Exploring Human Perception of Image Realism using Professional and Layperson Computational Cognitive Models

Proposal of Research and Related Works

Jack Danner
University of Illinois Urbana-Champaign, Department of Computer Science, Urbana IL, USA
CS 565 Human-Computer Interaction Spring 2021
jackld2@illinois.edu

ABSTRACT

The realism of a computer generated image can be measured by comparing it to a photographed reference image, using black-box machine learning models, or using other data-driven algorithms. However, these methods fail to take into account how a human actually perceives visual realism. There are studies on visual perception as well as computational models that incorporate attributes such as familiarity, aesthetic, color, light, and more in an attempt to model how humans gauge realism cognitively. However, to my knowledge there exist no computational cognitive models that attempt to emulate how a professional in computer generated imagery perceives realism and image quality. I believe that this is valuable information that can be compared to the cognitive model of an average person's judgement. I propose an experimental survey that will be used to create two separate computational cognitive models for professionals and laypersons to computer graphics. I then describe how these models can be compared and combined to explore what analysis techniques professionals find important but do not affect the perception of realism for a layperson.

1 INTRODUCTION

Ever since the creation of realistic rendering techniques and photo editing, there has been a need for computational models that detect and quantify image realism. Realism metrics are important for a wide variety of areas including but not limited to photo modification, evidence analysis, computer graphics research, game development, and other computer generated media [1].

There are many models that succeed in gauging realism, and they vary depending on the content of the image and application domain. For example, the presence of a reference image allows basic methods such as mean squared error (MSE)[1] to be used for close comparison. This approach would be considered *Objective* rather than *Subjective* because it is an automatic process that requires no human input. A subjective approach relies on human perception through experimentation with participants or computational cognitive modeling[4]. Machine learning is often used as a *black-box* approach and does not necessarily require reference images with its testing data. However, in a black-box approach the features of the image and their correlation is generally unknown, unlike a *white-box* approach. A white-box approach attempts to emulate how the human mind and visual system perceive the realism of an image[4]. These terms and comparisons will be further clarified in Section 2.

The research that I propose is inspired by the work of S. Fan et al. [1] who tested participants on their ability to deem an image real or fake. The survey also contained questions about image attributes which were then empirically modelled for correlation to realism. These correlating attributes such as *naturalness, oddness,* and *attraction* were automatically measured in multiple ways by parsing an image. Machine learning regression and classification were used to predict realism. This computational cognitive model was then compared to CNN-learned features in a black-box model. S. Fan et al. 's experiment showed that regression performed better for their human perception model than a black-box model with multilayer perceptron (MLP) chosen attributes. This is because their regression model was trained with human perception as the ground truth rather than the actual truth labels on the image. It was also shown that gamers, photographers, and graphic designers have higher sensitivity to fake images than a layperson (little exposure to computer generated images)[1].

The aspect of computer graphics and realism perception that this proposal focuses on is the distinction between what I will call *true realism* and *perceived realism.* For many domains in the computer graphics industry true realism (absolute photorealism with negligible reference image error) is not necessary. In fact, a lot of domains would benefit from a computational model that could accurately predict the perceived realism of a piece of media where a layperson or consumer’s ability to discern CGI from photograph is the ground truth. In other words, the model would be able to show how close a piece of media is to appearing photo real to a consumer. Absolute photorealism is attainable in computer graphics but it is computationally expensive and requires professionals with a vast knowledge of rendering techniques and a robust understanding of light transport. I believe that a consumer-based model could describe where complex light transport simulation leads to diminishing returns in more detail than ever before.

I predict that an exploration and comparison of separate computational cognitive models for CGI professionals and consumers respectively will reveal attributes of visual realism that are important to CGI professionals but not to consumers and laypersons. I propose using a modified version of the attribute survey used by S. Fan et al. to reflect what visual cues a CGI professional may use to gauge visual realism in an image they are presented with. For example, a professional may look for specular highlights, coherent shadows, ambient occlusion, and realistic light refraction. Since your average consumer usually cannot ascribe lack of realism to these technical attributes, the layperson would receive a generalized survey of attributes. Generalized attributes such as natural lighting, shadows, and color would be a parent category of two or more of the technical attributes presented to a professional participant. The dataset would comprise of CGI renders that exhibit commonplace objects and various examples of light transport. Each image would vary based on the quality and presence of technical attributes displayed in the image, and the attributes would be equally represented across the dataset. Another equal set of real photographs that contain similar scenes would be added alongside the renders.

Two equally-sized groups of participants would need to be sourced. The first group would consist of CGI professionals that have a background in computer graphics and are intimately familiar with all the terms presented on the survey. The second group would consist of consumers of computer generated media with only consumer level knowledge of the technical attributes in the professional survey. Just like S. Fan et al. the attribute correlation to realism for each survey would be determined. An examination of attribute correlations could then be made. Finally, a regression model could be constructed based on the technical attributes informed by the general attributes using a support vector machine (SVM). If this layperson-informed professional model were to perform almost identically to the strictly layperson model, it would lead to answers about what professional CG render techniques a layperson unknowingly finds important and which are too subtle for a layperson to deem detrimental to an image's realism. The rest of the paper is organized as follows. In Sec. 2 I describe related work that this proposal builds upon. Sec. 3 describes the attribute survey and the details of a possible dataset. Sec. 4 explains correlation analysis and the implementation of the computational cognitive models. Sec. 5 concludes with implications, relevant application domains, and future work.

2 RELATED WORKS

Related work topics span computer graphics, cognitive engineering, machine learning, and CG quality assessment.

2.1 QUALITY ASSESSMENT IN COMPUTER GRAPHICS

Lavoué's writing on quality assessment in computer graphics[4] defines some important terms that are key to understanding the use of a computational cognitive model for gauging realism. He defines *objective* vs *subjective* quality assessment as two distinct evaluations. A subjective approach relies on the opinion of observers while a subjective approach is an automated process. Objective approaches often use reference in their calculation while generally subjective approaches do not require reference material. Next, he defines *black-box metrics* as an approach to modeling quality where the underlying decision making and important features remain hidden and machine learning is used. *White-box metrics* are defined by a model that tries to emulate human visualization and perception.

Another distinction made is between *image artifacts* and *model artifacts*. In 3D computer graphics, objects are represented as meshes, essentially a long list of triangles in

3D space. These triangles are paired with material attributes such as texture coordinates, normal vectors, and any other useful information that is required to shade an object. Image pixels in a render are generated from this information using rasterization or ray tracing. Any artifact or imperfection that is a result of *shaders* (a function that tells how to shade a pixel based on 3D information and calculations) can be considered a model artifact. Artifacts that arise from calculating these pixel colors and modifying them with color correction, HDR, or any other process that does not require 3D information can be classified as image artifacts.

2.2 MEASURING PERCEPTION OF VISUAL REALISM

Rademacher et al.'s study on the perception of realism links some interesting factors to how humans gauge realism[2]. It was confirmed that the softness of shadows positively correlates with rated realism for the human visual system. It was also discovered that the smoother a surface in an image is, the less likely it is to be perceived as real. Further experiments with the number of objects and lights in a scene have no significant correlation with realism perception[2]. Surprisingly, all of Rademacher et al.'s experiments used real photographs of constructed scenes. The fact that real images can be perceived as fake proves that a successful model based on human perception would perform much differently than a true realism model.

2.3 COMPUTATIONAL COGNITIVE MODELING

Byrne details in *The Oxford Handbook of Cognitive Engineering* the considerations and challenges of interaction between a cognitive model and an external environment. Environments can be *static* or *dynamic* and *direct* or *distal* in their complexity[3]. Complex and dynamic tasks such as piloting or driving generally require simulated environments for participants that a cognitive model can interface with. Since a human judging an image is a static environment and interaction boils down to observing color values, there are no significant environmental interaction challenges when creating a computational cognitive model for realism. Computational cognitive models are used to answer quantitative questions about human performance and cognition in a task[3]. It is easy to hypothesise and test whether features such as shadow quality or texture detail in a CG render affect the visual perception of realism. However, a cognitive model of realism would provide the ability to quantify realism correlation of attributes like shadow softness down to the penumbra angle[2].

S. Fan et al.'s [1] paper on modeling human perception brings together previous work on measuring realism and integrates abstract visual attributes such as aesthetic, familiarity, emotion, and sentiment into their model[1][5]. In their experiment, they provided participants with a survey containing many attributes divided into categories of image composition. Participants were to annotate images with the attributes that applied to each test image and then state whether the image was real or fake. Later experiments also allowed participants to use a scale to assign a level of realism. They discovered significant correlation with realism for attributes in the familiarity, color, illumination, aesthetic, and human semantics categories. In order to automatically measure these attributes for a cognitive model, they used different algorithms and techniques for each attribute. Finally they developed an empirical model with an SVM using the correlating attributes. The SVM performed better than other objective methods and slightly out-performed a CNN that used black-box learned attributes. My proposal reuses the framework of their model and uses a different attribute survey and dataset.

3 EXPERIMENT

In this experiment, a model of CG professionals' perception will be compared to a layperson's perception. CG professionals will be asked questions broken into categories about advanced computer graphics analysis. These questions will represent attributes that can be automatically measured for a cognitive model. Laypersons will be asked a smaller set of questions that generalize a whole category into one question.

3.1 MODIFICATIONS OF SURVEY

S. Fan et al.'s survey style and correlation analysis should be used in this experiment. An online survey should contain 5 images with questions that pertain to chosen attributes. Since this experiment compares professionals to laypersons, the survey attributes from S. Fan et al.'s paper that will most likely not differ due to technical knowledge shown be removed. Attributes that relate to aesthetic, emotion, and familiarity should not be included in either survey. For simplicity, human semantics (human presence and expressions) will not be included in the survey and people will not be present in the dataset images. Human face perception is an incredibly complex topic on its own.

3.2 PARTICIPANT SELECTION AND CONSIDERATIONS

It is important to define what exactly a layperson and a CG professional are for this experiment. A layperson will simply be defined as a consumer of visual media such as games, art, simulations, and production graphics. A CG professional will be defined as a person who creates CG media and must judge their work's level of "realism" in the process. A CG professional could also be an individual highly familiar with computer graphics through education or career. They must know all of the attributes contained on the survey intimately. Professionals that only work with 2D or hyper-stylized media may not meet these requirements.

3.3 ATTRIBUTES

I will now introduce possible attribute categories and technical attributes that are worth exploring in this experiment. Note that all of these attributes will be a yes-or-no question with the exception of "**How realistic is this image?**" which will be a scale of 1-5. Layperson participants should be given instructions before the experiment to minimize misunderstandings about terminology. The definition for each attribute should be given.

Layperson General Attributes of Quality	CG Professional Attributes of Quality (Technical Attributes)			
Textures	Non-repetitive	Seamless	High Resolution	Depth
Illumination	Specular	Ambient Occlusion	Environment Light	Diffuse Lighting
Shadows	Hard Shadows	Soft Shadows		
Transparency	Correct Refraction	Caustics	Absorption	
Reflections	Same Environment	Correction Reflect	Perfect	
Objects	High-poly	No Clipping		
Color	HDR	Saturation	Temperature	
Image Quality	Anti-Aliasing	Resolution	Motion Blur	Bloom

Fig 1: List of proposed general layperson attributes and corresponding technical attributes

A layperson should be able to decide whether they feel the texture of an object is unrealistic. Objects that are smooth or contain no texture are generally perceived to be more unrealistic[2]. *Non-repetitiveness* will be defined as the absence of repeating patterns on a surface. *Seamlessness* is the absence of visible seams between textures. A professional's satisfaction with the scene's texture resolution and depth will be recorded as well.

Laypersons will be asked if objects are naturally lit in the scene. Shadows will be specified as a different category. In computer graphics, lighting is often broken down into three components: *ambient*, *diffuse*, and *specular*. These components were popularized upon the creation of the Phong reflection model[7] which simulates each component using only vector and color information at a given spot on an object. In some situations Phong shading can look photoreal, but for better realism path tracing and or photon mapping[7] are used to simulate realistic light transport at a major computational cost. The ambient component is the result or simulation of environmental light such as the sun. Diffuse lighting defines roughness of a surface, while specular defines reflectiveness.

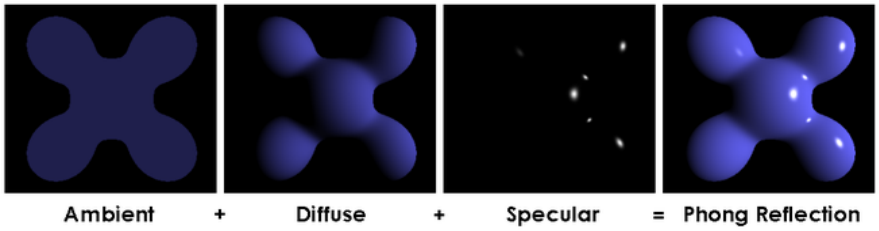

Fig 2: Visualization of the Phong reflection model's components.

Hard shadows will be defined as shadows that result from a single light source that is relatively far away, or a point light. Soft shadows will be the result of multiple or large light sources.

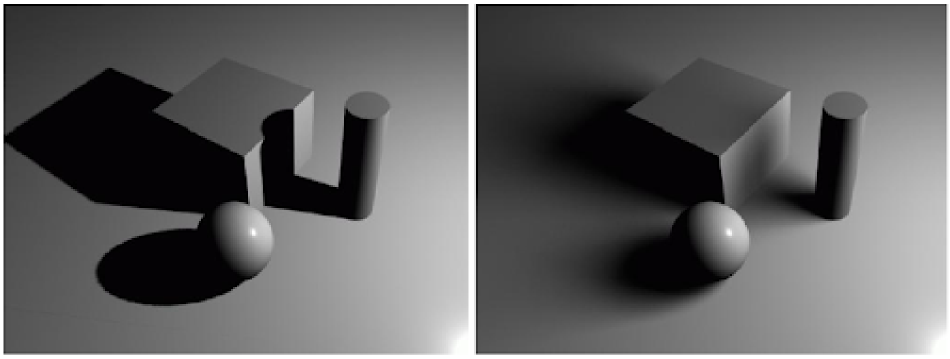

Fig 3: Hard vs soft shadows

The technical attributes for transparency will be broken down into caustics, correct refraction, and absorption. Caustics appear at curved transparent surfaces as seen in figure 4. Absorption will be the presence of objects that absorb some wavelength of light, such as colored glass. Finally, correct refraction will be determined by if light is being refracted in a physically accurate way. For example, a glass sphere will refract light upside down due to it's curve.

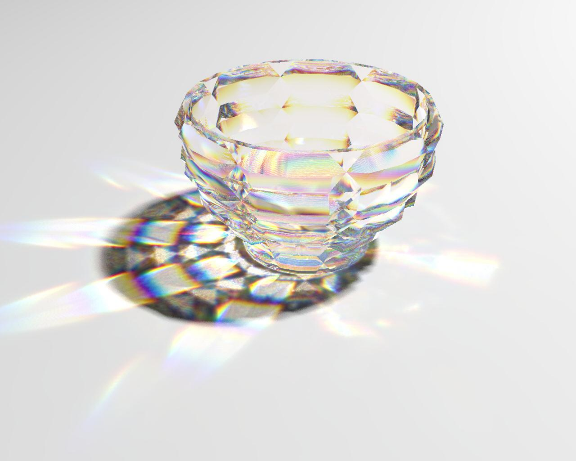

Fig 4: Example of caustics

Reflection will be similar to transparency in that correction reflection will be defined as physically accurate light rays in terms of distortion. Environment accuracy will be an attribute because often environment maps that are not accurate to the true environment are used for the sake of performance in many cases. Perfect reflection will be the presence of objects that reflect all light and have no texture or diffusion.

In terms of resolution and fidelity, object polygon count, high dynamic range, anti-aliasing, resolution, and other post-processing effects will be technical attributes.

3.4 DATASET

The dataset of this experiment would ideally consist of 1250 computer generated 3D renders and 1250 photographs that contain objects, shapes, textures, and environments that are familiar to the average person. These images should be curated by separate non-participating CG professionals in a way where attribute presence is as equally distributed as possible. No renders or photographs should contain humans or animals to avoid strong external factors that are not taken into account on the survey.

For the chosen photographs, it is important that there are no “dead give-aways” such as hyper-realistic cluttered environments or extremely high object counts. It is also important to choose some images that one could easily mistake for a computer render. Images with perfectly smooth surfaces and single light sources are a good example of this. The computer generated images should include ray-traced[6] production level renders, rasterized real-time computer graphics renders[7], and various reflection and light transport models such as phong shading[7] and path tracing[6].

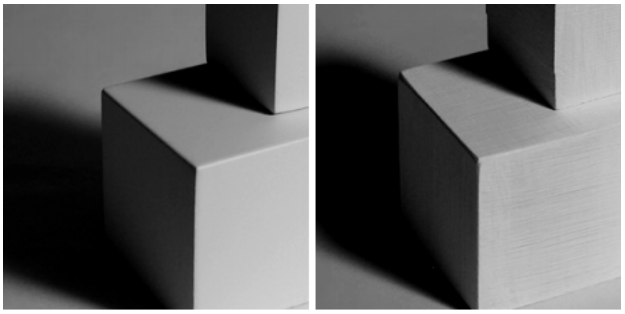

Fig 5: Both are real photographs of objects with different surface texture. The left was perceived as less real in experimentation[2].

4 MODELING PERCEPTION AND QUALITY

4.1 CORRELATION ANALYSIS

Like S. Fan et al.’s model, realism scores will be defined as the fraction of people who judged an image as real. Similar results can be inferred because the parameters of realism judgement are the same in the proposed experiment. Also like their model, signal detection theory will be used to compute a sensitivity index *(d’)* which shows signal separation from noise[1]. CG images are considered noise while real photographs are considered the signal. The resulting d’ values for each group of participants will show at what rate they correctly identify photos as real.

Just like previous work, Spearman’s rank-order correlations will be used to measure attribute correlation with realism nonparametrically[8]. These realism correlations (ρ_r) will be unique for each participant group. A second set of correlations (ρ_g) will be constructed for CG professional survey attributes. This second set will measure the correlation of professional attributes with their corresponding general layperson attribute instead of realism. These ρ_g values will demonstrate which technical attributes of each category correlate with a layperson's general view on the category. This will hopefully allow a quantitative description of a category. For example, during correlation analysis it could be discovered that there is a strong correlation between realism and realistic illumination for the layperson participants. The ρ_g values for the professional illumination attributes *specular, ambient occlusion, environmental light,* and *diffuse* will then show which technical aspects of the category a layperson is sensitive to.

To experiment further, there is a challenge in choosing features for the cognitive model. S. Fan et al. use exploratory factor analysis (EFA) to observe linear combinations of attributes that they grouped into conceptual "factors". However, this proposal focuses on exploring how a CG professional's expertise may introduce bias to their work and how they may underestimate the realism of the image. These conceptual factors will need to be formed by: (i) analyzing what general attributes correlate to realism for a layperson the most, and (ii) decide what technical attributes are not relevant. It is difficult to map out this process beforehand but I am confident that it could be done.

4.2 MODEL IMPLEMENTATION

Once the combinations of attributes for the cognitive model have been chosen, they would need to be measured automatically. I believe the best way to relay environmental information of the image to the cognitive model would be to design a smaller testset of images in a controlled environment using render software. Lighting, texture, color, and model information is readily available and can be passed straight to software. S. Fan et al.'s cognitive model was designed to operate on images without access to the 3D engine and scene environment so their empirical modeling relies solely on image processing algorithms. A cognitive model that uses render software as an environment lacks the ability to process photos and image files without a source, but this can actually be a benefit and is explained in Sec 5. This new dataset will require more layperson participants to get a ground-truth for realism and attribute correlations of the images.

An SVM regression model will be used with these automatic measurements of the cognitive model. A comparison could then be made to the ground truth realism of layperson participants. If the model performs identically to the participant ground truth then the model can be considered successful. The correlations of the general attributes with realism as well as parsed technical attributes could be examined and used for further modeling correction. It is possible that multiple iterations of the model would need to be made and different features would need to be tried to see novel results.

5 CONCLUSION

I believe that a successful computational cognitive model that predicts a consumer's perceived realism of a computer render using CG software parameters is useful in the domain of computer generated imagery. The human visual system is excellent at gauging whether an image is visually appealing or realistic but many of the underlying cognitive processes responsible remain a mystery. CG professionals have a vast knowledge of rendering techniques and a robust understanding of light transport that they use when creating and analyzing images. As a result, their perception of realism is different from the average consumer and their judgement of realism differs from a layperson. This may introduce bias when judging their own work and they may strive to achieve their own standard of realism rather than the target audience, leading to wasted resources.

There are many approaches that can be taken when attempting to make an image seem photo real, and the amount of resources and computational power needed are not equal. Real time computer graphics has always been a practice of tricking the eye for the sake of higher performance. For example, realistic reflections can be achieved with an environmental map[7] instead of simulating reflected light rays which is vastly more computationally expensive. Researchers could use a layperson-driven computational cognitive model for realism when comparing newly developed render techniques that strive to replace their more expensive counterparts and quantify a “loss” in realism.

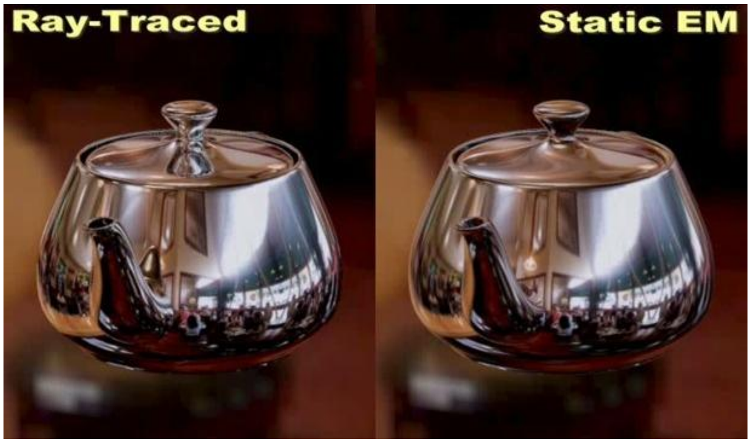

Fig 6: The left teapot portrays costly realistic light ray reflection, while the right uses a cubic environment map to shade reflections at a fraction of the cost.

It has been shown that real photographs can seem fake due to certain lighting factors[2]. Therefore, traditional realism models that use reference images may not be useful if the scene is inherently unrealistic to a consumer with smooth surfaces and single light sources. The proposed computational model would be useful for professionals to gauge their work for a target audience. It could save time and guarantee an unbiased quantitative value of visual appeal and realism.

5.1 FUTURE WORK

It is worth noting the lacking features of the model due to the dataset only using easily recognizable objects in scenes without animals and people. This means that aesthetic, CG characters, and emotional appeal are possible avenues to explore in the future. The model also judges images in a generalized way such that certain areas of the screen space or certain objects are muddled together. An object or fragment based model could be considered.

A successful model could provide real-time sentiment or realism analysis for a given viewport in a scene. Think along the lines of traditional real time graphics benchmarks such as frame rate counters and GPU utilization but for human perception. This could be especially useful for immersive simulations if absolute realism is desired. A developer could examine aspects of their scene and mark viewpoints or areas of the simulation that are lacking in realism. The cognitive model could also be automated in such a way that it scans a 3D environment and gives a 3D mapping of realism.

I also believe that a real-time cognitive model of an interactive environment viewport does not have to be limited to realism but could be refactored for many things such as fear, anxiety, or other emotions and sentiments. It would only require different attribute surveys and a different dataset.

REFERENCES

[1] S. Fan et al., "Image Visual Realism: From Human Perception to Machine Computation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 9, pp. 2180-2193, 1 Sept. 2018, doi: 10.1109/TPAMI.2017.2747150.

[2] Rademacher, Paul & Lengyel, Jed & Cutrell, Edward & Whitted, Turner. (2001). Measuring the Perception of Visual Realism in Images. 12th Eurographics Workshop on Rendering. 10.2312/EGWR/EGWR01/235-248.

[3] Byrne, Michael D. 2013. Computational Cognitive Modeling of Interactive Performance. In *The Oxford Handbook of Cognitive Engineering*, John D. Lee and Alex Kirlik, Eds. Oxford University Press

[4] Lavoué, Guillaume. Mantiuk, Rafał. 2014. Quality Assessment in Computer Graphics. In *Visual Signal Quality Assessment*, Chenwei Deng, Lin Ma, Weisi Lin, King Ngi Ngan, Eds. Springer. http://dx.doi.org/10.1007/978-3-319-10368-6 9

[5] E. Cetinic, T. Lipic and S. Grgic, "A Deep Learning Perspective on Beauty, Sentiment, and Remembrance of Art," in IEEE Access, vol. 7, pp. 73694-73710, 2019, doi: 10.1109/ACCESS.2019.2921101.

[6] Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. Physically Based Rendering: From Theory to Implementation (3rd. ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[7] Peter Shirley and Steve Marschner. 2009. Fundamentals of Computer Graphics (3rd. ed.). A. K. Peters, Ltd., USA.

[8] R. Bailey, Design of comparative experiments, vol. 25. Cambridge University Press, 2008.

[9] Chris Wyman et al. 2018 *ACM SIGGRAPH Introduction to DirectX Raytracing.* Retrieved April 21, 2021 from http://intro-to-dxr.cwyman.org/