

Abstract

With new technology, computer generated imagery (CGI) is closer than ever to becoming indistinguishable from photography. Architectural artists are keen on recreating the chaos which surrounds us; small details like finger prints, scratches, dirt etc. which is what lifts the visualisation from clinical and pure and into hyper realistic visualisations. However, with these realistic CGI it may mislead any potential client as the end result may be very different from what the artists had in mind.

The lack of validation of CGI surface materials may be one of the problems that ultimately leads to different visual appearance. In this project a novel approach was designed to validate real surface materials in two different environment with different levels of complexity (level of semantics), whilst keeping human perception in mind. The purpose of having two environments was to see, if there were any differences in visualising surfaces in a simple environment compared to a more complex environment and if the threshold of accepting what was similar were consistent in the two environments.

Three materials were processed in the project; a highly specular material (whiteboard), a diffuse material (post-it note) and a glossy material (table), which were all calibrated as close as possible to a real reference. Small deviations were made to the calibrated sample in order to get a range of samples that differed from the calibrated sample.

The results from the experiments shows that whilst the assessors were able to discern between the changes in the two environments, the low semantic environment were rated more in concordance between assessors and in general had more defined groupings for the specular and diffuse materials compared to the same materials implemented in the high semantic environment. The glossy material had very little groupings in both of the environment and indicates a larger range of acceptance for this type of materials.

Table of Content

1.	Introduction and Motivation.....	4
1.1	Human Perception and Computer-Generated Imagery.....	5
1.2	Psychophysics of Appearance	7
1.2.1	Perception of Geometric Shapes.....	7
1.2.2	Perception of Material/Texture.....	7
1.2.3	Perception of Illumination	8
1.3	Semantic Context.....	9
1.4	Cameras a Dimension of CGI.....	10
1.5	Summary.....	13
1.6	Hypothesis	14
2.	Analysis	15
2.1	Physical Properties of Materials and Light Interaction	15
2.1.1	Specular Component:	15
2.1.2	Directional Diffuse Component.....	16
2.1.3	Uniform Diffuse Component.....	16
2.2	BRDF's Models in Computer Graphics	16
2.3	Photometry and CG	18
2.4	Linear Workflow	19
3.	Procedure.....	20
3.1	Measuring an Approximate Reflectance of a Given Surface	20
3.2	Validation of Light Transportation.....	21
3.2.1	Validation of IES Files	22
3.2.2	Recreating the BRDF of a Given Sample Surface.....	27
3.3	Summary.....	29
4.	Implementation.....	30
4.1	Low semantic environment implementation	30
4.2	High Semantic Environment Implementation	34

4.2.1	Reflectance and Light Measurements.....	35
4.3	Experiment Design	38
5.	Results	40
5.1	Low Semantic Experiment	40
5.2	High Semantic Environment Test.....	42
5.3	Comparison of Semantic Environments	46
5.4	Visualization of BRDF's and Thresholds	47
6.	Discussion.....	48
7.	Conclusion.....	51
7.1	Future Development	52
8.	References	53
9.	List of Figures.....	56
10.	List of Tables.....	58
11.	Appendix 1, IES file example.....	59
12.	Appendix 2, IES Re-creator Program.....	60
13.	Appendix 3, Data Gathering Code	65
14.	Appendix 4, Instructions for the First Experiment.	66
15.	Appendix 5, Instructions for the Second Experiment.....	67

1. Introduction and Motivation

Creating photorealistic imagery has long been a holy grail for visualizing artists. In the process of creating such computer generated imagery (CGI), many aspects needs to be considered. Such as, how the lighting looks in this room or maybe if the floor should have more details etc. Many architectural visualizing artists uses various little tricks to create imperfections in their renderings such as "scratches in metal, splinters and chips in timber boards, even fingerprints" (Goss, 2013).

Photorealistic renders are slowly becoming more and more indistinguishable from actual photographs and may soon be completely realistic; so called "hyper-realistic renderings" (Goss, 2013). The future is exciting within this field, as a range of possibilities exists for creating materials and modifying lighting before any building has been built. But one major drawback of a architectural visualizations is the risk of creating "too stunning" images where *"The danger is that the client comes along at the end of it, sticks in a whole bunch of crap furniture and then the photographs of the building aren't as good as the render and everyone calls you out on it."* (Goss, 2013).

The risk can be overwhelming when one gives the artists a "carte blanche" to do whatever they want to make it look stunning. To avoid this it would be appropriate to use the 3D software to validate the renders in terms of reality, i.e. the materials, lighting, camera etc. could be efficiently calibrated and documented for a given render. This would not only help the architects, but also the client who will be able to confidently understand how a given project will look, based on measurements and solely by artistic interpretations. It is already happening and as Henry Goss, a professional architectural visualization artist, puts it:

"I see the whole industry heading in the direction where you have a single [digital] model [of a building]. This happens in a lot of the big commercial practices. They have a single model, which not only has all the architectural components, building services components and the structure and coordination of all things, but it's also testing lighting levels and testing environmental factors.

Computer-aided design has now reached a level where it's all becoming very integrated. The visualisation isn't purely visualisation anymore – you can actually use the same [digital] models with the same lighting rigs to test real-life environments and real-life situations. I use it to a small degree but people take it to a greater degree where they are really testing the actual lux levels in a space on a full environmental model." (Goss, 2013).

With this it's believed that one can formulate a workflow where, given photometric data, it is possible to recreate physically accurate materials and lighting in an environment ensuring the renders are physically validated. In this project I will focus on how human's perceive different surfaces in different environments with varied complexity.

1.1 Human Perception and Computer-Generated Imagery

Striving for visual realism in computer graphics (CG) has always been a strong goal for many researchers and artists. Throughout the inception of the term "photorealistic computer-generated imagery", there have been many attempts to utilize human perception to identify what exactly makes an image, both real and rendered, "realistic". In this section, I will investigate what previously has been done in this field.

Generally speaking, one can divide 3D rendered scenes into sub-elements (Ramanarayanan, Ferwerda, Walter, & Bala, 2007):

- Geometry, this is what (almost) every 3D model is built from, high polygonal models tends to generate more smooth curves and is in general used for offline rendering. A low polygonal model can be used for real-time renderings, where efficiency often is prioritised above appearance.
- Texture/material, Texture maps are used to give models its perception of colour (RGB map) or depth (bump/normal maps), furthermore maps such as reflectance maps, can enhance the perception of a i.e. scratchy surface or small details such as fingerprints on a glass surface making the object appear used. Materials can be used to simulate almost any surface using a Bidirectional Reflectance Distribution Function (BRDF). Depending on the BRDF, the rendered object will take completely different appearances i.e. human skin or a mirror.
- Lighting the scene is extremely important for any scene to appear aesthetically appealing, lights are often versatile can be manipulated by size, shadows, colour, intensity etc. All of these factors have a huge influence on the rendered scene.

This simplified list of elements is the basic building blocks common in 3D scenes. Many more processes are involved, e.g. rendering, rigging and animation. However, in the scope of this project, I will focus on a still-frame image excluding rigging and animation.

Rademacher, Lengyel, Cutrell, and Whitted (2001) investigated some of the afore-mentioned elements in a simplified scene which would either be a real photograph or a rendered scene, and they investigated how the number of lights, surface smoothness and shadow softness influenced the perceived realism. They asked users who either saw exclusively CG or photographs of their perception of the realism portrayed in the scene. Interestingly in some cases, users assessed the photographs unrealistic (which is per definition "photo-real") (Rademacher et al., 2001). This leads to the question of context. The users were presented with rather simple abstract scenes, which they could not put into a coherent context. Humans are neurologically programmed to scan for small details, i.e. human faces. This diversity in detail is very important for humans to have in a context (Reinhard, Efros, Kautz, & Seidel, 2013), meaning semantically rich CG renderings could be one of the majoring factors for creating CG images. A study by Elhelw, Nicolaou, Chung, Yang, and Atkins (2008) used abstract context to examine where the borders of perceived realism were broken. Using photographic references showing different static scenes from a bronchoscopy, they employed a self-reported questionnaire

to evaluate different levels of rendered quality. Five levels of stimuli were used sorted by visual quality (1-lowest, 5-highest):

1. Texture detail of the stimuli was affected by a Gaussian blur (7x7 kernel size).
2. Texture detail of the stimuli was affected by a Gaussian blur (3x3 kernel size).
3. Original texture image.
4. Original texture image with added specular highlights.
5. Actual photograph (Elhelw et al., 2008).

From the initial questionnaire, category 1 and 2 stimuli were too obvious for the users and removed. Categories 3 and 4 were further tested using a two alternate forced choice method (2AFC), which essentially is a side-by-side comparison task.

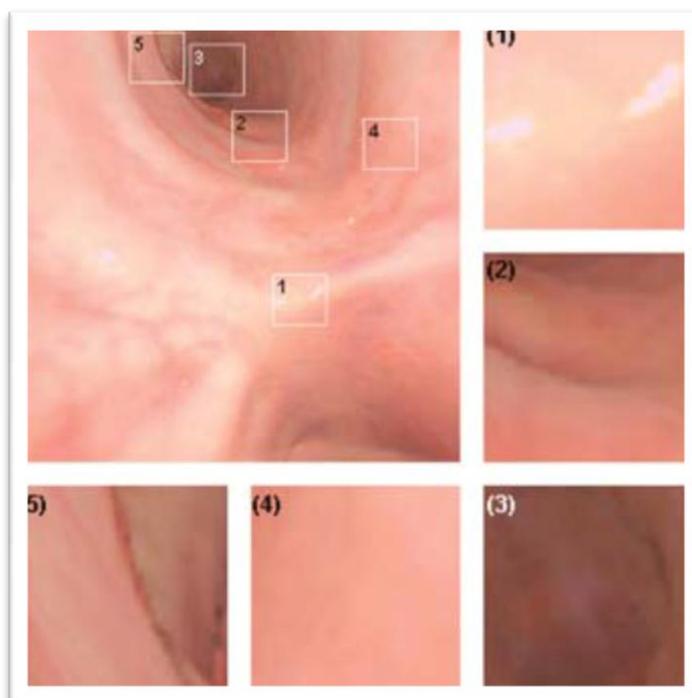


Figure 1: Rendered image of a bronchoscopy. The regions depict areas of interests where users based their realism decision making. Image from Elhelw et al. (2008).

Eye tracking was used to identify where the users showed the greatest interest in deciding whether it was real or not. The eye-tracking data provided details of where the users subconsciously made the decision, which were in areas with: light reflection/specular highlights (area 1 in figure 1), Contrast details (area 5) and 3D surface details (area 3). The results of the study showed that category 4 was in terms of realism the highest rated stimuli, and that users could not tell the difference between the reference and that specific stimuli, hinting that specular highlights could be linked to increased perception of visual realism (Elhelw et al., 2008). After each session the users were told to answer questions on which image aspects they made their decisions

on. The scores did not correspond to the eye-tracking data which underlines an important thing to keep in mind when using post-test questionnaires, namely that users are often not very good at recalling visual information and may be prone to bias such as the recency effect (Aldridge, Davidoff, Ghanbari, Hands, & Pearson, 1995).

1.2 Psychophysics of Appearance

Why do things look the way they do? Why are certain objects more or less perceived as rough or as shiny? Which influence do the afore-mentioned 3D sub-elements have when humans are evaluating visual stimuli? From a perceptual standpoint, many variables must be taken into consideration when humans make sense of a visual stimulus. In this regard, one could use the afore-mentioned 3D-sub-elements as a starting point in exploring the potential of "cheating" the Human Visual System (HVS).

1.2.1 Perception of Geometric Shapes

Perceiving objects is one of the fundamental processes in the human visual cortex and is based on principles articulated by the members of the Gestalt school (Wolfe, 2009). Simply put, the HVS relies heavily on grouping patterns, such as similarity, proximity, parallelism, symmetry etc. Furthermore, we have processes that complete contours and objects even when they are partially hidden behind occluders (called relatability heuristic) (Wolfe, 2009). Exactly how the HVS reconstructs objects has previously been investigated, and it has been found that the HVS is not always able to recover shapes under certain conditions from a fixed viewpoint (Belhumeur, Kriegman, & Yuille, 1999). At Pixar, they use a pipeline where models are reduced stochastically in detail. The geometric data is slightly enlarged to compensate for the reduced polygons all whilst still keeping a perceptual minimal impact (Reinhard et al., 2013). This can be related to how humans perceive details. This perception of fine details can be described as visual acuity (Wolfe, 2009) measured in cycle/degree (cy/de) in grating patterns. In the HVS, a contrast sensitivity function describes how the relation of cy/de affects the overall perception of contrast in a given grating pattern ranging from low (0.1 cy/de) to a high spacing (100 cy/de), where the peak of acuity perception of fine detail is around 3-4 cy/de (Wolfe, 2009), this basically means one should keep the limits of the HVS in mind when creating models for background areas with lots of detail.

1.2.2 Perception of Material/Texture

From a historic perspective, little research has been done on material perception ((Beck, 1972) cited in (Ramanarayanan et al., 2007)), however recent research found there is a correlation between perception of material and realism (Ramanarayanan et al., 2007). Their research shows, that the less an individual can see in a reflective material, the more they are likely to perceive it as equivalent to a more reflective material with another reflection map (Ramanarayanan et al., 2007). This, however, is only valid when comparing the same objects. In a study by Olkkonen and Brainard (2010) it was found that specular reflections and glossy

reflections can be hard to match precisely with a referenced. In the study, users were instructed to match the reflective and diffuse components to match a reference with particular diffuse or specular/glossy levels. Their findings shows that whilst diffuse samples were matched veridically with the reference the more specular and glossy samples were not matched as close to the reference, indicating that humans are more likely to accept imperfections.

When comparing objects of different shape rendered with the same BRDF, it is noticeably different and not perceptually equivalent (Vangorp, Laurijssen, & Dutré, 2007). This principle has also been extended into dynamic scenes (Vangorp et al., 2009). It should be noted when using the visual equivalence metric that it is necessary to have scene knowledge such as illumination, geometry and materials (Reinhard et al., 2013).

1.2.3 Perception of Illumination

Very similar to the materials, it is important how both the direct, indirect and reflected lighting is perceived. Vangorp et al. (2007) used a method to predict how a change in incident illumination affects the appearance of an object, depending on its geometry and material, which can lead to shortcuts or approximation of perceived materials whilst keeping the compared objects perceptually equal. Computationally speaking, indirect lighting can add significant time to a render, especially in a scene with many reflective and glossy materials. One of the more heavy time consumers is the computation of visibility (Reinhard et al., 2013). It has been shown that high accuracy of the indirect lighting is not strictly necessary as it has little impact on the perceived scene, this is particularly true for the combination of glossy reflections and occlusion (Kozlowski & Kautz, 2007). When simulating indirect lighting often found in offline renderings, humans are not the best to pick up subtle changes in differences in indirect lighting which means high accuracy is not always required (Yu et al., 2009). Whilst analyzing a scene, humans tend to have difficulty noticing inconsistencies in the lighting directions, and thus humans have a hard time estimating illumination (Reinhard et al., 2013).

Ulbricht, Wilkie, and Purgathofer (2006) reviewed the previous research within the field of verifying the render algorithms used to generate physically based renders. They proposed that the process of verifying photo realistic renders should be split into three steps:

1. First, one has to prove the correctness of the light reflection model through comparisons with measured physical experiments.
2. Verify the light transportation simulation (rendering algorithm).
3. Generate a simulated image using the radiometric data provided by the rendering algorithm, it has to take the output device and viewing conditions into account as well. (Ulbricht et al., 2006)

These steps can provide an important notion to remember when doing physical comparison studies. In the first step, one has to be certain that the material which is being simulated is correctly displayed when rendered, this often has to do with the specific material's BRDF which describes how light is reflected on an opaque surface. This function is uniform for all materials and has been implemented in rendering engines used in most graphic software. One way to gather data on a materials BRDF, is to use scientific equipment such as the gonioreflectometre (Ulbricht et al., 2006). However, this procedure is very costly. Verification of the rendering algorithm could be done by comparing the rendered image with an actual photograph, or the comparison could be done by using a multistage validation procedure as done by Myszkowski and Kunii (2000) or by comparing illuminances in the reference versus the rendered image (Mardaljevic, 1999).

Drago and Myszkowski (2001) investigated the importance of the BRDF components and how those influence the light transportation model. They used BRDF measurements to recreate a complex scene, and this render was contrasted with a render based on artistic impressions made by a skilled artists. The renders were rated on how well the scenes were reproduced. In terms of image fidelity, the artistic rendered image was rated higher than the BRDF measured image, however in terms of light accuracy, the BRDF measured render was rated highest. One thing to keep in mind is that, as previously discussed, humans tend not to be very accurate when comparing the accuracy of indirect illumination. One of the main issues with the experiment was the quite complex environment with many reflections with particular dispersions, none of which the BRDF render or artistic render managed to visualize.

1.3 Semantic Context

As mentioned, semantic context can be an influential factor, especially when it comes to assessing whether something is real or not. The everyday scenes humans perceive comprise a myriad of stimuli coming from the smallest speck of grain to a large building consisting of hundreds of materials. All of these stimuli are what makes it real to humans. To achieve a result similar to this in the CG world is notoriously hard, due to the sheer amount of imperfections and little details that need to be simulated. Reinhard et al. (2013) proposed four approaches for recreating visually rich environments:

1. Modelling everything. A laborious task which mainly is being done by larger animation/vfx houses who have the budget and time to achieve this task. This approach, however, does not always equal a success for the given projects.
2. Physics-based methods, or simulation methods. This is mainly for recreating explosions, water, or anything particle based, where Newton's laws are fairly consistently modelled.
3. Image Based Rendering (IBR) is a fairly novel technique where environments are recreated based on imagery captured by camera. This can be achieved by the use of for example a LIDAR scanner. The main drawback is the sheer amount of data that needs to be captured.

4. Data-Driven Synthesis, where algorithms are used to simulate a never-ending stream of context based on little input. This could i.e. be textures which are expanded based on pixel similarity or a never ending, randomized, video stream (Reinhard et al., 2013).

These four approaches all have their advantages and disadvantages, but more importantly the optimal approach may differ from task to task.

From the study by Elhelw et al. (2008), eye-tracking data suggested that certain areas are scanned more when evaluating stimuli. This cognitive behaviour is constantly happening every second when we interpret reality. These perceptual processes, over which we do not exert explicit conscious control, transform and aggregate data into what ultimately becomes a subjective vision of reality. This vision of reality is ultimately based on each individual's perception of previous experiences and expectations (Habermas, 1984), meaning reality cannot be standardized into global metrics. However, the subjective realism factor can be influenced by limiting the semantic complexity of a given stimuli, simply by limiting the visuals. This was for example done in the experiment done by Vangorp et al. (2007) where users only had to assess one abstract element in a comparison design. This concept of semantic richness can be operationally defined into two levels:

- I. A semantically rich (high) environment, where multiple objects defines a scene with coherence between the objects.
- II. A semantically poor (low) environment, where only one object makes up the scene and cannot be put into a coherent context.

This definition will be used throughout this project.

1.4 Cameras a Dimension of CGI

Whenever dealing with CG renders, one might find the result clinical or sterile. One issue is not only the semantics of the scene, but also the complexity. The real world is very complex with many small but very important details. This could be fingerprints on a glass surface or small cracks in a brick wall. This may often be neglected in CG renders and produces a "too good" looking environment (Reinhard et al., 2013).

When taking pictures with a DLSR camera, depending on the settings used, artefacts can often be produced such as noise or lens distortions. A study by Madsen, Borg, and Paprocki (2012) integrated small artefacts in an augmented reality context where a physical and CG object were in respect to a comparison context. They investigated the camera feed and found an average noise ratio of the video stream, which were merged with the CG elements. This small noise artefact was important for the CG object to be integrated in a perceptually viable way (Madsen et al., 2012).

The same inaccuracies are important in regards to still frame photography, where e.g. the camera's ISO speed directly affects the picture quality. ISO is an expression for a camera film's sensitivity to light, where low values of ISO are equivalent to less sensitivity to light. A modern camera's ISO values are often within the interval of 100-6400 units. Changing the light sensitivity has a certain effect on the captured image,

especially noise is an apparent artefact, see figure 2. The amount of noise which is being produced varies from camera to camera, but a rule of thumb is to set ISO as low as possible.



Figure 2: Comparison between ISO levels on a 100% crop of a DSLR image. Left: ISO 200, right: ISO 1600; images from (Tybjerg, Tybjerg, & Tybjerg, 2011).

All cameras have a lens of some sort which focuses the incoming light into the camera's image sensor. When light propagates through different media such as water or glass, its speed changes and the resultant wavelength is based on the medium's refractive index. This effect is apparent in many camera lenses where multiple glass lenses focus the image and typically results in colour fringing around the edges of objects – this effect is called chromatic aberration. Chromatic aberrations can be divided into two categories: longitudinal and transverse chromatic aberrations. Longitudinal aberrations are most apparent when a lens is not able to focus all colours in the image sensor plane (Walree, 2013). This effect happens due to light refracting near the edges of a focusing lens, see figure 3.

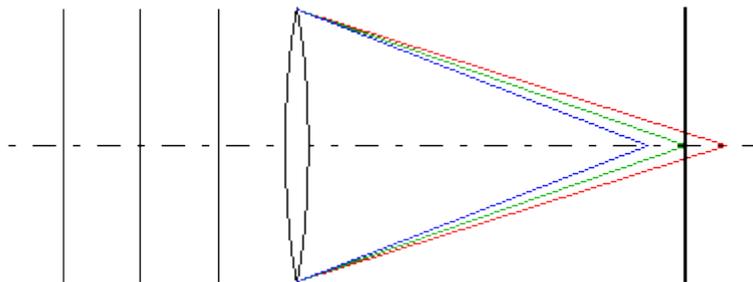


Figure 3: Longitudinal chromatic aberration. The focal points of the various colours do not coincide and only the green information is in focus (Walree, 2013).

Obliquely incident light leads to the transverse chromatic aberration, where a displaced colour foci is apparent as different colour fringing along the edges of objects. As opposed to longitudinal aberrations here, the colours coincide in the same plane, however they are being "shifted" due to the lens' physical refraction, see figure 4. In effect the size of the resulting image is different for each colour channel which is being shown as a distortion (Walree, 2013).

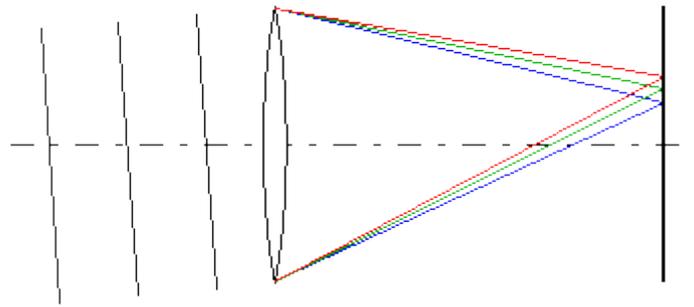


Figure 4: Origin of transverse chromatic aberration. The size of the image varies from one colour to the next (Walree, 2013).

One cannot avoid chromatic aberrations without using precisely calibrated equipment, although one can correct such artefacts using post-production software. To minimise the artefact, it is recommended to avoid shooting an object with a high contrast to its background. Furthermore, shooting with a fast camera aperture ($< f/4.0$) will due to the large lens exposure generate more aberrations (Walree, 2013).

Camera lens distortion is also an important factor to keep in mind when simulating a photographic reference. Due to the optical properties of the glass lenses, distortions of the picture can occur at different zoom levels, depending on the quality of the lens and the zoom level of the specific picture. For example, a typical Canon 18-55mm $f/3.5-5.6$ "kit lens" produces a so-called barrel distortion at wide-angle levels (Walree, 2013), see figure 5 for a graphical illustration of the distortion. Fortunately this distortion is easily neutralized by means of post-processing software such as Adobe Photoshop. This will help to match any 3D geometry to a background plate image (i.e. the photograph) ensuring exact geometric proportions.

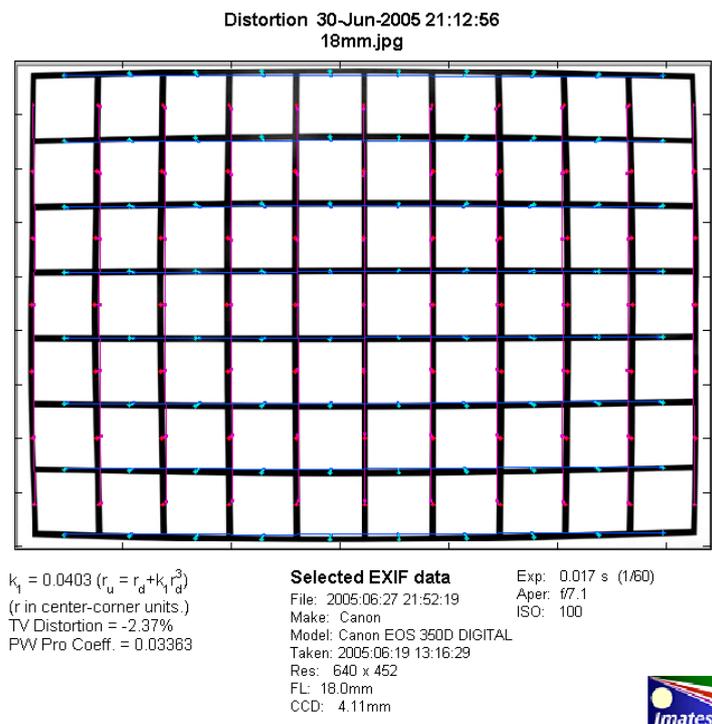


Figure 5: An example of barrel distortion created by a Canon 18mm-55mm lens at 18 mm.

As discussed when matching a real world reference with a simulated model, the camera produces significant artefacts that cannot be ignored when recreating the same scene in a 3D environment. Based on this analysis, the camera will be added to the three essential components of CGI and must be taken into consideration whenever dealing with recreation of photographic references, either by closely simulating the artefacts or by removing them.

1.5 Summary

During the analysis of previous work, it was found that investigation within the field of perception and how we perceive realism is an extensive subject. It was found that indirect lighting is very easily made inaccurate but still perceptually realistic (Reinhard et al., 2013; Yu et al., 2009). Reflections, especially specular highlights, seem to have huge impact on the perceived realism of a semantic rich scene (Elhelw et al., 2008); however, when assessing accuracy of the glossy reflection, humans tend to be imprecise (Olkkonen & Brainard, 2010). From a study of the components of CGI, it was found that the HVS is far from precise and can easily be tricked by either degradation of geometry (Reinhard et al., 2013) or by occluding geometry (Belhumeur et al., 1999). This means that a carefully picked scene may actually be full of geometric errors with little impact on the perceived end-result.

A discussion about how much is in the scene is important as well, and it is clear from analysis of Drago and Myszkowski (2001)'s experiment that the context of an assessed scene has an impact on the viewer. First of all, the viewer will have access to more stimuli to base their decision on, and this amount of stimuli may create a bias. To avoid this bias, one may divide the measurement into operationally defined categories: a **low semantic environment**, where only a sample is visible in the scene and a **high semantic environment**, where the sample is incorporated into a larger scene with auxiliary CG objects. Common for both of the environments is to appear as real as possible, i.e. using photometric validated environments and not artistically approximated.

Whenever holding a CG render as contrast to a photograph, it may appear "clinical" and "pure" and this may be due to small artefacts produced by the physical camera, which in contrast to a CG camera, is inaccurate when it comes to ISO, chromatic aberrations, distortions, etc. Thus it is proposed to include the camera as a fourth component of CGI. To ensure utmost similarity, it is necessary to mimic the camera as closely as possible by either simulating artefacts or removing them.

The steps proposed by Ulbricht et al. (2006) for recreating photorealistic renders were found to have a costly first step, where materials would have to be measured using precisely calibrated and expensive equipment. It is believed that a procedure can be made to recreate a surface as realistically as the expensive method using a gonioreflectometre.

With all of these factors influencing the realism perception, one can discuss the definition of "photorealism" and given the context, not all cases are viable to simulate photorealistic physical accurate results, but rather an image which is perceptually plausible. By perceptually plausible, it is meant as an image which **could**

have been real compared to a reference image. "Photorealism" will henceforth be defined as: physical inaccurate render which is perceived as a stimulus that could have been real. This is an operational definition and only valid for this particular project.

1.6 Hypothesis

Given the previously discussed limitations of the HVS, it is believed that human's may be willing to accept changes in a surface's material appearance and given the context of the surface this "threshold" may even vary, thus the main hypothesis for this project was formulated as:

Humans will be less likely to notice Bidirectional Reflectance Distribution Function changes of surfaces in a photorealistic high semantic environment compared to a low semantic environment.

In the following chapters I will analyse the components of a surface's BRDF and identify which of the components to vary. I will recreate a real surface in 3D as closely as possible and vary the afore-mentioned components.

The semantically rich environment will be recreated of an existing environment and will be as photorealistic as possible given the definition used in this project.

2. Analysis

In this chapter, analysis of how light interacts with surfaces and in which terms a surface can be described. Furthermore an analysis on how to recreate reality seen through a camera. Other key components on how photometry and CG software interacts and how to correctly show lighting in CG renders will be discussed.

2.1 Physical Properties of Materials and Light Interaction

Every surface has a BRDF that can be described using the components of a BRDF. The (basic) components consist of a specular, directional diffuse (glossiness) and uniform diffuse component. A visualisation of a sample BRDF can be seen in figure 6.

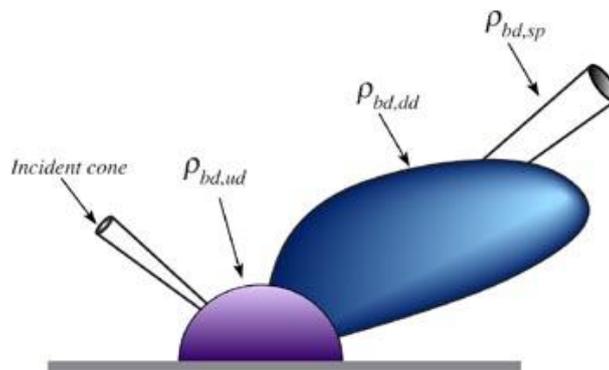


Figure 6: Components of a BRDF, where the white cone represents the specular component, the blue area is the glossy component and the purple area is the diffuse component, image from (CornellUniversity, 2011).

The components each have their own properties, such as the specular highlight, the glossy diffusion, etc. This model can be generalised across a lot of surfaces, however materials such as the human skin or a candle have yet more components, which is subsurface scattering of the incoming light and refraction. This general model assumes an isotropic, rough surface.

2.1.1 Specular Component:

The specular component is responsible for mirror-like reflections and can be described by the function:

$$dI_{r,sp} = |F|^2 \cdot e^{-g} \cdot S \cdot I_i$$

Where $|F|^2$ is the Fresnel reflectivity, which is a combined function of the material's refractive index and incident angle of the incoming light, g is a function of the surface roughness and S is a geometrical shadowing function (He et al., 1992). An example of such a function could be a perfect mirror where the incident reflection would be 100% preserved in the outgoing reflective angle. An aspect worth noticing is the Fresnel reflectivity which is often neglected in common materials in CG simulation software. This reflectivity function is, for example, absent in the standard material models found in Autodesk's Maya and actively has to be turned on using mental ray's MIA (mental ray architectural) materials.

2.1.2 Directional Diffuse Component

The directional diffuse component is the perception of reflected light that is spread once it hits the surface. It is responsible for highlights of light sources and the glossiness of a surface. It arises due to very small inaccuracies on every surface – this little roughness scatters the incoming light and creates the perception of a imperfect reflective surface (He et al., 1992). It should be noted that, from a perceptual standpoint, this is a property where accuracy (to a certain degree) is not of utmost importance (Olkkonen & Brainard, 2010).

2.1.3 Uniform Diffuse Component

The uniform diffuse component represents a non-directional portion of the incoming light which upon interaction with the material is scattered in all directions. This corresponds to a Lambertian reflection (Lambert material) and could arise from e.g. roughness on a surface. For many non-metallic materials, this is an important component and needs to be properly investigated before implemented in a 3D package. For metals with little or no roughness, the Lambertian reflection may be neglected (He et al., 1992).

2.2 BRDF's Models in Computer Graphics

Whilst creating CGI, there is almost always a need to simulate a material which is a composite of the aforementioned BRDF components. Chaos Group's V-Ray render engine uses an optimized shader which can use different implementations of the BRDF. Such implementations are the Phong, Blinn and Ward models. There are advantages and disadvantages to all of them.

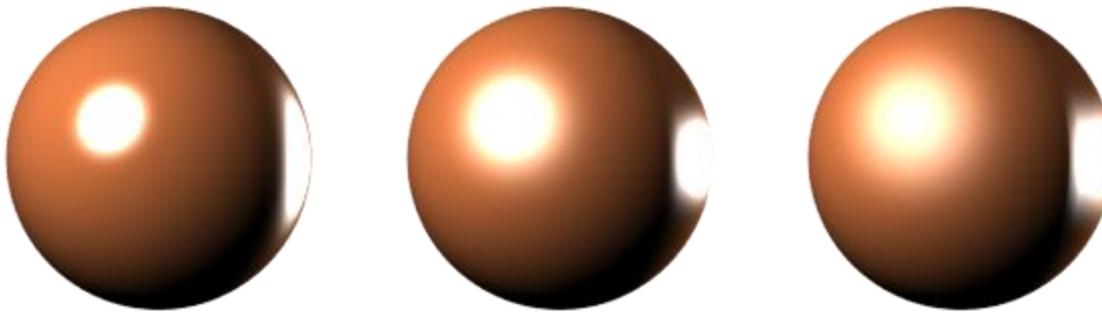


Figure 7: V-Ray's implementation of the Phong (left) Blinn (middle) and Ward (right) BRDF models (VisualDynamicsLLC, 2014).

The Phong model was formulated by Bui Tuong Phong in 1973 and uses three components; ambient, diffuse and specular. The ambient component is a simple combination of the light source colour and surface material colour. This was done to ensure any simulated material only reflects a certain wavelength, e.g. a blue surface does not reflect red light. This is also used in the diffuse component where the diffuse lighting is a cosine-weighted function of the incoming light vector and surface normal (Phong, 1975). The last component is the specular reflection which is the most complicated of the three. This component is dependent on the angle between the reflected light vector (R) and the view vector (V). R vector is computed as follows:

$$R = 2(N \cdot L)N - L$$

Where N is the surface normal and L is the vector pointing to the light source. This reflection vector is the vector expressing how the light would be reflected if the surface was a perfect mirror. The angle between the reflection vector and surface normal is equal to the angle between the light vector and surface normal.

Having computed the reflection normal, one may calculate the specular component using the view vector:

$$S = S_l \cdot S_m \cdot (R \cdot V)^\alpha$$

Where $S_l \cdot S_m$ is an expression for the light colour and surface colour and α is a specular exponent representing the directional diffuse component from the general model – the higher alpha is, the tighter the highlight will appear on the surface.

The Phong model is in general computationally expensive because of the expensive dot product calculations. This makes the Phong model less desirable in interactive applications (Van Verth & Bishop, 2008), but means little for offline rendering. The computations were drastically reduced by Jim Blinn who formulated an alternative way to calculate the reflection vector using a “half-angle” vector. This reduces the computations quite dramatically and is even found to be more precise when it comes to recreating BRDFs of samples (Ngan, Durand, & Matusik, 2005).

The Ward model was introduced in 1992 by Gregory J. Ward using an empirical model to fit surface reflectance data. It has become widely used in the computer graphics community (Walter, 2005) and in its raw form offers more control than the Phong and Blinn models. The Ward model supports anisotropic surfaces and is efficiently sampled using raytracing (Monte Carlo). Also this model simulates theoretical BRDF surface data well (Walter, 2005).

Ngan et al. (2005) analyzed a range of BRDF models and found that none of the afore-mentioned models were the best at representing reality. Other models such as Cook-Torrance, Ashikhmin-Shirley and the He models produced more realistic fits when it comes to a single specular lobe reflection. In general, anisotropic materials are very difficult to recreate using any of the afore-mentioned BRDF models.

V-Ray only supports the Phong, Blinn and Ward models, see figure 7, which affects the specular highlight quite dramatically. However, according to the lead developer at Chaos Group, more support is coming for the next generation of V-Ray (Koylazov, 2014). This includes support for the GGX BRDF¹ created by Sergey Shlyayev which has shown promising results.

¹ <http://www.shlyayev.com/rnd>

2.3 Photometry and CG

Within the field of light calculations and photometry using the correct terms is a vital part of understanding how light behaves in simulation software. In photometry, there are vital concepts to keep in mind and the most important are:

Luminous flux: is the core concept of light, which describes how much energy is being spread in all directions from the source. It is measured in candela per ster-radian which is the definition of lumen (Schubert, 2006).

Illuminance: is a term for the total amount of luminous flux incident on a surface per given unit area; perceptually, this is how bright a light appears on a given surface. Illuminance is measured in lumens per square metre (lm/m^2) which is the definition of lux (Schubert, 2006).

Luminance: is the luminous intensity emitted by a surface area of 1 cm^2 (or 1 m^2) of the light source. The unit of luminance is cd/m^2 or cd/cm^2 (Schubert, 2006).

Many software programs exist to simulate photometry in a correct and physical accurate way, for example, programs such as Relux², Velux Daylight Visualizer³ or Dialux⁴, uses backward raytracing or radiosity algorithms to achieve physically correct results (Iversen et al., 2013). However, these programs lack the possibility to effectively generate advanced materials and models. Other visualization software includes 3DS Max, Maya and Cinema4D and it is in general the individual artist who chooses his/hers preferred software. When it comes to rendering engines, some of the most popular amongst architects are V-Ray, Maxwell, Mental Ray (Villa, Parent, & Labayrade, 2010). Each of these rendering engines were tested on the perceived photorealism in a recent study by Villa et al. (2010). More specifically, they found that V-Ray produced superior images when it comes to subjective perception of similarity of rendering and photograph. Also the light atmosphere was rated higher than the other engines.

For this project, the main software used is Autodesk 3DS max with V-Ray render engine. Using photometric lights, it is possible to create lights which adhere to the inverse square law, which is vital in order to get photorealistic results. Using V-Ray specialized lights, one may choose from units of intensity in lumens which is luminous flux or cd/m^2 which is luminance. Furthermore, one may specify the units in Watts and Watts/m^2 , which is useful for simulating light bulbs.

² <http://www.relux.biz/>

³ <http://viz.velux.com/>

⁴ <http://www.dial.de/>

2.4 Linear Workflow

Whenever working with CG and especially when working with simulation of reality, it is important to have a linear workflow. A linear workflow is the process of working ensuring there will be no extra gamma correction happening when a scene is being rendered. The concept of linear workflow can be seen in figure 8.

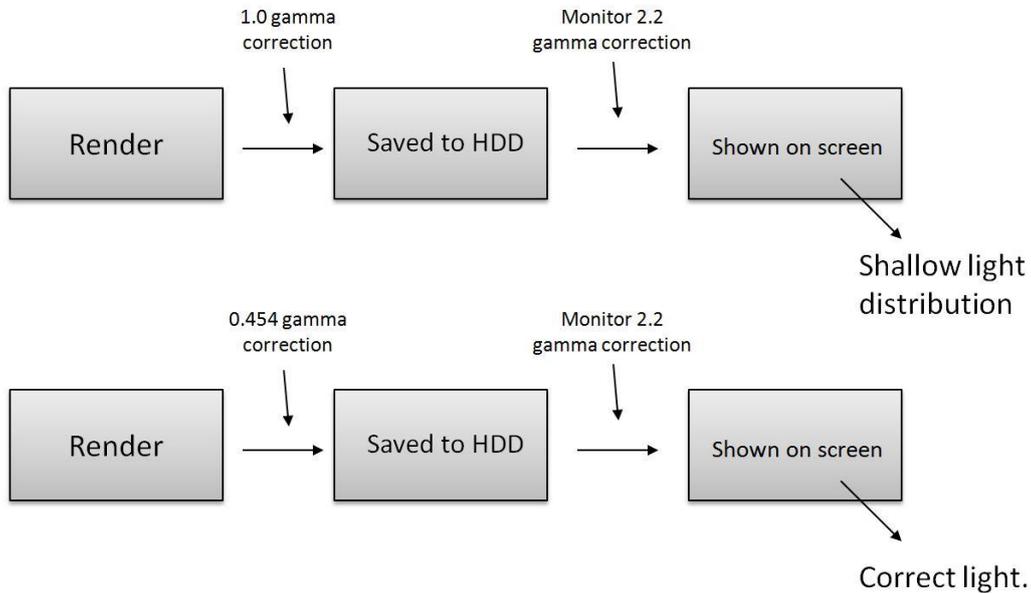


Figure 8: A chart overview of how proper linear workflow helps to achieve correct lit renderings.

To avoid a double gamma correction, one must set up the specific scene using the correct units (centimetres, metres, etc.) and make sure that colour swatches are properly handled by the render engine, either by manually using a gamma correction node or forcing gamma correction globally. This is also true for any texture used, where instead of treating the file with a normal gamma of 1.0 it is treated as an sRGB image that needs gamma correction to be displayed correctly, essentially telling the software to interpret the texture file as an under-gamma corrected image (normally 0.454). The renderer will then correctly use the gamma curve in the render process and the result is a nice even light distribution.

3. Procedure

Often when specialists are required to recreate reality in a photorealistic way, they would need precise data of the models they are supposed to build. For example, if a light source is determined to be photometrically validated in a 3D render, one would need a precise light distribution profile which is often described in an IES file.

The procedure described in this chapter is aimed at specialists who need to recreate certain elements in 3D using on-site measurements. This could be an office building, where the specialist would bring the measurement data and perform data gathering of a specific environment.

This procedure can be related to the first two steps in Ulbricht et al. (2006)'s three steps for verification of photo-realistic renders, where the reflection model is being compared to actual physical data and validation of the light transportation. I have divided the process into several sub-steps:

- Reflectance measurement and on-site photographic sample of material.
- Validate the light transportation in V-Ray.
- Validate IES profile of the given light source.
- Recreate a surface's BRDF in 3D for visualization purposes.

The aim of the process is to be confident with the parameters using V-Ray, recreating the surface as realistically as possible and to be able to confidently describe a surface in terms of BRDF.

3.1 Measuring an Approximate Reflectance of a Given Surface

The method takes a simplistic approach to recreating the reflectance instead of using complex gonioreflectometre readings. The first step is to actually create the reference for which the simulated surface should be recreated. This involves a physical surface with a given reflectance (sample), a small diffuse area light source where the illuminance is known and a physical camera that records the data in a 32bit (HDR, TIFF etc.) format. The sample will be taken in a dark room where the light source will illuminate the sample at a 45 degree incident angle, and the camera will record the resulting reflected surface on an opposite 45 degree outgoing angle, see figure 9 for a diagram of the process.

Secondly, one needs to calibrate the light source inside the 3D software. This needs to be done to ensure the artificial light source in the 3D software corresponds with the real one. This is done by the use of a luminance meter which will record the amount of light emitted from the light source. Using V-Ray lights one may input the luminance parameter such that it corresponds to the physical light source.

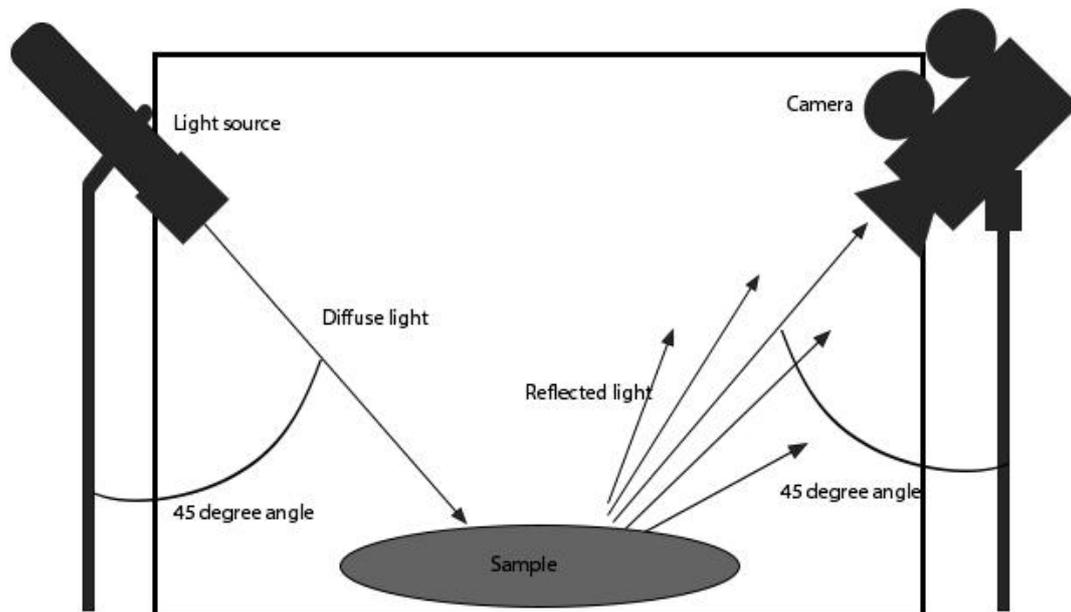


Figure 9: Setup for gathering a given sample's surface appearance.

The last step in the process is to photograph the actual sample. With the known light intensity and camera data, we can recreate the surface reflectance in the 3D software. This is done by tuning the glossiness and reflection until the real and simulated samples look alike (authors subjective assessment).

3.2 Validation of Light Transportation

To validate the light transportation of the V-Ray one may use an integrating sphere approach, where a light source is beamed into a sample placed inside a sphere with a nearly perfectly diffuse coat. The beam of light can be said to be parallel and will not hit the sphere directly but only be lit by the indirect illumination bouncing of the sample surface. If one were to measure the illuminance in the darkest area (i.e. just next to the parallel light source entrance), it is possible to determine if the final renderings need any type of correction. If, for example, the readings showed a deviation in lux as the glossiness was varied, one would have to correct this in order to get correct light distribution. The method for measuring illuminance in V-Ray was to use an illuminance pass which renders an image with lux values for a given pixel. To get correct results one has to use an un-biased rendering algorithm, which in V-Ray is called "brute-force". The renders were sampled with a brute-force subdivision at 512 sample and 5 bounces. Even though the resolution was very low (128x128) it still took around 2 hours per render, due to the extremely precise algorithm. Some noise was apparent in the renders. To remain consistent between sampling of the measurements the renders were scanned for lowest pixel value and which then was recorded.

Glossiness \ Reflectivity	0	0.2	0.5	0.7	1
0	0 lux				
0.25	14.3 lux	14.1 lux	13.6 lux	13.5 lux	13.5 lux
0.5	28.9 lux	27.9 lux	27 lux	26.7 lux	27.4 lux
0.75	43.5 lux	41.7 lux	40.2 lux	40 lux	40.9 lux
1	57.2 lux	55.5 lux	53.3 lux	52.9 lux	53.1 lux

Table 1: Measurements of the illuminance levels of the sphere, the values for reflectivity and glossiness have been normalized.

From the values we can see a small decline in lux values as the glossiness decreases on the sample, yet spikes a very little as the glossiness parameter approaches one. It was decided that due little variation no correction was necessary.

3.2.1 Validation of IES Files

In 1986, the Illuminating Engineering Society of North America (IESNA) published one of the first industry standards for the electronic dissemination of photometric information for architectural lighting fixtures (also



Figure 10: A sample IES file.

known as "luminaries") and other light sources (Ashdown, 1998). The standard was created in order to accurately describe photometric properties of luminaries and ensured a standard for describing light distribution of light sources with a possibility for simulation software to read such files and accurately recreate the light distribution, see figure 10. One of the major drawbacks of IES file format is that the light distribution is being described as far-field photometry, essentially resulting in a point source light where luminaries often have a volume (Labayrade, 2010). This can fortunately be adjusted by V-Ray by applying a physical shape to the IES file which distributes the luminance in accordance with the shape and area of the particular shape. The resulting new shape of the IES light creates significantly softer shadows which is expected of any luminaire with a volume, see figure 11.

The file is structured by a fixed set of horizontal and vertical angles in a web-like structure. The intensity values of the angles are described in sets as candela values, per angle. This means the first set of candela values is an expression for all vertical angles at first horizontal angle, then the next set of candela values is an expression for all vertical angles at second horizontal angle and so forth until the max horizontal value has been reached. Along the luminous intensity at a given set of angles, information about the manufacturer, intensity (measured in cd/m^2), units (metres or feet) is included in the file (Ashdown, 1998). An example of an IES file can be seen in Appendix 1, IES file example.



Figure 11: A comparison of the far field photometry problem in a standard IES file (right) and the approximate grid divided area light shape used to counter this (left).

To use IES profiles in a 3D application and to use them to recreate a BRDF in a photometrically correct way, one needs to validate the method of visualizing the IES profile. In order to validate the method, one may use an IES file, and using measurement, recreate the same IES profile and see if any major deviations occur. The method is carried out in 3DS max using V-Ray as the primary render engine.

To recreate an IES file, a very wide angle of view is required (the specific IES file range from 0 to 90 degrees in light distribution.) A camera was set up in the origin of an empty scene with a target pointing directly upwards. In the same place, the light emitter was set to emit the IES light in the same direction. To capture the light in a homogenous manner, a hemisphere with radius of 1 metre was created to cover the entire hemisphere of illuminance. To avoid any perspective distortion, the camera was set as an orthogonal camera and distance was matched with the hemisphere radius. When rendered, this gives a 2D projection of the lit hemisphere, see figure 12. Using V-Ray to render the scene with an illuminance pass returns the pixel values as a percentage of max illuminance of a given pixel (measured in lux). For example, if the max

illuminance threshold was set as 10000 lux then a pixel at (x,y) which returns a value of 0.876 means the candela value of the specific pixel is 8760 cd.



Figure 12: initial IES file beamed on a hemisphere. Seen directly from below.

This information is useful as one can use the x,y coordinates to recreate the angles from the origin at which the x,y coordinates lie. This can be used to recreate the IES format where the angles are expressed as candela values per angle; essentially we are doing transformation from (x,y,i) to (φ,θ,i) where "i" is intensity, φ is the polar angle and θ is the azimuthal angle in the x-y plane. This set of angles is used to express the vector of a point on the sphere.

The transformation needed to go from (x,y,i) to (φ,θ,i) is straight forward as the method uses a hemisphere with an orthographic perspective and a constant radius (Edwards & Penney, 2008). Essentially the transformation requires the angles of the IES file which one may calculate the corresponding coordinate set by the use of spherical coordinates, where in terms of Cartesian coordinates:

$$x = r \cos(\theta) \cdot \sin(\varphi)$$

$$y = r \sin(\theta) \cdot \sin(\varphi)$$

Where r is radius ($r \in [0, \infty]$), $\theta \in [0, 2\pi]$ and $\varphi \in [0, \pi]$.

Given the spherical angles, we can now calculate the x,y coordinates corresponding to this specific set of angles.

An implementation of this process has been done in C++ using OpenCV⁵ libraries, where the function above can be implemented by :

```
float intensityCalc(Mat& img, float angleT, float angleP, int width, int height, int imgDepth){
    //calculate intensity of a given pixel per theta/phi
    double x, y;
    y = sin(angleP)*sin(angleT);
    x = cos(angleP)*sin(angleT);

    // check for image bit depth. Return different typecasted lookup functions, uchar for 8
    bit, ushort for 16 bit and float for 32 bit.

    if (imgDepth == 0 || imgDepth == 1){
        if (x < 0) return (img.at<uchar>(x*width + width, y*height + height - 1)); //special
        exception where x would be minus one at angles (90,180). Return pixel next to it.
        if (y < 0) return (img.at<uchar>(x*width + width - 1, y*height + height)); //special
        exception where y would be minus one at angles (270,90). Return pixel next to it.
        // return the pixel value at x,y times the radius then scaled to fit the dimensions
        of the image and minused one due to 0 index of the Mat construct.
        return (img.at<uchar>(x*width+width-1, y*height+height-1));
    }
    else if (imgDepth == 2 || imgDepth == 3){
        if (x < 0) return (img.at<unsigned short>(x*width + width, y*height + height - 1));
        if (y < 0) return (img.at< unsigned short>(x*width + width - 1, y*height + height));
        return (img.at< unsigned short>(x*width+width-1, y*height+height-1));
    }
    else if (imgDepth == 4 || imgDepth == 5){
        if (x < 0) return (img.at<float>(x*width + width, y*height + height - 1));
        if (y < 0) return (img.at<float>(x*width + width - 1, y*height + height));
        return (img.at<float>(x*width+width-1, y*height+height-1));
    }
}
```

The function takes six parameters: the image which is being processed, theta and phi in radians, the height and width for the image and the depth of the image ranging from 1-5. The function returns the pixel value at the given set of angles which is being parsed at run time.

The coordinate system used in OpenCV starts in the top left corner where x positive is in the right direction and y positive is down, see figure 13. This means the (x,y) coordinate system used in the calculation function has to take into account a translation of the pixels. For example, for an input image of 1024x1024, the centre of 0,0 would actually be 512,512. This is taken care of by adding the radius of the circle in pixel values to pixel lookup index. The radius of the circle is for a square image equal to half of the width/height of the image and in the example the radius would be 512, which is added in the lookup call in the `intensityCalc` function ensuring the coordinates are correctly evaluated.

⁵ <http://opencv.org/>

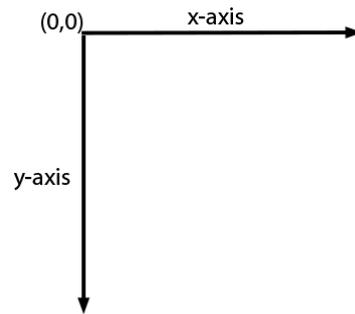


Figure 13: OpenCV's coordinate system, starts from the top left corner of the image and have a downwards positive y axis and a right hand positive x-axis.

Precision in the calculation is quite important. 32-bit images can hold pixel values above the maximum value which is viewable on a low dynamic range display. To preserve the data as much as possible it was decided to implement bit depth based calculation to get as dynamic images as possible albeit making the program slightly slower. The input image can contain pixel values in float, unsigned integers or unsigned chars format. This means a potential six-decimal resolution (for the float format).

This `intensityCalc` function is being used in a loop running through the entire image to calculate the intensity for every specified angle:

```
for (double i = 0; i <= numHangles; i += hAnglesStep) //begin loop through image using angles as
constraints
{
    for (double j = 0; j <= numVangles; j += vAnglesStep)
    {
        if (counter < 10){ //ensure we are not above 10 entries = 80 chars

            //intensity calc based on the angle given by i and j, multiplied with max lux to
            get a pseudo lumen.
            myfile << (intensityCalc(gray_image, sind(j), sind(i), halfWidth,
            halfHeight)/pixMult) * maxLum << " ";
            counter++;
        }
        else{
            myfile << endl << " "; //escape if limit has been reached
            myfile << (intensityCalc(gray_image, sind(j), sind(i), halfWidth,
            halfHeight)/pixMult) * maxLum << " ";
            counter = 0; //reset count for new line
        }
    }
    counter = 0; //reset count for new line
    myfile << endl;
}
}
```

In the loop, we run through the parsed image at steps given by the user. `numHangles` and `numVangles` denote the max theta and max phi angles, respectively, and `hAnglesStep`, `vAnglesStep` denotes the step size between each calculated sample. The smaller the step size, the more calculations will happen resulting in a finer resolution of the resulting data. In order to fit the IES standard, one must make sure that each line has

a maximum of 132 characters (Ashdown, 1998), and this is taken care of by the `counter` variable. Each time a calculation is done, it adds to this variable whilst checking if it is above 10 entries. If it is, then create a new line in the file.

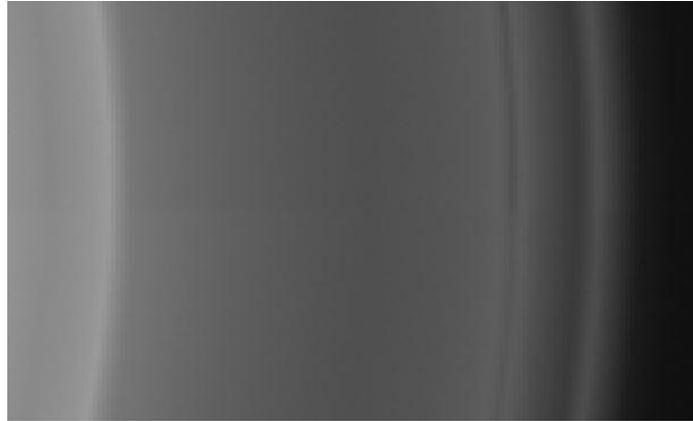


Figure 14: A zoomed comparison between the original IES (top part) and the recreated IES file (lower part). There is a very subtle difference in the overall illuminance levels.

The pixel value returned from the main calculation function is multiplied by a scalar, `pixMult`, depending on the bit depth of the loaded image (Which would be 256 for 8 bit, 65536 for a 16 bit and 1 for a 32-bit image).

An example IES file was processed through the program and the computed data was stripped from the file and compared to the theoretical data given by the same IES file, the average error was for this specific file was $\pm 0.9\%$. Given the difference is so small, it is assumed that humans will not detect any difference in a render where the recreated IES file is used in context. The difference, illuminance-wise, can be seen in figure 14.

This method validates the use of V-Ray and the use of an illuminance pass to measure the correct values of lux incoming on a perfectly diffuse surface, the code in its entirety can be seen in Appendix 2, IES Recreator Program.

3.2.2 Recreating the BRDF of a Given Sample Surface

With V-Ray validated for light transportation, we may now incorporate rebuilding an approximate BRDF of a surface. Again using the same method for recreating an IES file, we can now insert a material which is being refracted upon the hemisphere. This is done by using a parallel light source beaming into the sample. The resulting refractive reflection is then visible on the hemisphere which gives an approximate BRDF shape for the given sample.



Figure 15: The reflected three materials, top: diffuse, middle: hybrid, bottom: specular. Note the specular highlights appear the same, but in fact contains pixel which is out of range for normal displays.

Due to the nature of backwards ray-tracing computation used in V-Ray, one must use bi-directional path tracing to effectively receive any light information out of reach of the camera (ChaosGroup, 2010a). To effectively receive light which is reflected upon any material, one needs a propulsion medium which contains the energy information. One such phenomenon are caustic photons. Caustics are often used in the CGI world as light transportation through mediums such as water, glass or in general any refractive material. In this case, the system requires caustic photons to be shot from the light source and conserve the energy of the light until scattered to an insignificant level. This is exactly what happens on the surface of the material which is being measured. The caustic photons hit the surface and scatter the light on the hemisphere, where the resulting pixel value of a given x and y is an expression for the luminous flux emitted from the sample.

In order to use this method in V-Ray, we need an unbiased render method or a so called brute-force method (ChaosGroup, 2010a). As this is an exact method, the amount of samples required are quite intense. Another drawback is due to the nature of the brute-force method it is not possible to simulate caustics from a point light source reflected in a mirror (ChaosGroup, 2010a). A biased method such as photon mapping or irradiance cache will allow this process, but at a reduced accuracy (ChaosGroup, 2010b).

To test this, three samples were chosen to be recreated using the proposed method. The three "simple" materials were an almost reflective material, a perfectly diffuse material and a hybrid. The samples were processed as described earlier where a circular light source beamed at a 45-degree angle onto the sample which reflects light onto the same white hemisphere, resulting in a light pattern resembling the BRDF of that given sample, See figure 15.

These three samples of the materials were passed through the afore-mentioned program to recreate the “IES” of the given surface material. The resulting pixel values can be seen as a vector expression which can be used to recreate the shape of the materials BRDF. The “IES” file was parsed in the 3D program Blender where a script for recreating shapes of IES files has been made⁶ by Blenderartists.org users "Rickyblender" and "Simonced".

The three sample's shapes were recreated using this specific script, all using the same scale max luminosity scale specified in the IES recreation program.

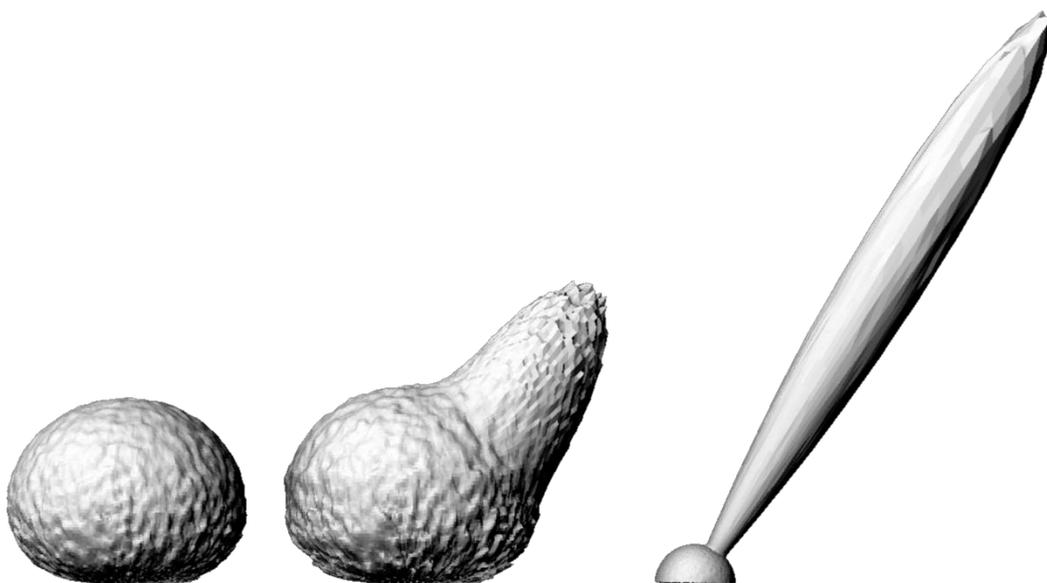


Figure 16: The three material samples' BRDF. Left: diffuse, middle: hybrid, right: specular.

In figure 16 the three sample's BRDF has been recreated and one can see the components of a BRDF. These components will be the varied factors in the final experiments.

3.3 Summary

In the previous chapters I discussed a method for capturing a given surface and how it may be recreated in a given 3D software program. Furthermore, V-Ray had its light transportation validated using brute-force algorithm, where it was found that no illumination correction was necessary. A program was designed and implemented which is able to reproduce IES files may also be used to recreate BRDF's of a given surface. This was done using three simple sample materials, a diffuse, a hybrid and a specular surface. These materials all contain each of the principal components of the basic BRDF model. It's been decided to use real surfaces with similar properties in two level of semantic context.

⁶ Source code is available from: <https://code.google.com/p/blenderiesreader/>

4. Implementation

In this chapter I will describe how the two environments were constructed and implemented. Three samples similar to the three test samples identified in the previous chapter was found and implemented using the procedure described in chapter 3. Procedure. In the end of the chapter I will discuss how these two environments can be tested in perceptual context.

4.1 Low semantic environment implementation

For a low semantic environment experiment, three material samples were chosen and processed through the procedure. The samples were chosen based on whether they would fit into a semantically rich environment and also due to the properties of the given material. The samples should fit the three main components of a given BRDF which are: a diffuse, a specular and a hybrid with glossy reflections. The diffuse sample was chosen to be an ordinary Post-it note which is commonly found in office environments. For the hybrid sample, it was chosen to be the grey area of a table, where an obvious glossy reflection is visible. The specular component chosen was a large whiteboard that resembles a mirror-like reflection and has a very well-defined highlight, see figure 17. Besides the three afore-mentioned samples, a mirror reference was also recorded. Each sample will be simulated in five editions. One sample is as close as possible to the real world stimulus sample (calibrated), the others will have a deviation from the calibrated in the primary BRDF component, i.e. the diffuse component for the diffuse sample, the glossy component for the hybrid component and the highlight for the specular material.

To ensure consistent stimuli, the physical samples were all treated for lens distortion and noise. Furthermore all the images were cropped to fit inside four times the diameter of the specular mirror reference. The images' positions were also slightly adjusted to ensure the sample had a straight orientation.



Figure 17: Recording images of the samples, this is the specular whiteboard sample being recorded, the ambient room lighting was turned off during sample recording.

The light source for this measurement was a simple torchlight. To ensure the light source was emitting evenly and as uniformly as possible, a couple of modifications had to be made. An extension was made out of a white plastic tube material extending the light by around 5 cm (diameter of the light source). The extender would cover the outer front of the light. A thin piece of white paper was placed in front of the original light opening and the extender was wedged on keeping the thin paper layer in place. At the end of the extension, another thin layer of paper was placed to make it even more diffuse. The extender and paper were wrapped tightly with black tape to ensure no light spillage or any side emittance. The resulting light is a much more diffuse and uniformly spread light that will be more consistent with what can be recreated in 3ds max.

The diffuse light source was measured at 5800 cd/m^2 using a luminance metre. To verify the measurement, one may hold this measurement against an illuminance metre reading, where the relationship between the two is $\text{luminance} \cdot \pi = \text{illuminance}$ (Labayrade, 2010). The illuminance metre reading was done by directly aiming the light source above the reader at a distance of around one centimetre. The metre reading was 16100 lux giving an error rating of 13%. The error may be due to the formula assuming a spherical light source, where in this measurement was a flat emitting surface, also the light may not have been perfectly diffuse.

The scene was recreated in Autodesk 3DS Max 2014, where the parameter for the light source were input as a V-Ray mesh light, where the mesh is a circle with the same radius ($r = 2,7$). The light source had an input of 5800 cd/m^2 .

Each stimulus was matched with the photographic reference as closely as possible. This was either done by a completely procedural solution, texture map or a combination of both. The diffuse material was a pure texture map of the existing Post-it. It was calibrated by altering the diffuse level (ranging from 0-255) in the

software. The calibrated surface was set at a certain level and then altered by deviating from this value by a percentage of the total range, e.g. if the calibrated surface was set at a diffuse level of 110, and then deviations by 10 and 20% were done in both ranges resulting in values 60, 85, 110, 135 and 160.

The hybrid model cannot be said to be contained within one parametre only, therefore it was decided to split this stimulus into two segments, the glossiness and reflectivity. The component is very dependent on both the specular and the directional diffuse BRDF component. Thus, equally to the diffuse component, the reflectivity was deviated from a calibrated sample by $\pm 10\%$ and $\pm 20\%$. Furthermore the glossiness was varied very little in terms of input values (ranging from 0 - 1), where the deviations were $\pm 2\%$ and $\pm 4\%$. The glossiness varied by this amount as the author found it to be suitable for a noticeable difference in the high deviations. This material was recreated using procedural diffuse input and a texture for the specular component (scratches, similar to the reference) and also a procedural noise for generating splotches of more reflective areas.

The specular stimulus recreation was completely procedural, and based on the specular highlight from the reference. A bump map was created to break the light in the same manner. This was achieved by a speckle map which was carefully matched in size and pattern. Looking at the reference of the whiteboard, it can be seen that the specular layer is like a coat of a white material. Meaning it both had diffuse properties and specular properties. The resulting changing variable was the highlight glossiness which was altered with $\pm 1\%$ and $\pm 2\%$ from the calibrated sample. The change resulted in a softer/tighter sheen which was very important for that stimulus.

All the stimuli were created to look like the reference and processed through the BRDF re-creator program described in section 3.2.1, an overview of the calibrated samples can be seen in figure 18.

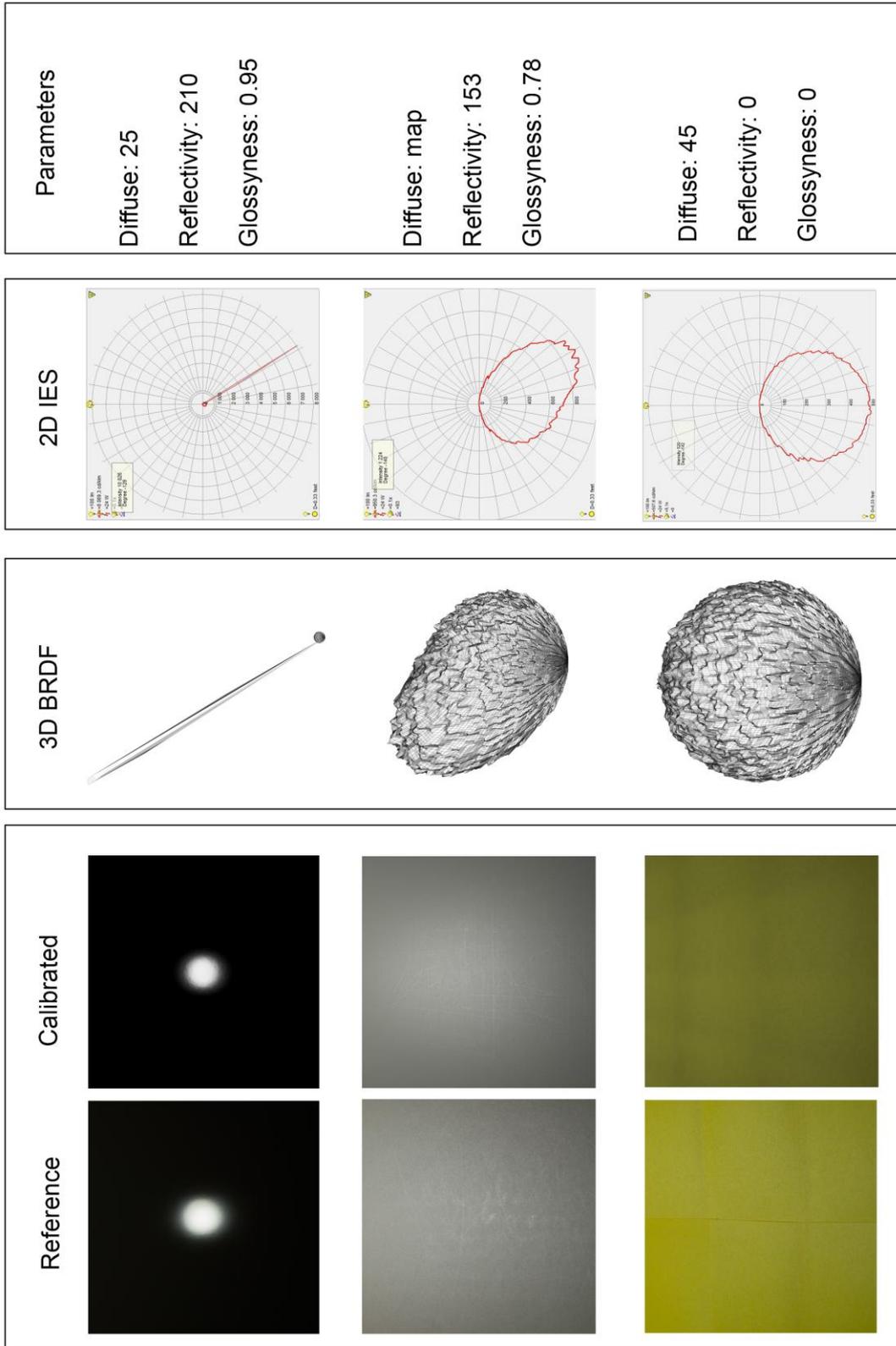


Figure 18: A composite showing the reference and the rendered calibrated sample, a 3D BRDF of the calibrated surface, 2D IES file of the surface and the parameter used in V-Ray.

4.2 High Semantic Environment Implementation

For a high semantic environment, it was decided to follow the first approach identified by Reinhard et al. (2013), stating that modelling everything is a viable option for recreating reality. The environment was a meeting room, with basic furniture and luminaries. The dimensions of the room were $L \times W \times H = 8.65\text{m} \times 4.65\text{m} \times 2.75\text{m}$, and the furniture found in the room was simple. There were multiple tables at the centreline of the room which were placed to form one big table. The table width was 1.4 m and the total length was 6.6 m (the big table consisted of six tables with a length of 0.7 m and two tables 1.2 m long). The table height was 0.72 m. The ceiling consisted of suspended white acoustic tiles with the size 60x60 cm. A total of 10 luminaries were fitted in the ceiling grid, positioned every third row and third column, only three luminaries were turned on when the picture was taken, see figure 19. The luminaries were recessed in the ceiling, each luminaries had three fluorescent light bulbs. The luminaries were made by Fagerhult⁷ and has publicly made an IES file available. Common furniture like office chairs, cupboards and radiators were also present in the room. For the purpose of demonstration, only one half of the meeting room was fully modelled. The reference picture was taken using a Canon 550D, 18mm lens at f/4.5, shutter speed 1/10 second, ISO 100, white balance 3200k.



Figure 19: The reference image taken at the meeting room.

⁷ <http://www.fagerhult.com/Products/Recesso/Recesso/#lightdistribution>

4.2.1 Reflectance and Light Measurements

Following the first step in Ulbricht et al. (2006) proposed method, to effectively achieve a realistic and validated result, one must gather some information about the scene environment. This includes measurements of the luminance in the room and how the illuminance is reflected on various surfaces.

The reflectance of a given surface is a value representing the percentage of light being reflected on the basis of the specific surface's BRDF. A surface with a "high reflectance" will reflect more light than a surface with a "low reflectance". Reflectance can intuitively be connected with the colour of a given surface meaning that brighter(/white) colours will be more reflective than darker(/black) colours. Other factors that influence the reflectance of a given surface could be if the surface is rough, has translucent properties or is refractive. To effectively recreate the room in 3D, the most important (i.e. biggest surfaces) had their reflectance measured. To measure the reflectance, one needs a surface given reflected light (luminance) and a reference sample with known reflectance. Once you have the measurements of a sample's luminance, the ratio can be calculated, which is described as (Roien, 2013):

$$\rho_{sample} = \frac{L_{sample} \cdot \rho_{reference}}{L_{reference}} \quad (1)$$

The reflective surfaces and reference were measured using a luminance metre, which is calibrated by the supplier. This instrument measures the reflected light in candela per square metre (cd/m^2). At every sample there was a sample on a surface with a known reflectance. This surface is a white diffuse surface with a 95.1% reflectance. An example could be a measurement of the floor. The measured luminance for a point sample and reference was $9,8 cd/m^2$ and $33,2 cd/m^2$, respectively. This gives the rho of the floor by using (1):

$$\rho_{floor} = \frac{9,8 cd/m^2 \cdot 0,951}{33,2 cd/m^2} = 0,28$$

A list of the most important items can be seen in table 2.

Reflectance of surfaces	Wall	Ceiling	Floor	Table	Cupboards	Whiteboard
Reference plate measurements (cd/m^2)	44,4	31,5	33,2	96,4	111	55,8
Sample measurements (cd/m^2)	42,1	27,6	9,8	57,7	59,4	53,1
Rho	90%	83%	28%	57%	51%	95%

Table 2: Reflectance measurements of the meeting room.

Some of the measurements presented in table 2 are from a study by Roien (2013), where the same room was used for similar purposes. These measurements can directly be used as a guideline for the reflectivity parametre used with V-Ray materials.

The office room was rebuilt using Autodesk 3ds max software with V-Ray 3.0 render engine using a linear workflow. A free target camera was created and using 3ds Max's camera perspective match tool. With this

tool it is possible to match perspective with a given reference image. When the perspectives matched, the modelling could begin. All the geometry were matched from reference photography and recreated as exactly as possible, however keeping in mind that background geometry can be less precise compared to the foreground geometry as discussed in section 1.2.1 Perception of Geometric Shapes. The calibrated render can be seen in figure 20.



Figure 20: The recreated 3D render of the meeting room scene.

Luminaries in the scene were recreated using V-Ray IES lights which read the provided IES file to recreate the light distribution. To ensure the lighting was the same as the reference, six illuminance measurements using a V-Ray light metre were conducted. These measurements were compared to the sampled measurements on site and compared. The comparison can be seen in table 3.

Illuminance measurements	Table 1	Table 2	Table 3	Whiteboard Vertical	Whiteboard	End wall
On-site measurements (lux)	275	293	287	133	230	250
V-Ray illuminance readings (lux)	327	275	216	206	294	318
Difference	+16%	-6%	-32%	+35%	+27%	+27

Table 3: Illuminance measurements done at the meeting room.

The differences between the physical readings and V-Ray illuminance readings were moderately different, especially the whiteboard and table 3 measurements, this may due to incorrectly reflectance input parameters

used on the large surfaces, such as the walls. The measurements are not ideal, but it was decided to continue albeit not having perfectly calibrated illuminance levels.



Figure 21: False colour map showing illuminances at different areas in the scene. Red color's denominated an area with more illuminance (~ 1000 lux) whereas dark blue areas are less illuminated (~ 100 lux).

The data can also be visualized in an illuminance map with false colour overlay, this is shown in figure 21. To simulate the physical camera as closely as possible, V-Ray exposure control was used on the shot camera. This enables the entry of data available from the shot and to recreate the exposure as closely as possible. The materials from the low semantic environment were transferred to this scene and implemented on the respective surfaces, i.e. the specular was applied to the shiny whiteboard area on the left, the diffuse material to the Post-it notes located on the table and the hybrid material was applied to the gray area of the tables. The choice of materials were based on the findings by Elhelw et al. (2008), where the focussing area of assessors were in areas with more contrast in areas with light reflections and specular highlights. The diffuse sample was a relatively small sample compared to the overall size of the room, but it was placed in an area such that it created strong contrast. The largest surface recreated was the table surface, which was, due to the camera angle, very sensitive to the light. The specular surface was placed relatively peripheral compared to the other stimulus, but the surface reflections had strong contrast with the reflected environment. Two changes were made to the stimuli in the transition between the low semantic environment and the rich semantic environment. Firstly the specular material got its Fresnel index of reflection changed to better match the photographic reference, which is a change from the first test where the index of reflection was not

relevant (as there were no environment to be reflected). Secondly, the hybrid diffuse colours were changed to fit the scene, in the sense that the diffuse colour of the table was a bit too greenish, compared to the photographic reference. The change was done as an aesthetic choice and has no impact on the independent variables which are being changed in the test (level of glossiness).

With this relatively richer semantic environment compared to the first test, it is impossible to split the materials into separate areas of the stimuli, meaning all of the three materials will be visible in the stimuli at all times. As a consequence of this, there will only be one independent variable which is being changed from stimulus to stimulus. For example, if the diffuse component has an increase of 20%, the specular and hybrid samples will remain at calibrated levels.

4.3 Experiment Design

The first experiment featured the material samples that was photographed and simulated through the procedure. The assessors only judged the visual stimuli of the material sample against a photographic reference. The assessors evaluated the material's appearance compared to the reference in a within subject design. The experiment was split into three parts, one part for the diffuse component, one part for the hybrid and one part for the specular component. The assessors were informed that visual aesthetics were not important in this context and should not be taken into account when assessing the sample.

In the second experiment assessors evaluated the same materials used in the first experiment, the difference lied in the semantic richness of the environment. Now the assessors had to pinpoint exactly where the difference was and assess based on this interpretation. Compared to the first test, the structure of the second experiment was a bit different; It was not split into any parts but presented as 17 different stimuli and the reference picture was not a photographic reference but the calibrated 3D reference (which also was hidden amongst the other stimuli). The choice of not having a photographic reference in this experiment was due to the desired outcome of the data. With a photographic reference there was a risk of getting data on how different the stimuli was from a photograph, rather than how differently the deviated samples were perceived. For each sample the user had to assess the image on a scale from 1-7 compared to the reference image, where 1 was "relatively very different" and 7 was "identical". It was important for the assessors to understand that this assessment had to be viewed as a relative comparison, meaning that even little change in the image might qualify as a "relatively very different" rating. To introduce the assessors to the scale and especially the relativity of the assessment scale a short familiarization test was run before the actual test. The familiarization contained three stimuli from each test (not the reference), the stimuli were typical extremes for the specific test. The assessors were not able to see the reference at this point, but were able to shuffle through the stimuli as many times as they wanted.



Figure 22: Experiment setup.

The experiments will show how effective the procedure is for recreating materials exclusively and if the calibration is perceptually close to the reference. The second experiment will feature one or more objects in a high semantic environment (office meeting room), where assessors will rate the same materials in the same manner as in the first test in a within subject design. A comparison between the two tests will be done to evaluate if the environment will influence the perception threshold of the given surfaces. The assessors viewed each image on a Dell U2713H 27" TFT LED screen which was calibrated to Adobe sRGB colour space at 300 cd/m^2 , see figure 22 for the experiment setup.

5. Results

5.1 Low Semantic Experiment

The experiments were conducted at Aalborg University, Copenhagen, using convenience sampling. Before each assessor performed the test, a randomized script was run on the stimuli to shuffle the order of appearance see Appendix 3, Data Gathering Code for an overview of the script. This was done for each of the three tests (diffuse, hybrid and specular). The users were asked to fill in a small questionnaire about themselves (N = 19, 10 males, 9 females), the mean age was 29 (SD = 4.52) years old. Every assessor had self-reported normal or corrected to normal vision. After the short questionnaire, the assessors were told to read a short introduction to the test and how to do the assessments, the introduction can be seen in Appendix 4, Instructions for the First Experiment. Before the assessors started the real test, a short familiarization was run in order to make the assessors aware of what kind of stimulus the test contained. After the familiarization, the assessors were shown the actual interface for the test where the reference was to the right side of the screen and the stimuli was on the left. There was intentionally put a border between the reference and the stimuli in case of the event where a comparison was very similar, such that assessors were not able to directly discern between the images. The null and alternative hypotheses can be formulated as:

$h_{0[\text{diffuse, hybrid, specular}]}$: *There is no perceptual difference between the [diffuse/specular/hybrid] samples in a low semantic context.*

$h_{a[\text{diffuse, hybrid, specular}]}$: *There is a perceptual difference between the [diffuse/specular/hybrid] samples in a low semantic context.*

The results for the diffuse parameter showed a tendency for how the diffuse parameter affects the perception of the material. The calibrated diffuse stimulus was rated highest of the reproduced stimulus (M = 4.053, SD = 1.393) and the lowest rated stimulus was the sample with 20% more diffuse (M = 1.842, SD = 1.214). The reference was the highest rated stimulus with an average mean of 6.368 (SD = 0.765).

All the rankings were processed using Friedman's rank test for k-correlated samples. It is assumed the stimulus is correlated within each subject, i.e. the diffuse parameter is correlated in terms of percentage deviations from the calibrated stimulus (Howell, 2008). The Friedman's process is analogous to a standard one-way repeated measure analysis of variance (ANOVA), but the main difference is the statistics are applied to rankings compared to raw data (Howell, 2008).

The analysis shows a significant difference between the hidden reference and all the stimuli but the calibrated one, which suggests the hidden reference was easy to spot and is different from the stimuli. The calibrated stimulus is significantly different from the stimuli with 10% ($p = 0.013$) and 20% ($p = 0.001$) added diffuseness.

All the rankings were pair-wise compared using Nemenyi's procedure (post-hoc), which gives an overview of the tendencies for the data by calculating the adjusted value of α by division of all the comparisons. Also

this process normalizes the ranks based on the amount of samples (Trawiński, Smętek, Telec, & Lasota, 2012), see table 4. The table shows where the differences become noticeable for the assessors. In the case of diffuse stimuli, it appears to be whenever there is added diffuseness.

Sample (Low semantic)	Sum of ranks	Mean of ranks (0-6)	Groups
Diffuse Plus 20	33.000	1.737	A
Diffuse Plus 10	41.500	2.184	A B
Diffuse Minus 20	66.500	3.500	B C
Diffuse Minus 10	68.000	3.579	B C
Diff Calibrated	79.500	4.184	C D
Diff Reference	110.500	5.816	D

Table 4: Groupings of the diffuse stimuli in a low semantic environment.

A notable difference is happening with the 20% increase of the diffuse parametre which only can be grouped together with the 10% added diffuse stimulus. With these results, one may reject the h_0 and accept the alternative hypothesis h_a : there is a perceptual difference between the diffuse samples in a low semantic context.

The hybrid test showed little difference between the chosen stimuli. The calibrated stimulus was again the highest scored recreated stimuli ($M = 4.789$, $SD = 1.273$), but the overall results show the difference between stimuli is very subtle and hard to detect. The only stimulus which could be grouped from the other was the reference which was significantly different from the other stimuli ($X_Q = 59.709$, $X_{0.05}(9) = 16.919$, $p = <0.0001$). Therefore one can reject the null hypothesis for the hybrid samples and accept the alternative hypothesis $h_{a[\text{hybrid}]}$.

The results' indications are twofold. First of all the reference and the presented stimuli are significantly different meaning the implementation of the material could be flawed and, secondly, the differences between the stimuli are negligible and may be deviated even more from the calibrated stimuli. Pair wise comparison and grouping of hybrid stimuli can be seen in table 5.

Sample (Low semantic)	Sum of ranks	Mean of ranks (0-10)	Groups
HybridGlossyPlus 4	69.500	3.658	A
HybridGlossyMinus 2	84.500	4.447	A
HybridReflectMinus 10	90.500	4.763	A
HybridReflectMinus 20	94.000	4.947	A
HybridGlossyMinus 4	101.000	5.316	A
HybridGlossyPlus 2	101.000	5.316	A
HybridReflectPlus 10	102.500	5.395	A
HybridReflectPlus 20	107.500	5.658	A
HybridCalibrated	110.000	5.789	A
HybridRef	184.500	9.711	B

Table 5: Groupings of the hybrid stimuli in a low semantic environment, numbers in the "Sample" column denotes deviation percent.

The specular stimuli, again, had a significant difference ($X_Q = 63.483$, $X_{0.05}(5) = 11.07$, $p = <0.0001$) amongst the groups. Here the hidden reference was again rated highest ($M = 6.632$, $SD = 0.684$) and is significantly different from the other stimuli. The calibrated stimulus had a below average mean score of 3.342 ($SD = 0.961$). The grouping of the data using the same method as in the diffuse and hybrid stimuli shows a significant difference between the maximum added glossy specular component and the subtracted glossiness pair. The grouping can be seen in table 6.

Sample (Low semantic)	Sum of ranks	Mean of ranks (0-6)	Groups
SpecPlus 2	29.000	1.526	A
SpecPlus 1	46.000	2.421	A B
SpecCalibrated	63.500	3.342	B
SpecMinus 1	73.500	3.868	B
SpecMinus 2	76.000	4.000	B
SpecRef	111.000	5.842	C

Table 6: Groupings of the specular stimuli in a low semantic environment numbers in the "Sample" column denotes deviation percent.

Again one may reject the null hypothesis and accept the alternate hypothesis, $h_{a[\text{specular}]}$.

To test for the level of agreement between the assessors, one may compute Kendall's coefficient of concordance (W). This coefficient is similar to normal correlation coefficients such as Pearson's r or Spearman's rho correlation coefficient, but with Kendall's W one may correlate k assessors.

Using the procedure for all assessors we get a W value of 0.5531, which is concordant with a significance level of 0.05 ($p = <0.001$, $df = 21$). As Kendall's W in itself is arbitrary value ranging from 0 to 1 it is recommended by (Howell (2008)) to converted into an average Spearman correlation, which is done by:

$$r_{s=} = \frac{kW - 1}{k - 1} \quad (2)$$

Where k is the total amount of assessors and W is Kendall's W . Given the data for the assessors this gives:

$$r_{s=} = \frac{kW - 1}{k - 1} = \frac{19 \cdot (0.5531) - 1}{19 - 1} = 0.528$$

This coefficient can be compared with the assessors in the high semantic environment test to see if the assessors generally agree with their ratings.

5.2 High Semantic Environment Test

Similar to the low semantic environment test, this test was performed using convenience sampling. The assessors were asked to fill in a small questionnaire about themselves ($N = 19$, 7 males, 12 females), the mean age was 33 ($SD = 11.3$) years old. Every assessor had self-reported normal or corrected to normal

vision. After the self-report, the users were instructed to read a test introduction to establish the procedure. The instruction can be read in Appendix 5, Instructions for the Second Experiment.

Before the assessors began the actual experiment, they went through a small-scale familiarization test similar to the one in the first test scenario.

After the familiarization, the assessors were shown the actual test interface, where the screen was split into two with the reference on one side and the assessment on the other. The users were instructed on how to use the interface and asked if they had any questions; if not, then the testing could start.

The null and alternative hypotheses can be formulated as:

$h_{0[\text{diffuse, hybrid, specular}]}$: *There is no perceptual difference between the [diffuse, hybrid, specular] samples in a high semantic context.*

$h_{a[\text{diffuse, hybrid, specular}]}$: *There is a perceptual difference between the [diffuse, hybrid, specular] samples in a high semantic context.*

The analysis of the diffuse set of samples using Friedman's test shows a significant difference between the samples ($X_Q = 23.847$, $X_{0.05}(4) = 9.488$, $p = < 0.0001$). This means one may have to reject the h_0 and accept the $h_{a[\text{diffuse}]}$. Using Nemenyi's procedure to group the rankings, we can see a grouping between the 20 % subtracted diffuseness and the other deviations and calibrated image. There is a significant difference between the diffuse with 20% subtracted diffuseness and the calibrated ($p = 0.033$), and the diffuse with 10% added diffuseness ($p = < 0.0001$). Furthermore there is a significant difference between the diffuse with 10% added diffuseness and the sample with 10% subtracted ($p = 0.021$). The groupings can be seen in table 7.

Sample (High semantic)	Sum of ranks	Mean of ranks (0-5)	Groups		
DiffuseMinus20	37.500	1.974	A		
DiffuseMinus10	45.000	2.368	A	B	
DiffusePlus20	62.500	3.289	A	B	C
Calibrated	65.500	3.447		B	C
DiffusePlus10	74.500	3.921			C

Table 7: Groupings of the diffuse stimuli in a high semantic environment.

The results from the hybrid samples show little variance in the assessment across the whole sample pool. The Friedman test shows no significant difference between the assessments ($X_Q = 15.520$, $X_{0.05}(8) = 15.507$, $p = 0.055$) and one may not reject the null hypothesis $h_{0[\text{hybrid}]}$ meaning there is no difference between the stimuli, see table 8.

Sample (High Semantic)	Sum of ranks	Mean of ranks (0-9)	Groups
HybridSpecularMinus20	67,000	3,526	A
HybridSpecularPlus20	76,500	4,026	A
HybridSpecularMinus10	92,500	4,868	A
HybridGlossyPlus2	97,500	5,132	A
HybridGlossyPlus4	100,500	5,289	A
Calibrated	101,000	5,316	A
HybridSpecularPlus10	101,500	5,342	A
HybridGlossyMinus2	106,500	5,605	A
HybridGlossyMinus4	112,000	5,895	A

Table 8: Groupings of the hybrid stimuli in a high semantic environment.

Similar for the specular stimuli, there was found little variance between the assessments. The Friedman test gives a p value of 0.295 ($X_Q = 4.915$, $X_{0.05}(4) = 9.488$). Two assessors had apparently missed two different assessments from the specular stimuli and subsequently had their assessment removed from the calculations, see table 9. Given the fact there is no statistical difference, one may not reject the null hypothesis $h_{0[\text{specular}]}$.

Sample (High Semantic)	Sum of ranks	Mean of ranks (0-5)	Groups
SpecularPlus1	44,500	2,618	A
SpecularPlus2	47,000	2,765	A
SpecularMinus1	51,000	3,000	A
SpecularMinus2	51,500	3,029	A
Calibrated	61,000	3,588	A

Table 9: Groupings of the specular stimuli in a high semantic environment.

Using formula (2) as in the low semantic test, one may calculate Spearman's Rho to see if the rating happened in concordance. The two assessors who missed one assessment had to have their ratings excluded from this test. Using the remaining assessors, we get a W value of 0.1966, which is concordant with a significance level of 0,05 ($p = <0.001$, $df = 16$), and calculated into a Spearman correlation coefficient, we get a value of 0.146.

Specular	Minus 1	Minus 2	Plus 1	Plus 2
R^2	0.018	0.000	0.164	0.390
F	0.669	0.000	7.043	23.040
P	0.419	1.000	0.012	< 0.0001

Table 10: Comparison between the two semantic environments for the specular stimuli. R is how they correlate. F and P values for the comparison and significant differences.

Diffuse	Minus 10	Minus 20	Plus 10	Plus 20
R^2	0.166	0.088	0.834	0.677
F	7.140	3.454	180.500	75.615
P	0.011	0.071	<0.0001	<0.0001

Table 11: Comparison between the two semantic environments for the diffuse stimuli. R is how they correlate. F and P values for the comparison and significant differences.

Hybrid	Glossy Minus 2	Glossy Minus 4	Glossy Plus 2	Glossy Plus 4	Reflect Minus 10	Reflect Minus 20	Reflect Plus 10	Reflect Plus 20
R²	0.371	0.363	0.184	0.286	0.249	0.034	0.141	0.035
F	21.246	20.547	8.111	14.430	11.951	1.269	5.906	1.304
Pr > F	< 0.0001	< 0.0001	0.007	0.001	0.001	0.267	0.020	0.261

Table 12: Comparison between the two semantic environments for the hybrid stimuli. R is how they correlate. F and P values for the comparison and significant differences.

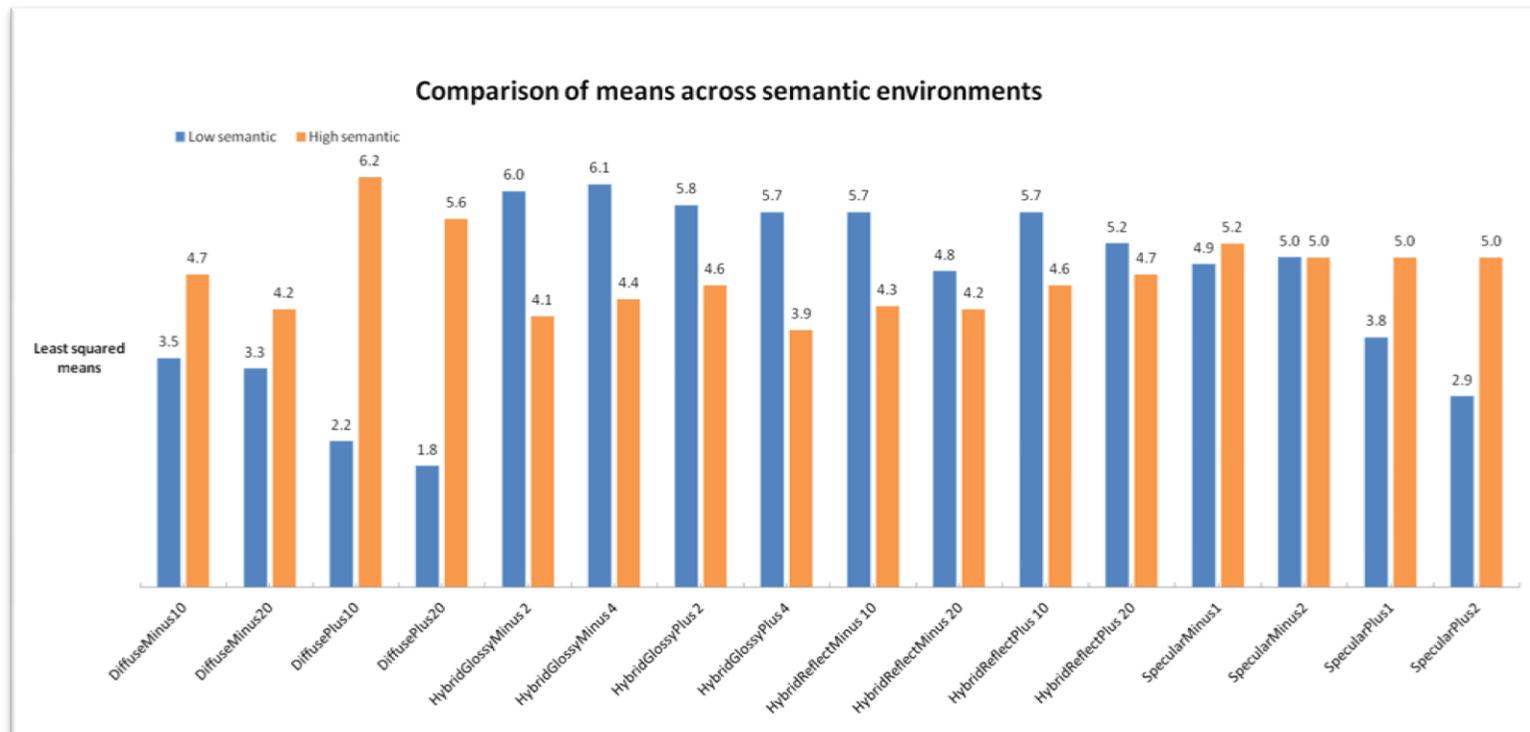


Figure 23: Comparison between each pair of stimuli in the two semantic environments.

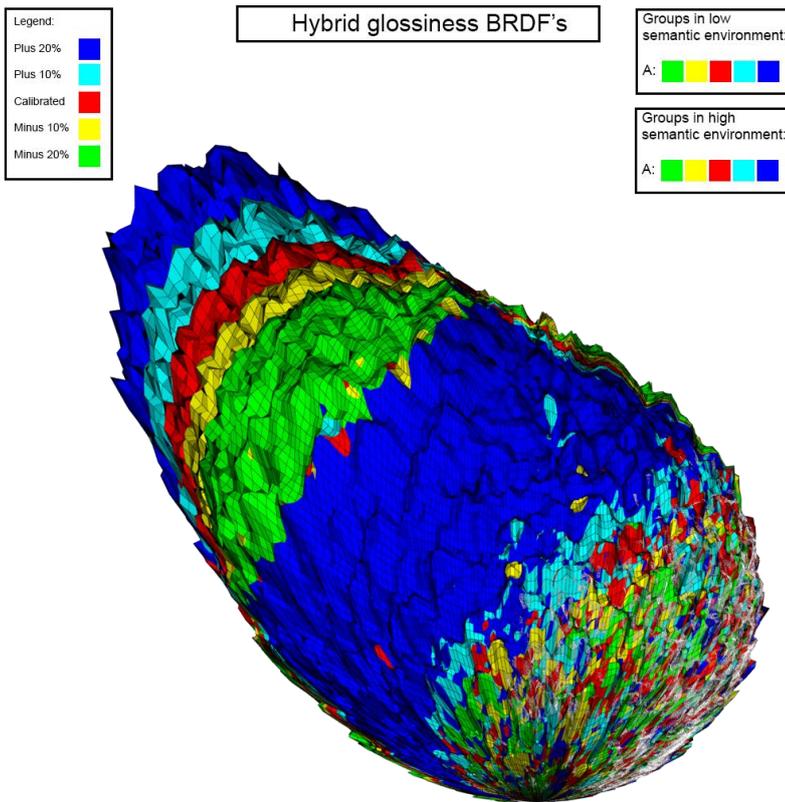


Figure 24: Visualizations of the hybrid stimuli with glossiness varied.

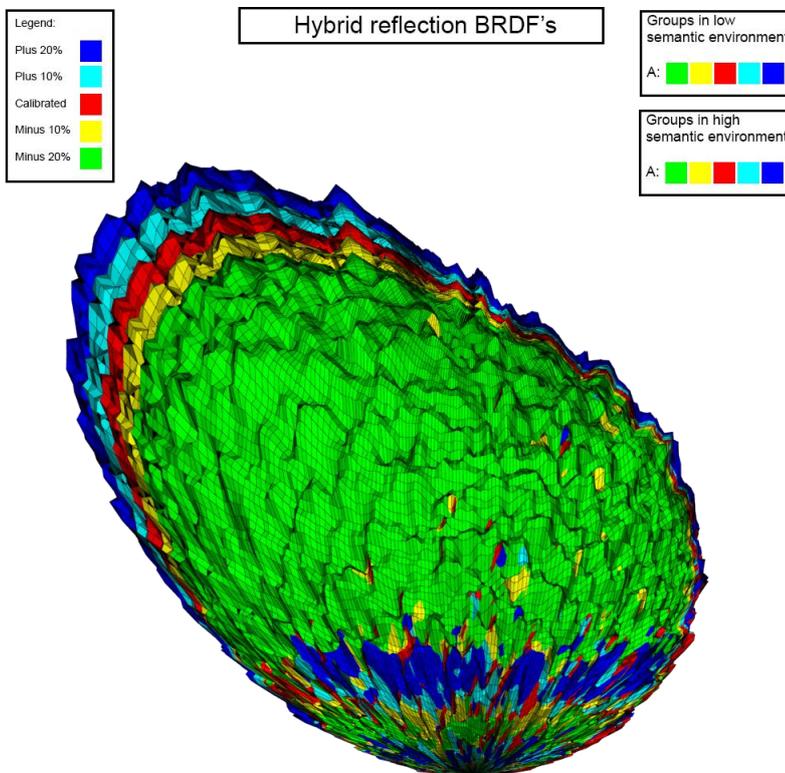


Figure 25: Visualizations of the hybrid stimuli with reflection varied.

5.3 Comparison of Semantic Environments

A pair-wise comparison between the two environments was made using a two-way ANOVA with Tukey’s HSD post hoc analysis. The results show there is a difference between the two environments. If we compare per stimuli pairs, we see that in the diffuse component, the minus 20% percent component was not rated significantly different between the two environments ($F(1,18) = 3.454, p = 0.071$), but the other stimuli pairs were significantly different, see table 11. The hybrid stimuli showed a difference between every stimulus except the extremes of the reflection deviations (Reflect $\pm 20\%$), in general stimuli in the low semantic environment were rated higher than the richer semantic environment, see figure 23 and table 12.

The specular component is significantly different between two stimuli pairs, namely the added specularity (plus 1 and plus 2). ($F(1,18) = 7.403, p = 0.012$) and ($F(1,18) = 23.040, p = <0.0001$), respectively. For a full range of comparisons please refer to table 10.

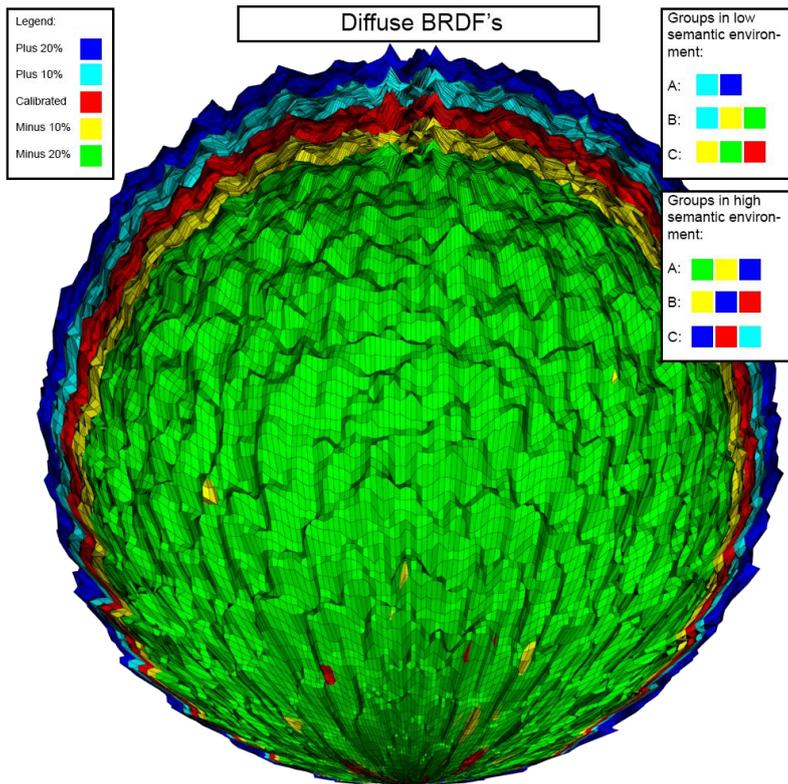


Figure 26: Visualisation of the diffuse BRDF's

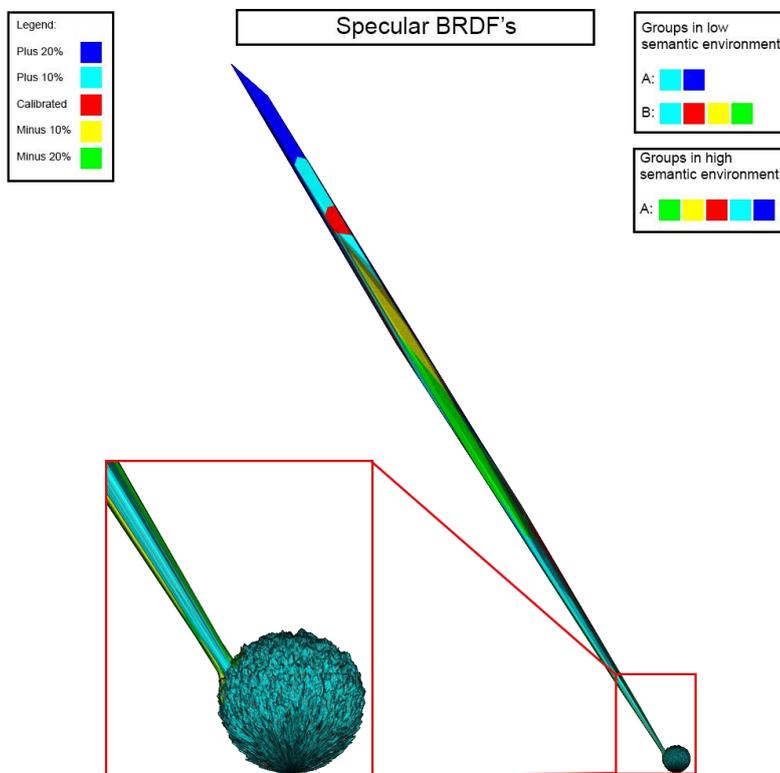


Figure 27: Visualizations of the specular stimuli with the different BRDF's shown. The zoom on the left is of the diffuse part of the BRDF's.

5.4 Visualization of BRDF's and Thresholds

All of the simulated surfaces were processed through the procedure of recreating the given surface's BRDF. This was done to have a visualize representation of the BRDF's used in this project and to identify where in the BRDF the critical difference occurred. The resulting BRDF's were layered into the same spot and rendered as a "cut through".

Based on the grouping in the results we can see that the difference in the hybrid BRDF's is not enough to create a difference for the assessors, this is valid for both the change in reflection and glossiness, see figure 24 and 25.

The diffuse stimuli showed more grouping where in the low semantic environment the added diffuseness were ranked significantly lower than the calibrated stimuli (group A), however in the high semantic environment the same stimuli with added diffuseness were ranked such that they could be grouped (group C) as the highest rated stimuli, see figure 26. The specular BRDF's showed a significant difference in the low semantic environment where the added specularity resulted in a lower similarity score. In the high semantic environment, there was very little difference amongst the ratings resulting in no difference between the stimuli, see figure 27.

6. Discussion

The threshold where humans perceive a difference in surface reflectance is a blurry line and very context dependent. The results from the experiments carried out in this thesis shows a tendency for where this threshold occurs in very specific environment. From the first experiment using a low semantic environment it was discovered that the calibrated stimuli were ranked as highly as the hidden reference for the diffuse stimuli. The hybrid low semantic test showed little difference between the rankings and consequently all of the stimuli can be grouped together besides the hidden physical reference. More spread data was collected from the specular stimuli where a tighter highlight were more likely to be rated as less similar than softer highlight, again the hidden reference was correctly identified.

The results from the low semantic test indicates that humans have a hard time to differentiate between the glossiness levels and are somewhat more sensitive to more areas of contrast, such as in the specular component where a tighter sheen were significantly different from the other stimulus. This can be related to the results by Ramanarayanan et al. (2007). Where the less one could see in a particular reflection the more likely you are to accept another reflection as the same.

Using the same materials in a richer semantic context showed that the diffuse stimuli still were discernible from each other, where the sample with 10% added diffuseness were rated the highest but still can be grouped together with the calibrated and 20% added samples. The two other stimuli showed no grouping between the samples and especially the specular samples were rated similar.

Comparing the two environments there seems to be a tendency of tolerance difference. This is apparent if one were to look at the diffuse sample and the specular samples. The diffuse sample for the low semantic environment had distinguishing groups where the stimuli with added diffuseness were rated significantly lower than the calibrated sample, this was not the case in the richer semantic environment where the same stimuli with added diffuseness were rated in the same grouping as the calibrated.

In the specular sample there were a broader range of tolerance of the specular sheen in the high semantic environment where the samples were more less rated equally, whereas in the low semantic environment a clear distinction happened the more tight the sheen (plus 1 and plus 2) appeared. In fact comments during the high semantic test of how hard it was to notice the specular changes indicate either that; the specular difference is very hard to detect or the actual sample may have been placed in peripheral area of the stimuli.

In regards to the glossiness stimuli groups there seems to be no difference between the environments and the estimated deviations of $\pm 20\%$ reflection and $\pm 4\%$ of the glossiness may be increased even more. This seems to coincide with the study by Olkkonen and Brainard (2010), where similar results were found in terms of glossiness perception.

Talking about the confidence of the assessors, it is clear from the Kendall's W 's that assessors in the low semantic environment experiment were more concordant ($W = 0.553$, $r_s = 0.528$) compared to assessors in the rich semantic environment experiment ($W = 0.196$, $r_s = 0.146$). The level of concordance indicates that in the high semantic environment it was difficult to assess the images and they may have been rated more or less random amongst the assessors. There seems to be concordance in the low semantic environment, which indicates confidence in the data gathered. The difference between the W 's suggests there is a perceptual difference of the same materials when they are in different levels of semantic richness.

The familiarization event was common for both of the experiments. This process is prone to bias as the user's may notice what to look for. This was especially apparent for the high semantic environment, where due to the nature of the stimuli, only a small area differed between the samples. The familiarization was however deemed important in order to establish the stimuli and relativity of the assessment scale to the assessors. It's a fine balance between actively telling the assessors what to look for and making sure the data gathered is measuring the intended construct. For this project it was a conscious choice to include the familiarization due to the risk of not recording correct data.

During the rich environment experiment many of the assessors commented on the highlights on the wall seen in the background of the stimuli. They all had comments on a colour difference between the images, in particular the highlight on the wall. The assessors mentioned that on the left side of the screen it appeared more whitish whereas on the right hand side of the screen it appeared slightly tinted towards a red hue. The fact that this occurred throughout the test and even though the images are exactly equal in the particular spot there is obviously something wrong. It is believed to be the screen which unfortunately did not seem to be correctly calibrated, which might be due to a shift of rooms between the calibration and actual data gathering. When this fact became apparent during the experiment, it was balanced as much as possible by switching the images (i.e. reference was put from the left side to right side and vice versa for the stimuli) after the tenth assessor.

This balancing act does unfortunately not diminish the fact that assessors may have focused on this particular area to see for a change, resulting in an assessment based on a different factor than the intended.

The proposed method for recreating given materials on the spot seems to be working as the results from the first low semantic environment where only the material were present, returned decent results in the form of the calibrated stimulus which in the diffuse category could be grouped together with the reference and in the hybrid rated highest amongst the stimuli. Even though it was rated relatively low in the specular category it could still be grouped together with the highest rated stimulus. The process is still very limited to a few materials and the recreation of the image relies on the artist's skill. But the results from the experiments shows there is, within a certain interval, some minor perceptual differences yet still perceived as the same. This also occurs when we port the materials into a richer semantic environment where this threshold seems even larger.

This experiment cannot ensure construct validity because of the relatively novel approach, however, it would be interesting to see how this process holds in other environments or in general with more materials that are not exclusively dealing with the main BRDF components.

It is the author's belief that the experiment measures the intended construct and face validity is ensured, however the internal consistency of the method and procedure cannot be validated completely based on the authors own interpretations. One may calculate Cronbach's alpha to have an expression for the internal consistency (Lopez, 2007). Calculating the Cronbach's alpha for two experiments, we get an alpha value of 0.64 for the low semantic environment and 0.821 for the high semantic environment. In general it is recommended to have an alpha value of at least 0.7 (Lopez, 2007), meaning the experiment may not be internally reliable.

The assessment scale used in the experiments may have had poorly chosen words for anchors, i.e. the low score of 1 had a relative "very different" anchor whereas the other end anchor point, 7, had a absolute term "identical". In hindsight the 7 anchor point should have been phrased as "very similar".

7. Conclusion

In this project I have investigated a novel approach to sample surfaces found in common areas and recreate those surfaces in a 3D environment. The process used the principal components of a given surface's BRDF to find where the threshold of how much the component had to deviate from a calibrated version before it became noticeably different. The assessment was made in two experiments, one where only the given sample was visible and one where the same sample was put into context. These stimuli were presented to assessors who rated the similarity compared to a reference. The results show that there is a difference in the threshold between the two scenarios and that threshold is larger in a richer semantic environment scenario. This difference is only valid for this particular environment and samples, but it would be interesting to see this taken further and create a range of materials which can be assessed in the same manner.

The process of taking a photographic reference at an angle of 45 degree's light incident angle and recreate material surfaces based on this proved to be a very difficult task. As the results shows from the first, low semantic test there was a significant grouping of all of the hidden references except the diffuse where the calibrated stimulus could be grouped together with the reference. This shows that the calibrated surface materials were not recreated to be exactly equal. It is the author's belief that no matter how close you get to the exact sample, there will always be tiny details portrayed in the reference which will be perceived as different from any 3D created surface.

This process also went through the process of validating V-Ray and 3dsmax as an instrument for creating photometrically calibrated renders. This was done partly through the procedure described by Ulbricht et al. (2006), where the steps taken to validate the environment were done by measuring data from the real scene in the form of reflectance measurement and illuminance data. V-Ray was also tested in form of a integrating sphere where adjustment of parameters were conducted in order to have confidence in the light transfer model. The last step, calibration of the viewing device, was unfortunately not correctly implemented and may have potentially skewed the results by having a badly calibrated screen.

Overall the goal of this particular project was to investigate where the threshold of accepting changes to a material in a low semantic environment and then see if this had any change when the same materials were applied in a scene in a richer semantic environment. The results shows that humans are more likely to accept changes in the directional diffuse component of the BRDF, whereas when increasing the specular highlights they were more sensitive to small changes. The diffuse component showed a tendency in both scenarios to be more sensitive to changes.

The hypothesis of this thesis was:

Humans will be less likely to notice Bidirectional Reflectance Distribution Function changes of surfaces in a photorealistic high semantic environment compared to a low semantic environment.

Given the results from the two experiments, it indicates that humans are less likely to notice changes in the specular component and diffuse component when the surface is exposed in high semantic environment compared to a low semantic environment. The hybrid component showed no difference between environments but is expected to be more flexible than the other two components.

7.1 Future Development

It would be interesting to see other researchers using the proposed method to see how robust this method really is. Other possibilities could be to use the proposed method in a combination with post processing techniques, such as rotoscoping. For example, using the scene from the high semantic environment one could change the table's surface by compositing the simulated materials onto the physical reference and then run the experiment as described.

In its current implementation it is fairly limited to simple surface materials and offline renderings. New novel possibilities for human interaction with 3D reality such as the Oculus Rift⁸ are currently blooming in these years. The immersive possibilities are extremely interesting, especially if one were to recreate photometrically calibrated architectural environments which are operating at an interactive level. This probably requires a very powerful simulation engine, but with the release of Unreal Engine 4 in early 2014 it seems more and more possible to create such interactive installations. With this interactive approach the samples would of course need to work from every angle. The proposed procedure has currently only been processed using at a 45 degree angle. The obvious next step would be to vary the angles of incident light in order to validate this procedure for more than one view, such as in interactive installations.

⁸ <http://www.oculusvr.com/>

8. References

- Aldridge, R., Davidoff, J., Ghanbari, M., Hands, D., & Pearson, D. (1995). *Recency effect in the subjective assessment of digitally-coded television pictures*. Paper presented at the Image Processing and its Applications, 1995., Fifth International Conference on.
- Ashdown, I. (1998). Parsing The IESNA LM-63 Photometric Data File. Retrieved 14/04, 2014, from <http://lumen.iee.put.poznan.pl/kw/iesna.txt>
- Beck, J. (1972). *Surface color perception* (Vol. 2): Cornell University Press Ithaca, NY.
- Belhumeur, P. N., Kriegman, D. J., & Yuille, A. L. (1999). *The bas-relief ambiguity*. International journal of computer vision, 35(1), 33-44.
- ChaosGroup. (2010a). GI methods for V-ray. Retrieved 24/04, 2014, from <http://help.chaosgroup.com/vray/help/150SP1/gimethods.htm>
- ChaosGroup. (2010b). Progressive path tracing with V-Ray. Retrieved 25/4, 2014, from http://help.chaosgroup.com/vray/help/150SP1/tutorials_pathtracing.htm
- CornellUniversity. (2011). Components of the BRDF. Retrieved 27/5, 2014, from <http://www.graphics.cornell.edu/~westin/multimedia-paper/node5.html>
- Drago, F., & Myszkowski, K. (2001). *Validation proposal for global illumination and rendering techniques*. Computers & Graphics, 25(3), 511-518.
- Edwards, C. H., & Penney, D. E. (2008). *Calculus Early Transcendentals* (Seventh Edition ed.): Pearson Education International.
- Elhelw, M., Nicolaou, M., Chung, A., Yang, G.-Z., & Atkins, M. S. (2008). *A gaze-based study for investigating the perception of visual realism in simulated scenes*. ACM Transactions on Applied Perception, 5(1), 1-20. doi: 10.1145/1279640.1279643
- Goss, H. (2013). The addition of real-world imperfections is taking architectural visualisation to the next level. In R. Bryant (Ed.). <http://www.dezeen.com/2013/08/12/henry-goss-on-architectural-visualisations/>: Dezeen.
- Habermas, J. (1984). *The theory of communicative action, Vol. I*. Boston: Beacon.
- He, X. D., Heynen, P. O., Phillips, R. L., Torrance, K. E., Salesin, D. H., & Greenberg, D. P. (1992). *A fast and accurate light reflection model*. Paper presented at the ACM SIGGRAPH Computer Graphics.
- Howell, D. (2008). *Fundamental Statistics for the Behavioral Sciences* (Sixth Edition ed.): Thomson Wadsworth.
- Iversen, A., Roy, N., Hvass, M., Jørgensen, M., Christoffersen, J., Osterhaus, W., & Johnsen, K. (2013). *Dagslysberegninger i praksis* (pp. 55).
- Koylazov, V. (2014). Custom BRDF's for Vray (Forum post). Retrieved 20/5-2014 <http://forums.chaosgroup.com/showthread.php?77413-Custom-BRDF-s-for-Vray>

- Kozłowski, O., & Kautz, J. (2007). *Is accurate occlusion of glossy reflections necessary?* Paper presented at the Proceedings of the 4th symposium on Applied perception in graphics and visualization.
- Labayrade, R. (2010). *Use of manufacturers' luminaire data in light transport simulation programs*. *Light & Engineering*, 18(4).
- Lopez, M. (2007). *Estimation of Cronbach's alpha for sparse datasets*. Paper presented at the 20th Annual Conference on Computin Qualifications-NACCQ. Samuel Man and Noel Bridgeman (Eds).
- Madsen, C. B., Borg, M., & Paprocki, M. M. (2012). *Perceptual Evaluation of Photo-Realism in Real-Time 3D Augmented Reality*.
- Mardaljevic, J. (1999). *Daylight simulation: validation, sky models and daylight coefficients*.
- Myszkowski, K., & Kunii, T. L. (2000). *A case study towards validation of global illumination algorithms: progressive hierarchical radiosity with clustering*. *The Visual Computer*, 16(5), 271-288.
- Ngan, A., Durand, F., & Matusik, W. (2005). *Experimental analysis of BRDF models*. Paper presented at the Proceedings of the Sixteenth Eurographics conference on Rendering Techniques.
- Olkkonen, M., & Brainard, D. H. (2010). *Perceived glossiness and lightness under real-world illumination*. *Journal of vision*, 10(9), 5.
- Phong, B. T. (1975). *Illumination for computer generated pictures*. *Communications of the ACM*, 18(6), 311-317.
- Rademacher, P., Lengyel, J., Cutrell, E., & Whitted, T. (2001). *Measuring the perception of visual realism in images* *Rendering Techniques 2001* (pp. 235-247): Springer.
- Ramanarayanan, G., Ferwerda, J., Walter, B., & Bala, K. (2007). *Visual equivalence: towards a new standard for image fidelity*. Paper presented at the ACM Transactions on Graphics (TOG).
- Reinhard, E., Efros, A. A., Kautz, J., & Seidel, H.-P. (2013). *On Visual Realism of Synthesized Imagery*.
- Roien, M. G. (2013). *Development of a camera-based lighting control*. (MSc.Eng), Technical University of Denmark.
- Schubert, E. F. (2006). *Human eye sensitivity and photometric quantities: Light-Emitting Diodes*, 2nd ed. Cambridge, UK: Cambridge University Press.
- Trawiński, B., Smętek, M., Telec, Z., & Lasota, T. (2012). *Nonparametric Statistical Analysis for Multiple Comparison of Machine Learning Regression Algorithms*.
- Ulbricht, C., Wilkie, A., & Purgathofer, W. (2006). *Verification of physically based rendering algorithms*. Paper presented at the Computer Graphics Forum.
- Van Verth, J. M., & Bishop, L. M. (2008). *Essential mathematics for games and interactive applications: a programmer's Guide*: CRC Press.
- Vangorp, P., Condon, T. S., Ferwerda, J. A., Bala, K., Schoukens, R., & Dutré, P. (2009). *Visual equivalence in dynamic scenes*. CW Reports.
- Vangorp, P., Laurijssen, J., & Dutré, P. (2007). *The influence of shape on the perception of material reflectance*. *ACM Transactions on Graphics*, 26(3), 77. doi: 10.1145/1276377.1276473

- Villa, C., Parent, E., & Labayrade, R. (2010). *Calibrating a display device for subjective visual comfort tests: selection of light simulation programs and post-production operations*. Paper presented at the Proceedings of CIE Light Efficiency Conference, Vienna, Austria, 9p.
- Walree, P. v. (2013). Photography and photographic optics. Retrieved 10/03, 2014, from <http://toothwalker.org/optics/chromatic.html>
- Walter, B. (2005). *Notes on the Ward BRDF*. Program of Computer Graphics, Cornell University, Technical report PCG-05-06.
- Wolfe, J. M. (2009). *Sensation & Perception* (Second Edition ed.): Sinaur Associates, Inc.
- Yu, I., Cox, A., Kim, M. H., Ritschel, T., Grosch, T., Dachsbacher, C., & Kautz, J. (2009). *Perceptual influence of approximate visibility in indirect illumination*. ACM Transactions on Applied Perception, 6(4), 1-14. doi: 10.1145/1609967.1609971

9. List of Figures

Figure 1: Rendered image of a bronchoscopy. The regions depict areas of interests where users based their realism decision making. Image from Elhelw et al. (2008).....	6
Figure 2: Comparison between ISO levels on a 100% crop of a DSLR image. Left: ISO 200, right: ISO 1600; images from (Tybjerg, Tybjerg, & Tybjerg, 2011).	11
Figure 3: Longitudinal chromatic aberration. The focal points of the various colours do not coincide and only the green information is in focus (Walree, 2013).	11
Figure 4: Origin of transverse chromatic aberration. The size of the image varies from one colour to the next (Walree, 2013).	12
Figure 5: An example of barrel distortion created by a Canon 18mm-55mm lens at 18 mm.	12
Figure 6: Components of a BRDF, where the white cone represents the specular component, the blue area is the glossy component and the purple area is the diffuse component, image from (CornellUniversity, 2011).15	15
Figure 7: V-Ray's implementation of the Phong (left) Blinn (middle) and Ward (right) BRDF models (VisualDynamicsLLC, 2014).	16
Figure 8: A chart overview of how proper linear workflow helps to achieve correct lit renderings.....	19
Figure 9: Setup for gathering a given sample's surface appearance.	21
Figure 10: A sample IES file.	22
Figure 11: A comparison of the far field photometry problem in a standard IES file (right) and the approximate grid divided area light shape used to counter this (left).....	23
Figure 12: initial IES file beamed on a hemisphere. Seen directly from below.	24
Figure 13: OpenCV's coordinate system, starts from the top left corner of the image and have a downwards positive y axis and a right hand positive x-axis.	26
Figure 14: A zoomed comparison between the original IES (top part) and the recreated IES file (lower part). There is a subtle difference in the overall illuminance levels.	27
Figure 15: The reflected three materials, top: diffuse, middle: hybrid, bottom: specular. Note the specular highlights appear the same, but in fact contains pixel which is out of range for normal displays.	28
Figure 16: The three material samples' BRDF. Left: diffuse, middle: hybrid, right: specular.	29
Figure 17: Recording images of the samples, this is the specular whiteboard sample being recorded, the ambient room lighting was turned off during sample recording.	31
Figure 18: A composite showing the reference and the rendered calibrated sample, a 3D BRDF of the calibrated surface, 2D IES file of the surface and the parametre used in V-Ray.	33
Figure 19: The reference image taken at the meeting room.	34
Figure 20: The recreated 3D render of the meeting room scene.	36

Figure 21: False colour map showing illuminances at different areas in the scene. Red color's denominate an area with more illuminance (~ 1000 lux) whereas dark blue areas are less illuminated (~ 100 lux).	37
Figure 22: Experiment setup.	39
Figure 23: Comparison between each pair of stimuli in the two semantic environments.	45
Figure 24: Visualizations of the hybrid stimuli with glossiness varied.	46
Figure 25: Visualizations of the hybrid stimuli with reflection varied.	46
Figure 26: Visualisation of the diffuse BRDF's.	47
Figure 27: Visualizations of the specular stimuli with the different BRDF's shown The zoom on the left is of the diffuse part of the BRDF's.	47

10. List of Tables

Table 1: Measurements of the illuminance levels of the sphere, the values for reflectivity and glossiness have been normalized.....	22
Table 2: Reflectance measurements of the meeting room.....	35
Table 3: Illuminance measurements done at the meeting room.....	36
Table 4: Groupings of the diffuse stimuli in a low semantic environment.....	41
Table 5: Groupings of the hybrid stimuli in a low semantic environment, numbers in the “Sample” column denotes deviation percent.....	41
Table 6: Groupings of the specular stimuli in a low semantic environment numbers in the “Sample” column denotes deviation percent.....	42
Table 7: Groupings of the diffuse stimuli in a high semantic environment.....	43
Table 8: Groupings of the hybrid stimuli in a high semantic environment.....	44
Table 9: Groupings of the specular stimuli in a high semantic environment.....	44
Table 10: Comparison between the two semantic environments for the specular stimuli. R is how they correlate. F and P values for the comparison and significant differences.....	44
Table 11: Comparison between the two semantic environments for the diffuse stimuli. R is how they correlate. F and P values for the comparison and significant differences.....	44
Table 12: Comparison between the two semantic environments for the hybrid stimuli. R is how they correlate. F and P values for the comparison and significant differences.....	45

11. Appendix 1, IES file example

This is an example IES file. Created in the format (Ashdown, 1998):

```

IESNA:LM-63-1995<Keyword 1>
02 <Keyword 2>
03 ...
04 <Keyword n>
05 TILT=<file-spec> or <INCLUDE> or <NONE>
06 <lamp-to-luminaire geometry>
07 <# of pairs of angles and multiplying factors>
08 <angles>
09 <multiplying factors>
10 <# of lamps> <lumens per lamp> <candela multiplier>
    <# of vertical angles> <# of horizontal angles> <photometric type>
    <units type> <width> <length> <height>
11 <ballast factor> <future use> <input watts>
12 <vertical angles>
13 <horizontal angles>
14 <candela values for all vertical angles at first horizontal angle>
15 <candela values for all vertical angles at second horizontal angle>
16 ...
17 <candela values for all vertical angles at nth horizontal angle>

```

The following is an actual IES file formatted as IESNA:LM-63-1995 photometric file.

```

IESNA:LM-63-1995
[TEST]      ABC1234 ABC Laboratories
[MANUFAC]   Aardvark Lighting Inc.
[LUMCAT]    SKYVIEW 123-XYZ-abs-400
[LUMINAIRE] Wide beam flood to be used without tilt
[LAMPCAT]   MH-400-CLEAR
[LAMP]      Metal Halide 400 watt
[BALLASTCAT] Global 16G6031-17R
[BALLAST]   400W 277V MH
[MAINTCAT]  4
[OTHER]     This luminaire is useful as an indirect flood
[MORE]      and to reduce light pollution in down light
[MORE]      applications.
[SEARCH]    POLLUTION SPORTS INDIRECT
[BLOCK]
[LUMCAT]    TENNISVIEW 123-XYZ-abc-400
[LUMINAIRE] Wide beam flood for indirect applications.
[ENDBLOCK]
TILT=INCLUDE
1
13
0 15 30 45 60 75 90 105 120 135 150 165 180
1.0 .95 .94 .90 .88 .87 .98 .87 .88 .90 .94 .95 1.0
1 50000 1 5 3 1 1 .5 .6 0
1.0 1.0 495
0 22.5 45 67.5 90
0 45 90
10000 50000 25000 10000 5000
10000 35000 16000 8000 3000
10000 20000 10000 5000 1000

```

12. Appendix 2, IES Re-creator Program

This program was created in Microsoft Express 2014, using OpenCV 2.4.8 (needs to be installed on the computer). The program takes an image as input (needs to be in the same folder as the program) creates an IES file based on the pixel values in the image.

Instructions:

This programme works by converting pixel data to IES data. The program takes a square image and processes the pixel values by using spherical coordinates.

To use this properly:

- 1): render an image in a square format, i.e. 1024x1024. To get the most precise results use a 32bit float format.
- 2): put the image in the **same** directory as the programme.
- 3): Run the program. follow the prompts on screen. Current limitations for max angles are: vertical 90, horizontal: 360. angles step denotes the precision of the computations. The lower the step size the more precise the IES file will be.
- 4): An text file (can be renamed to .ies instead of .txt to work) will be created in the same folder as the image with the same name as you specified in the programme.

The program uses OpenCV 2.4.8 which needs to be installed on the computer.

```

1  #include <opencv2/opencv.hpp>
2  #include <stdio.h>
3  #include <iostream>
4  #include <math.h>
5  #include <fstream>
6  #include <string>
7  #include <time.h>
8
9  using namespace cv;
10 using namespace std;
11 #define PI 3.14159265
12
13 double sind(double angle) //convert an angle in radians to degree's
14 {
15     return angle * PI / 180.0f;
16 }
17
18 float doMathPhi(float x, float y, int height, int width){ //calculate phi angle based on x,y coords.
19     Used for debugging
20     y = y / height; //scale the coords to fit within the unit circle
21     x = x / width;
22     return atan2f(y, x); //arctan of y/x
23 }
24
25 float doMathTheta(float x, float y, int height, int width){ //calc theta of given x,y coord.
26 Used for debugging
27     x = x / width; //scale the coords to fit within the unit circle
28     y = y / height;

```

```

29     float z = sqrt((pow(x, 2) + pow(y, 2))); //calc hypotenuse
30     if (z >= 1){ return acosf(1); } // make sure if z is bigger than 1 then return max    value
31     else { return (acosf(z));
32     } // return arccos of hypotenuse
33 }
34
35 float intensityCalc(Mat& img, float angleT, float angleP, int width, int height, int
36 imgDepth){ //calculate intensity of a given pixel per theta/phi
37     float x, y;
38     y = sin(angleP)*sin(angleT);
39     x = cos(angleP)*sin(angleT);
40     //cout << "x and y is: " << x*width + width-1<< " " << y*height + height-1<< endl;
41
42     // check for image bit depth. Return different typed casted lookup functions, uchar for    8 bit,
43     ushort for 16 bit and float for 32 bit.
44     if (imgDepth == 0 || imgDepth == 1){
45         if (x < 0) return (img.at<uchar>(x*width + width, y*height + height - 1)); //special exception
46         where x would be minus one at angles (90,180). Return pixel next to it.
47         if (y < 0) return (img.at<uchar>(x*width + width - 1, y*height + height)); //special exception
48         where y would be minus one at angles (270,90). Return pixel next to it.
49
50         // return the pixel value at x,y times the radius then scaled to fit the dimensions of the
51         image and minused one due to 0 index of the Mat construct.
52         return (img.at<uchar>(x*width + width - 1, y*height + height - 1));
53     }
54     else if (imgDepth == 2 || imgDepth == 3){
55         if (x < 0) return (img.at<unsigned short>(x*width + width, y*height + height - 1)); //special
56         exception where x would be minus one at angles (90,180). Return pixel next to it.
57         if (y < 0) return (img.at<unsigned short>(x*width + width - 1, y*height + height)); //special
58         exception where y would be minus one at angles (270,90). Return pixel next to it.
59
60         // return the pixel value at x,y times the radius then scaled to fit the dimensions of the
61         image and minused one due to 0 index of the Mat construct.
62         return (img.at<unsigned short>(x*width + width - 1, y*height + height - 1));
63     }
64     else {
65         if (x < 0) return (img.at<float>(x*width + width, y*height + height - 1)); //special exception
66         where x would be minus one at angles (90,180). Return pixel next to it.
67         if (y < 0) return (img.at<float>(x*width + width - 1, y*height + height)); //special exception
68         where y would be minus one at angles (270,90). Return pixel next to it.
69
70         // return the pixel value at x,y times the radius then scaled to fit the dimensions of the
71         image and minused one due to 0 index of the Mat construct.
72         return (img.at<float>(x*width + width - 1, y*height + height - 1));
73     }
74 }
75 }
76
77 const string currentDate() { //get the date function
78     time_t now = time(0); //create a template for time
79     struct tm tstruct; //time struct struct which contains the date format members
80     char buf[80]; //char array containing the date
81     tstruct = *localtime(&now); // get the time and store it in the struct
82     strftime(buf, sizeof(buf), "%Y-%m-%d", &tstruct); //format into year, month, day format
83
84     return buf;
85 }
86
87 void writeToFile(){ //begin main calculations and write to a file
88     String fileName;
89     cout << "Please enter file name with extension (file should be in same directory as this
90     program)." << endl;
91     cin >> fileName;
92

```

```

93 Mat imgMat = imread(fileName, CV_LOAD_IMAGE_ANYDEPTH); //load the image with any depth
94 while (imgMat.empty()) //if the image doesn't exists in the directory, prompt the user
95 {
96     cout << "Cannot load image! try again (spelling?, remembered the extension?" << endl;
97     cin >> fileName;
98     imgMat = imread(fileName, CV_LOAD_IMAGE_ANYDEPTH);
99 }
100 cout << "File loaded, yay!" << endl;
101 //get the file depth from the image
102 int fileDepth = imgMat.depth();
103
104 float pixMult; //assign a multiplication factor depending on the image depth. uchar for 8 bit,
105     ushort for 16 bit and float for 32 bit.
106 if (fileDepth == 0 || fileDepth == 1){
107     pixMult = 255.f;
108 }
109 else if (fileDepth == 2 || fileDepth == 3){
110     pixMult = 65536.f;
111 }
112 else {
113     pixMult = 1.f;
114 }
115
116 short h = imgMat.size().width, w = imgMat.size().height;
117 short halfHeight = h / 2;
118 short halfWidth = w / 2;
119 short maxLum;
120 cout << "Enter the maximum luminosity value (pixel val of 1 = max lux): ";
121 cin >> maxLum; // user input for max lux value
122 /*Scalar intensity; //debugging
123 Point2d pt;*/
124
125 float numVangles, numHangles, vAnglesStep, hAnglesStep; //get user info of max angles and the
126     increments wanted.
127 cout << "Please enter max vertical angle: " << endl;
128 cin >> numVangles;
129 cout << "Please enter max horizontal angle: " << endl;
130 cin >> numHangles;
131 cout << "Please enter the step size of vertical angles: " << endl;
132 cin >> vAnglesStep;
133 cout << "Please enter the step size of horizontal angles:" << endl;
134 cin >> hAnglesStep;
135
136 int lastindex = fileName.find_last_of("."); //find the extension of the file
137 string strippedFileName = fileName.substr(0, lastindex); // strip the raw filename from the file
138     and toss the extension
139 ofstream myfile(strippedFileName + "_IES.txt"); //create a file
140 int counter = 0, anglCounter = 0; //create counters to make sure we are within the character
141     limits of the IES standard.
142
143 if (myfile.is_open()){ //if it is open
144     //enter some header lines in the IES format.
145     myfile << "IESNA:LM-63-1995\n";
146     myfile << "[TEST] This is a text file containing all the angles corresponding to the IES file
147 format.\n[DATE]          " << currentDate() << "          "\n[MANUFAC]
148 SBI\n[LUMCAT]\n[LUMINARIES]\n[MORE]\n[LAMP]\n[BALLAST]\n[OTHER] By Anders Lumbye\n";
149     myfile << "TILT=NONE\n";
150     myfile << "1 " << maxLum << " 1 " << (numVangles / vAnglesStep) + 1 << " " << (numHangles
151 / hAnglesStep) + 1 << " 1 1 -0.33 0 0\n";
152     myfile << "1 1 24\n";
153
154     for (float i = 0; i <= numVangles; i += vAnglesStep){ //write all the vertical angles out
155         if (anglCounter < 20){ //max 20 entries per line
156             myfile << " " << i;

```

```

157         anglCounter++;
158     }
159     else{
160         myfile << endl << " "; //escape if limit has been reached
161         myfile << i;
162         anglCounter = 0; //reset count for new line
163     }
164
165 }
166 myfile << endl;
167 anglCounter = 0;
168 for (float j = 0; j <= numHangles; j += hAnglesStep){
169     if (anglCounter < 20){ //max 20 entries per line
170         myfile << " " << j;
171         anglCounter++;
172     }
173     else{
174         myfile << endl << " "; //escape if limit has been reached
175         myfile << j;
176         anglCounter = 0; //reset count for new line
177     }
178
179 }
180
181 myfile << endl;
182
183 for (double i = 0; i <= numHangles; i += hAnglesStep) //begin loop through image using angles
184     as constraints
185 {
186     for (double j = 0; j <= numVangles; j += vAnglesStep)
187     {
188         if (counter < 10){ //ensure we are not above 10 entries = 80 chars
189             //intensity calc based on the angle given by i and j, divided by max value to get a
190             percentage, then multiplied with max lux to get a pseudo lumen.
191             myfile << (intensityCalc(imgMat, sind(j), sind(i), halfWidth,
192                 halfHeight,fileDepth)/pixMult) * maxLum << " ";
193             counter++;
194         }
195         else{
196             myfile << endl << " "; //escape if limit has been reached
197             myfile << (intensityCalc(imgMat, sind(j), sind(i), halfWidth,
198                 halfHeight,fileDepth)/pixMult) * maxLum << " ";
199             counter = 0; //reset count for new line
200         }
201     }
202     counter = 0; //reset count for new line
203     myfile << endl;
204 }
205
206 }
207
208 cvNamedWindow("image:", CV_WINDOW_AUTOSIZE);
209 imshow("image:", imgMat);
210 cout << "DONE!" << endl;
211 myfile.close();
212 cvWaitKey();
213
214 //debuggin module:
215
216 //while (1){
217
218 //     cout << "Please enter x: ";
219 //     cin >> inputX;
220 //     cout << "Please enter y: ";

```

```
221 // cin >> inputY;
222 // cout << "Phi (horizontal angle) is: " << 90.f-sind(doMathPhi(inputX - halfHeight, inputY -
223 //         halfWidth, halfWidth, halfHeight)) << endl; //minus 90.f to rotate 90 degrees
224 // cout << "Theta (vertical angle) is: " << 90.f-sind(doMathTheta(inputX - halfHeight, inputY -
225 //         halfWidth, halfWidth, halfHeight)) << endl;
226 // pt.x = inputX;
227 // pt.y = inputY;
228 //
229 // //if (inputX <= halfWidth && inputY <= halfHeight){
230 // //   inp = gray_image.ptr<uchar>(pt.x, h-pt.y);
231 // //   cout << "top left quad" << endl;
232 // //}
233 // //else if (inputX > halfWidth && inputY < halfHeight){
234 // //   inp = gray_image.ptr<uchar>(w-pt.x, pt.y);
235 // //   cout << "top right quad" << endl;
236 // //}
237 // //else if (inputX > halfWidth && inputY > halfHeight){
238 // //   inp = gray_image.ptr<uchar>(pt.x, h - pt.y);
239 // //   cout << "lower right quad" << endl;
240 // //}
241 // //else if (inputX < halfWidth && inputY > halfHeight){
242 // //   inp = gray_image.ptr<uchar>(w - pt.x, pt.y);
243 // //   cout << "lower left quad" << endl;
244 // //}
245 // intensity = gray_image.at<uchar>(pt.y,pt.x);
246 // //uchar* inp = gray_image.ptr<uchar>(pt.x,pt.y);
247 // // cout << "Coords are (x,y): " << pt.x << " " << pt.y << endl;
248 // cout << "pixel value is: " <<intensity.val[0] << endl;
249 // cout << "pseudo lumen val is: " << (intensity.val[0]/255.0f) * maxLum << endl;
250 // waitKey(5);
251 // circle(gray_image, pt, 3, Scalar(0, 255, 255), 2);
252 // //cvShowImage("image1:", img_grey);
253 // waitKey(5);
254 // imshow("image:", gray_image);
255 // waitKey(1);
256 //} //end debug
257
258
259 } //end writeToFile
260
261 int main(){
262     writeToFile(); //call the calculation function
263     return 0;
```

13. Appendix 3, Data Gathering Code

Code used in the data gathering process. This code generates a text file for the assessor and stores the entered assessment for the given stimuli. The program was created using Microsoft Visual Basic scripting language, which was integrated together with Microsoft Powerpoint 2007.

```
Private Sub CommandButton1_Click()
' Purpose: Creates a text file and stores the survey results
' in the text file.

Dim objFSO As Scripting.FileSystemObject
Dim objTS As Scripting.TextStream

' Create the text file.
Set objFSO = New Scripting.FileSystemObject
Set objTS = objFSO.OpenTextFile("D:\Copy drev\Copy\AAU SBI
Rendering\test1\Survey_Results.txt", _
ForAppending, True)
' Write the results to the text file and then
' close the file.
objTS.WriteLine "Rating = " & Me.TextBox1.Text & "      " & Me.Label1.Caption
objTS.Close
' go to next stimuli
ActivePresentation.SlideShowWindow.View.Next

'clear the text box for the next stimuli
Me.TextBox1.Text = ""

End Sub
```

Randomize script which was run before each assessor began the experiments:

```
Sub sort_rand()
Dim i As Integer
Dim myvalue As Integer
Dim islides As Integer
islides = ActivePresentation.Slides.Count
For i = 1 To ActivePresentation.Slides.Count
myvalue = Int((i * Rnd) + 1)
ActiveWindow.ViewType = ppViewSlideSorter
ActivePresentation.Slides(myvalue).Select
ActiveWindow.Selection.Cut
ActivePresentation.Slides(islides - 1).Select
ActiveWindow.View.Paste
Next
End Sub
```

14. Appendix 4, Instructions for the First Experiment.

Dansk

Denne test går ud på at rangere billederne du ser på skærmen i forhold til en reference som vises på en anden skærm. Testen er opdelt i tre dele, der omhandler opfattelsen af forskellige overflader.

Du skal vurdere hvor godt billederne ligner referencen i forhold til fremtoningen på en skala rangerende fra 1-7, hvor en vurdering på 1 svarer til "relativ meget forskellig" og en vurdering på 7 svarer til "identisk". Du skal se bort fra strukturelle forskelle, som f.eks. ridser der ikke ser ens ud, men i stedet kigge på billedet som helhed.

Billederne vil forekomme meget ens og der vil være lille forskel i mellem dem. Hvis det er muligt så brug så meget af skalaen som muligt. Der er kan enten være 6 billeder pr. test eller 10 pr. test.

Du skal afgive din besvarelse i den lille boks, hvor du skriver din vurdering mellem 1 og 7. Når du er tilfreds med din besvarelse så tryk på "Next Slide".

English

In this test you are assessing a range of images in terms of appearance against a reference image. The test is split into three different small tests, where you will assess surfaces of common materials.

You should not take into consideration any structural differences, i.e. scratches, but assess the image as a whole. You should assess each image on a scale from 1-7 compared to a reference image, where 1 is "relatively very different" and 7 is "identical".

The pictures will appear very similar, with little difference between them. If possible try to use as much of the scale as possible. For each test there will be either 6 or 10 images that needs assessment.

You should enter your assessment in the small box in the lower right corner, where you can type your evaluation in numbers between 1 and 7. When you are satisfied with you assessment you may proceed by pushing the "Next Slide" button.

15. Appendix 5, Instructions for the Second Experiment

Dansk

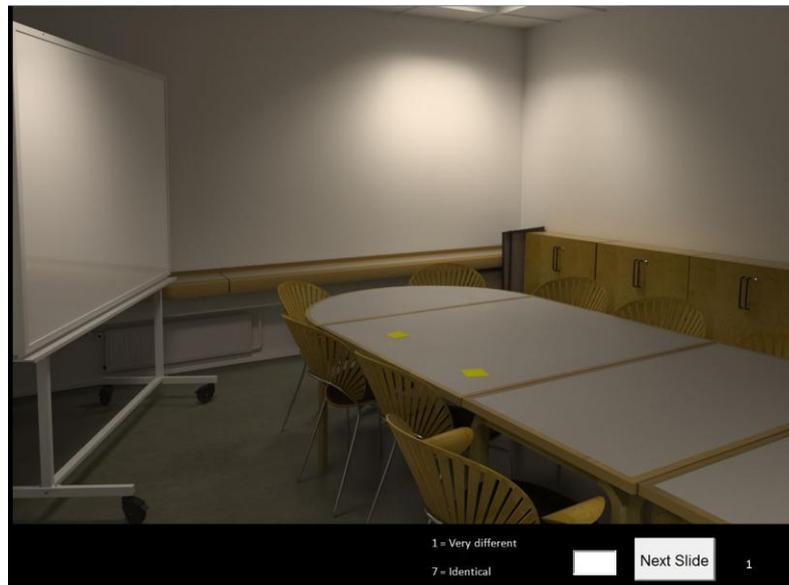
Denne test går ud på at rangere billederne du ser på skærmen i forhold til en reference. Testen består af i alt 17 billeder, som omfatter, hvordan mennesker opfatter forskellige overflade materialer.

Billederne vil forekomme meget ens, og der vil være en meget lille forskel i mellem dem. Hvis det er muligt så brug så meget af skalaen som muligt.

Du skal vurdere, hvor godt billederne ligner referencen i forhold til fremtoningen på en skala rangerende fra 1-7, hvor en vurdering på 1 svarer til "relativt meget forskellig" og en vurdering på 7 svarer til "identisk". Bemærk du skal vurdere en relativt forskellighed i forhold til en reference.

Du skal ikke vurdere hvordan billedet ser ud æstetisk, men tænke mere på, hvordan billedet fremtoner som helhed.

Du skal afgive din besvarelse i den lille boks, hvor du skriver din vurdering mellem 1 og 7. Når du er tilfreds med din besvarelse så tryk på "Next Slide". Du kan se, hvor langt du er nået nede i højre hjørne af billedet.



Example from the test / et eksempel fra testen.

English

In this test you are assessing a range of images in terms of appearance against a reference image. The test consist s of 17 different images, where you will assess the appearance of common surface materials.

You should assess each image on a scale from 1-7 compared to a reference image, where 1 is "relatively very different" and 7 is "identical". Please note the assessment is a relative comparison to a reference.

The pictures will appear very similar, with very little difference between them. If possible try to use as much of the scale as possible.

You should enter your assessment in the small box in the lower right corner, where you may type your evaluation in numbers between 1 and 7. When you are satisfied with you assessment you may proceed by pushing the "Next Slide" button.