# INFLUENCE OF VISUAL FIELD CONGRUENCE AND REALISM THROUGH GLOBAL ILLUMINATION ON EGOCENTRIC DEPTH PERCEPTION IN VIRTUAL AND AUGMENTED ENVIRONMENTS

by

Matthew McQuaigue

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2024

Approved by:

Dr. Kalpathi Subramanian

Dr. Paula Goolkasian

Dr. Zachary Wartell

Dr. Aidong Lu

# ABSTRACT

MATTHEW MCQUAIGUE. Influence of Visual Field Congruence and Realism through Global Illumination on Egocentric Depth Perception in Virtual and Augmented Environments. (Under the direction of DR. KALPATHI SUBRAMANIAN)

This research investigated how the similarity and realism of the rendering parameters of background and foreground objects affected egocentric depth perception in indoor virtual and augmented environments. I refer to the similarity of the rendering parameters as visual 'congruence'. Across three experiments, participants performed a perceptual matching task where they manipulated the depth of a sphere to match the depth of a designated target peg. Each experiment served to evaluate the influence of different levels of realism and congruence between environment objects on depth perception. In the first experiment, the sphere and peg were both virtual and unrealistic. The background would switch between virtual and real. In the second experiment, the sphere was virtual (unrealistic) and the peg was real. In the third experiment, the sphere was virtual (realistic), and the peg was either real or virtual (realistic). Realistically rendered objects used physically based rendering and global illumination from the real environment compared to traditional simplistic rendering. In all experiments, depth perception accuracy was found to depend on the levels of realism and congruence between the sphere, pegs, and background. My results demonstrate the use of PBR techniques and baked global illumination applied to virtual targets and manipulated objects resulted in a reduction in depth estimation errors compared to the non-PBR condition. Moreover, I found that the visual congruence between virtual and real elements plays an important role in depth judgments with real targets influencing accurate depth perception when interacting with PBR-manipulated objects. Interaction effects between target and manipulated object affect depth perception switching from overestimations to underestimations depending on rendering conditions. My findings contribute to the growing body of knowledge on AR and VR depth perception which can validate the efforts

for developing advanced rendering techniques for augmented environments. Results can help determine the design and development choices for future AR/VR applications, particularly those requiring precise depth judgments and seamless real-world integrations of virtual objects.

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr Kalpathi Subramanian for supporting this work and other work throughout my journey as an undergraduate and graduate student. I would like to thank my committee members Dr. Paula Goolkasian, Dr. Zachary Wartell, and Dr. Aidong Lu for their support, guidance, and encouragement throughout this research journey. Their expertise and patience have been invaluable in creating this dissertation. I would also like to thank Dr. Erik Saule for his support in my research, professional development, and overall academic journey.

I am increadibly thankful to the University of North Carolina at Charlotte for providing the resources and environment for efficient academic growth and freedom of exploring interesting research ideas. Thank you for the finacial support from GASP and Graduate School Summer Funding for being instrumental in the completion of this dissertation. Their funding enabled me to complete my research goals and finish this dissertation.

I would like to thank my friends and family who have supported me along this journey. By offering encouragement, time, and understanding when needed. Their friendship provided the motivation and comfort during a challenging time. Their presence during this process made the journey more enjoyable.

Lastly, I am grateful to my partner for their unwavering support, understanding, time, and patience during the long and difficult process of writing this dissertation. Their love and encouragement have been a beacon of happiness and joy throughout the dissertation process.

To everyone mentioned above, thank you for being a part of this journey.

TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ACRONYMS

ANOVA An acronym for Analysis of Variance.

AR An acronym for Augmented Reality.

BRDF An acronym for Bidirectional Reflectance Distribution Function.

FOV An acronym for Field of Vision.

GI An acronym for Global Illumination.

HMD An acronym for Head Mounted Display.

NDC An acronym for Normalized Device Coordinates.

OSTHMD An acronym for Optical See-Through Head Mounted Display.

PBR An acronym for Physically Based Rendering.

ToF An acronym for Time of Flight.

VE An acronym for Virtual Environment.

VI An acronym for Virtually Impoverished.

VR An acronym for Virtual Reality.

# CHAPTER 1: INTRODUCTION

Virtual Reality (VR) and augmented reality (AR) applications have become popular devices adopted in different disciplines of research while also becoming more accessible for mainstream and occupational use. The ability to perceive depth is one of the most important behaviors in virtual environments, and it is currently a popular field of study. Consistently throughout past research, misestimations of distance in virtual environments have been reported [1].

To accurately represent the environment in future visual systems and to improve a range of human behaviors influenced by depth judgments, such as reaching for an object at a perceived distance, moving virtual objects to match a target (perceptual matching), or understanding the distance relationship between groups of objects, it is important to understand the source of the misestimation. These applications have become effective ways to train new employees on their day-to-day tasks. Blümel [2] states that these technologies have become important in the ecosystem of these companies.

One major area of popularity for augmented and virtual systems is that of medical imaging and medical research. Medical applications can benefit from VR/AR systems by providing more efficient and intuitive ways to understand information about their patients and the tasks they are performing in the form of 3D imaging data. Gsaxner et al. [3] developed a markerless Image-to-face registration for untethered AR in head and neck surgery. This allows the physicians wearing the AR to be untethered and see medical visualizations precisely on the patient. VR as a clinical tool to train users to perform tasks has been the subject of many studies. According to Sutherland, et al. [4] two groups of cardiologists

performed the same training in learning carotid artery procedures. One group learned the traditional way and the other used VR. The group training with VR overall performed much better due to a significant reduction on objectively assessed procedure errors.

Since the popularity of research in medical imaging and systems has grown, the amount of tools and methods of viewing/learning information has grown as well. More importantly, medical systems must be extremely accurate to convey correct information for the clinical user to complete their task. Understanding depth perception in AR compared to VR environments and how different aspects of the visual field affect depth perception performance is important when designing a system to be used accurately and reliably.

Virtual environments provide the ideal space to study perception because of the features that can be controlled through the system. Researchers can influence the 3D space to study individual aspects that affect the user's perception of depth such as manipulating individual cues to measure and understand its influence. The misperception of distance measured in these virtual environments (VEs) is in the form of an egocentric distance underestimation or overestimation. An egocentric judgment is When an observer measures the distance to an object from their self in their field of vision. The reason for this distance miscalculation is still not completely understood. Multiple studies have developed methods to help improve the error of misperceptions.

A question that has sparked significant debate is whether advanced rendering techniques and/or the use of realistic VEs play a role in lessening misperceptions. It was believed that realism in VEs was associated with distance estimation and influenced misperceptions according to Knapp et al. [5]. However, previous studies were conducted before consumer headsets became widely available. It's crucial to reassess these findings considering advancements in VR/AR technologies for a better understanding of depth perception in modern immersive environments.

This dissertation attempts to understand the issue by using AR/VR environments to randomize different depth cues and capture improvements in the observer's performance while comparing against different levels of foreground manipulated object, foreground target, and background graphical fidelity. Doing so allows for direct analysis of depth estimation performance compared between the graphical detail and depth cues of a matching object and a target.

To begin to understand the issue of depth perception in AR and VR, we discuss the issues of depth perception in augmented and virtual environments and the recent research in solving this problem. In addition, background knowledge of depth perception and its simplified fundamental components will be described. Fundamental knowledge of depth perception in the real world is critical to digesting and analyzing work within this area. Also, an understanding of virtual/augmented environments is needed along with the hardware that is used to interact with them. My work tries to answer how depth cues and their alignment with background/foreground graphical fidelity is a possible solution to increase depth perception accuracy. To test this solution, a perceptual matching task framework has been designed to test a multitude of depth cues and rendering cues along with different levels of graphical rendering quality. Graphical rendering quality can occur on a spectrum where higher fidelity graphics mean that aspects of the rendering (lighting, shadow, texturing) more closely match the real world, and lower fidelity graphics are further from realism.

# CHAPTER 2: BACKGROUND

## 2.1 DEPTH PERCEPTION IN THE PHYSICAL WORLD

Depth perception within virtual environments has been a growing area of research. This is due to the mass increase in commercial and consumer use of augmented and virtual reality devices. As the popularity of the devices grows, so do the human efforts to improve the virtual experience to enhance occupational domains such as architecture, construction, medicine, teaching, etc. One major concern when using such devices for professional use is the accurate estimation of distance. It is an important and essential aspect of interaction that makes the adoption of these systems useful. But, the process of creating or mimicking the real world is a complex task and computationally expensive. It is not possible to create a virtual or augmented environment that utilizes all mechanisms of the human vision.

Visual depth perception both egocentric and exocentric have been the topic of active research in the past decade with the goal of mitigating as much misperception as possible. What makes this task even more difficult is that research has found the misperceptions to differ between subjects. To conduct research on this topic, one must understand how the human visual system uses visual characteristics and mechanisms in the depth perception process. One must also know how to construct virtual environments and test human depth perception efficiently.

## 2.2 VISUAL CUES

Subconsciously, the human visual system relies on mechanisms called depth cues to perceive depth. These cues can be split into three categories: proprioceptive cues, visual cues, and binocular cues.

### 2.2.1 PROPRIOSEPTIVE CUES

In general, when a human looks at an object nearby, the eyes reflexively react to the stimuli using three different response mechanisms. These response mechanisms is made up of convergence, accommodation, and pupillary constriction. Convergence is when the eyes perform an inward rotation to fixate on a point within the field of vision (FOV) and perform a fusion of the two image vantage points. Fusion plays an important role in depth perception because the differing images (stereo disparity) of the same object allow for the extraction of depth information. These differing images in stereo disparity are due to interpupillary distance creating a left and right image of the object.

Accommodation is another mechanism that describes the change in focal length from the eye lens to focus on the object when the object is near or far. Accommodation happens when the eyes shift focus from multiple objects to a new focal point. An example of this would be looking out a window and then down at a book. Pupillary constriction extends the depth of field in an act to maintain a clear image.

### 2.2.2 MONOCULAR CUES

Monocular cues are depth cues that the human interprets from one image or one eye to perceive depth. From a young age, monocular cues are learned unconsciously. This also includes pictorial cues which humans use to judge depth from a static non-moving picture. The following section introduces different monocular visual cues.

- Relative size of an object is an important depth perception cue. Size decreases with distance and increases the closer it gets. If two objects are similar in size, the object that is slightly bigger will appear closer to the observer. This applies to both 3D scenes and 2D images. For this to be accurate, the observer must have information about what the actual dimensions of the object are to judge how far it is. This idea is a part of the absolute size or familiar size of an object. If you don't know exactly how small or large an object is, it will appear farther or closer. The familiarity of the size of a car allows humans to accurately judge its distance away from their position.

- Occlusion of an object by another makes it possible to determine a relative distance to another object.

- Elevation of an object's position in relation to the horizon can help with depth. The closer the object is to the horizon, the farther away it is while those farther from the horizon are seen as closer.

- Texture gradient gives essential monocular cues. When looking at an object from a distance, the texture becomes less apparent the farther away it is. The foreground of a scene has more detailed textures than the background.

- Motion parallax helps with distance estimates because objects that are close to the observer move across the field of view (FOV) much more quickly than objects from a farther distance away.

- Atmospheric perspective/Aerial perspective is how the atmosphere effects an object as the distance from the observer and object increases. The more atmosphere placed between the observer and the object, the more the details of the object decreases and the colors become saturated losing contrast between the object and the background.

### 2.2.3 BINOCULAR CUES

Binocular vision is the ability to perceive depth within the field of vision by the use of two eyes. The overlapping of vision comes from the left and right eye being horizontally different which allows for slightly different viewpoints. The result of the different viewpoints provide an overlap in vision and interpreted depth. Stereopsis is the impression of depth that the observer gets from a scene viewed with binocular vision and allows for exocentric depth extraction of objects. The differences viewed in binocular vision is called binocular disparity seen in Figure 2.1 and is one of the main functions that allow for the brain to process and calculate depth. The retinal disparity is the separation of objects from the left and right eye. Retinal disparity provides relative depth information between two objects but not absolute depth. Objects closer to each other will have small retinal disparity, while objects farther away will have greater retinal disparity. The disparity also helps with partial occlusion of objects and judging the distances apart.

## 2.3 DEPTH PERCEPTION METHODOLOGY

To understand and test a user's perceived depth in VEs complex experiments and measurements are used throughout perception research. This wide variety of experiments has observers estimate the target distance of an object from a stationary position or judge the distance between two objects. As stated by Jamiy et al. [1]

> "Measuring distance perception is challenging, it cannot be assessed directly because it is a conscious event, and for that reason, the different experimental techniques include quantifying judgments reported by a subject."

Experiments used to test isolated depth cues are usually performed on one of the three subdivisions of the observer's 3D space.

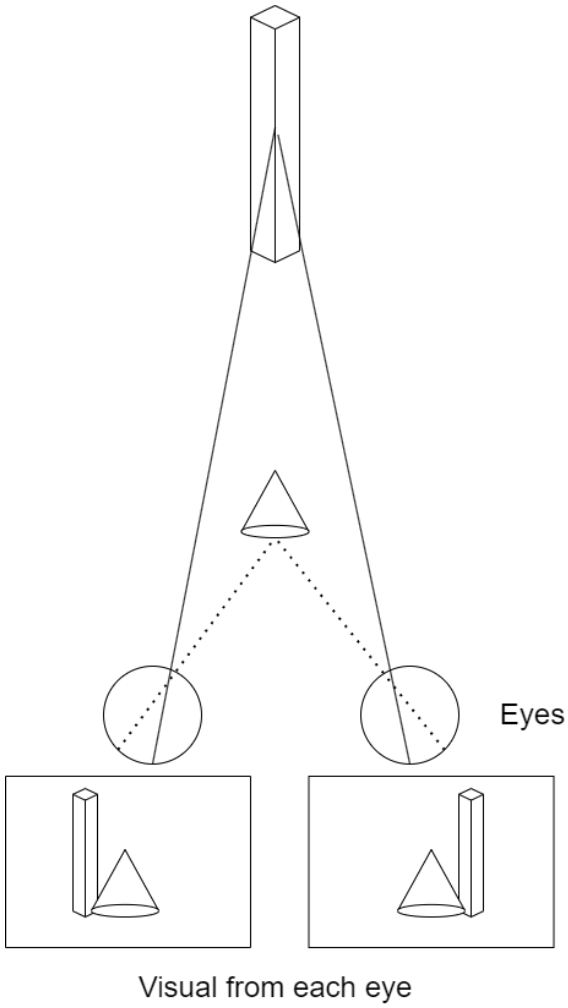



Figure 2.1: Binocular disparity where the horizontal difference between the two eyes perceive objects from different perspectives that get fused together. The two eyes perceive slightly different images which the brain combines to create a sense of depth and distance.

- Personal space: distance between the observer and 2m
- Action space: between 2m and 30m of the observer
- Distant space: greater than 30m from the Observer

### 2.3.1 DEPTH PERCEPTION TASKS

Throughout past research, different depth measurement methods have been developed as a way for the subject to convey their perception of the VE to the researcher. The measure-

ment methods are made up of tasks that are used in different combinations in the research literature. Examples of subject tasks are:

- Verbally stating the distance observed from an internal scale e.g. 15m (verbal report)
- Using a slider, controller, or user interface to move an object to a desired Depth (perceptual matching)
- Reporting relative position between two objects e.g. A is twice as far as B (relative verbal report)
- Reaching to the perceived distance of an object (reaching)
- Throwing an object to a perceived distance (blind throwing)
- blind walking or triangulation (directed action)

In blind walking the subject must locate the position of an object by walking the perceived distance with their eyes closed. Triangulation is done in steps by having the subject face the object. The subject then turns and moves to another designated spot. At this new spot, they must relocate the object as seen in Figure 2.3.

Based on the type of cues or systems being tested, researchers will select a subset of these tasks to use in their study to judge distance estimations. The results of an experiment can change based on the task chosen. Different tasks used for testing for depth perception are more or less accurate than others.

In order to assess an individual's depth perception in virtual environments, it is necessary to render both the environments and the objects within them. This involves simulating depth cues and interactions through rendering. Depending on the specific tasks and depth cues being tested, it is crucial for developers to explore and evaluate different rendering techniques and software in order to efficiently create an appropriate environment.

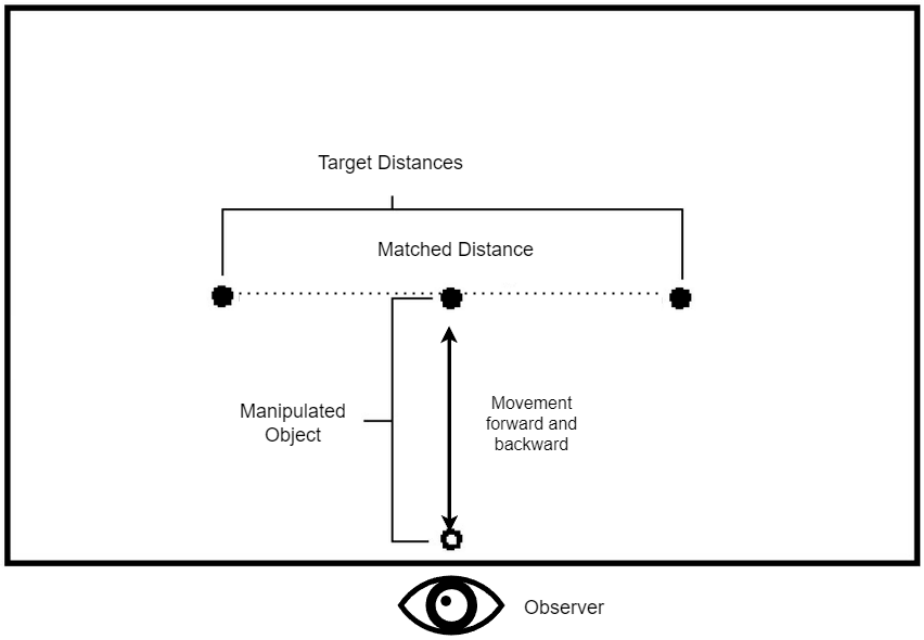



Figure 2.2: Perceptual matching task where an observer moves the object in front of their FOV to match the depth of a set or single target moving the object only forward and backward.

## 2.4 3D MODELING AND RENDERING PIPELINE

Rendering and modeling involve creating 3D objects using vertices, edges, and faces, which form primitive mesh objects, and displaying them on a 2D screen. The commonly used primitive for the representation of objects in images is triangles. It is the simplest polygon for representing complex 3D meshes. Some rendering workflows use quad representations for primitives which is a convex object with four vertices. For triangles, each vertex describes a 3D position using $x, y, z$ coordinates. A mesh can be made of millions of triangles sharing multiple vertices and edges connected. Each vertex is coupled with more information to describe how it is colored. This usually includes attributes such as:

- Normal: A unit vector orthogonal to the face of the primitive. This is used for lighting calculations and surface interactions.

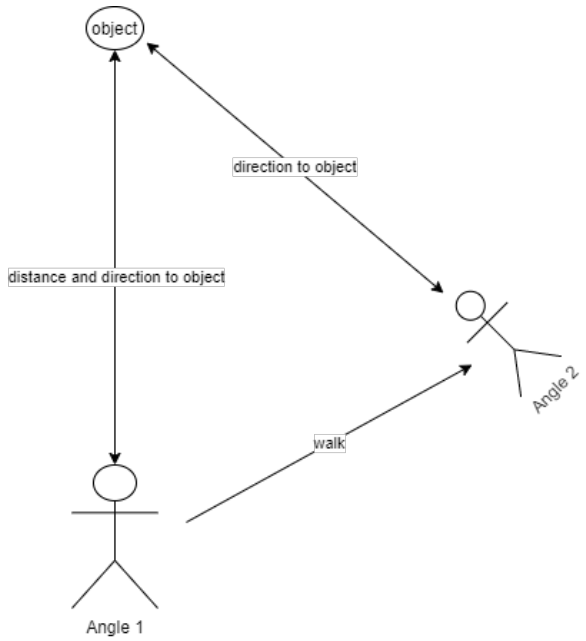



Figure 2.3: Subject performing triangulated walking for a perceived object by having the subject face the object. The subject then turns and moves to another designated spot. At this new spot, they must relocate the object

- Color: This is a 3D vector describing red, blue, and green information.
- Texture UVs: Uv (x, y) axis coordinated describing image pixel values for coloring a specified vertex.

These values along with the vertex information for the mesh are passed through the rendering pipeline to color and shade the object. The information is usually passed as an array buffer interleaved with descriptors on how to access certain portions of the information for a given vertex. This structure is described in Figure 2.4.

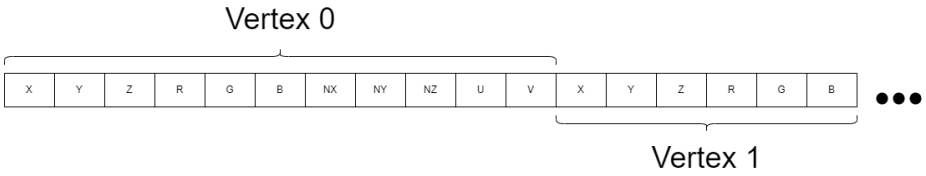



Figure 2.4: Interleaved vertex buffer containing position, color, normal, and texture coordinate information for a list of vertices. For each piece of information, a byte offset is described to designate the beginning for vertices data.

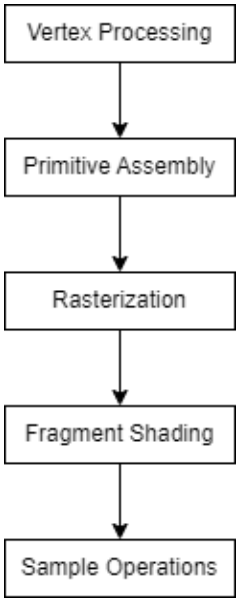



Figure 2.5: The rendering pipeline

### 2.4.1 RENDERING PIPELINE

When rendering an object it goes through several stages within a rendering pipeline to create a 2D image from 3D data representation as seen in Figure 2.5. These pipeline stages can sometimes be longer or shorter depending on the rendering software used. The most important stages for this work are vertex processing and fragment shading. In the first stage, the mesh goes through vertex processing. This stage performs multiple levels of transformations to go from the local objects' 3D space to a screen mapping where the vertices are mapped to pixels within the window coordinates. These transformations are in the order of:

$$ModelTransformation \rightarrow ViewTransform \rightarrow Projection \rightarrow Clipping \rightarrow ScreenMapping$$

The model, view, and projection transforms can be written as:

$$T = P \times V \times M$$

Here, $T$ represents the combined transformation matrix. When transforming a point or vertex from object space to screen space, you would multiply the coordinates of each vertex by this combined transformation matrix $T$.

For a given Vertex:

$$\begin{bmatrix} X_{object} \\ Y_{object} \\ Z_{object} \\ 1 \end{bmatrix}$$

The modeling transformation (M) transforms object coordinates to world coordinates and is represented as:

$$\begin{bmatrix} X_{world} \\ Y_{world} \\ Z_{world} \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & t_x \\ y_1 & y_2 & y_3 & t_y \\ z_1 & z_2 & z_3 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} X_{object} \\ Y_{object} \\ Z_{object} \\ 1 \end{bmatrix}$$

The view transformation (V) transforms world coordinates to camera or view coordinates and is represented as:

$$\begin{bmatrix} X_{view} \\ Y_{view} \\ Z_{view} \\ 1 \end{bmatrix} = \begin{bmatrix} r_x & u_x & f_x & 0 \\ r_y & u_y & f_y & 0 \\ r_z & u_z & f_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} X_{world} \\ Y_{world} \\ Z_{world} \\ 1 \end{bmatrix}$$

where the vectors $r$, $u$, and $f$ represent the camera's right vector, the camera's up vector, and the camera's forward vector.

The projection transformation (P) transforms camera coordinates to clip coordinates and is represented as:

$$\begin{bmatrix} X_{proj} \\ Y_{proj} \\ Z_{proj} \\ W_{view} \end{bmatrix} = \begin{bmatrix} \tan^{-1}\left(\frac{FOV_x}{2}\right) & 0 & 0 & 0 \\ 0 & \tan^{-1}\left(\frac{FOV_y}{2}\right) & 0 & 0 \\ 0 & 0 & -\frac{Z_{far}+Z_{near}}{Z_{far}-Z_{near}} & -\frac{2*Z_{far}*Z_{near}}{Z_{far}-Z_{near}} \\ 0 & 0 & -1 & 0 \end{bmatrix} * \begin{bmatrix} X_{view} \\ Y_{view} \\ Z_{view} \\ 1 \end{bmatrix}$$

The full matrix transform can be multiplied together to be:

$$\cdot \begin{bmatrix} X_{proj} \\ Y_{proj} \\ Z_{proj} \\ W_{view} \end{bmatrix} = \begin{bmatrix} \tan^{-1}\left(\frac{FOV_x}{2}\right) & 0 & 0 & 0 \\ 0 & \tan^{-1}\left(\frac{FOV_y}{2}\right) & 0 & 0 \\ 0 & 0 & -\frac{Z_{far}+Z_{near}}{Z_{far}-Z_{near}} & -\frac{2*Z_{far}*Z_{near}}{Z_{far}-Z_{near}} \\ 0 & 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} r_x & u_x & f_x & 0 \\ r_y & u_y & f_y & 0 \\ r_z & u_z & f_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x_1 & x_2 & x_3 & t_x \\ y_1 & y_2 & y_3 & t_y \\ z_1 & z_2 & z_3 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} X_{object} \\ Y_{object} \\ Z_{object} \\ 1 \end{bmatrix}$$

Once these transforms are performed, clipping is used to discard vertices that are outside the visible region of the viewing frustum. For a homogeneous coordinate

$$\begin{bmatrix} X, Y, Z, W \end{bmatrix}$$

The vertices are compared and clipped inside the region:

$$X < W || X > -W$$
$$Y < W || Y > -W$$
$$Z < W || Z > -W$$

If the vertices fall outside the region, they are discarded or modified to lie on the clipping plane by using algorithms such as the Cohen-Sutherland algorithm or the Cyrus-Beck algorithm. Next, perspective division is used to map the clip space coordinates to normalized

device coordinate (NDC) space [-1,1] using:

$$\begin{bmatrix} X_{ndc} \\ Y_{ndc} \\ Z_{ndc} \end{bmatrix} = \begin{bmatrix} \frac{X_c}{W} \\ \frac{Y_c}{W} \\ \frac{Z_c}{W} \end{bmatrix}$$

To map these NDC coordinates to the screen space coordinates, the viewport offset and dimensions are used to convert from the [-1,1] range to the range of pixels on the screen. For a screen with an origin at (x,y) and a given width and height, the screen pixel can be calculated as:

$$X = (X_{ndc} + 1) * (width/2) + x$$
$$Y = (Y_{ndc} + 1) * (height/2) + y$$

Now that the pixel location for the vertices is calculated and the primitive type is determined in primitive assembly, the output is used as input into the rasterization stage. In this stage, it is determined which pixels are covered by which triangle. This is solving a visibility issue about which primitive should be shown on the screen. This process conducts hidden surface removal through depth testing. A depth buffer that contains the depth at each pixel is used to determine which fragments are closest to the camera. For each fragment, its the z value contained within the passed depth buffer is compared to the z value of the primitive coming through the pipeline. If the z value of the primitive is smaller than the z value in the depth buffer (closer to the screen), it is overwritten. If it is greater than the value in the depth buffer, the z value in the depth buffer is kept. After each primitive has gone through this rasterization stage, correct visibility is determined for the closest values in each fragment.

In the fragment shading stage, fragments are then used for each pixel that is part of the

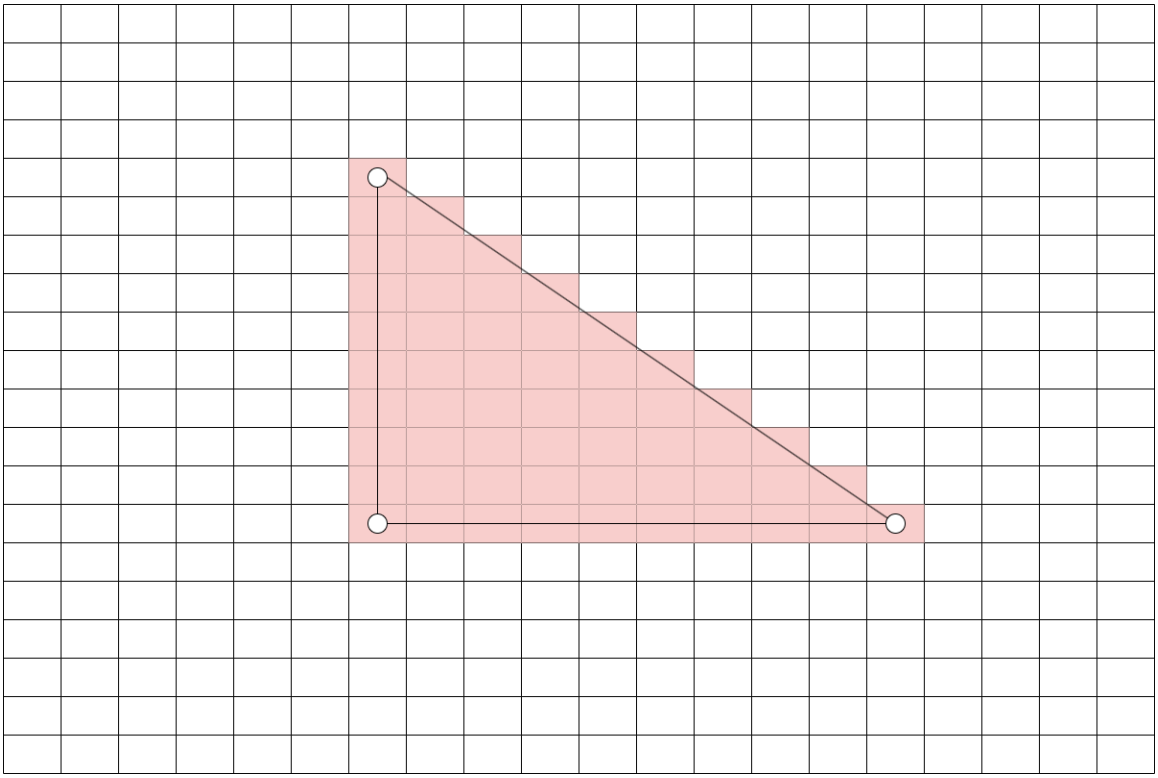


Figure 2.6: The shading of pixels in the fragment shader based on which pixels are covered by the triangle

triangle shown on the screen. These fragments are generated by interpolating across primitive vertex attributes using barycentric coordinates. The fragments generated are which pixels get colored.

The color of the pixel is determined by multiple different possible calculations. A texture image could be used to color the pixel based on the interpolated UV coordinate passed with the vertex attribute. A basic color can be applied based on a predetermined color passed with the vertex attribute as seen in Figure 2.6. The fragment could also use calculations based on the 3D positions of simulated light to shade the object dependent on its orientation and distance from the light source. This process is called shading.

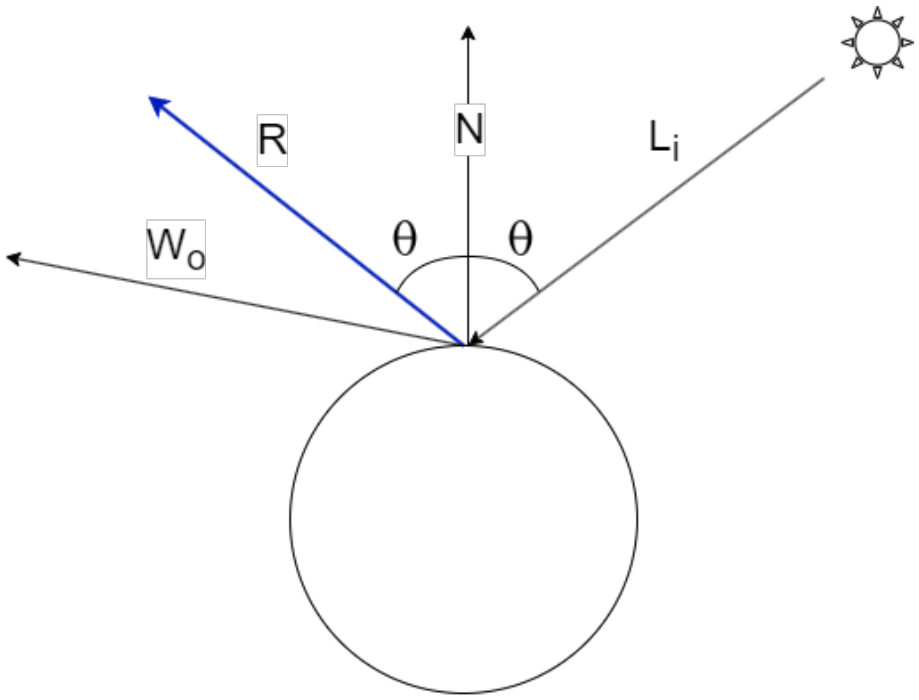



Figure 2.7: The local lighting model where $L_i$ is the incoming light, $W_o$ is the viewing direction, N is the surface normal, and R is the reflectance vector.

### 2.4.2 SHADING

Shading is the process of calculating the color of a pixel corresponding to the surface of an object based on its material properties, position/distance from a given simulated light source, and the direction the object is being observed. The lighting and material calculations used are referred to as shading models. Based on the shading model used, objects can be represented realistically or unrealistically as seen in Figure 2.8. Some shading models are simple by using minimal calculations that interpolate light applied to vertices. These simple models are called local lighting models. Local lighting models are empirical-based because the mathematical models approximate appearances. As seen in Figure 2.7 these models have fewer parameters determining color such as ambient, diffuse, and specular coefficients to determine the amount of incoming light from one direction. These models only calculate direct illumination from the light source not considering indirect illumination from light bouncing off other objects. Also, material surfaces are represented by uniform

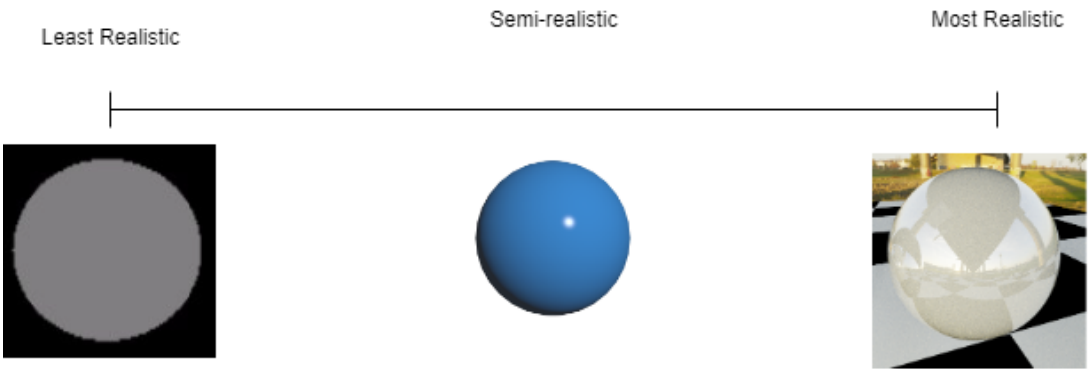



Figure 2.8: Scale representing different levels of realism for a 3D sphere based on the rendering model used. The left sphere uses ambient shading only. The middle sphere is using a local lighting model. The right sphere uses path-traced global illumination.

roughness due to only using the single normal vector.

To achieve more realistic representations of objects, more realistic calculations of how light moves and interacts with materials on a surface should be used. Physically-based rendering is used to simulate how light realistically interacts with an object's surface. Global Illumination is used for simulating how light moves throughout an environment. When using these two methods to shade an object, renderers can achieve highly realistic virtual objects.

### 2.4.3 GLOBAL ILLUMINATION

Global illumination is the process of simulating how light propagates throughout a virtual environment by following light rays reflecting and refracting off surfaces. This allows for surfaces to be illuminated by direct and indirect lighting. Global illumination is the global contribution of both direct and indirect light rays evaluated at a surface point. The model for this light transport is described and solved using the rendering equation expressed by Kajiya et al. [6] below.

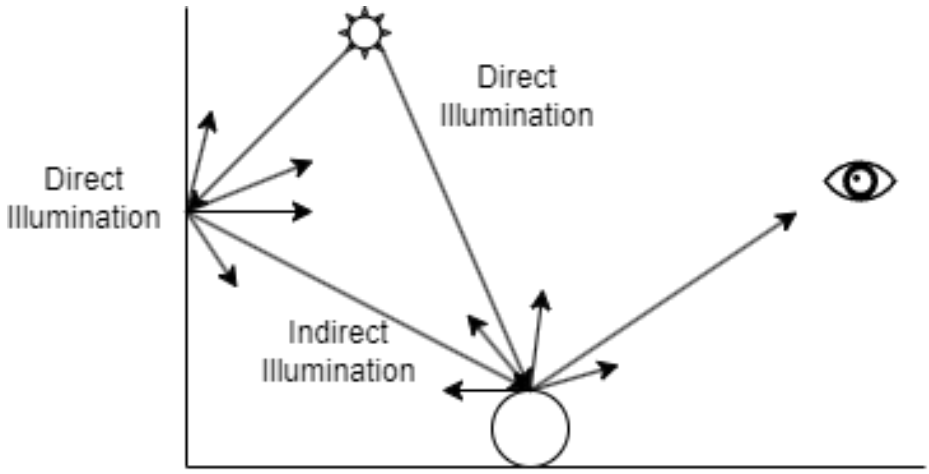



Figure 2.9: Global Illumination model calculating direct and indirect lighting within a scene.

$$L_o(x,\omega_o) = L_e(x,\omega_o) + \int_\Omega f_r(x,\omega_i,\omega_o) L_i(x,\omega_i)(\omega_i \cdot n) d\omega_i \tag{2.1}$$

$L_o$ is the outgoing radiance that is begin calculated and perceived by the viewer, $L_e$ is the emitted radiance if the surface of the object emits light, $f_r$ is the BRDF to be evaluated for the raito of incoming to outgoing light, $L_i$ is the incoming radiance, $\omega_i$ and $\omega_o$ are the incoming light and outgoing reflection directions, and $n$ is the surface normal that is orthognal to a surface. $\Omega$ represents that the equation is evaluating the hemisphere over the surface around the normal for all possible incoming direct and indirect light directions. A BRDF is a bidirectional reflectance distribution function. A BRDF is a mathematical function that describes the ratio of incoming light as a directional vector and outgoing light to the view direction vector for a given point on a surface. Using the rendering equation, accurate evaluation of how light interacts with all objects within a environment ca be evaluated and compared to the real world. As seen in Figure 2.10, this is computed for all incoming directions which is signified by the $\int \omega$ in the equation.

Solving the rendering equation requires a recursive traversal as the radiance at a point on a surface is determined by the outgoing radiance from other objects in the scene where

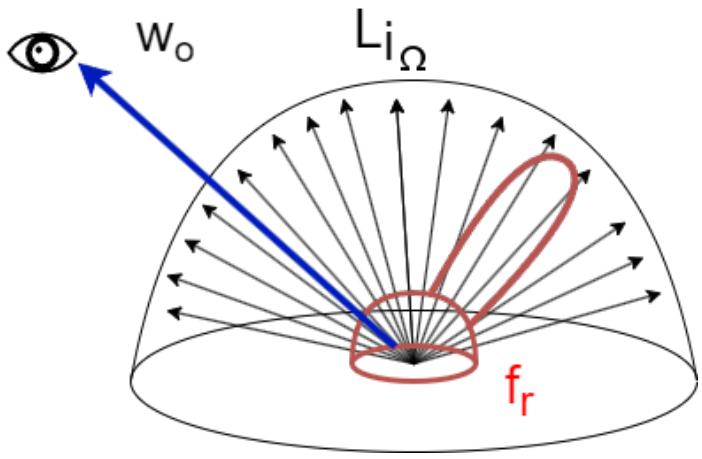



Figure 2.10: Diagram of the rendering equation. For the outward view direction $\omega_o$, all incoming light $L_i$ is evaluated over a hemisphere $\omega$ at an incident position on a surface using a given BRDF $f_r$.

the traveling ray is reflected. When following a ray throughout a scene, it bounces multiple times in different directions depending on the BRDF of the object at a given intersection. These bounces continue until a bounce limit is reached, calculating direct illumination and indirect illumination at surface interactions as it travels. This creates high complexity in rendering computation due to the integral of incoming rays and recursive branching rays seen in Figure 2.9. Monte Carlo algorithms have been developed to solve the rendering equation integral. According to Glassner [7] monte Carlo methods approximate the solution by randomly sampling rays within the hemisphere of an incident position on a surface. One of the most popular Monte Carlo methods is Path Tracing.

#### 2.4.3.1 PATH TRACING

Path tracing simulates the path of light in a backward direction. This means that a ray is followed from the view position to the light position. For each pixel in an image, a ray is shot into the environment. The closest intersection is calculated for the ray and the intersection position is determined and stored. There are different ray intersection formulas for different types of primitives. For instance, a ray-triangle intersection for mesh representations or a ray-sphere and ray-cube intersection for spheres and cubes respectively. At the

intersection point on an object's surface, the direct illumination is calculated by sampling the light source with a light ray. If the light ray intersects another object, it means it is in shadow and doesn't receive direct illumination. Next, a new random direction is sampled based on the BRDF of the surface at the intersection. Recursively, the newly generated ray is traced. This continues until the ray reaches a specified distance or bounce count. Then the radiance from all paths is accumulated to determine the pixel color. This process is similar to Figure 2.9 but in the opposite direction.

As more samples for each pixel (more rays) are calculated, the pixels converge to an accurate image based on the lighting in an environment. Convergence speeds can also vary based on the number of geometry, complex materials, and number of light sources in a scene. Various techniques have been developed to help with the speed limitations, such as importance sampling, next event estimation, radiance caching, RESTIR for multiple light sources, and machine learning methods. Due to long convergence times, early images can be noisy if rays don't converge to a light source. Methods for denoising such as spatio-temporal guidance filtering by using information from previous frames to predict the illumination of a pixel in the next frame can help with noisy images.

Since global illumination is used to simulate realistic lighting in an environment, objects and their surface must be realistically simulated to accurately reflect and refract light during light transport. There are BRDFs that are physically based and not physically based. Physically based BRDFs accurately reflect and refract light closely to the level of the real world to make the surface of an object look realistic. The parameters and BRDFs that mimic these interactions are what is termed physically based rendering.

### 2.4.4 PHYSICALLY BASED RENDERING

Physically-based rendering (PBR) simulates how light realistically interacts with a surface. For interactions to be considered physically based, they must be:

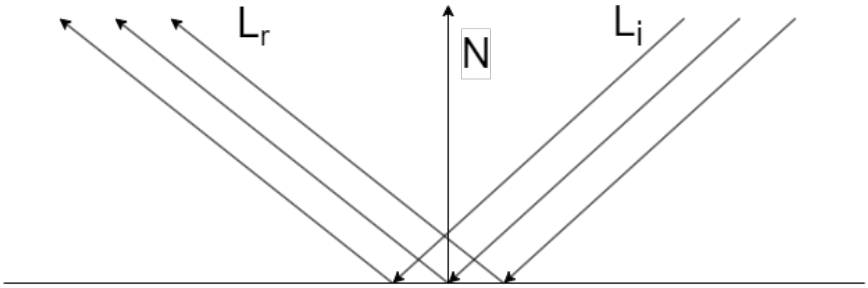
- Using a physically based BRDF
- Energy Conserving
- Use a microfacet surface model within the BRDF

#### 2.4.4.1 MICROFACETS

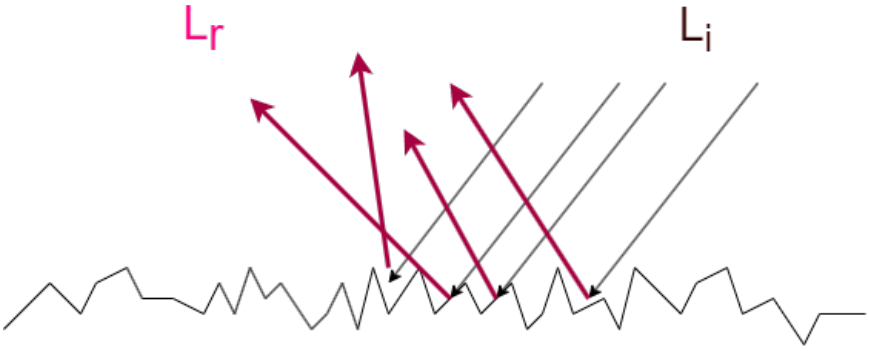
PBR models are based on the theory of microfacets [8]. This theory describes how light interacts with the surface of an object focused on mimicking the interactions at a microscopic level. Real surfaces have many randomly oriented facets called microfacets that act as reflectors of light. These microfacet reflectors are perfectly reflective. Based on the microscopic properties of the surface, the microfacets differ in alignment as seen in Figure 2.11. The more aligned the microfacets are, the more mirror-like the surface appears. Due to the surface of an object at a microscopic level never having perfectly aligned microfacets, light rays will scatter in random directions the more that the microfacets are unaligned.

#### 2.4.4.2 ENERGY CONSERVATION

Physically based rendering models must also be energy-conserving. This means that the total energy that is reflected by a surface at an incident position cannot be more than the energy received. To ensure that energy conservation is true, some rays are treated as reflections from the surface, while some are treated as refractions that get absorbed within the subsurface of the object. Refractions that are absorbed in the subsurface sometimes exit the surface in a random direction. These rays can be considered diffuse reflectance rays. BRDFs must take into account that there should be energy conservation through adjustments in the formulas. One method is through normalization. The radiance from all incoming light should not exceed 1.0. BRDFs can control the diffuse and specular coefficients by making the diffuse contribution $1 - K_s$ where $K_s$ is the percentage of specular reflectance.




(a) Mirror surface representative of perfectly flat microfacets creating specular reflections




(b) Noisy surface with random microfacet distribution creating diffuse reflections

Figure 2.11: Two different representations of microfacet surface geometry with a noisy and smooth distribution creating specular and diffuse rays.

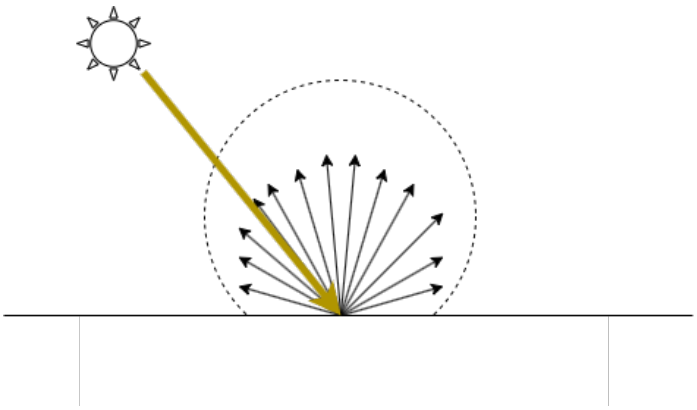


Figure 2.12: Lambertian BRDF showing a uniform diffuse scattering of rays.

#### 2.4.4.3 BIDIRECTIONAL REFLECTANCE DISTRIBUTION FUNCTIONS

Based on the BRDF used, different material types can be simulated such as metallics and plastics. One of the most common BRDFs used for diffuse reflectance is the Lambertian BRDF seen in Figure 2.12. Lambertian BRDF is used to represent diffuse materials where light is scattered randomly after an intersection. Given the formula:

$$f_r(\omega_i, \omega_o) = \frac{\rho}{\pi} \tag{2.2}$$

where $\rho$ is the diffuse albedo of the surface, Lambertian defines a uniform matte or diffuse surface reflectance.

A simple combination of a Lambertian diffuse BRDF can be combined with a simple specular BRDF to create dielectric materials. PBR BRDFs that are energy conserving use more advanced mathematical models, such as the Cook-Torrance BRDF [9] or the Disney principled BRDF [10], to describe how light interacts with surfaces.

It combines the microfacet theory with the Fresnel equations to describe the specular and diffuse reflection of light from a surface.

The Cook-Torrance BRDF expressed by Cook et al. [9] is:

$$f_r(l,v) = \frac{F(v,h)G(l,v,h)D(h)}{4(n\cdot l)(n\cdot v)} \tag{2.3}$$

where $F(v,h)$ is the Fresnel term from Schlick et al. [11], which describes the fraction of light that is reflected at the surface. $G(l,v,h)$ is the geometric attenuation factor using the Smith model from Smith et al. [12], which calculates how many of the micrfacets self-shadow and occlude each other. $D(h)$ is the microfacet distribution function from Beckmann et al. [13], which describes the statistical distribution of microfacet normals. $n$ is the surface normal which is a vector orthogonal to the surface. $l$ is the direction of light hitting the surface. $v$ is the view direction. $h$ is the half-vector between the light direction and the view direction. Combining the Cook-Torrance BRDF and the Labertian BRDF, an accurate representation of specular and diffuse lighting is achieved using the following equation:

$$f_r(l,v) = k_d * f_{labertian} + k_s * f_{cook-torrance} \tag{2.4}$$

The expanded equation is:

$$f_r(l,v) = k_d * \frac{\rho}{\pi} + k_s * \frac{F(v,h)G(l,v,h)D(h)}{4(n\cdot l)(n\cdot v)} \tag{2.5}$$

Combining the final BRDF into the rendering equation is:

$$L_o(x,\omega_o) = L_e(x,\omega_o) + \int_\Omega k_d\frac{\rho}{\pi} + k_s\frac{F(v,h)G(l,v,h)D(h)}{4(n\cdot l)(n\cdot v)}L_i(x,\omega_i)(\omega_i\cdot n)d\omega_i \tag{2.6}$$

The Cook-Torrance BRDF provides a physically based and intuitive way to model the appearance of a wide range of materials, from smooth and glossy surfaces to rough and matte ones. By adjusting the surface roughness parameter $\alpha$ and the base reflectance $F_0$, one can control the specular and diffuse reflection properties of the material.

## 2.5 GLOBAL ILLUMINATION AND RENDERING ADVANCEMENTS IN AUGMENTED AND VIRTUAL REALITY

Advanced rendering techniques can play a significant role in creating realistic objects for augmented and virtual reality. However, due to the computational complexity of lighting models in VR and AR, researchers have been working on developing efficient ways of rendering objects on such hardware. These advancements have made it possible to produce high-quality graphics in VR and AR while maintaining a smooth experience for users. One application of photorealistic rendering in AR is accurate lighting simulation and efficient computation. Santos et al. [14] presented real-time ray tracing for AR showing improvements in computational efficiency and advanced rendering techniques. They created a new rendering pipeline focused on ray tracing for AR environments. Fradet et al. [15] developed Light4AR, a shadow estimator for multiple light sources for more PBR AR environments. Their work used captured real cast shadows based on the estimation of light source positions. Using the captured lighting from environment models and shadow directions from a source object, lighting parameters can be calculated for rendering virtual objects, matching the lighting directions within the real environment.

Pessioa et al. [16, 17] proposed a global illumination solution through dynamic image-based lighting. At every frame, environment maps were created for each virtual object in the scene, with varying levels of glossiness. This allowed for consistent simulation of illumination effects like color bleeding and specular reflections for the virtual objects, even when the scene changed. Alhakamy et al. [18, 19] introduced CubeMap360 for interactive

global illumination rendering, by developing a method to extract lighting information from the scene and use it for rendering virtual objects in AR/MR.

Creating point cloud renderings in AR can be important for using real-world geometry for global illumination calculations for virtual objects in AR. Marroquim et al. [20] developed an efficient point-based rendering method using image-reconstruction. Continuous surfaces from scattered one-pixel projections were generated with a new algorithm. Using pull-push interpolation with integration of elliptic boxfilters can render large point scenes. Rohmer et al [21] created a complete two-stage process introduced to acquire and enhance the real-time camera images using depth sensors from a mobile device. The process involves working on a recorded 3D point cloud of the actual environment, which contains high dynamic range color values.

Recent work by Liu et al. [22] and Plopski et al. [23] have explored real-time lighting estimation for AR using differentiable screen-space rendering and reflectance and light source estimation techniques, respectively, further advancing the state-of-the-art in photorealistic AR rendering.

## 2.6 VIRTUAL ENVIRONMENTS AND HARDWARE

Virtual environments (VE) need to be perceived through a device where virtual objects are rendered to a screen. This is what creates the difference between a real environments and a virtual ones. The device can be a virtual reality headset or an augmented reality headset. These headsets are called head-mounted displays (HMD) and optical see-through head-mounted displays (OSTHMD) respectively. Each of these devices has its set of restrictions and characteristics that differ from reality like field of view (FOV), distorted or fewer depth cues, lower rendering quality, etc. These devices have their own hardware system that allows the user to view and interact with the VE through head tracking, eye tracking, and hand tracking.

### 2.6.1 VIRTUAL REALITY (HMD) HARDWARE

Virtual reality Headsets or HMDs come in a multitude of different brands and specifications that are developed by different companies. The most popular brands of HMDs are Meta Quest and HTC Vive. To see the VE, a virtual image is rendered on optical lenses that are computer-generated to immerse the user in a pre-developed environment. Usually, these HMDs show stereoscopic imagery which has the potential to display a different image in each eye. Also when viewing VE through HMDs, there is less Field of vision (FOV) than that of our real-world vision. Without an HMD our FOV is around 180 degrees. This drops to around 110 degrees in consumer-grade HMDs. Although a lower field of view does lower immersion in the VE, it does not affect depth perception which was previously thought and disproven by Knapp et al. [24] in the directed tasks within the action space. The resolution of recent HMDs is 1600x1200 pixels per eye. Usually, this resolution is described as arc-minutes per pixel per degree which is 10-20 pixels/(1arcmin/pixel). HMDs have peripherals such as head tracking controlled by their positioning system, eye tracking, and hand/controller tracking.

### 2.6.2 AUGMENTED REALITY (OSTHMD) HARDWARE

Augmented reality headsets or OSTHMDs feature instant and unhindered views of the real world with superimposed computer-generated imagery (CGI) on the lens of the device. This is done by projecting the CGI through a partially reflective mirror and viewing the world directly. A popular OSTHMD used is the Microsoft Hololens 2. They come with an onboard processor, display, and sensors for different tasks. Unlike HMDs, AR systems such as the Hololens 2 must achieve accurate positioning of virtual objects in real-world and match the physical surroundings. This happens through its sensors including 4 cameras and a time-of-flight (ToF) range camera which calculates depth information. The localization and mapping of the real environment around the device is calculated through real-time

SLAM. SLAM is simultaneous localization and mapping which constructs and/or updates a map of unknown environments while keeping track of user positioning. The mapped physical environment is represented as a triangle mesh which allows for the accurate placement of virtual objects within the real environment.

## 2.7 DEPTH PERCEPTION IN VIRTUAL REALITY

Virtual reality environments have been gaining widespread popularity not only within casual consumer use cases but also as a tool in academia for studying human behavior. These behaviors can range from redirected walking to 3D volume interactions. One critical behavior within virtual environments is depth perception. Human behaviors involving depth perception include many actions such as throwing an object to a perceived target or moving around multiple obstacles. These actions within the real world require accurate perception of one's environment and recognition of many depth cues. Given the importance of depth perception in the real world, virtual environments distance perception has been thoroughly studied.

### 2.7.1 DEPTH PERCEPTION ISSUES IN VIRTUAL REALITY

Recent studies have reported an underestimation or overestimation of egocentric distance estimations [1, 25], and their capability to transfer to behavior in real environments [26]. Misperceptions also seem to vary by viewer [27, 28]. Waller et al. [29] and Swan et al. [30] examined estimations with VR Head Mounted Displays (HMD) with tasks such as blind walking. They reported that distances in VEs were consistently underestimated.

Some studies looked at comparing the differences in distance estimates between the VEs and real environments. Feldstein et al., [31] performed an extensive review of earlier works (spanning 40 years) on this problem; while their review did not show a significant difference, many factors increased the discrepancy between the two environments. It was

found that distance judgments in real environments are generally accurate. They state that the purpose of the studies to identify what factors are responsible for distance compression observed in VE. They argue that the lack of underestimation of distances was because large-screen immersive displays were used instead of an HMD which has a larger vertical FOV. Also, previous work used visually directed tasks while this one used imagined walking which may account for the similarity found. Plumert et al. [26] conducted experiments and found that people's time-to-walk estimates were similar across real and virtual environments up to 60 feet and underestimations were found after that distance. The lack of underestimation was due to the use of large-screen immersive displays instead of an HMD, which has a larger horizontal field of view (FOV). Visually directed tasks might also account for the similarity between real and virtual environments. In contrast, Feldstein et al's work reports an average ratio of virtual to real distance estimates of 77% using HMDs. More recently, Jamiy et al. [32] studied whether feeding real-world images or live video into an HMD would influence distance estimation. They found that distance compression with live video averaged 80.2% accuracy and the real-world images averaged 81.4%.

Virtual environments block out all aspects of the outside world. Developers need to build the environment to mimic the dimensions of the real world to give the user access to depth cues that can be used while interacting in this virtual space. This is a product of immersion in VR where the user is provided with the feeling of presence in the environment they see. To give the user a feeling of immersion and access to depth cues, characteristics should be supplied in the environment such as field of vision, stereoscopic rendering, virtual objects, and illumination of the visual environment. In the VE, detail provided by lighting and geometry influences the egocentric perception of an object being observed. The fewer visual cues provided within the VE, the more egocentric distance misperception is influenced.

Research has tried to manipulate virtual environments to influence the incorrect sense of space. Peer et al. [33] evaluated the effectiveness of correcting perceived distance and

then manipulated the sense of space to appear more distant or closer to the user. The difference from past work is that in this paper the author's personalized warps of the virtual environment on an individual basis based on the performance from the blind throwing task. This space warping was calculated by using a relative percent error between performance in the real world and performance in the virtual environment. The virtual objects were offset in the vertex shader to match the real world in the user's gaze direction.

Cidota et al. [34] how visual effects on 3D virtual objects in a virtual environment provide improved depth perception. Their experiment aimed to measure the impact of the blur and fade effects on depth perception accuracy. They concluded that the visual effects had a negative impact on depth perception.

While researchers are investigating the software side of the depth perception issue, other work has been done on the mechanical side related to the HMD device itself and its variables. Willemsen et al. [35] conducted experiments aimed at determining if mechanics related to the HMD is linked to the distance distortions. They focus on the mass and moments of inertia relating to wearing the HMD. They tested users wearing a normal HMD and a mock HMD matching mass and moments of inertia of the real HMD. The results conclude that these aspects of the HMDs do not explain the underestimation of depth recorded in virtual environments. Willemsen also did more work [36] on stereo viewing conditions and their effect on distance perceptions. This involves manipulating stereo viewing conditions in the HMD. The manipulations are both measured and fixed interpupillary distances. Using measured IPD did not improve overall performance compared to having a fixed IPD of 60.5mm. Similarly, Ponto et al. [28] improved by introducing a calibration process to determine individual viewing parameters accurately. The parameters of this model are the center of projection (PO) and the binocular disparity (BD). Their 20-subject study was able to improve multiple aspects of perception including, depth acuity, distance estimation, and perception of shapes.

### 2.7.2 VR RENDERING QUALITY

In the past, realism in VEs was thought to possibly be linked to distance estimation and an element for identifying misperceptions. Since the resolution in which we see and interact with objects in the real world is of unmatched quality, the change to unrealistic rendering had an effect on the user's perceived space. Virtual reality environments in early depth perception studies were simple. This was because GPUs were not as powerful as they are now. Early studies 'high-fidelity rendering' conditions were not as photo-realistic as they would be with today's GPUs. Modern hardware can create much more complex meshed and shading models. Due to this increase in computational power, there has also been an increase in screen resolution and environment texture resolution.

#### 2.7.2.1 SCREEN/ENVIRONMENT RESOLUTION

Thompson et al [37] compared wireframe graphics with panorama images and found that there is no difference in depth estimations. Iuliu et al. [38] investigated the influence of texture fidelity in VR and found that high-fidelity textures had a positive influence on precision but no significant influence on accuracy. Vienne et al. [39] found that bias in depth perception is proportional to the accommodation-vergence conflict size. The conflict had less influence on bias when multiple depth cues were available in a rich virtual CAVE environment. Buck et al. [40] results showed significant distance compression in the Vive Pro with higher resolution, similar to the original Vive. This suggests display resolution does not impact distance compression.

#### 2.7.2.2 ENVIRONMENT QUALITY

Knapp et al. [5] believed that the increase in rendering environment complexity and quality of graphics would improve perceived depth in VEs using VR HMDs. However,

research findings that tested Knapp's hypothesis were not conclusive. This conflicted with the work by Phillips et al. [41] who created multiple VEs, one with a photorealistic VE and one with a non-photorealistic VE. Testing participants in blind walking and verbal responses showed significantly better results in the photorealistic environment. Kunz et al. [42] conducted experiments, where two graphical quality versions of a lab room were constructed. The low-quality version had low-resolution polygonal models of virtual objects and low-spatial-frequency generic textures on objects within the virtual environment such as walls, ceilings, and floors. The high-quality rendering had high-spatial-frequency generic textures and realistic lighting. In the first experiment, a blind walking task was used. Higher-quality and low-quality rendering had similar results with higher-quality rendering performing slightly better. In the second experiment, the methodology was the same except participants used verbal reporting. This experiment resulted in significantly more accurate results with the high-quality rendering version. Gerig et al. [43] found that additional rendered depth cues (texture gradient, shadows, aerial perspective) perform slightly worse than participants without in a VR reaching task. They believe this is because the additional cues were not only unnecessary but distracting. Vaziri et al. [44] found a small but significant difference between real-world and video/non-photorealistic rendering conditions in distance underestimation, but no significant difference between regular video and non-photorealistic rendering/video.

## 2.8 DEPTH PERCEPTION IN AUGMENTED REALITY

Augmented Reality (AR) is the method of rendering virtual objects superimposed onto the real world from an optical see-through head-mounted display (OSTHMD) [1]. Given that the device is see-through, users, unlike VE in HMD, are able to use real-world depth cues. In AR systems, depth cues such as lighting models, shadow quality/methods, and texture gradient calculations are crucial for blending virtual items seamlessly into the real envi-

[1] Video see-through HMD AR and video, see-through tablet AR is also available.

ronment. However, transferring these properties to the virtual objects in real-time can be challenging, as they are generated by the environment itself and begin to reach the limitations of the hardware in terms of performance and capabilities. To measure distances in the action space, researchers often use OSTHMD and blind walking, throwing, perceptual matching, and other tasks to examine virtual objects and judge their depth in the surrounding environment. A large part of the work in depth perception has focused on egocentric depth judgments which focus on the distance estimation from an observer to a perceived distance [1, 30, 45–55].

### 2.8.1 DEPTH PERCEPTION ISSUES IN AUGMENTED REALITY

Many early studies were conducted in AR environments and found that egocentric depth perception was underestimated. [56] [30] Swan et al. [30] found that the results were consistent with previous studies that believed this was due to restricted FOV, but the misestimations were smaller than previously found. In a more recent study, Ping et al. [45] compared depth estimation between VR and AR and found that AR estimation is more accurate than VR but still underestimated with an error that grows with distance. Ishio and Miyao [47] revisited old research that found convergence and accommodation in the human eyes occur in response to a virtual image. They found that accommodation and convergence are important and should be considered when using binocular see-through smart glasses. Adams et al. [57] compared depth perception and distance estimation in two AR headsets. Their results showed that video see-through AR displays may cause more underestimation in comparison to OSTHMDs. They also found that drop shadows and object height impact perceived depth. Krajancich et al. [58] developed a gaze-contingent stereo rendering algorithm that significantly improved the alignment of virtual objects with physical objects when compared to standard rendering using calibrated IPD.

Most work in AR VEs devoted to depth perception has investigated egocentric depth

judgments [1] [45] [46] [47] [48] [49] [50] [51] [59] [30]. Such studies typically test the user by having them perform blind walking or triangulation by walking tasks to walk toward previously seen objects. To test the influence of stimulus on depth perception, AR research has used distance judgments to on and off-ground objects [50], walking tasks or reaching tasks [49], perceptual matching tasks [53] [46], direct pointing [54], verbal reports of depth [55], or combinations of these methods. With the use of different testing methods for distance judgments in AR, researchers have developed different methods of solving or identifying the problems with underestimation in AR. For example, recent work tested the relationship of different shading models and rendering characteristics that contributed the most to depth perception [59].

### 2.8.2 AR RENDERING QUALITY

AR environments are quite complex because the background is of the highest fidelity. Ping et al. [52] conducted multiple studies that tested combinations of rendering techniques to study how they contributed to depth perception performance. They concluded that realistic lighting models that use environment lighting and shading applied to realistic objects can help the participant relate the real-world depth cues with the virtual object's position. This was also found in a study by Diaz et al. [46] where drop shadows/cast shadows and shading from objects played a large factor in determining the depth of an AR object. In isolation and combination, these cues improved the participant's depth performance. Shading models can play an important role in a user's perception of depth in AR because the virtual object appears alongside real-world objects. Ping et al. [52] continued exploring the effect of opacity and three shading models on aligning virtual objects while also comparing the relationship of color and size in distance perception. Their conclusion showed that users perform better in depth matching tasks with green and yellow virtual spheres rather than blue ones. In the three shading models, users had significantly more accuracy in the depth matching with the Cook-Torrance shading model, in comparison to the Half-Lambert and

Blinn-Phong models.

# CHAPTER 3: PROBLEM AND MOTIVATION

## 3.1 PROBLEM

In both VR and AR, there are conflicting results in understanding if graphical fidelity has an influence on depth perception. There are multiple aspects that could be affecting results along with fidelity. It was believed that realism in VEs was associated with distance estimation and influenced misperceptions. However, previous studies were conducted before consumer headsets became widely available. It's crucial to reassess these findings considering advancements in VR/AR technologies for a better understanding of depth perception in modern immersive environments. Murgia and Sharkey [60] conducted a study concluding depth estimation was underestimated in both poor and rich cue environments with greater underestimation in the poor cue environment. Egocentric depth perception is not affected by image quality according to Thompson et al [37]. Benjamin [42] also reported similar findings. However, work by Díaz et al. [61] and Phillips et al. [41] investigated different rendering styles of volumes and environments respectively and contradicted earlier findings. Díaz et al. compared multiple shading models (No shading, Phong shading, Half-angle slicing, Direction occlusion shading) on volume data sets and found that Direction Occlusion shading performed significantly better than No Shading in user tasks, suggesting that advanced volume illumination improved depth perception. Phillips et al's work had similar results, and found that people tend to underestimate distance by a greater margin, when comparing renderings based on line drawings to its high-fidelity counterpart. Earlier studies only compared certain subsets of aspects that pertain to higher fidelity graphics (only background, only textures, only shading models).

D´ıaz et al. [61] stated, "If depth perception was just related to stereoscopy (creating or enhancing the illusion of depth in an image by means of depth perception), we would have expected a similar performance among the compared shading approaches. Instead, the results of both tasks show significant differences between them. Thus, we can state that depth perception is not just due to stereoscopy but also the shading plays a role..." This quote was stated when investigating how shading affects the perception of volumetric models. Earlier studies only compared certain subsets of aspects that pertain to higher fidelity graphics (only background, only textures, only shading models).

## 3.2 EXPERIMENTS

The experiments presented in this work aim to better understand how depth perception performance is influenced by depth cues and graphical attributes applied to three aspects of a participant's visual field.

- foreground target (pegs): stationary objects whose distance the participant is trying to match.
- foreground manipulated object (sphere): an object whose position the participant controls and whose distance must be matched to the target
- background and peripheral objects: all other points in the participant's visual field excluding the target and manipulated object are termed the background.

In AR and OSTHMD displays, any of the three visual field aspects can be real or virtual because the light field of any real part comes directly from the participant's physical environment. There are 8 possible combinations of conditions, but the combination of real-real-real involves no display technology, therefore an AR perceptual matching experiment has 7 possible combinations.

To my knowledge, no prior perceptual matching experiment has systematically explored the 7 combinations we have identified using a common display system. To address this gap in knowledge, we have designed a set of experiments that will allow us to investigate how depth estimation errors change when switching between these different combinations while holding all other variables constant (FOV, display resolution, tracking latency, pixel illumination levels, pixel contrast ratio, etc.). At present, one can only attempt to answer these questions by comparing and reconciling results across different prior works, each of which used different display technology, and then trying to intuit in what way the differing display technologies are responsible for multiple prior works' contradictory results. This is not satisfactory.

By varying the 7 combinations using a common display, can one answer questions such as:

- How do depth estimate errors change between the "virtual manipulated object"-"virtual target" condition and a "virtual manipulated object"-"real target" condition?
- Do the above results change if the background is virtual or real?
- Do the above results change if the virtual background is a virtual representation of the real physical background versus a virtual background that is a uniformly illuminated white sphere (rendered as an environment map)?

We can also begin to understand how the visual congruence between the three aspects of the visual field affects depth perception accuracy. For example, how does depth accuracy compare between all aspects being virtual and some aspects are real and others are virtual? We also explore different levels of visual fidelity effects on depth perception. Visual fidelity refers to the accuracy and quality of visual representation. The virtual background/foreground targets have the lower visual fidelity and the real background/foreground targets have the highest visual fidelity.

A perceptual matching framework was designed for testing multiple depth cues and rendering methods to the background, foreground, and manipulated object across two experiments.

We chose an AR OSTHMD, the Microsoft HoloLens, for this exploration. In particular the two experiments provide a direct comparison between virtual-to-virtual and virtual-to-real perceptual matching performance.

- Virtual-to-virtual: refers to a virtual object matched to a virtual target (pegs).
- Virtual-to-real: refers to a virtual object matched to a real target.

We also explore how physically based rendering (PBR) and baked global illumination (GI) applied to depth cue information and graphical attributes affect depth perception in a perceptual matching task for indoor augmented reality environments. Independently manipulating the PBR attributes applied to the manipulated object and target and creating lighting conditions that match the real world allows for testing how more realistic depth cues play a role in depth estimation performance. We also compare my experiments to understand the congruence of graphical fidelity across three visual field aspects: background, foreground, and foreground-manipulated object. The first two experiments compared virtual-to-virtual and virtual-to-real perceptual matching tasks. The tasks systematically varied the background and foreground depth cues and rendering conditions to understand the interaction effects between subsets of the visual field being real or virtual and how it affects participant depth perception performance. In those experiments, the virtual aspects of the target and manipulated object were unrealistic and rendered with non-PBR techniques.

In the third experiment, we address the gap in knowledge on whether all three aspects of the visual field are realistic or real help performance due to the congruence of depth

cue and environment information. Using PBR methods and global illumination, the manipulated objects and targets can match the fidelity of the real environment. By analyzing the full spectrum of visual field congruence from fully virtual and unrealistic to partially virtual and fully realistic, we aim to understand the interaction between aspects of the virtual environments. We also determine whether high-fidelity graphics rendering through advanced PBR methods can improve depth perception accuracy, in comparison to the simpler rendering techniques in my experiments one and two. Results also aim to establish PBR rendering guidelines for AR applications and the balance for rendering complexity for perceptual performance benefits. Below is the general procedure we followed for each experiment.

## 3.3 EXPERIMENTAL HARDWARE

The Hololens 2 was used to render the virtual stimulus (background, pegs, sphere). All participants used the same Hololens to view the 3D environment. The virtual renderings were created in Unity [62], a popular game engine. Once the environment was constructed on Unity, the program was compiled and deployed onto the Hololens 2. All participants conducted the experiments in the same room under the same lighting conditions. We have full control over the position of objects in the environment and the customization of the object's properties.

The participants used a Microsoft Xbox controller that was connected through Bluetooth to manipulate the sphere and submit a response. The left analog stick was used to move the sphere forward and backward (away from and toward the participant respectively). The A button was used to submit a trial response.

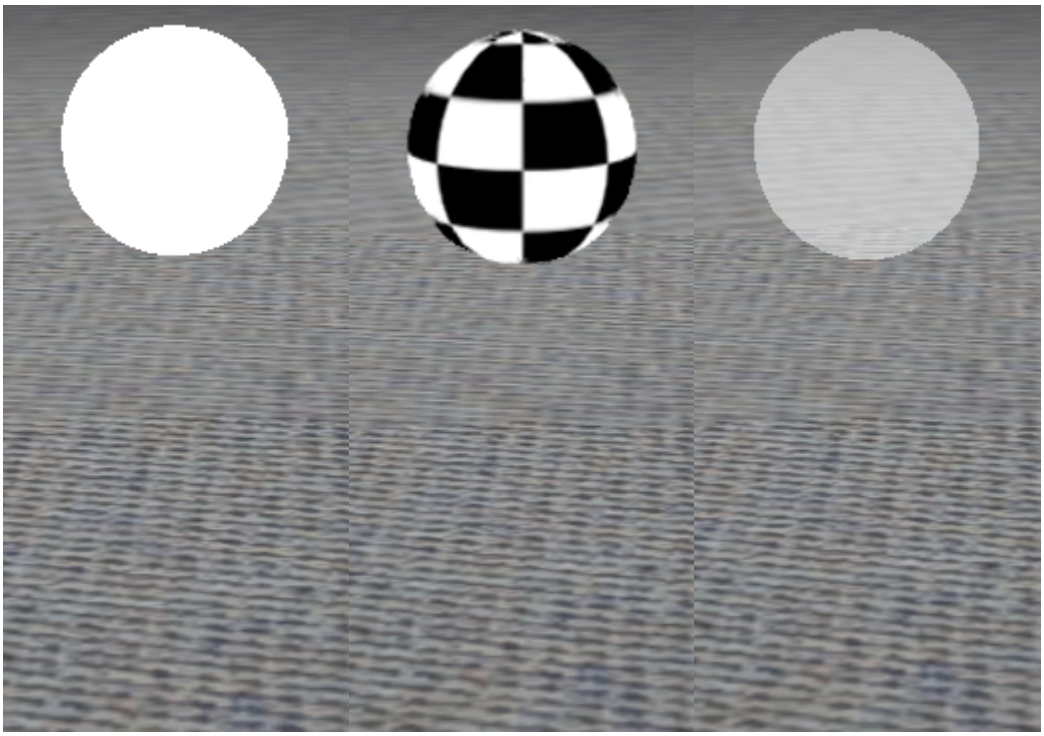
## 3.4 PERCEPTUAL MATCHING TASK

The perceptual matching task was designed to investigate depth and rendering cues when applied to the background, foreground target, and foreground manipulated object. Participants, who were seated, were asked to use a joystick to move a virtual sphere to match the location of a pair of pegs that were positioned at one of three locations (close, middle, or far) from them. The perceptual matching task was adapted from Diaz et al's [46] work, where they used virtual spheres and varied the conditions as follows: drop shadows, texture gradient in a checkerboard pattern, shading, opacity, and a control case with no depth cues. The task allows for the isolation of head movement and other possible confounds that may affect results, such as motion sickness.

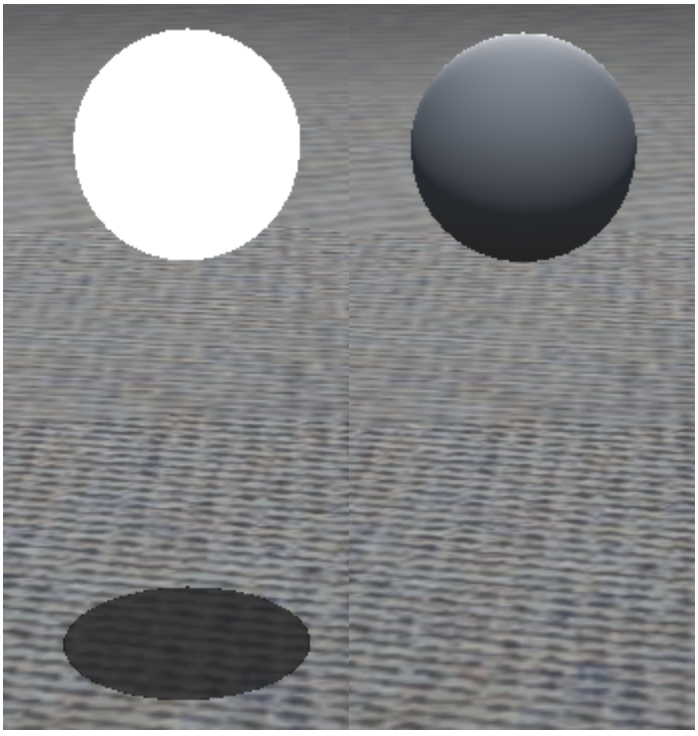
Figure 3.1 shows the five sphere conditions that we investigated. Three of the conditions added depth cues to the sphere while the other two (opacity and white) were used for comparison.

- *Texture gradient* is the gradual change in the appearance of objects so that the closer they are the more distinct the texture elements. The texture becomes less apparent the farther away it is.
- *Shadows* are a dark area that appears when an opaque object occludes a given light source for a specific surface. It is an important monocular depth cue.
- *Object shading* is the process of determining the color of an object's surface based on the relative 3D position to a light source within a 3D environment.

Billboarding, in which the virtual object is rendered perpendicular and facing the camera, from Diaz e tal. [46] was not used because the participants were stationary. The sphere was 1m away from the participant in the starting position and 0.7m off the ground to be centered with the peg's height. The sphere had a diameter of 0.4m.



(a) Control (b) Textured (c) Opacity



(d) Shadow (e) Shading

Figure 3.1: Sphere Experimental Conditions.

Each experiment had three pairs of pegs. The pegs are 0.2m in width, 0.4m in depth, and 1m in height. Each pair has one on the left and one on the right of the participant's front FOV. From the center of the FOV, the pegs are offset by 1m to the right and 1m to the left. The sets of pegs are spaced equidistant apart starting from 2.44m - 5.64m. The distances were chosen to replicate previous work from Diaz et al. [46], where these distances represent an equal interval throughout the action space.

There were two peg conditions (rich and impoverished). The rich pegs had added depth cues while the impoverished condition did not, as shown in Fig 6.1.

The pegs are modified versions from Swan et al. [30], located on both sides of the participant's FOV. Swan et al. pegs were adapted to avoid 2D solvable geometry, such as overlapping objects and same-size comparisons. Since the peg and sphere are completely different shapes, participants must rely on information from the isolated depth cues to judge the depth rather than the inherent attributes of the shapes.

To log participant performance data, the Universal Window Platform (UWP) was used to write data to a JSON file describing rendering and depth information of the environment along with metrics such as signed distance judgment error, time to complete, and distance from the user.

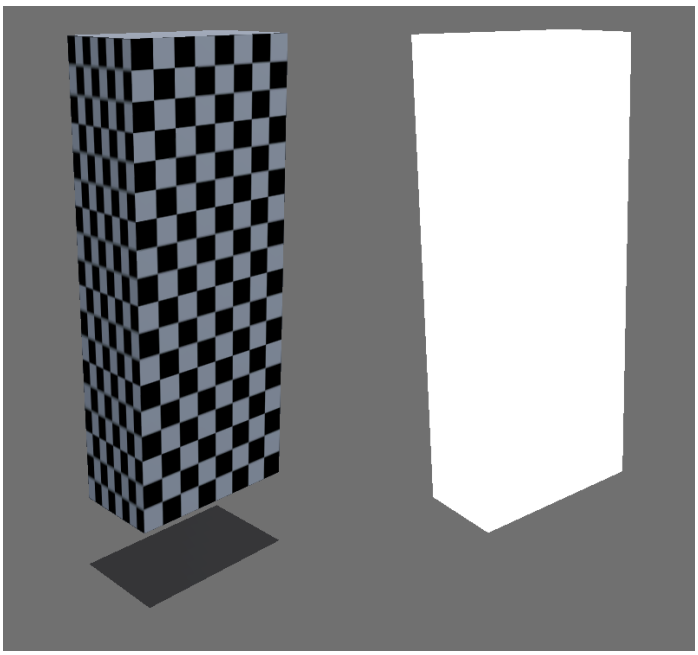
## 3.5 EXPERIMENTAL PROCEDURE

Students were recruited from the University of North Carolina at Charlotte SONA system, which allows psychology undergraduate students to participate for research credit. Each participant was awarded one research credit for 45-60 minutes of their time. Participants were at least 18 years old and had stereoscopic vision. After filling out an informed consent sheet, they had to pass an anaglyph stereo test, where they had to identify a set of shapes from colored stereo images.

After receiving directions about the task and how to operate the hardware, the participants were given 10 practice trials to familiarize themselves with the task. Participants were run individually in 45-minute sessions held in a room that was 7.0104m $\times$ 9.7536m. Participants were seated in the center of the room against the front wall opposite to the screen shown in Fig 3.

On each trial, participants would see one of the randomly rendered scenes with different depth cues applied depending on the experiment. They were instructed to move the sphere to match its center with the center of a specific pair of pegs. Participants were also instructed to stay stationary with minimal head movement. A word *Near*, *Middle*, or *Far* was displayed in the center of the participant's field of vision on each trial to denote the target set of pegs. *Near* corresponds to the depth 2.44m, *middle* corresponds to 4.04m, and *far* corresponds to 5.64m. Peg sets were randomly chosen and appeared an equal number of times at each of the three distances. In the top left of the participant's FOV was a number that displayed how many trials were completed. The sphere would always start 1m from the participant but have different variable speeds to avoid the confound of memorizing movement timing. This velocity was randomized between v=3.0 m/s and v=5.0 m/s. The movement was then calculated using the following formula: movement = z * Joystick Axis * v. z is the current z position of the object and Joystick Axis is the magnitude of directional movement of the joystick in the positive z or negative z direction. Pulling backward on the joystick results in values from [-1, 0) and pushing forward results in values from (0, 1]. After 75 trials had been completed, participants were asked to take a 5-minute break to minimize eye fatigue and possible motion sickness.

This research was conducted in accordance with the ethical guidelines set forth by the Office of Research Protections and Integrity, and all participants provided informed consent. Approval for this study was granted under IRB protocol IRB-22-1025.



(a) Rich (b) Impoverished

Figure 3.2: Peg Experimental Conditions.

# CHAPTER 4: VIRTUAL-TO-VIRTUAL STUDY

In this experiment, the virtual spheres were matched to one of the three pairs of virtual pegs presented in the foreground. The depth cues and graphical details described by Howard [63] and Diaz et al [46] were applied to the sphere and pegs (texture, shading, drop shadow, opacity) and viewed against VR, AR, and no background conditions. By splitting the scene into three different aspects for independent rendering: manipulated object, foreground frame of reference, and background, we could test multiple depth cues and rendering methods to determine if they had a singular effect or interacted across different aspects of the scene. In particular, having a virtual or real background was used to see if higher levels of information and fidelity can help with depth perception through interaction effects.

Previous depth perception studies typically used virtual-to-real perceptual matching because, in augmented reality, users are mapping virtual objects to some position in the real world. In VR environments, there is only virtual space so all perceptual matching tasks are virtual to virtual. That being said, AR still requires making virtual-to-virtual distance judgments when not matching to a real-world object.

We focused on using a virtual sphere and virtual pegs with different depth cues and a background that can be virtual or real to test the following hypotheses.

**H1**: The AR background is expected to provide the most accurate perceptual matching compared to the VR and the No Background conditions because of the realistic background.

**H2**: Depth cues when added to the sphere are expected to improve accuracy in the

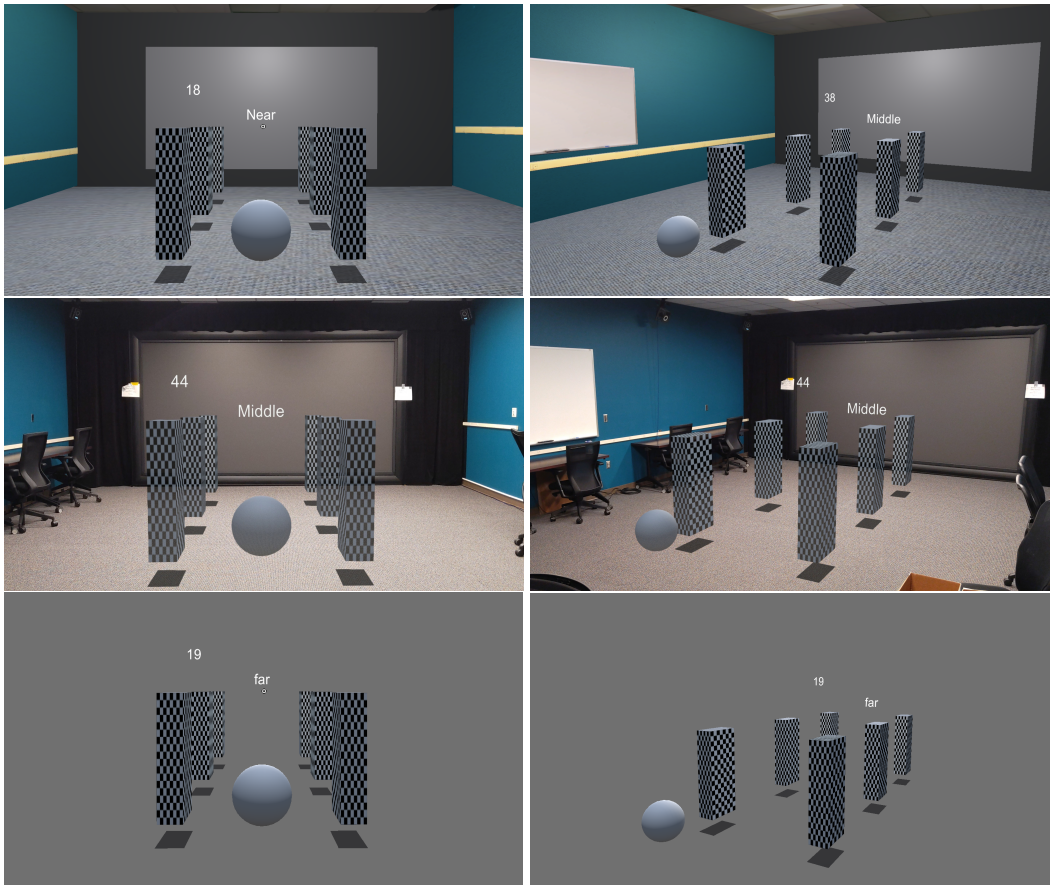



Figure 4.1: Experimental setup of 3 trials showing different background conditions from the POV of the participant and from the side; *Top Row:* Experiment 1 with a shaded sphere, rich pegs, and VR background, *Middle Row:*. Experiment 1 with a shaded sphere, rich pegs, and an AR background, *Bottom Row:* Experiment 1 with a shaded sphere, rich pegs, and no (VI) background.

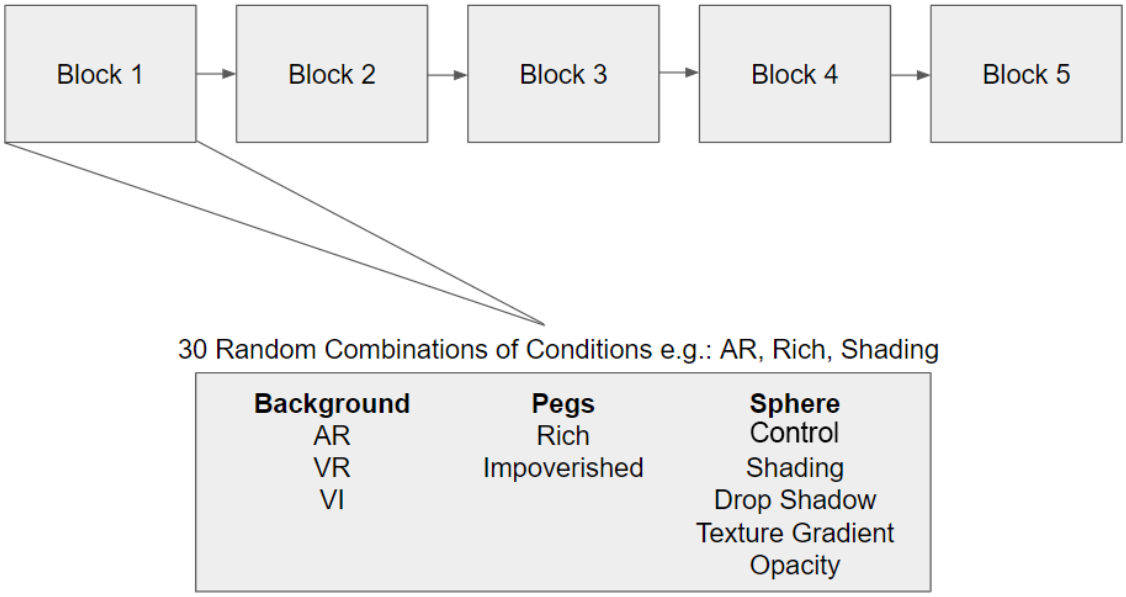



Figure 4.2: Virtual-to-Virtual Experimental Design

perceptual task when compared to the control sphere.

**H3**: The rich pegs are expected to improve performance relative to the impoverished pegs.

**H4**: Interaction effects are expected among the background, peg, and sphere conditions.

## 4.1 METHODOLOGY

### 4.1.1 PARTICIPANTS

Thirty-one participants (20 males, 11 females) (White: 18, Black: 6, Asian: 6, Hispanic: 1) took part in this experiment. Each participant completed 150 trials for a total of 4,650 trials in the whole experiment.

### 4.1.2 EXPERIMENTAL CONDITIONS

During each trial, participants are presented with a randomized combination of rendering conditions for the background (3 conditions), the pegs (2 conditions), and the sphere (5 conditions). The 3 background conditions are called VR (“Virtual Reality"), VI (“Visually Impoverished"), and AR (“Augmented Reality"), They are illustrated in Fig. 6.5. In the VR condition, the background is a virtual rendering of the physical room with texture mapping, lighting, and shading. In the VI condition, the background is rendered with a uniform, blank gray environment map. In the AR condition, no background pixels are rendered, so the user sees the physical room through the OSTHMD display.

Each of these conditions represents a different level of graphical fidelity, having a different level of congruence with the rendering conditions of the pegs and sphere. The VR background condition emulates a VR headset on the HoloLens OSTHMD. Switching to an actual VR headset would add undesired covariates into the experiment because the VR headset would differ from the HoloLens in FOV, resolution, focal distance, weight, etc.

The physical room’s lighting level is fixed for all participants so that the VR and VI conditions’ virtual backgrounds generally occlude the details of the physical room and remain sharp and in focus.

Table 4.1: Two different conditions of the virtual pegs and their depth cues.

| Peg Conditions | |
|---|---|
| Rich | Impoverished |
| Lambertian Shading | Ambient Shading |
| Texture Gradient | None |
| Drop Shadows | None |

The virtual pegs have 2 rendering conditions, “rich" and “impoverished". Table 4.1 lists the rendering attributes of each condition. The two conditions are shown in Fig. 5.3.

The sphere has 5 rendering conditions labeled as follows:

- *Control*: Ambient shading only. This is the control condition with minimal depth cues.
- *Drop Shadow*: Ambient shading plus a drop shadow.
- *Opacity*: Ambient shading plus transparency using alpha blending (alpha = 0.5).
- *Shading*: Lambertian shading.
- *Texture*: Ambient shading plus a checkerboard texture image.

### 4.1.3 ANALYSIS

The experiment used a repeated measure design. Within each block of 30 trials, five sphere conditions, 3 background conditions, and 2 peg conditions were randomly presented at the three distances (near, middle, and far). At the beginning of each block, the combination was re-randomized for a new order. In total, there were 150 trials for each participant.

The distance estimation errors from each participant were averaged across the 5 replications for each of the 30 experimental conditions. The data from any given trial were trimmed if the response speed exceeded 2.5 times the standard deviation from each participant's mean and if the distance error was greater than 1 meter.

The mean trimmed distance estimation errors were analyzed with a $3 \times 2 \times 5$ repeated measures of analysis of variance (ANOVA), which tested for the main and interaction effects of the background (virtual reality, augmented reality, virtually impoverished), peg (rich impoverished), and sphere conditions (control, shading, drop shadow, texture gradient, opacity).

For the 31 participants, the mean distance estimate errors were calculated for each of the

30 experimental conditions across 150 total trials. A significance level of 0.05 was used for all statistical tests and the Greenhouse-Geisser correction was made to the p-value where appropriate to protect against possible violations of assumptions of sphericity. Bonferroni post-hoc pair-wise tests were used when the main effects were found to be significant.

### 4.1.4 RESULTS

On average 1.3% of the trials were trimmed from a given participant. Two participants were removed; one for failing the stereo vision pretest, and another for not following instructions to move the sphere on every trial. For the remaining 31 participants, the mean trimmed distance estimate errors were calculated for each of the 30 experimental conditions across 150 total trials.

#### 4.1.4.1 DISTANCE ESTIMATIONS

All main effects were found to be significant. The analysis of responses to different backgrounds showed considerable differences in distance estimation errors among the three AR, VR, and VI backgrounds ($F(2,60) = 96.8$, $p < 0.001$, $\eta_p^2 = 0.763$). The means were 0.079m, -0.002m, and 0.031m respectively. Follow-up Bonferroni tests showed statistically significant differences among all three background conditions p's < 0.001. Both the AR and VI conditions resulted in overestimated distance errors while the VR condition had a slight underestimation. The VR condition was the most accurate and the AR condition had the largest distance errors.

The main effect of sphere conditions was also significant showing an impact on distance estimations ($F(4, 120) = 4.9$, $p = 0.008$, $\eta_p^2 = 0.141$). Table 4.2 presents the average distance error rates for the sphere conditions. Follow-up Bonferroni tests showed that the texture gradient condition resulted in larger error rates in comparison to the control ($p = 0.001$), and opacity condition ($p < 0.001$). Surprisingly, all sphere conditions averaged an

Table 4.2: Breakdown of sphere conditions distance estimations errors

| Sphere Condition | Mean | Standard Deviation |
|---|---|---|
| Control | 0.027 m | 0.020 m |
| Drop Shadow | 0.031 m | 0.017 m |
| Opacity | 0.020 m | 0.022 m |
| Shading | 0.040 m | 0.022 m |
| Texture | 0.061 m | 0.023 m |

overestimation.

The two peg conditions had a strong and opposite effect on perceived depth ($F(1, 30) = 107.8$, $p < 0.001$, $\eta_p^2 = 0.782$). The rich pegs resulted in underestimated depth judgments (M = -0.31) while the impoverished pegs (M = 0.102) were associated with an overestimation of depth.

The peg conditions were also found to interact significantly with both background conditions ($F(2, 20) = 6.85$, $p = 0.002$, $\eta_p^2 = 0.186$) and sphere conditions ($F(4, 120) = 39.059$, $p < 0.001$, $\eta_p^2 = .566$). As shown in Fig 5a, the background conditions with impoverished pegs showed an overestimation while all background conditions with rich pegs showed an underestimation.

Similarly, for the impoverished pegs, all estimations for sphere conditions were overestimations; but for the rich pegs, the sphere depth judgments were much more accurate with some slight underestimations. Surprisingly, the interaction between pegs and drop shadows resulted in the largest misestimation.

As shown in Fig 5, the third two-way interaction between background and sphere was also significant ($F(8,240) = 2.23$, $p = 0.026$, $\eta_p^2 = 0.069$). In the AR condition, all sphere conditions resulted in an overestimation of depth. The VI background condition was also associated with overestimation but to a much lesser extent than in AR. With the VR background, however, depth estimations were quite accurate across all sphere conditions except

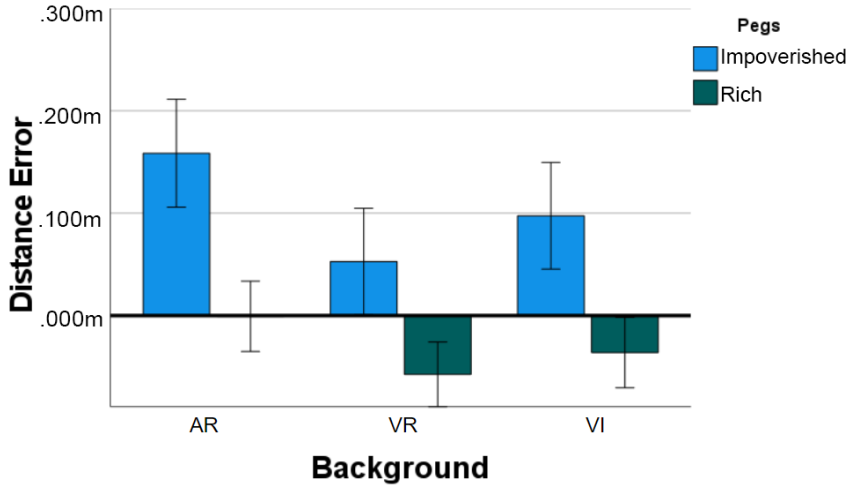
the sphere texture condition. The only interaction effect that was not significant was the three-way interaction between background, sphere, and peg ($F(8, 240) = 1.858$, $p = 0.086$, $\eta_p^2 = 0.058$)

#### 4.1.4.2 DISCUSSION

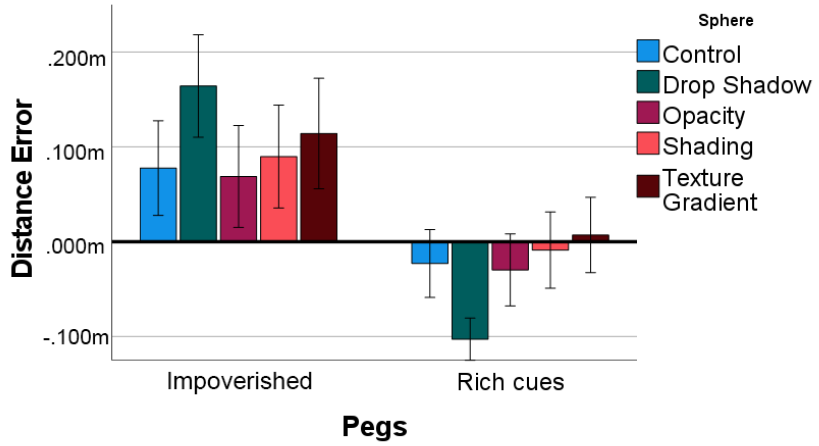
In the AR condition, depth judgment is more accurate than VR for the rich pegs. For the impoverished pegs, AR and the VI conditions become less accurate than the VR condition. These effects can be seen in Fig 4.3a. This partially supports H1. A possible explanation is that accurate depth estimation requires congruence among the three independent rendering variables of the scene. That is the rendering conditions of the foreground pegs, sphere, and background must match. Past research shows that AR is usually more accurate than VR when performing virtual-to-real matching. That would mean that the pegs and background are congruent, but not the manipulated sphere (virtual-real-real). There was only congruence of the pegs and sphere (virtual-virtual-real). Also, VR and VI conditions switched to an underestimation when more cues were applied in the rich pegs. This seems to be a pattern where when more depth cues are applied to a target, the more underestimation.

There was no support for H2 when looking at background and sphere interaction effects. As seen in Fig 4.3c many of the sphere conditions performed worse or similar compared to the control condition for a given background condition. In VR, drop shadow had a positive influence and in the VI background, only opacity improved judgments. This is the opposite of what was reported by Diaz et al [46]. The main difference between my study and Diaz et al's study is that my study used virtual pegs to indicate the target distance while they used physical markers to indicate the target distance. Also, Diaz et al had the physical markers directly below the manipulated object; therefore drop shadows made distance matching very easy.

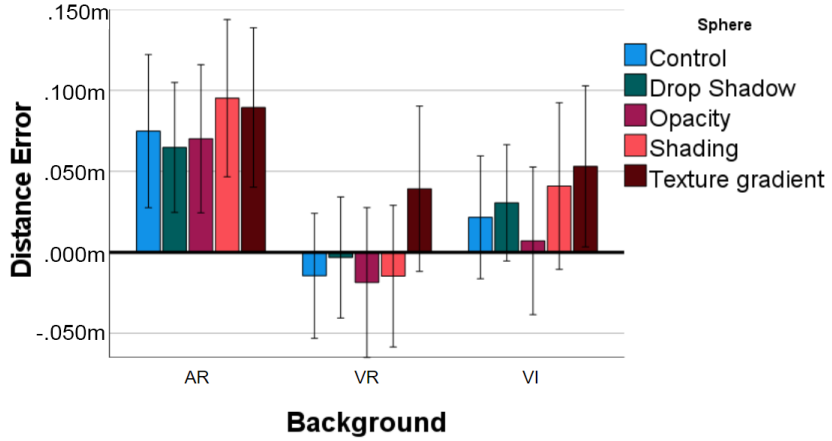
For all of the sphere conditions, the VR and VI conditions outperformed the AR condi-




(a) Peg and Background Distance Estimates




(b) Sphere and Pegs Distance Estimates.




(c) Background and Sphere Distance Estimates

Figure 4.3: Distance estimates for two-way interactions

tion. This again hints at some rendering congruence needed between the environment and sphere. In Fig 4.3b, interestingly, when the pegs switched from impoverished (having no cues) to rich (having depth cues), average estimations became an underestimation. This underestimation was also more accurate than the impoverished condition. This aligns with H3 and further supports the idea that objects rendered more realistically with more depth cues yield better distance estimations.

H4 is also supported by the significant two-way interaction effects. For the background and peg interaction, the AR background with rich pegs was the most accurate. This was a significant increase in accuracy compared to AR with no depth cues (AR Impoverished M = 0.158 m, AR Real M = -0.007 m). The VR condition switched from overestimation to underestimation when the pegs went from impoverished to rich. It seems more depth cues present in the pegs yield more underestimation. When the pegs had rich cues, almost all sphere conditions switched to an underestimation. This aligns more closely with past research. This experiment verified that more realistic pegs yield underestimation. It seems that the more realistic the virtual pegs look, the closer the results get to the results of experiments with a virtual manipulated object and physical targets (virtual-to-real). Drop shadow had the largest error in both peg conditions, which is odd because the drop shadow only needs to be aligned with the drop shadow of the peg. The VR background also prompted underestimation for sphere interactions while AR prompted overestimation. This and previous interactions suggest that more depth cues have a higher chance for underestimation when congruence occurs and overestimation when congruence does not occur. The VR condition may also have been more accurate because of the sphere and background congruence.

# CHAPTER 5: VIRTUAL-TO-REAL STUDY

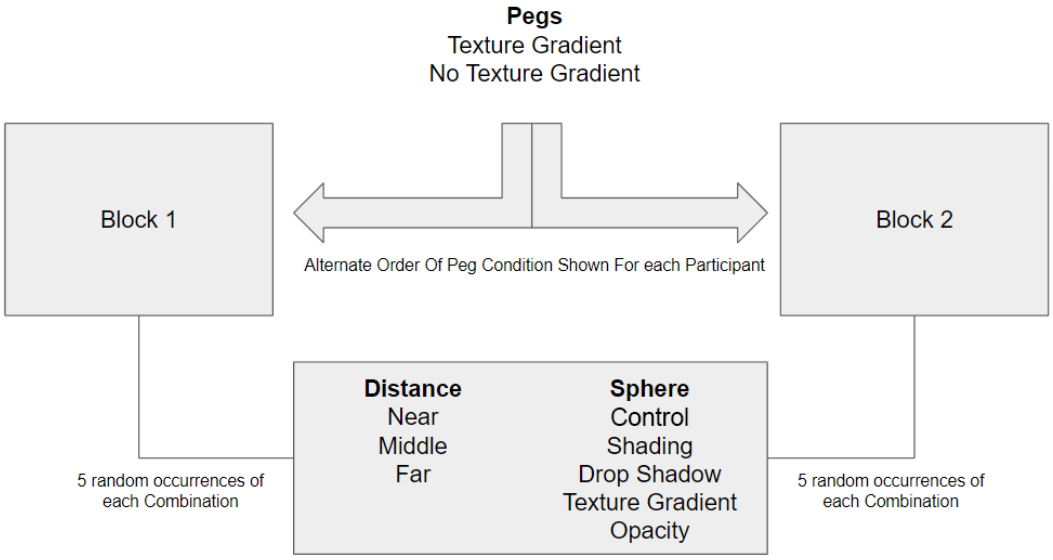



Figure 5.1: Experiment Design for Virtual-to-Real Task.

Virtual-to-real depth perception tasks are very common in the AR research area. Augmented reality has the added difficulty of mapping virtual objects to their real-world positions. All objects are immersed in the real world through the HoloLens 2 spatial mapping to make sure the objects are in the correct position. This also means users must make comparisons of a virtual object to a real object to accurately determine its position in the real world. The virtual-to-real task design seen in Fig 5.2 is an expansion of the first experiment described in section 6.1.3. The background is always real (the AR condition in Experiment 1) along with real pegs rather than virtual pegs. All sphere conditions are the same.

From the results of my virtual-to-virtual experiment, the following hypotheses were proposed:


Figure 5.2: Setup up for experiment 2 displaying trials from the POV of the participant and from the side; *(Top to Bottom)*

**H1**: The real background and real pegs are expected to prompt an underestimation of sphere and distance conditions

**H2**: All sphere rendering conditions are expected to have a large distance error due to incongruence between peg and sphere realism

**H3**: Depth estimations are expected to be least accurate for the far pegs and most accurate with the near pegs

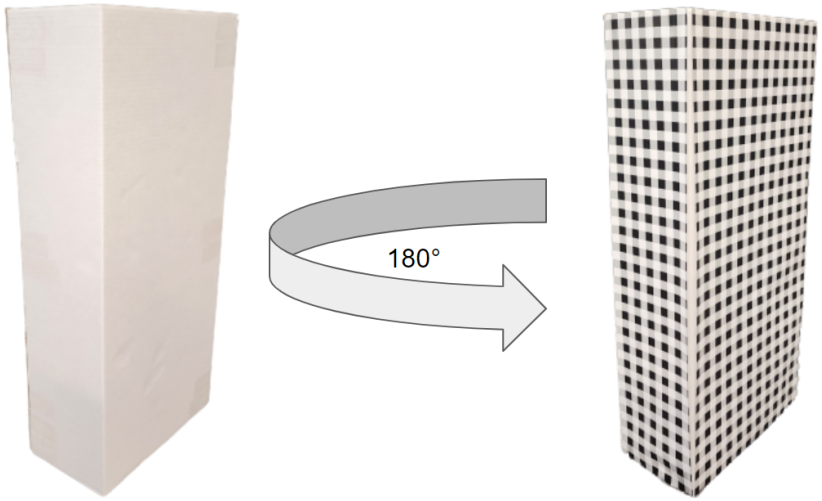



Figure 5.3: Different textures are applied to opposite sides of the real pegs to easily switch conditions by rotating the pegs 180 degrees.

## 5.1 METHODOLOGY

### 5.1.1 PARTICIPANTS

Twenty-nine participants (10 females and 19 males) (White: 15, Black: 3, Asian: 6, Hispanic: 5) were recruited from the University of North Carolina at Charlotte SONA system. Each participant completed 150 total trials. That is a total of 4,350 trials across the whole experiment.

### 5.1.2 EXPERIMENTAL CONDITIONS

This experiment used real pegs that were the same dimensions as the virtual pegs in Experiment 1. Each of the pegs had two sides with checkerboard wrapping paper (representing a texture gradient) on one side and a blank white texture (no texture gradient) on the other (Fig. 5.3). This allowed us to rotate the pegs 180 degrees to switch between no texture and texture conditions. To make sure the pegs were placed in the correct positions

(the same as in the virtual-to-virtual Experiment 1), virtual markers were rendered on the floor at the same locations as the virtual pegs in Experiment 1. At the halfway point in each participant's experiment, the pegs were rotated 180 degrees to show the opposite condition. The real replicas of the pegs were moved to the virtual marker positions and then the markers were hidden. The sphere had the same five rendering conditions and size as in Experiment 1 (Table 4.2).

This experiment, unlike the first, did not use a virtual background. Experiment 2's background condition always matched Experiment 1's AR background condition, where the real room's background was visible. This was to test the performance of a virtual-to-real distance comparison when both the pegs and background were real. In Experiment 1, the virtual pegs were tested with a virtual background and a real background.

As seen in Fig 5.1, experiment 2 used a repeated measure design with the two peg conditions (with and without texture gradient) blocked and presented in counterbalanced order across participants. Within each block of 75 trials, the 5 sphere conditions were randomly presented at each of the three distances (near, middle, and far) 5 times. At the beginning of each block, the combination was re-randomized for a new order. In total, there were 150 trials for each participant. Experiment 2 differed from Experiment 1 by expanding the number of trials at each of the near, middle, and far distances from 3 to 5 and adding distance into the design as a factor.

After 75 trials had been completed, participants were asked to take a five-minute break to minimize eye fatigue and possible motion sickness. Then the peg condition was switched by rotating the pegs 180 degrees. After the break, the participant continued with the remaining 75 trials.

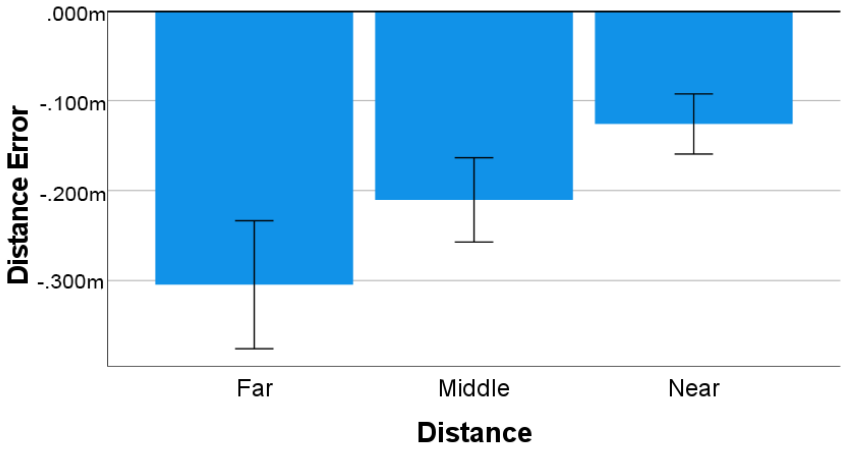
### 5.1.3 ANALYSIS

The experiment used a repeated measure design with the two peg conditions (real textured and real nontextured) blocked and presented in counterbalanced order across participants. Within each block of 75 trials, five sphere conditions were randomly presented at each of the three distances (near, middle, and far) 5 times. At the beginning of each block, the combination was re-randomized for a new order. In total, there were 150 trials for each participant.

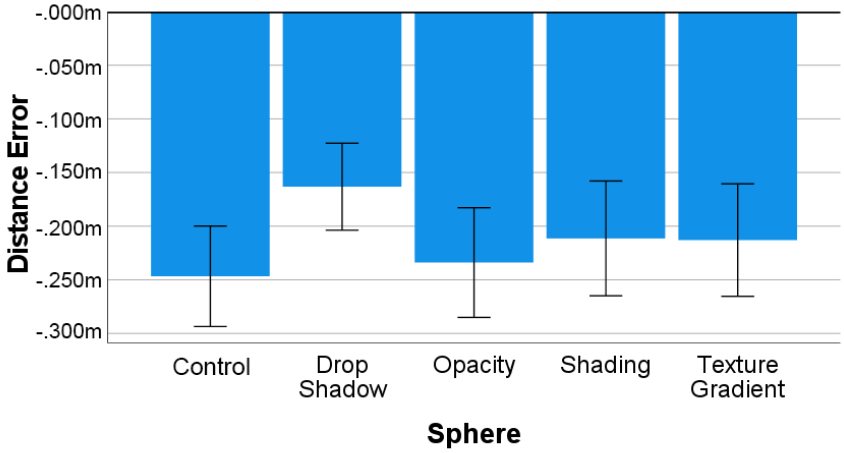
The distance estimation errors from each participant were averaged across the 5 replications for each of the 30 experimental conditions. The data from any given trial were trimmed if the response speed exceeded 2.5 times the standard deviation from each participant's mean and if the distance error was greater than 1 meter.

The mean trimmed distance estimation errors were analyzed with a $3 \times 2 \times 5$ repeated measures of analysis of variance (ANOVA), which tested for the main and interaction effects of the distance (near, middle, far), peg (real textured and real nontextured), and sphere conditions (control, shading, drop shadow, texture gradient, opacity). On average 1.3% of the trials were trimmed from a given participant due to low response speed and distance error greater than 2.5 times the standard deviation.

For the 29 participants, the mean distance estimate errors were calculated for each of the 30 experimental conditions across 150 total trials. A significance level of 0.05 was used for all statistical tests and the Greenhouse-Geisser correction was made to the p-value where appropriate to protect against possible violations of assumptions of sphericity. Bonferroni post-hoc pair-wise tests were used when the main effects were found to be significant.




(a) Mean estimations for distance main effect




(b) Mean estimations for sphere main effect

Figure 5.4: Distance estimations for the main effects of the peg distance and sphere experimental conditions

### 5.1.4 RESULTS

#### 5.1.4.1 DEPTH ESTIMATIONS

The distance of the pegs had a significant effect ($F(2, 56) = 36.254$, $p < 0.001$, $\eta_p^2 = 0.564$) on participant's depth judgments. As shown in Fig. 5.4a, there was a significant linear trend showing increased underestimation with distance ($F(1,28) = 39.69$, $p < 0.001$, $\eta_p^2 = .586$). The near condition was significantly more accurate compared to the middle and far conditions. The far condition was 0.179 m less accurate on average than the near condition. However, the peg conditions with or without the texture did not influence the degree of underestimation ($F(1, 28) = 2.081$, $p = 0.160$, $\eta_p^2 = 0.069$).

As in Experiment 1, the sphere conditions had an impact on participant performance ($F(4, 112) = 14.449$, $p < 0.001$ $\eta_p^2 = 0.34$). Follow-up Bonferroni tests (at the $p < 0.05$) showed that the control (No Cue) sphere was less accurate than all conditions except opacity. Opacity's mean difference with drop shadows and texture gradient was significant, however. Drop shadow had the lowest mean underestimation while the control had the highest (Fig. 5.4b). Sphere shading and texture gradient had very similar performances.

Sphere conditions were also found to interact with peg distance ($F(8, 224) = 9.937$, $p < 0.001$, $\eta_p^2 = 0.262$). The two-way interaction appears to be due to the fact that adding cues to the sphere had a greater effect with distance. Figure 6.7b shows at the near distance, the effect of the sphere conditions is slight but the effect grows with distance.

None of the other interactions were significant. Peg distance was not found to interact with peg condition ($F(1, 34) = 2.569$, $p = 0.112$, $\eta_p^2 = 0.084$); or with sphere conditions ($F(4, 112) = 0.006$, $p = .547$, $\eta_p^2 = 0.019$). The three way interaction of distance $\times$ pegs $\times$ sphere ($F(8, 224) = 0.367$, $\eta_p^2 = 0.037$) was not significant.

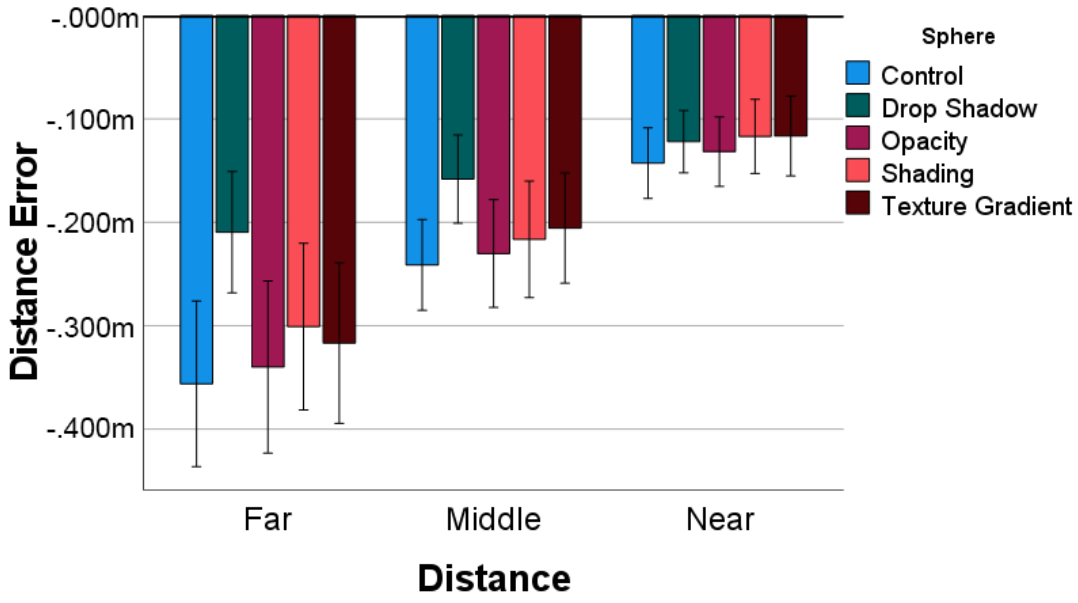



Figure 5.5: Interaction effects of mean distance errors for distance and sphere

#### 5.1.4.2 DISCUSSION

All sphere conditions prompted an average underestimation error. This is also true for all three distance conditions. This supports H1 and previous research that used virtual-to-real perceptual matching tasks. Each sphere condition also had an estimation error ($M < -0.10$) for all conditions regardless of distance, supporting H2. Using real pegs rather than the virtual pegs as in Experiment 1, the foreground and the background conditions were congruent in Experiment 2, resulting in larger estimation errors than in Experiment 1. Perhaps congruence with sphere and peg is a more important factor in depth estimation than background and peg congruence. The 6 pegs provide a frame of reference for the sphere's movement and help to define the scene's foreground.

Supporting H3, the results showed that the peg at further distances has greater error. This is true for all sphere conditions and the textured peg. The increase in error based on distance was a linear trend for all conditions. This is consistent with past research that used

verbal response tasks [64] where participants' distance underestimations increased linearly with distance.

In Experiment 2, the peg condition did not have an effect as it did in Experiment 1. It is possible that the surface properties of the plain white peg, such as real shading and very subtle texture gradients of the plain white paper, added some depth cues that made the plain white peg and textured peg more equivalent than intended. That being said, the checkerboard pattern did provide a very small amount of accuracy discrepancy in tasks. A new experiment is needed where the sphere is also realistically rendered to understand if the congruence of all three components will make improvements that are as accurate as the VR virtual-to-virtual condition in Experiment 1.

# CHAPTER 6: REALISTIC-TO-REALISTIC/REAL STUDY

The aim of this experiment is to better understand how depth perception performance is influenced by realistic depth cues and graphical attributes applied to a manipulated object. We use PBR methods and global illumination to create realistic virtual objects that match the real environment. As researchers continue to advance real-time global illumination techniques for AR, my findings serve as a validation of their efforts and underscore the importance of their work in creating more perceptually accurate and immersive AR experiences. By demonstrating the significant improvements in depth perception accuracy achieved through global illumination, my study highlights the potential of these techniques to improve the way users perceive and interact with virtual content in AR environments.

Our experiment tests the difference in performance for depth judgments in a realistic-to-realistic and realistic-to-real perceptual matching task. It is important to compare both virtual-to-virtual and virtual-to-real perceptual matching conditions for comparing congruence because AR involves making both kinds of egocentric judgments. In a perceptual matching task, a participant manipulates an object to match the distance with a stationary target. The target is a designated pair of vertical rectangular pegs and the manipulated object is a sphere as seen in Figure 6.6.

To my knowledge, there are no prior perceptual matching experiments that systematically explore the influence of PBR and global illumination on depth perception. Therefore, in this experiment, we compared the effects of PBR on depth perception to experiments one and two, which used less sophisticated rendering techniques within the same perceptual matching task. This allows for the systematic adjustment of the lighting realism of the

virtual sphere and pegs while holding all other variables constant (FOV, display resolution, tracking latency, pixel illumination levels, pixel contrast ratio, etc.).

By varying the different aspects of the PBR rendering, one would like to answer questions such as:

- How do depth estimate errors change when the sphere's visual fidelity is congruent with that of the target peg?
- Do the above results change based on objects being more metallic or dielectric?
- With realistically calculated global illumination, do realistic shadow and shading directions improve depth estimation?

We can also begin to understand how the visual congruence between the three aspects of the visual field affects depth perception accuracy. For example, how does depth accuracy differ when all aspects are virtual compared to when some aspects are real and others are virtual?

If the added realism better aligns cues between real and virtual elements, we would expect improved depth accuracy when comparing PBR to non-PBR conditions and when using higher levels of PBR fidelity. We designed this experiment to test the following hypotheses.

**H1**: The PBR virtual-to-virtual matching was expected to be more accurate than my previous experiments because of enhanced graphical congruence.

**H2**: Added reflections on materials were expected to increase accuracy compared to diffuse conditions because of additional position information.

**H3**: Shadows were expected to cause higher errors in estimations due to multiple shadows being rendered.

**H4**: Interaction effects were expected among the peg, sphere, and distance conditions.

Testing these hypotheses across systematic PBR variant combinations reveals whether advanced rendering of virtual objects provides measurable depth perception benefits over less sophisticated renderings. Practical insights are provided into balancing graphical cost and accuracy tradeoffs. The implementation described in this section forms the foundation for a robust and controlled experimental setup, enabling us to isolate and study the factors influencing depth perception in AR. By leveraging state-of-the-art rendering techniques and advanced AR hardware like the HoloLens 2, this research contributes to a deeper understanding of how users perceive and interact with virtual objects in augmented reality environments.

## 6.1 METHODOLOGY

### 6.1.1 GLOBAL ILLUMINATION CALCULATIONS

To accurately capture the real-world environment and enable seamless integration of virtual objects, a 3D reconstruction of the physical laboratory space was created. The HoloLens 2 features a high-resolution display (2048 × 1080 per eye), a 52° diagonal field of view, and advanced depth sensing capabilities through its time-of-flight (ToF) depth sensor. The virtual lab depth and color were captured using the HoloLens 2 to scan the environment with its long throw depth camera and RGB color camera through coordinate space mapping. The depth sensor projects a structured light pattern onto the environment and measures the time it takes for the light to reflect back, allowing it to calculate the distance to each point in the scene. The measured depth information is in the depth camera space in the form of a 3D point. This point must be transformed into the RGB camera space to get the necessary color value for the 3D point. The Hololens 2 provides access to the sensor stream information and their transformations. This process involves several linear

transformations, which take the following form:

$$M = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \tag{6.1}$$

where $R$ represents a rotation matrix and $T$ represents the translation vector.

We will need to transform the depth camera coordinates to the rigNode origin by using the depth camera's extrinsic transform, and then multiply that by the inverse of the RGB camera' extrinsic transform to map to RGB space.

$$M_{depth-to-rgb} = M_{rgb}^{-1} \cdot M_{depth} \tag{6.2}$$

$$= \begin{bmatrix} R_{rgb}^{-1} & -R_{rgb}^{-1} \cdot T_{rgb} \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R_{depth} & T_{depth} \\ 0 & 1 \end{bmatrix} \tag{6.3}$$

Let $P_{depth}$ represent a 3D point in the depth camera coordinate system. Applying $M_{depth-to-rgb}$ to a 3D point, $P_{depth}$, we get

$$P_{rgb} = T_{depth-to-rgb} \cdot {P_{depth}}^{T} \tag{6.4}$$

Using the above transformation, we can associate an illumination value for each 3D point when we generate a dense 3D point cloud of the environment to approximate the geometry of the physical lab space. The raw point cloud data was then processed using a Truncated Signed Distance Function (TSDF), $\phi(x,y,z)$, to reconstruct a smooth, continuous 3D mesh using the Open3D library [65] and scripts from Hololens 2 research mode [66]. The TSDF is a volumetric representation that stores the signed distance to the nearest surface at each voxel in a 3D grid. The sign of the distance indicates whether the voxel is inside (negative)

or outside (positive) the surface. The TSDF value at a point $(x,y,z)$ is calculated as:

$$\phi(x,y,z) = sign(x,y,z) * \min(||(x,y,z) - (x_p,y_p,z_p)||) \tag{6.5}$$

Next, the completed mesh seen in Figure 6.4 was imported into the Unity scene where we began global illumination calculations. The manipulated sphere only moves along the z-axis, thus it is sufficient to collect lighting at these positions. Light probes were placed along the z-axis corresponding to the sphere's path. Light probes are virtual objects used to capture and store information about lighting in a scene. They are used for creating more accurate and realistic lighting for dynamic environments. They are placed in positions where you want to store global illumination information. At each probe location, Unity samples the indirect illumination from the environment. The lighting information is stored in the form of spherical harmonics, which permits the storage of directional lighting information in a compact form.

At runtime, when a participant moves the sphere along the axis, lighting values are interpolated from nearby probes and applied to the manipulated sphere. The interpolation weights are calculated using the inverse distance weighting (IDW) method, which assigns higher weights to closer probe locations:

$$w_i = 1/(d_i)^p$$

where $w_i$ is the weight for the i-th probe, $d_i$ is the distance between the sphere and the $i$-th probe and $p$ is a power parameter that controls the influence of distance on the weights.

Additionally, there are reflection probes positioned along the sphere path. In total, there

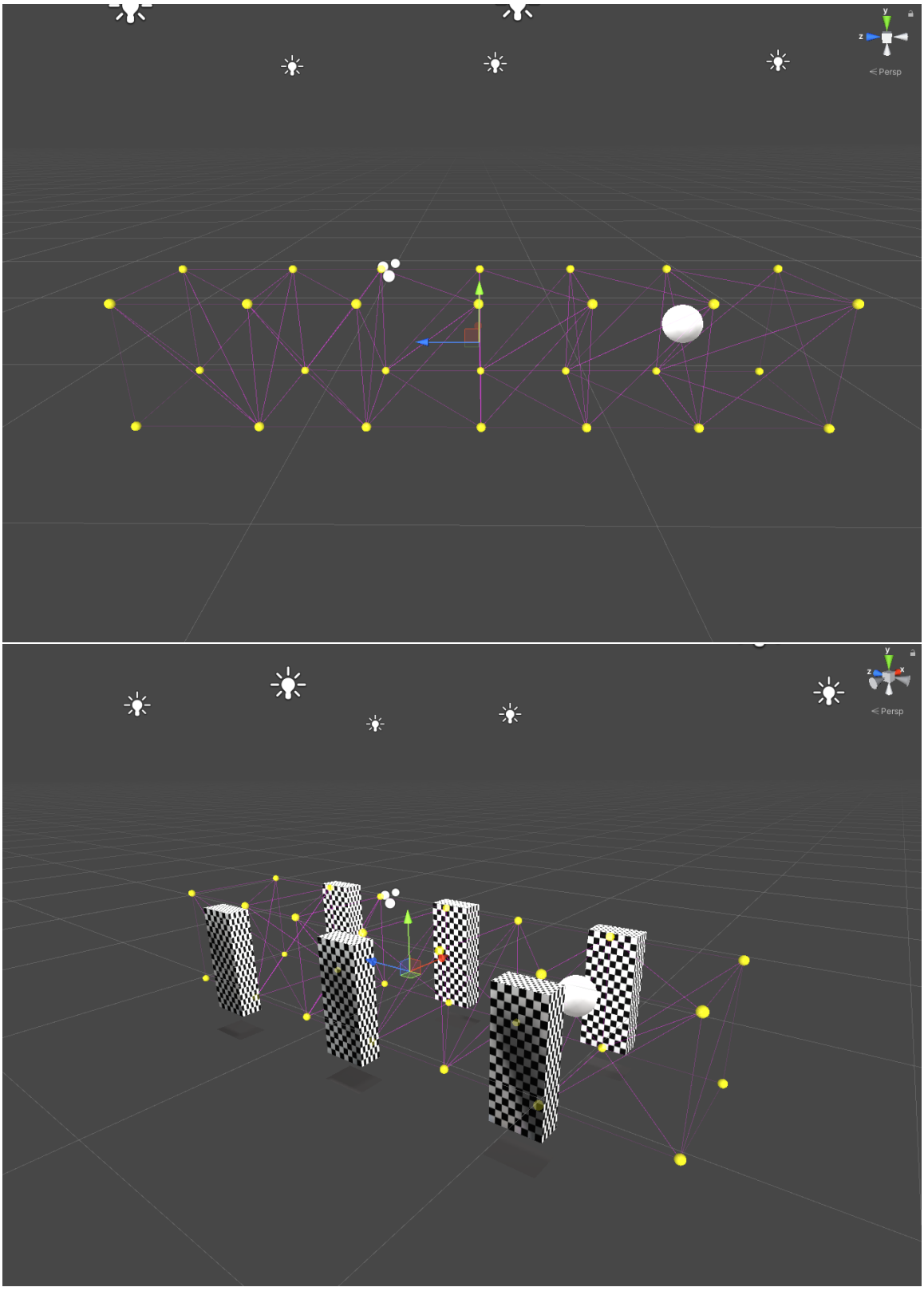



Figure 6.1: Light Probe volume within the travel path of the manipulated sphere.

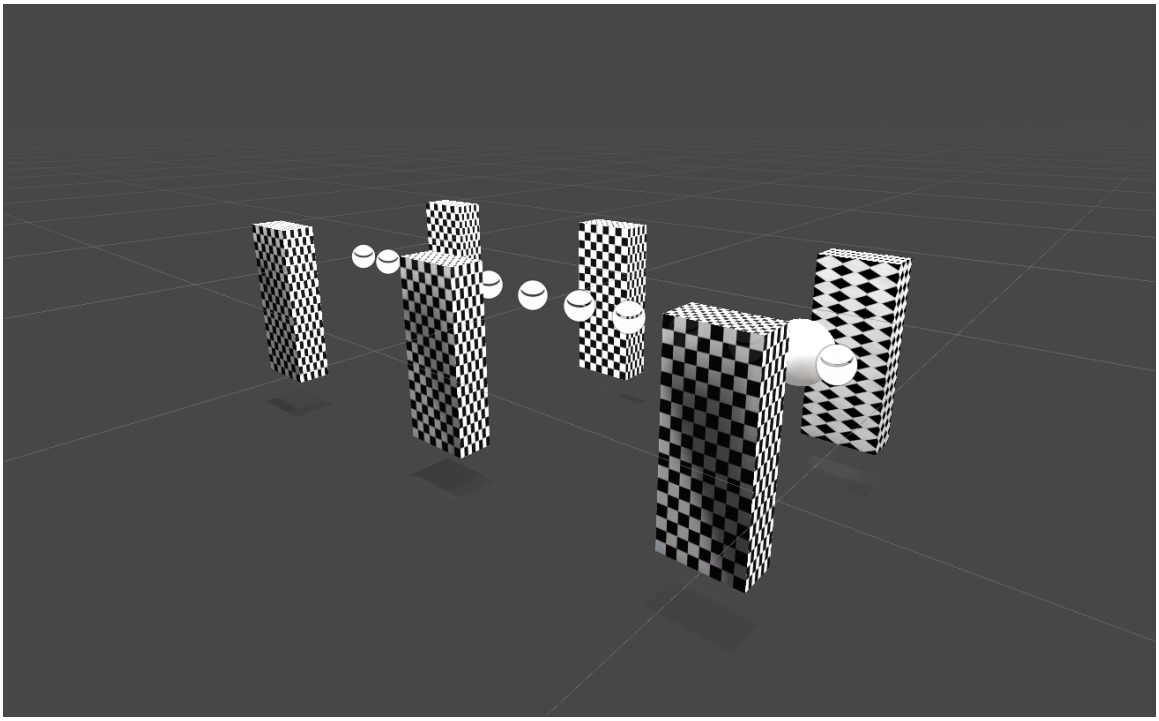


Figure 6.2: Reflection Probes along the z-axis.

are 11 reflection probes. There is a probe at the distance of each pair of pegs with 4 probes spaced equidistant between each pair as seen in fig 6.2. Reflection probes are objects that capture and store reflection illumination information in a scene. The reflection information is stored as a cubemap. The cube map stores the incoming light from six orthogonal directions (positive and negative x, y, z) and is used to look up the appropriate reflection direction based on the viewing angle and surface normal. An example generated reflection probe from our environment is seen in Figure 6.3. At runtime, the reflection information is used on objects in the scene to simulate mirrors, metallic objects, and glossy dielectric objects. Each reflection probe is static and interpolated between the two closest probes for reflection calculation on the manipulated object. The reflection intensity is determined by the smoothness of the object.

Capturing lighting information in light and reflection probes is a more performant way of approximating global illumination because the lighting is precomputed rather than at runtime for every frame. Precomputing global illuminaion is appropriate for the HoloLens

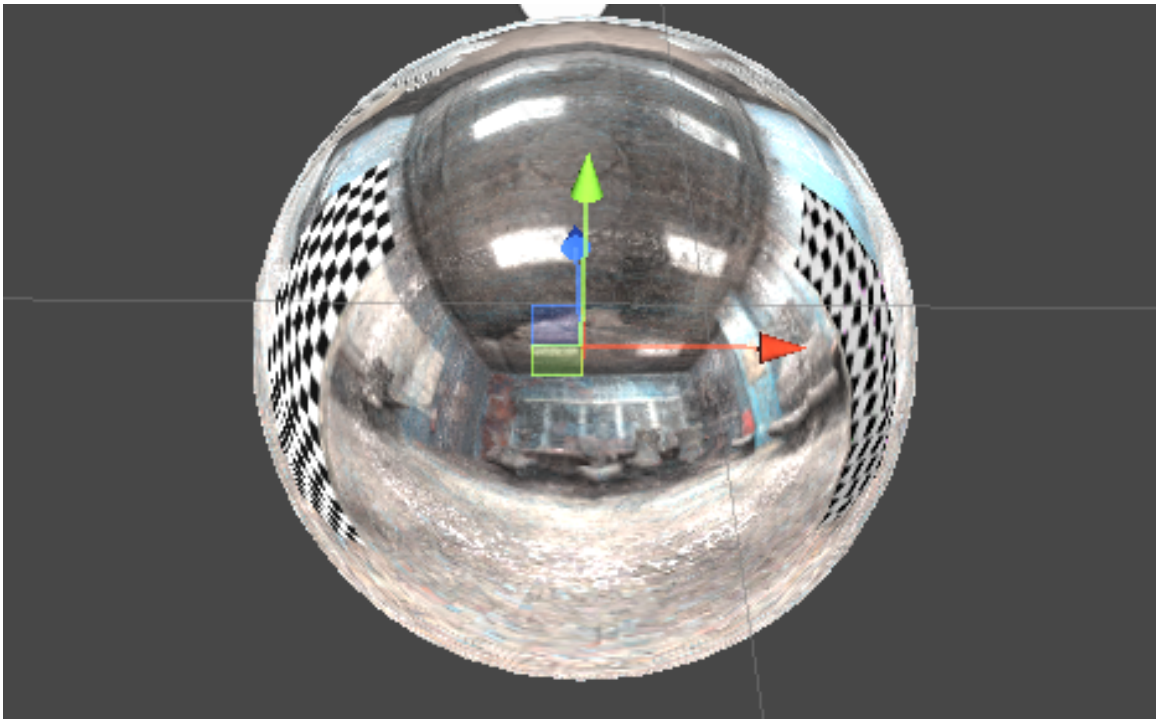


Figure 6.3: An example generated reflection probe from the TSDF geometry of the real environment.

as it is a mobile device with limited computational resources; this helps us maintain smooth and interactive framerates with the ability to apply global illumination on the manipulated objects.

### 6.1.2 PARTICIPANTS

Thirty-two participants (18 males, 14 females) (White: 18, Black: 4, Asian: 6, Hispanic: 4) took part in this experiment. Students were recruited from the University of North Carolina at Charlotte's SONA system, which allows psychology undergraduate students to participate for research credit. Each participant was awarded one research credit for 45-60 minutes of their time. Participants were at least 18 years old and had stereoscopic vision. After filling out an informed consent sheet, they had to pass an anaglyph stereo test, where they had to identify a set of shapes from colored stereo images. This research was conducted in accordance with the ethical guidelines set forth by the Office of Research

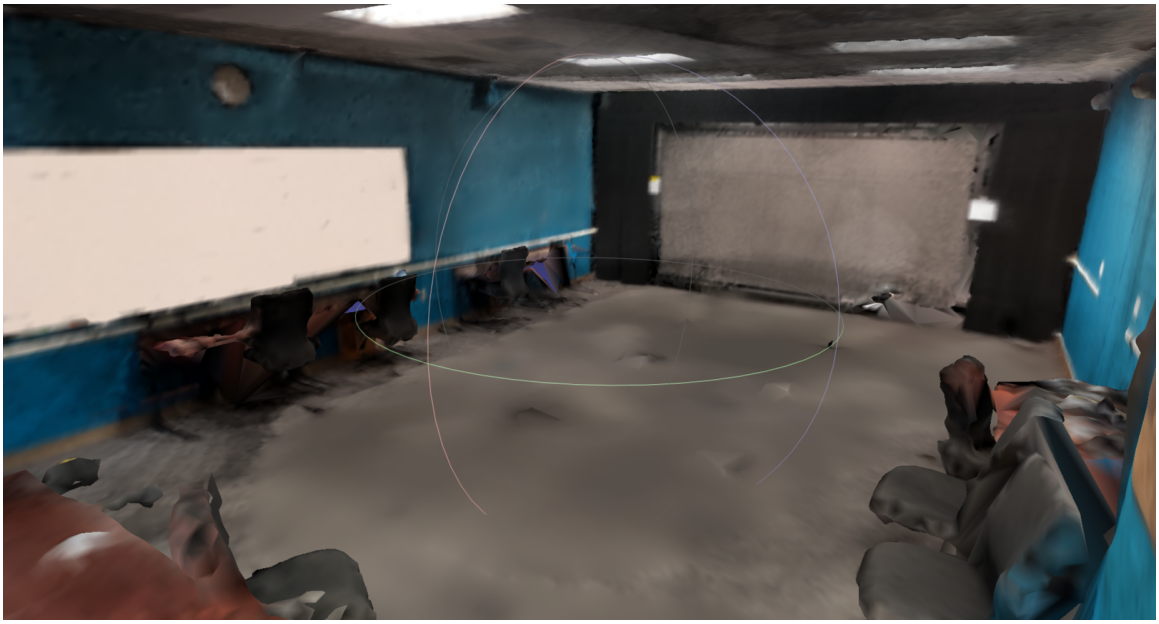


Figure 6.4: Final geometry from TSDF calculation.

Protections and Integrity, and all participants provided informed consent. Approval for the study was granted under IRB protocol IRB-22-1025.

### 6.1.3 EXPERIMENTAL CONDITIONS

There were two peg conditions, Real and Virtual; and they are illustrated in figure 6.5. The virtual pegs were included in the baked global illumination calculation to match the graphical fidelity of the real pegs as closely as possible. The surface has a checkerboard pattern and is a diffuse surface. The real pegs have a checkerboard pattern matching the virtual peg. Each of these peg conditions represents a different level of graphical fidelity, having a different level of congruence with the rendering conditions of the sphere. The physical room's lighting level is fixed for all participants so that the virtual objects remain sharp and in focus. Due to the virtual sphere being rendered with global illumination based on the virtual rendering of the real environment, the virtual peg closely matches the real peg in visual fidelity.

The sphere has 5 rendering conditions labeled as follows:

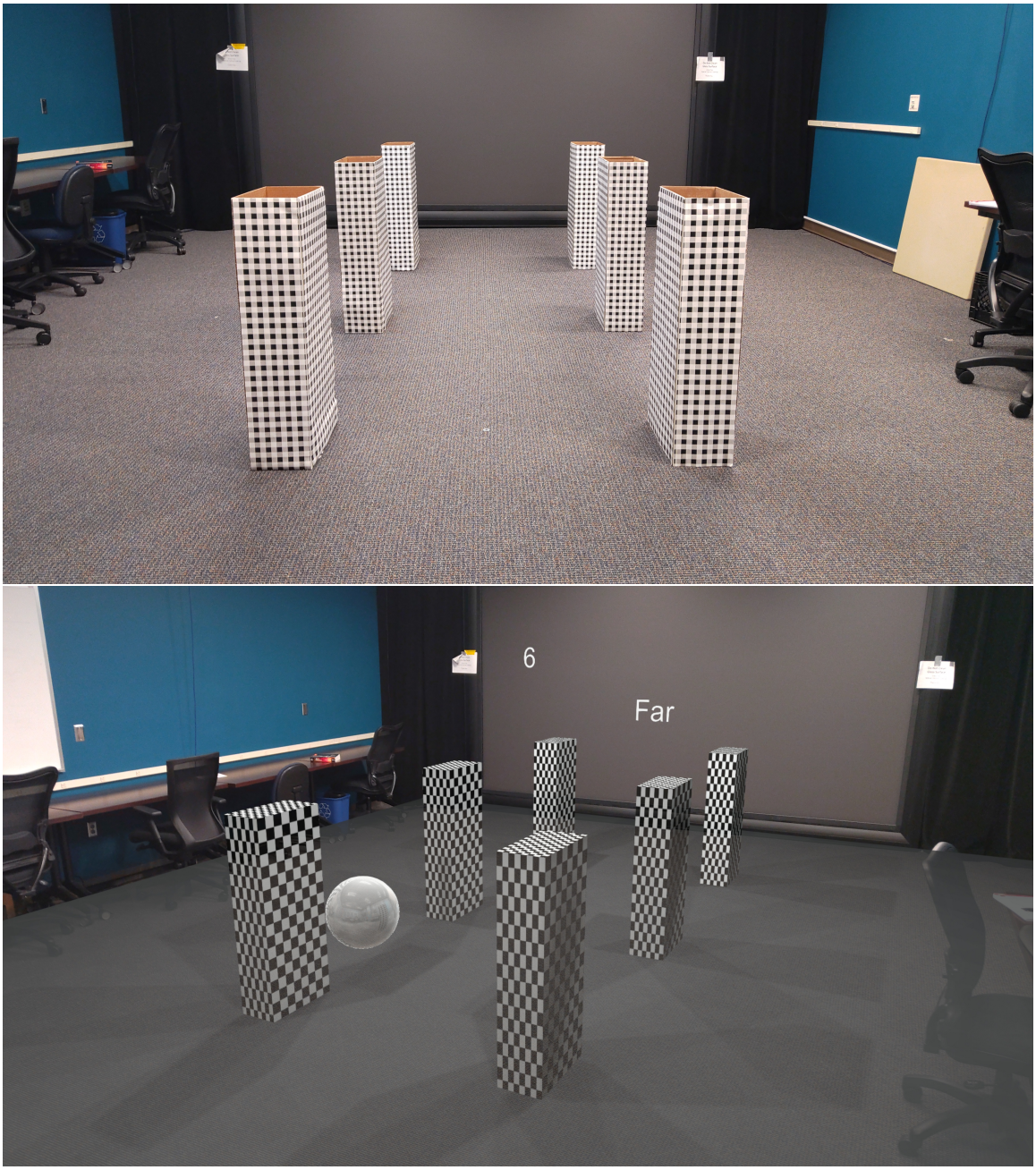



Figure 6.5: Experimental setup *Top:* Real pegs setup from the POV of the participant, *Bottom Row:* Side view of the experimental setup with virtual pegs and sphere.

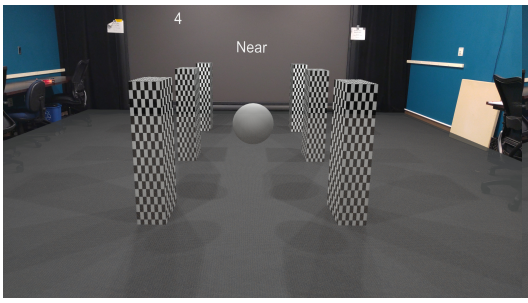
- *diffuse material* a sphere with uniform reflection of light from all incident locations on the surface. The object has no specular highlights or reflections.
- *glossy material* a sphere with the metallic coefficient set to 0.0 and smoothness coefficient set to 1.0 giving a slight mirror look.
- *textured diffuse material* a sphere with uniform reflection of light from all incident locations on the surface. The albedo color is determined by a checkerboard texture.
- *metallic material* a sphere with the metallic coefficient and smoothness coefficient set to 1.0 giving a mirror look.
- *shadows* a sphere with diffuse shading and rendered shadows on the floor.

Figure 6.6 shows these five sphere conditions. The sphere was 1m away from the participant in the starting position and 0.7m off the ground to be centered with the peg's height. The sphere had a diameter of 0.4m.

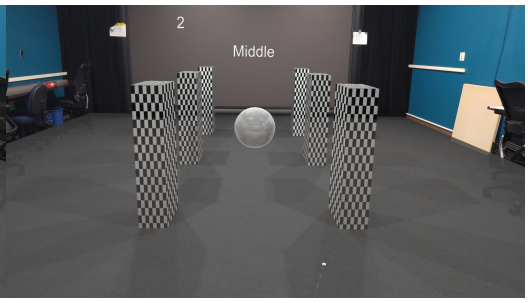
### 6.1.4 ANALYSIS

The experiment used a repeated measure design with the two peg conditions (real and virtual) blocked and presented in counterbalanced order across participants. Within each block of 75 trials, five sphere conditions were randomly presented at each of the three distances (near, middle, and far) 5 times. At the beginning of each block, the combination was re-randomized for a new order. In total, there were 150 trials for each participant.

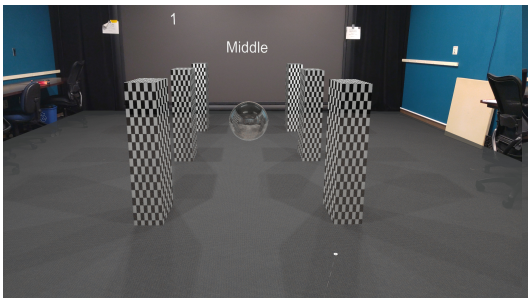
The distance estimation errors from each participant were averaged across the 5 replications for each of the 30 experimental conditions. The data from any given trial were trimmed if the response speed exceeded 2.5 times the standard deviation from each participant's mean and if the distance error was greater than 1 meter.



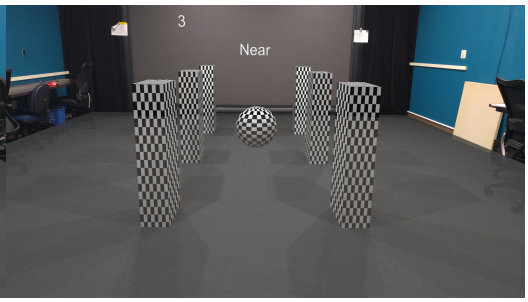

(a) Diffuse (Control)



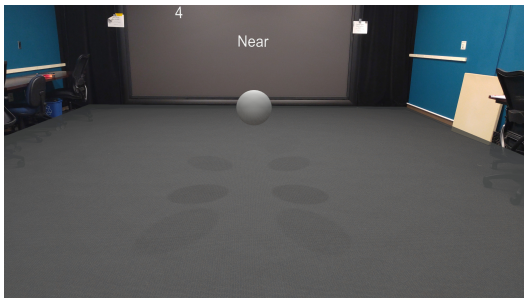

(b) Glossy




(c) Metallic




(d) Texture




(e) Shadow

Figure 6.6: Sphere Experimental Conditions.

The mean trimmed distance estimation errors were analyzed with a $3 \times 2 \times 5$ repeated measures of analysis of variance (ANOVA), which tested for the main and interaction effects of the distance (near, middle, far), peg (real and virtual), and sphere conditions (diffuse, shadow, metallic, glossy, texture). On average 1.3% of the trials were trimmed from a given participant due to low response speed and distance error greater than 2.5 times the standard deviation.

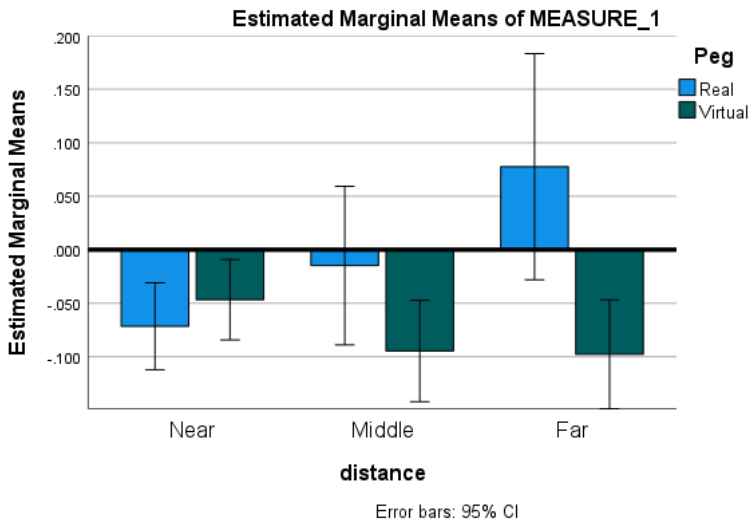
For the 32 participants, the mean distance estimate errors were calculated for each of the 30 experimental conditions across 150 total trials. A significance level of 0.05 was used for all statistical tests and the Greenhouse-Geisser correction was made to the p-value where appropriate to protect against possible violations of assumptions of sphericity. Bonferroni post-hoc pair-wise tests were used when the main effects were found to be significant.

### 6.1.5 RESULTS

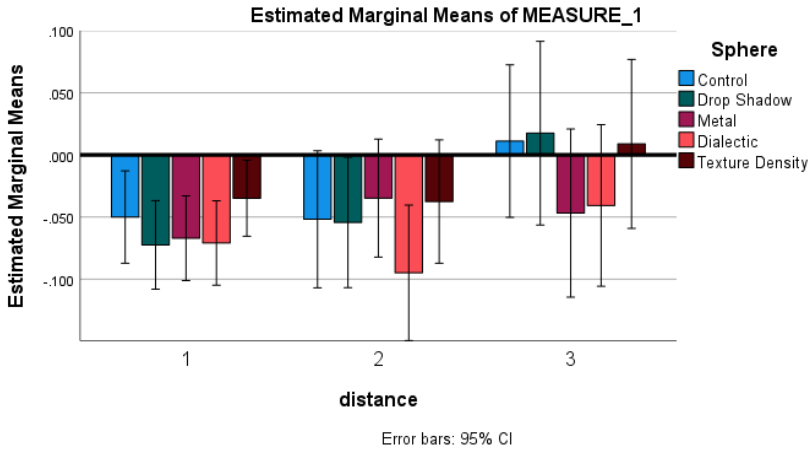
#### 6.1.5.1 ESTIMATION ERRORS

As expected, target distance was found to affect the accuracy of participants' depth judgments ($F(2, 62) = 4.160$, $p = .037$, partial $\eta_p^2 = .118$); but the kind and degree of estimation error depended upon the whether the targets were real or virtual ($F(2, 62) = 26.593$, $p < 0.001$, partial $\eta_p^2 = .462$). As shown in Figure 6.7a, virtual targets were associated with underestimations which increased with target distance while real targets were associated with underestimations when near and overestimation errors when targets were placed at far locations. The peg main effect shown in Fig 6.8c was also significant ($F(1, 31) = 4.293$, $p = .047$, partial $\eta_p^2 = .122$).

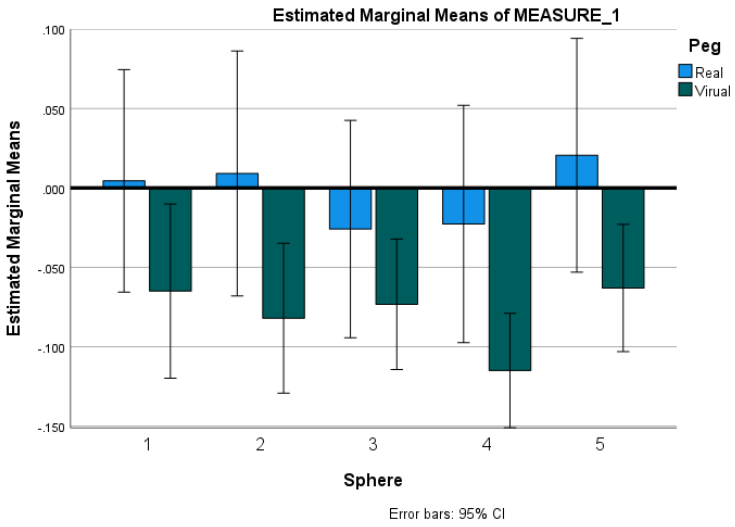
The 5 sphere conditions presented in Fig 6.8b were also found to vary in their degree of influence on depth estimations, ($F(4, 124) = 6.349$, $p < 0.001$, partial $\eta_p^2 = .170$)). Post-hoc tests showed that the glossy dialectic sphere differed from the two sphere conditions with




(a) Distance and Peg Interaction Estimates




(b) Sphere and Distance Interaction Estimates.




(c) Peg and Sphere Interaction Estimates

Figure 6.7: Distance estimates for two-way interactions in meters

depth cues—drop shadow and texture; and it was the only one of the sphere conditions to differ significantly from the control diffuse sphere.

Additionally, the sphere condition was found to interact with the target distance ($F(8, 248) = 3.883$, $p = 0.004$, partial $\eta_p^2 = .111$)). As shown in Fig 6.7b, it appears as though estimation errors were much reduced among all 5 sphere conditions when the targets were far rather than positioned in the near or middle locations. The remaining two interaction effects, the two-way peg-by-sphere interaction and three-way interaction of distance, peg, and sphere were not significant.

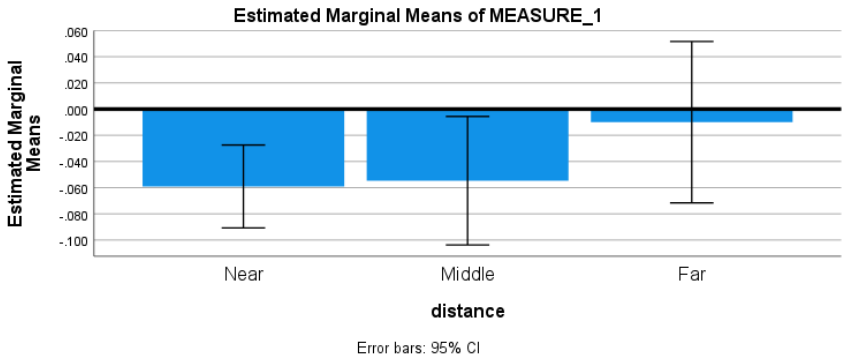
Overall, the results show significant main effects of all three manipulated variables (distance, peg, and sphere) and two interactions where distance interacted separately with peg and sphere conditions.

#### 6.1.5.2 DISCUSSION

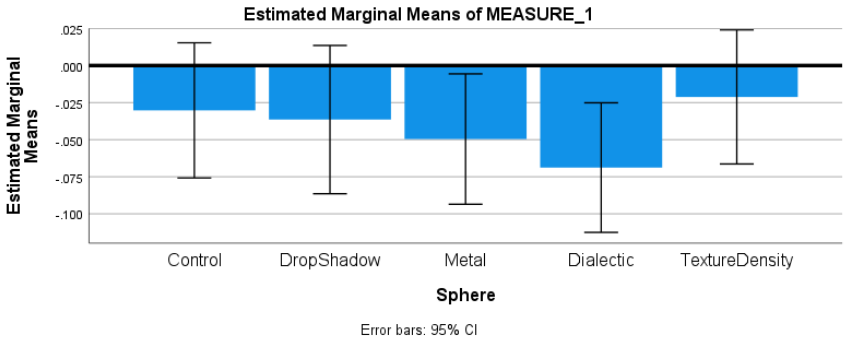
There is partial support for H1. All conditions had more accuracy than experiment 2. This includes interaction effects and main effects. This could be due to more congruence of objects in the visual field. In experiment 2, there was less congruence due to the virtual object being unrealistic and the target/background being real.

The differences observed between depth performance when matching the virtual spheres against the real pegs compared to global illumination rendered virtual pegs emphasize that virtual-to-real matching is more accurate in the middle of action space. While rendering costs may limit feasible complexity, maintaining realistic rendering of basic graphical attributes can boost depth judgments in virtual-to-real tasks compared to tasks that use a less realistic manipulated object.

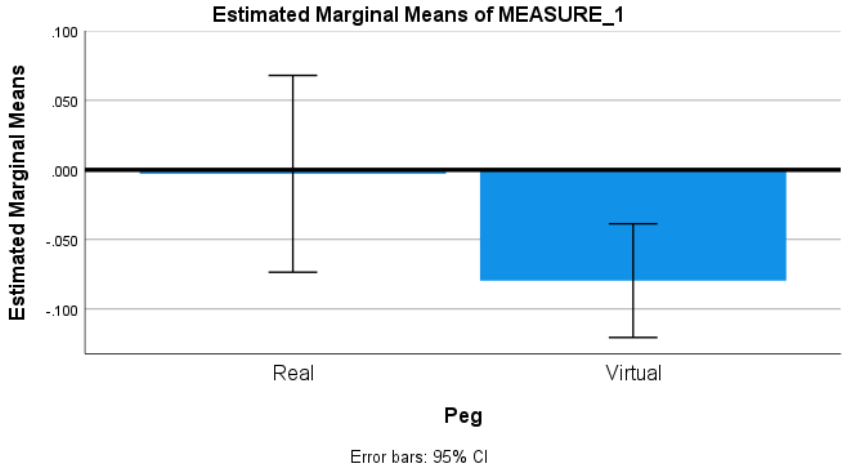
H2 was not supported as the glossy dialectic sphere material and metallic sphere material resulted in a slightly higher error compared to the diffuse control sphere and spheres with




(a) Distance Estimates




(b) Sphere Estimates




(c) Peg Estimates

Figure 6.8: Distance estimates for main effects in meters

depth cues like shadows and texturing. Since the reflections applied to the sphere were interpolated, it may have been providing information to the participant to prompt inaccurate placements if relying mostly on reflections.

In support of H3, the addition of shadows did increase error where the shadowed sphere enabled slightly less depth performance compared to the control diffuse case. This indicates that the realistic calculation of shadows from multiple light sources may not help with virtual object placement for objects high above the shadow plane.

The interaction effects (H4) also showcase that both low-level material factors and high-level position and environment elements combine in complex ways to determine perceptual outcomes. The experiment manipulated isolated variables, but in practice comprehensively integrating virtual objects with convincing depth cues into the real world scene depends on balancing these multiple competing factors. There are noticeable differences in distance estimation errors when the foreground frame of reference for the perceptual matching task is placed in the virtual world together with the sphere or placed separately in the real world. And importantly, the interaction of peg condition by distance shows that the differences between the two peg conditions become greater with increasing distance. At the near distance, both peg conditions show similar underestimations of depth, at the middle position, however, depth judgments with real pegs are quite accurate, while virtual pegs show greater underestimations. It is at the farthest distance that the differences between the peg placement are most evident, with overestimated distance errors in the real peg condition vs underestimated depth errors when virtual pegs are used.

When comparing distance with sphere interaction, all sphere conditions for the near and middle positions were underestimated. The far condition appeared more accurate only because half the trials averaged an overestimation (Real) and the other half averaged an underestimation (virtual). This averaged out the results for the far condition to seem more accurate when comparing its interaction with sphere conditions.

However, with continuing improvement in processing power and rendering efficiency, the experiment showcases meaningful improvements toward seamless integration of virtual graphics that maintain geometrically accurate depth relationships in the viewer's surroundings. Sophisticated modeling of physical interactions between light, materials, and geometry that characterize existing reality now shows promise for critical depth tasks when overlaying artificial objects into real-world scenes.

# CHAPTER 7: CONCLUSIONS

## 7.1 COMPARISON OF EXPERIMENTS

This study followed the two-part approach of experiments one and two, exploring the effects of graphical congruence and depth cues on depth perception in virtual environments. Our previous experiments analyzed perceptual matching performance and distance estimation errors when systematically varying attributes of the background, foreground targets, and manipulated spheres rendered with traditional non-physically based methods. By using virtual and real pegs for virtual-to-virtual and virtual-to-real matching along with different levels of realism across all studies, a direct comparison of performance related to congruence and target type can be made. We also adopted the same distances and sizes of pegs. The environment and headset used in the experiments are the same.

Our first experiment analyzed a virtual-to-virtual matching task and found that combinations of realistic backgrounds and non-realistic foreground elements led to overestimated distance judgments. These estimations became more accurate and switched to underestimations when additional cues were added to the foreground pegs to better align them graphically with the background. The interaction effects suggested that congruence between the graphical fidelity of the scene elements plays an important role in depth perception. Results showed that congruence between the target and sphere was important. Both the sphere and pegs were rendered with non-PBR and non-global illumination techniques.

The second experiment utilized a virtual-to-real paradigm with virtual spheres matched to real peg targets. The sphere was rendered with the same techniques as experiment one.

As expected, all conditions prompted some degree of distance underestimation, with errors increasing at further distances. The mismatch between the rendering style of the basic virtual spheres compared and real peg targets again highlights the role of graphical congruence in accurate distance judgments.

The third experiment rendered the manipulated sphere using PBR and GI to match the realism of the real pegs as closely as possible. This experiment had more accurate depth judgments compared to the second experiment. All aspects of the visual field were more congruent. Also, similar to experiment one, changes in target rendering affected perception.

### 7.1.1 COMPARISON OF RESULTS

If rendering fidelity matters, my PBR enhancements should show measurable improvements over the non-realistic sphere and peg conditions from past experiments. By holding the experimental approach constant while narrowing the realism gap between real and virtual visual elements, my work clarifies whether rendering advances can overcome depth judgment challenges previous authors encountered.

#### 7.1.1.1 VIRTUAL-TO-REAL TASK COMPARED TO REALISTIC-TO-REALISITIC/REAL TASK

The results showed significant effects depending on the distance, sphere material properties, and whether real or photorealistic virtual pegs were used as targets. Unlike the consistent underestimation seen previously with virtual-to-real matching, where the virtual sphere was rendered with unrealistic rendering methods (Figure 5.4b), the combination of the real-world environment and advanced rendered virtual elements applied to the manipulated object led to a balance of both over and underestimations depending on the specific peg condition and distance (Figure 6.7a). The middle condition within the action space was very accurate.

The third experiment also led to more accurate depth judgments (Figure 6.8) for all main effects compared to experiment two (Figure 5.4). The real pegs and realistically rendered pegs matched with realistically rendered spheres leading to smaller errors compared to the previous experiment because of higher visual congruence between the manipulated object and target. In experiment two, realistic pegs and unrealistic spheres may have caused a conflict in understanding environmental information that increased the effects of distance misestimation. The virtual object and real target receive different lighting information. Removing this conflict decreased underestimation.

Unlike experiment three's results, experiment two's real pegs did not have a significant impact on depth perception. Both previous peg conditions were real with only a texture gradient change. The contextual switch between a real and virtual target in this experiment shows that the type of target is important for determining depth.

#### 7.1.1.2 VIRTUAL-TO-VIRTUAL TASK COMPARED TO REALISTIC-TO-REALISTIC/REAL TASK

Experiment three and experiment one both show that there is a significant change in depth perception dependent on the full congruence of the visual field. In experiment one, there was more accuracy when the target had virtual depth cues and was rendered to match the virtual object and virtual background (Figure 4.3c  4.3a). Some of the conditions prompted underestimation. The AR condition was mostly overestimated. In experiment three, with more realistic depth cues applied to the target and manipulated object creating full congruence similar to experiment one's VR condition, depth performance was more accurate with some over and underestimation (Figure 6.7).

Surprisingly, the congruence between the virtual manipulated object and virtual target in experiment three was not as accurate as in experiment one which also had a virtual manipulated object and virtual target congruence (Figure 4.3b  6.7c. This may be because

experiment one had some trials where all aspects of the visual field were virtual and thus congruent.

The interaction effects demonstrate the combined influence of environment depth, rendering congruence, and material properties in estimating distances in augmented reality matching tasks. These experiments provide practical insights into utilizing rendering simulations to integrate virtual content with realistic congruence to the surrounding physical environment.

## 7.2 LIMITATIONS

While the first two studies contribute valuable insights to the field, it is essential to recognize their inherent limitations. A comprehensive understanding of the research findings necessitates a critical examination of the constraints and potential factors that may impact the generalizability and robustness of my conclusions. In this section, we describe the limitations of my study, providing transparency and context for the interpretation of the results.

One limitation was the use of an OSTHMD to simulate VR conditions. We chose to do this to avoid the confounds that using multiple types of headsets would introduce, as the HoloLens cannot give the same optical conditions as a VR HMD. One way to improve on this research is to use a newer headset that provides the ability to switch between a pass-through AR mode and an enclosed VR mode.

Also, all pegs were in static positions. This could induce some learning effect on participants as they would place a sphere in the same position multiple times in one experiment. We tried to mitigate this by introducing variable speeds in sphere movement, and randomized order in target instructions. An easy solution for virtual targets is to use a continuous random variable to control the peg distances, but this solution is difficult for real targets, which requires changing the physical pegs' distances continuously between trials.

We also did not render the peripheral geometry (desks and chairs) in the physical room in the VR condition. Though this was a slight oversight, the small FOV in the HoloLens would not have allowed the participants to see such geometry when fixated on the sphere. Also, participants remained seated and were instructed not to move their heads, which kept the pegs in their field of view during the experiment.

Each participant was tested for stereopsis using an anaglyph stereo test. While advan-

tageous for simplicity and accessibility, anaglyph tests exhibit limitations, including color distortion and potential discomfort for some individuals.

Some limitations of the third study are with the rendering techniques used for environment creation. Our environment used pre-computed calculations for this specific experiment and setting. Using baked global illumination and reflections doesn't allow for dynamic changes in lighting conditions. In more realistic situations, lighting information changes and it would be important to investigate how these changes influence depth perception. The reflection probes were also static and interpolated as the object moved. Having real-time reflections can provide more accurate cues of the environment for participants. Also, my scanned environment could be a higher resolution creating more accurate lighting interactions from background material properties and complex geometry. Methods in the related work section could be used for these calculations. We also investigated a limited range of PBR material properties. Our conditions focused on metallic, glossy, and diffuse materials. Investigating different BRDFs and adjusting a wider range of properties affecting material appearance could be an area of future work.

## 7.3 GENERAL CONCLUSION

We presented three experiments to gain insights into how depth perception performance is affected by changes in the depth cue information and graphical attributes of the background, foreground, and manipulated objects in virtual and augmented reality environments. I have built upon earlier work creating a robust experimental framework to test a multitude of experimental conditions, and directly compare performance between virtual-to-virtual and virtual-to-real matching.

Experiment one showed that depth estimations can switch from over- to underestimations depending on the background condition and its congruence with pegs or the sphere. Specifically, participants showed greater overestimation in scenes with more realistic backgrounds with nonrealistic pegs. This result switched to underestimation when more depth cues were applied to the pegs, specifically when congruent with the VR background.

Experiment two further supports the underestimation effect resulting from peg and background congruence in AR environments. Also, it reveals that congruence between objects in the scene is an important factor that affects distance estimation. Additionally, results supported the idea that more depth cues on pegs prompt underestimation in VR Access to depth cues differing from the sphere has a higher chance for underestimation when congruence occurs for the peg and background and overestimation when congruence does not occur.

With my use of both virtual-to-virtual and virtual-to-real perceptual matching in the same testing framework across experiments, we also found a difference in depth perception performance based on the peg condition in an AR environment. There was mostly overestimation for virtual pegs and underestimation for real pegs when the manipulated sphere is unrealistic. The results indicate that participants' estimation errors increased with distance, and that texture gradient, drop shadow, and shading may have an effect on dis-

tance estimation accuracy.

Our third experiment explored how physically based rendering techniques applied to depth cue information and graphical attributes affect depth perception in a perceptual matching task for indoor augmented reality environments. By independently manipulating the PBR attributes applied to the manipulated object and target, we were able to test how more realistic depth cues play a role in depth estimation performance. Comparing this experiment with the two previous experiments also allowed us to analyze the effect of graphical fidelity congruence across the background, foreground, and foreground manipulated objects.

This study found significant main effects for distance, peg type, and sphere rendering on depth estimation errors. Two-way interactions between distance and peg type as well as distance and sphere rendering were also significant. However, there was no interaction between peg type and sphere rendering when accounting for distance. These results suggest PBR enhancements can measurably improve depth perception over non-realistic rendering in certain cases. When it comes to estimating distances, the placement of the foreground frame of reference in the perceptual matching task has a significant impact on the accuracy of results. Additionally, the interaction between the peg condition and distance reveals that the disparities between the two peg conditions become more pronounced as the distance increases.

Furthermore, the differences observed between depth performance when matching the virtual spheres against the real pegs, compared to global illumination rendered virtual pegs, also indicate that virtual-to-real matching is more accurate when using PBR-based manipulated objects.

Comparisons to the previous experiments provide further insights. Switching between real and virtual targets had an important effect, indicating target type matters for determin-

ing depth in addition to rendering quality. However, the persisting misestimation shows limitations remain even with high-fidelity graphics.

Overall, these findings have implications for the design of VR and AR environments and the understanding of how humans perceive distances for different background/foreground contexts. Our contributions help to better understand the factors that affect depth perception in AR and VR background/peg congruence and highlight the importance of considering depth cue and graphical attribute combinations when designing virtual environments. This research demonstrates that advanced physically based rendering techniques can provide depth perception benefits in augmented reality applications, but should be balanced with considerations of cost, target type, and environment. While my experiment uses precomputed advanced rendering techniques, these results serve to validate the efforts of creating real-time global illumination and accurate lighting calculations in AR and its potential to increase depth perception accuracy. While photorealism is not a complete solution, deliberate PBR enhancements focusing on depth cue congruence offer measurable improvements. These findings can help guide the development of AR systems that require precise spatial judgments in fields like healthcare. Further research should explore additional cue combinations and identify optimal fidelity trade-offs.

## 7.4 FUTURE WORK

The current study provides valuable insights into the impact of physically based rendering (PBR) and global illumination (GI) on depth perception in augmented reality (AR) environments. However, further topics to expand upon to gain a better understanding of the aspects that contribute to depth perception in virtual environments.

One important direction for future work is to investigate the impact of various lighting conditions on depth perception in AR environments rendered with GI and PBR techniques. This could involve exploring the effects of different light intensities, colors, and shadow

qualities on depth judgment accuracy. Additionally, comparing the performance of real-time GI techniques, while optimizing performance using foveated.

Another area of interest is the comparison of depth perception in indoor and outdoor AR environments. Conducting comparative studies on the impact of PBR and GI in these different lighting conditions can help understand the effect of change in illumination conditions, spatial scales, and complex geometries on depth perception. Understanding the effect these aspects have on depth perception can help with rendering optimizations to maximize fidelity and perception performance related to occlusions, reflections, and atmospheric perspective.

Investigating the impact of various optimization techniques, such as level-of-detail (LOD) simplification, texture compression, and adaptive shading, on depth perception for larger environments could help understand the impact of performance optimizations on depth perception. Additionally, evaluating the effectiveness of foveated rendering techniques in maintaining accurate depth perception could also be an effective area of study. Testing different shadow techniques, lighting models, and BRDFs can help understand which lighting models help depth perception the most.

Adaptive and personalized depth perception is another promising direction for future work. Investigating the potential of adaptive rendering techniques that dynamically adjust depth cues based on user performance, preferences, and perceptual abilities could lead to a calibrated AR experience. This can be related to the time taken to learn and improve at a given depth perception task.

In conclusion, the future work outlined in this section highlights research directions that can be pursued to further advance the understanding of depth perception in AR. By investigating different lighting conditions, rendering optimizations, shadow techniques, and adaptive AR experiences researchers can contribute to the development of more perceptu-

ally accurate and efficient AR experiences. These efforts will not only enhance the quality of AR applications but also create new opportunities and innovations in the field of AR research and development.

# REFERENCES

[1] F. El Jamiy and R. Marsh, "A survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality and mixed reality," *IET Image Processing*, vol. 13, 01 2019.

[2] E. Blümel, "Global challenges and innovative technologies geared toward new markets: Prospects for virtual and augmented reality," *Procedia Computer Science*, vol. 25, pp. 4–13, 2013. 2013 International Conference on Virtual and Augmented Reality in Education.

[3] C. Gsaxner, A. Pepe, J. Wallner, D. Schmalstieg, and J. Egger, *Markerless Image-to-Face Registration for Untethered Augmented Reality in Head and Neck Surgery*, pp. 236–244. 10 2019.

[4] J. Sutherland, J. Belec, A. Sheikh, L. Chepelev, W. Althobaity, B. Chow, D. Mitsouras, A. Christensen, F. Rybicki, and D. La Russa, "Applying modern virtual and augmented reality technologies to medical images and models," *Journal of Digital Imaging*, vol. 32, 09 2018.

[5] J. Knapp and J. Loomis, *Visual Perception of Egocentric Distance in Real and Virtual Environments*, vol. 11, pp. 21–46. 06 2003.

[6] J. T. Kajiya, "The rendering equation," in *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '86, (New York, NY, USA), p. 143–150, Association for Computing Machinery, 1986.

[7] A. S. Glassner, *Principles of Digital Image Synthesis*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994.

[8] K. E. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfaces∗," *J. Opt. Soc. Am.*, vol. 57, pp. 1105–1114, Sep 1967.

[9] R. L. Cook and K. E. Torrance, "A reflectance model for computer graphics," *ACM Trans. Graph.*, vol. 1, p. 7–24, jan 1982.

[10] B. Burley, "Physically-based shading at disney," 2012.

[11] C. Schlick, "An inexpensive brdf model for physically-based rendering," *Computer Graphics Forum*, vol. 13, no. 3, pp. 233–246, 1994.

[12] B. Smith, "Geometrical shadowing of a random rough surface," *IEEE Transactions on Antennas and Propagation*, vol. 15, no. 5, pp. 668–671, 1967.

[13] P. Beckmann and A. Spizzichino, "The scattering of electromagnetic waves from rough surfaces," 1963.

[14] A. L. D. Santos, D. Lemos, J. E. F. Lindoso, and V. Teichrieb, "Real Time Ray Tracing for Augmented Reality," in *2012 14th Symposium on Virtual and Augmented Reality*, (Rio de Janiero, Brazil), pp. 131–140, IEEE, May 2012.

[15] M. Fradet, P. Hirtzlin, P. Jouet, A. Laurent, and C. Baillard, "Light4AR: a Shadow-based Estimator of Multiple Light Sources in Interactive Time for More Photorealistic AR Experiences," in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 304–309, Oct. 2021.

[16] S. A. Pessoa, G. de S. Moura, J. P. S. M. Lima, V. Teichrieb, and J. Kelner, "A Global Illumination and BRDF Solution Applied to Photorealistic Augmented Reality," in *2009 IEEE Virtual Reality Conference*, pp. 243–244, Mar. 2009. ISSN: 2375-5334.

[17] S. Pessoa, G. Moura, J. Lima, V. Teichrieb, and J. Kelner, "Photorealistic rendering for Augmented Reality: A global illumination and BRDF solution," in *2010 IEEE Virtual Reality Conference (VR)*, pp. 3–10, Mar. 2010. ISSN: 2375-5334.

[18] A. Alhakamy and M. Tuceryan, "CubeMap360: Interactive Global Illumination for Augmented Reality in Dynamic Environment," in *2019 SoutheastCon*, pp. 1–8, Apr. 2019. ISSN: 1558-058X.

[19] A. Alhakamy and M. Tuceryan, "Real-time Illumination and Visual Coherence for Photorealistic Augmented/Mixed Reality," *ACM Computing Surveys*, vol. 53, pp. 1–34, May 2021.

[20] R. Marroquim, M. Kraus, and P. R. Cavalcanti, "Efficient Point-Based Rendering Using Image Reconstruction,"

[21] K. Rohmer, J. Jendersie, and T. Grosch, "Natural Environment Illumination: Coherent Interactive Augmented Reality for Mobile and Non-Mobile Devices," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 2474–2484, Nov. 2017. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[22] C. Liu, L. Wang, Z. Li, S. Quan, and Y. Xu, "Real-Time Lighting Estimation for Augmented Reality via Differentiable Screen-Space Rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, pp. 2132–2145, Apr. 2023.

[23] A. Plopski, T. Mashita, K. Kiyokawa, and H. Takemura, "Reflectance and light source estimation for indoor AR Applications," in *2014 IEEE Virtual Reality (VR)*, pp. 103–104, Mar. 2014. ISSN: 2375-5334.

[24] J. M. Knapp and J. M. Loomis, "Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments," *Presence*, vol. 13, no. 5, pp. 572–577, 2004.

[25] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert, "The perception of egocentric distances in virtual environments - a review," *ACM Comput. Surv.*, vol. 46, Dec. 2013.

[26] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker, "Distance perception in real and virtual environments," *ACM Trans. Appl. Percept.*, vol. 2, p. 216–233, July 2005.

[27] A. Peer and K. Ponto, "Perceptual space warping: Preliminary exploration," in *2016 IEEE Virtual Reality (VR)*, pp. 261–262, 2016.

[28] K. Ponto, M. Gleicher, R. G. Radwin, and H. J. Shin, "Perceptual calibration for immersive display environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 691–700, 2013.

[29] D. Waller and A. Richardson, "Correcting distance estimates by interacting with immersive virtual environments: Effects of task and available sensory information," *Journal of experimental psychology. Applied*, vol. 14, pp. 61–72, 04 2008.

[30] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric depth judgments in optical, see-through augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 3, pp. 429–442, 2007.

[31] I. T. Feldstein, F. M. Kölsch, and R. Konrad, "Egocentric distance perception: A comparative study investigating differences between real and virtual environments," *Perception*, vol. 49, no. 9, pp. 940–967, 2020.

[32] F. El Jamiy, A. Ramaseri Chandra, and R. Marsh, "Distance accuracy of real environments in virtual reality head-mounted displays," 08 2020.

[33] A. Peer and K. Ponto, "Mitigating incorrect perception of distance in virtual reality through personalized rendering manipulation," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 244–250, 2019.

[34] M. A. Cidota, R. Clifford, S. Lukosch, and M. Billinghurst, "Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality," pp. 172–177, 09 2016.

[35] P. Willemsen, M. Colton, S. Creem-Regehr, and W. Thompson, "The effects of head-mounted display mechanics on distance judgments in virtual environments," pp. 35–38, 01 2004.

[36] P. Willemsen, A. Gooch, W. Thompson, and S. Creem-Regehr, "Effects of stereo viewing conditions on distance perception in virtual environments," *Presence*, vol. 17, pp. 91–101, 02 2008.

[37] W. Thompson, P. Willemsen, A. Gooch, S. H. Creem-Regehr, J. Loomis, and A. Beall, "Does the quality of the computer graphics matter when judging distances in visually immersive environments?," *Presence: Teleoperators and Virtual Environments*, vol. 13, pp. 560–571, 2004.

[38] A. Lucaci, M. Bach, P. Jensen, and C. Madsen, "Influence of texture fidelity on spatial perception in virtual reality," pp. 244–251, 01 2022.

[39] C. Vienne, S. Masfrand, C. Bourdin, and J.-L. Vercher, "Depth perception in virtual reality systems: Effect of screen distance, environment richness and display factors," *IEEE Access*, vol. 8, pp. 29099–29110, 2020.

[40] L. Buck, R. Paris, and B. Bodenheimer, "Distance compression in the htc vive pro: A quick revisitation of resolution," *Frontiers in Virtual Reality*, vol. 2, 12 2021.

[41] L. Phillips, B. Ries, V. Interrante, and M. Kaeding, "Distance perception in npr immersive virtual environments, revisited," pp. 11–14, 01 2009.

[42] B. R. Kunz, L. Wouters, D. Smith, W. Thompson, and S. H. Creem-Regehr, "Revisiting the effect of quality of graphics on distance judgments in virtual environments:

A comparison of verbal reports and blind walking," *Attention, Perception, and Psychophysics*, vol. 71, pp. 1284–1293, 2009.

[43] B. K. W. F. R. R. W. P. Gerig N, Mayo J, "Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements," 2018.

[44] K. Vaziri, P. Liu, S. Aseeri, and V. Interrante, "Impact of visual and experiential realism on distance perception in vr using a custom video see-through system," in *Proceedings of the ACM Symposium on Applied Perception*, SAP '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[45] J. Ping, Y. Liu, and D. Weng, "Comparison in depth perception between virtual reality and augmented reality systems," pp. 1124–1125, 03 2019.

[46] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir, "Designing for depth perceptions in augmented reality," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 111–122, 2017.

[47] H. Ishio and M. Miyao, "Importance of visual distance adjustment for ar application of binocular see-through smart glasses," in *2019 14th International Conference on Computer Science Education (ICCSE)*, pp. 1086–1088, 2019.

[48] C. Jerome and B. Witmer, "The perception and estimation of egocentric distance in real and augmented reality environments," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, 09 2005.

[49] J. II, L. Kuparinen, S. Rapson, and C. Sandor, "Visually perceived distance judgments: Tablet-based augmented reality versus the real world," *International Journal of Human-Computer Interaction*, vol. 33, 12 2016.

[50] C. S. Rosales, G. Pointon, H. Adams, J. Stefanucci, S. Creem-Regehr, W. B. Thompson, and B. Bodenheimer, "Distance judgments to on- and off-ground objects in aug-

mented reality," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 237–243, 2019.

[51] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger, "Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr?," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 585–592, 2019.

[52] J. Ping, B. H. Thomas, J. Baumeister, J. Guo, D. Weng, and Y. Liu, "Effects of shading model and opacity on depth perception in optical see-through augmented reality," *Journal of the Society for Information Display*, vol. 28, no. 11, pp. 892–904, 2020.

[53] J. E. Swan, G. Singh, and S. R. Ellis, "Matching and reaching depth judgments with real and augmented reality targets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1289–1298, 2015.

[54] C. J. Lin and B. H. Woldegiorgis, "Egocentric distance perception and performance of direct pointing in stereoscopic displays," *Applied Ergonomics*, vol. 64, pp. 66–74, 2017.

[55] S. E. Kirkley and M. A. Siegel, "Augmented reality performance assessment battery (arpab): object recognition, distance estimation and size estimation using optical see-through head-worn displays," 2003.

[56] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, (New York, NY, USA), p. 9–14, Association for Computing Machinery, 2008.

[57] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer, "Depth perception in augmented reality: The effects of display, shadow, and position," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801, 2022.

[58] B. Krajancich, P. Kellnhofer, and G. Wetzstein, "Optimizing depth perception in virtual and augmented reality through gaze-contingent stereo rendering," *ACM Trans. Graph.*, vol. 39, nov 2020.

[59] J. Ping, B. H. Thomas, J. Baumeister, J. Guo, D. Weng, and Y. Liu, "Effects of shading model and opacity on depth perception in optical see-through augmented reality," *Journal of the Society for Information Display*, vol. 28, no. 11, pp. 892–904, 2020.

[60] A. Murgia and P. Sharkey, "Estimation of distances in virtual environments using size constancy," vol. 8, pp. 67–74, 01 2009.

[61] J. Díaz, T. Ropinski, I. Navazo, E. Gobbetti, and P.-P. Vázquez, "An experimental study on the effects of shading in 3d perception of volumetric models," *Vis. Comput.*, vol. 33, p. 47–61, Jan. 2017.

[62] Unity Technologies, "Unity."

[63] I. Howard, *Perceiving in depth. Volume 1 Basic Mechanisms*. 03 2012.

[64] A. Beall, J. Loomis, J. Philbeck, and T. Fikes, "Absolute motion parallax weakly determines visual scale in real and virtual environments," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 2411, 08 2002.

[65] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3d: A modern library for 3d data processing," 2018.

[66] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. St{"uhmer, T. J. Cashman, B. Tekin, J. L. Sch{"onberger, B. Tekin, P. Olszta, and M. Polle-

feys, "HoloLens 2 Research Mode as a Tool for Computer Vision Research," *arXiv:2008.11239*, 2020.

[67] L. Qian, X. Zhang, A. Deguet, and P. Kazanzides, "Aramis: Augmented reality assistance for minimally invasive surgery using a head-mounted display," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* (D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, eds.), (Cham), pp. 74–82, Springer International Publishing, 2019.

[68] I. Sicaru, C. Ciocianu, and C.-A. Boiangiu, "A survey on augmented reality," 12 2017.

[69] H. Ghandorh, R. Eagleson, and S. de Ribaupierre, "An investigation of head motion and perceptual motion cues' influence on user depth perception of augmented reality neurosurgical simulators," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 557–558, 2018.

[70] J. Wang, H. Suenaga, K. Hoshi, L. Yang, E. Kobayashi, I. Sakuma, and H. Liao, "Augmented reality navigation with automatic marker-free image registration using 3-d image overlay for dental surgery," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1295–1304, 2014.

[71] D. Douglas, C. Wilke, J. Gibson, J. Boone, and M. Wintermark, "Augmented reality: Advances in diagnostic imaging," *Multimodal Technologies and Interaction*, vol. 1, p. 29, 11 2017.

[72] M. Livingston, A. Dey, C. Sandor, and B. Thomas, *Pursuit of "X-Ray Vision" for Augmented Reality*. 09 2012.

[73] S. Palmisano, B. Gillam, D. Govan, R. Allison, and J. Harris, "Stereoscopic perception of real depths at larger distances," *Journal of vision*, vol. 10, p. 19, 06 2010.

[74] R. Konrad and T. Kong, "Depth cues in vr head mounted displays with focus tunable lenses," 2015.

[75] F. El Jamiy and R. Marsh, "Distance estimation in virtual reality and augmented reality: A survey," pp. 063–068, 05 2019.

[76] B. Li, R. Zhang, A. Nordman, and S. A. Kuhl, "The effects of minification and display field of view on distance judgments in real and hmd-based environments," in *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, SAP '15, (New York, NY, USA), p. 55–58, Association for Computing Machinery, 2015.

[77] A. Peer and K. Ponto, "Evaluating perceived distance measures in room-scale spaces using consumer-grade head mounted displays," in *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 83–86, 2017.

[78] E. Peillard, F. Argelaguet, J. Normand, A. Lécuyer, and G. Moreau, "Studying exocentric distance perception in optical see-through augmented reality," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 115–122, 2019.

[79] J. Ping, D. Weng, Y. Liu, and Y. Wang, "Depth perception in shuffleboard: Depth cues effect on depth perception in virtual and augmented reality system," *Journal of the Society for Information Display*, vol. 28, no. 2, pp. 164–176, 2020.

[80] M. A. Cidota, R. M. S. Clifford, S. G. Lukosch, and M. Billinghurst, "Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality," in *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 172–177, 2016.

[81] R. Wang, Z. Geng, Z. Zhang, R. Pei, and X. Meng, "Autostereoscopic augmented reality visualization for depth perception in endoscopic surgery," *Displays*, vol. 48, pp. 50–60, 2017.

[82] M. Kersten-Oertel, S. J. Chen, and D. L. Collins, "An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 391–403, 2014.

[83] C. J. Lin and B. H. Woldegiorgis, "Interaction and visual performance in stereoscopic displays: A review," *Journal of the Society for Information Display*, vol. 23, no. 7, pp. 319–332, 2015.

[84] M. Fischer, C. Leuze, S. Perkins, J. Rosenberg, B. Daniel, and A. Martin-Gomez, "Evaluation of different visualization techniques for perception-based alignment in medical ar," in *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 45–50, 2020.

[85] F. Heinrich, K. Bornemann, K. Lawonn, and C. Hansen, "Depth perception in projective augmented reality: An evaluation of advanced visualization techniques," in *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[86] R. Wang, Z. Geng, Z. Zhang, and R. Pei, "Visualization techniques for augmented reality in endoscopic surgery," in *Medical Imaging and Augmented Reality* (G. Zheng, H. Liao, P. Jannin, P. Cattin, and S.-L. Lee, eds.), (Cham), pp. 129–138, Springer International Publishing, 2016.

[87] H. Choi, B. Cho, K. Masamune, M. Hashizume, and J. Hong, "An effective visualization technique for depth perception in augmented reality-based surgical navigation," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 12, no. 1, pp. 62–72, 2016.

[88] T. Sielhorst, C. Bichlmeier, S. M. Heining, and N. Navab, "Depth perception – a major issue in medical ar: Evaluation study by twenty surgeons," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (R. Larsen,

M. Nielsen, and J. Sporring, eds.), (Berlin, Heidelberg), pp. 364–372, Springer Berlin Heidelberg, 2006.

[89] J. J. Gibson, *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.

[90] B. Faludi, E. Zoller, N. Gerig, A. Zam, G. Rauter, and P. Cattin, *Direct Visual and Haptic Volume Rendering of Medical Data Sets for an Immersive Exploration in Virtual Reality*, pp. 29–37. 10 2019.

[91] M. S. Banks, J. C. A. Read, R. S. Allison, and S. J. Watt, "Stereoscopy and the human visual system," *SMPTE Motion Imaging Journal*, vol. 121, no. 4, pp. 24–43, 2012.

[92] M. Kalia, N. Navab, S. Fels, and T. Salcudean, "A method to introduce evaluate motion parallax with stereo for medical ar/mr," pp. 1755–1759, 03 2019.

[93] M. Kytö, A. Mäkinen, J. Häkkinen, and P. Oittinen, "Improving relative depth judgments in augmented reality with auxiliary augmentations," *ACM Transactions on Applied Perception*, vol. 10, pp. 1–22, 02 2013.

[94] D. A. Bowman and R. P. McMahan, "Virtual reality: How much immersion is enough?," *Computer*, vol. 40, no. 7, pp. 36–43, 2007.

[95] I. Howard, *Perceiving in depth. Volume 1 Basic Mechanisms*. 03 2012.

[96] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. Stühmer, T. J. Cashman, B. Tekin, J. L. Schönberger, P. Olszta, and M. Pollefeys, "Hololens 2 research mode as a tool for computer vision research," 2020.

[97] F. Osti, G. M. Santi, and G. Caligiana, "Real time shadow mapping for augmented reality photorealistic rendering," *Applied Sciences*, vol. 9, p. 2225, 05 2019.

[98] J. T. Kajiya, "The rendering equation," *SIGGRAPH Comput. Graph.*, vol. 20, p. 143–150, aug 1986.

[99] A. Alhakamy and M. Tuceryan, "Ar360: Dynamic illumination for augmented reality with real-time interaction," in *2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT)*, pp. 170–174, 2019.

[100] A. Alhakamy and M. Tuceryan, "Cubemap360: Interactive global illumination for augmented reality in dynamic environment," in *2019 SoutheastCon*, pp. 1–8, 2019.

[101] C. Liu, L. Wang, Z. Li, S. Quan, and Y. Xu, "Real-time lighting estimation for augmented reality via differentiable screen-space rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, pp. 1–1, 01 2022.

[102] T.-M. Li, M. Aittala, F. Durand, and J. Lehtinen, "Differentiable monte carlo ray tracing through edge sampling," *ACM Trans. Graph.*, vol. 37, dec 2018.

[103] S. Liu, W. Chen, T. Li, and H. Li, "Soft rasterizer: A differentiable renderer for image-based 3d reasoning," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7707–7716, 2019.

[104] C. Zhang, L. Wu, C. Zheng, I. Gkioulekas, R. Ramamoorthi, and S. Zhao, "A differential theory of radiative transfer," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 227:1–227:16, 2019.

[105] hferrone, "HoloLens Research Mode - Mixed Reality."

[106] A. Jakl, "Benefits and Parameters of Shadow in Augmented Reality-Environments," p. 42.

[107] M. Haller, "Media Technology and Design."

[108] F. Osti, G. M. Santi, and G. Caligiana, "Real Time Shadow Mapping for Augmented Reality Photorealistic Rendering," *Applied Sciences*, vol. 9, p. 2225, May 2019.

[109] L. Gruber, J. Ventura, and D. Schmalstieg, "Image-space illumination for augmented reality in dynamic environments," in *2015 IEEE Virtual Reality (VR)*, pp. 127–134, Mar. 2015. ISSN: 2375-5334.

[110] J. Dibene and E. Dunn, *HoloLens 2 Sensor Streaming*. Nov. 2022.

[111] E. Heitz, "Sampling the ggx distribution of visible normals," *Journal of Computer Graphics Techniques (JCGT)*, vol. 7, pp. 1–13, November 2018.

[112] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, "Microfacet models for refraction through rough surfaces," in *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, EGSR'07, (Goslar, DEU), p. 195–206, Eurographics Association, 2007.

[113] C. Wheatstone, "Contributions to the physiology of vision.–part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838.