDGEMENTS

This research was partially funded under contract number W911NF-18-C-0029 from the U.S. Army Combat Capabilities Development Command Soldier Center (CCDC SC) SFC Paul Ray Smith Simulation & Training Technology Center (STTC). The STTC supports the research and development of next-generation simulation technologies, to include augmented reality, in order to increase the effectiveness of Soldier training. The authors also wish to thank Dr. Darryl Ahner, Director of the Air Force Institute of Technology STAT Center of Excellence for his mentorship and guidance.

REFERENCES

Air Force Institute of Technology STAT Center of Excellence. (n.d.). *Best Practices and Test Planning Guides.* Retrieved May 18, 2020, from https://www.afit.edu/STAT/statdocs.cfm?page=1126

Breen, D. E., Whitaker, R. T., Rose, E., & Tuceryan, M. (1996). Interactive Occlusion and Automatic Object Placement for Augmented Reality. https://www.cs.drexel.edu/~david/Papers/eg96.pdf

Dong, S., Feng, C. & Kamat, V. (2013). Real-Time Occlusion Handling for Dynamic Augmented Reality Using Geometric Sensing and Graphical Shading. 10.1061/(ASCE)CP.1943-5487.0000278.

Du, C., Chen, Y., Ye, M. & Ren, L. (2016). Edge Snapping-Based Depth Enhancement for Dynamic occlusion Handling in Augmented Reality. http://cs.uky.edu/~cdu226/assets/DynamicOcclusion_ISMAR2016.pdf

Fischer, J., Regenbrecht, H., & Baratoff, G. (2003). Detecting dynamic occlusion in front of static backgrounds for AR scenes. 10.1145/769953.769971

Jorge, J., Anjos, R., & Silva, R. (2019). Dynamic occlusion Handling for Real-Time AR Applications. 10.1145/3359997.3365700.

Maidi, M., Ababsa, F., & Mallem, M. (2010). Handling occlusions for robust augmented reality systems. 10.1155/2010/146123

Mathew, N. (2018, January 28). Why is Occlusion in Augmented Reality So Hard?. https://hackernoon.com/why-is-occlusion-in-augmented-reality-so-hard-7bc8041607f9

Martin, M. & Shen, Y. (2014). The Effects of Game Design on Learning Outcomes, Computers in the Schools, 31:1-2, 23-42, DOI: 10.1080/07380569.2014.879684

Naval Postgraduate School Simulation Experiments and Efficient Designs (SEED) Center. (n.d.). *Software Downloads.* Retrieved May 18, 2020, from https://nps.edu/web/seed/software-downloads

Parmar, Arundhati (2019, December, 8). You know VR, AR, but what is spatial computing? MedCityNews. https://medcitynews.com/2019/12/you-know-vr-ar-but-what-is-spatial-computing-and-why-that-matters-magic-leap-dreem-and-others-seek-to-gain-more-from-our-brains/.

Pavan, K. (2018, December 7). Occlusion is important in AR. Medium. https://medium.com/selerio/occlusion-is-important-in-ar-9a672ff9ca2d

Sandström, D. (2018). Dynamic occlusion of Virtual Objects in an 'Augmented Reality' Environment (Dissertation). http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-71581

Shah, N. (2018). Realtime Object Occlusion In Augmented Reality Environments. Retrieved from https://pdfs.semanticscholar.org/0329/d772efe01b5d818dfc22e4a357d03324a01e.pdf

Walton, D, R. & Steed, A. (2017). Accurate real-time occlusion for mixed reality. 10. 10.1145/3139131.3139153

Xu, C. Li, S., Wang, J., Peng, T. & Xie, M. (2008). Occlusion Handling in Augmented Reality System for Human-Assisted Assembly Task. 5315. 121-130. 10.1007/978-3-540-88518-4_14.

Yuan, T., Tao, G., & Cheng, W. (2010). Real-Time Occlusion Handling in Augmented Reality Based on an Object Tracking Approach. Sensors. 10. 10.3390/s100402885.