CS667 **Lecture 15: Perceptually Based Image Synthesis** 31 March 2005

Mark Rubelmann Lecturer: Jim Ferwerda

1 Introduction

The three basic steps involved in generating photo-realistic images are modeling, rendering, and imaging. Modeling can be considered to include building the geometry of the scene being rendered, plus assigning material properties to all the surfaces. Rendering is the process of simulating light transport to determine the luminance values at every point in the scene. The final step, imaging, is needed to convert the calculated luminances to color values that can be displayed on-screen. At each of these steps, insight into human perception can dramatically increase the realism of a rendered scene or help speed up the rendering process.

2 Modeling: Psychophysics of Material Appearance

2.1 Gloss Space

Color is a very well understood material property and is well defined within the context of computer graphics as various color spaces. Thanks to colorimetry, these color spaces are given in terms of human perception and allow for meaningful color dimensions, uniform scales, predictable appearance, color matching, color differences, and just-noticeable differences(JND). Gloss is a material property that is just as important as color and its real-world causes are also well understood. In computer graphics, however, there has not been an analogous gloss space defined until recently. Accurate gloss measurements can be captured by BRDFs but have most commonly been approximated by simple reflection models such as Ward. One problem with using approximations such as these is that the parameters are not perceptually meaningful and, as a result, the appearance of the surface is difficult to predict.

In order to define a gloss space, perceptually meaningful axes were necessary. To come up with these axes, an experiment was performed where subjects were shown a pair of images with varying surface parameters and asked to indicate how similar the apparent gloss was. See Fig. (**??**).

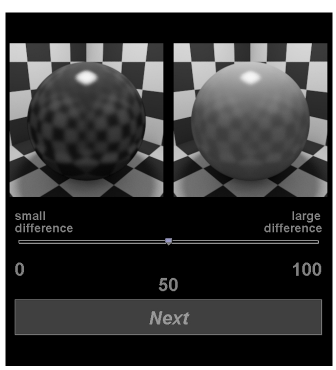

Figure 1: Experiment 1 - Defining the dimensions of gloss space

The material properties for the spheres shown in the images were based on measured BRDFs of a wide range of latex paints. These BRDFs were then approximated by isotropic Ward reflection parameters and rendered. The Ward model has three parameters that define the surface reflection - diffuse reflection(ρ_d),

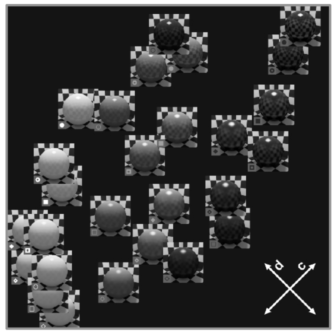

Figure 2: Experiment 1 results

specular reflection(ρ_s), and highlight size(α). Each of these parameters was given three different values to make a set of 27 images.

After having the viewers in the experiment decide on the relative similarity of the pairs of images, a distance between each pair of images could be calculated. With that data, a multidimensional scaling algorithm was used to sort all the data points in 2D, preserving the distance between all points. Once that was plotted, two meaningful axes were determined, c and d. The c axis is the contrast gloss(or the contrast of the reflected image), while d is the distinctness-of-image gloss(or the sharpness of the reflected image). See Fig. (**??**).

2.2 Scaling Gloss Space

The next step to formally defining gloss space was coming up with a way to scale visual gloss. A second experiment was performed where subjects were shown three images - one with no gloss, one with full gloss, and one test image that was somewhere in between. They were asked to decide what the gloss percentage was, based on the two base images. See Fig. (**??**).

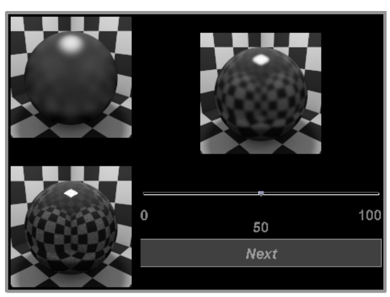

Figure 3: Experiment 2 - Scaling gloss space

The gloss parameters c and d were then defined in terms of ρ_d, ρ_s, and α by finding a best fit to the standard Ward model.

$$c = \sqrt[3]{\rho_s + \frac{\rho_d}{2}} - \sqrt[3]{\frac{\rho_d}{2}} \tag{1}$$

$$d = 1 - \alpha \tag{2}$$

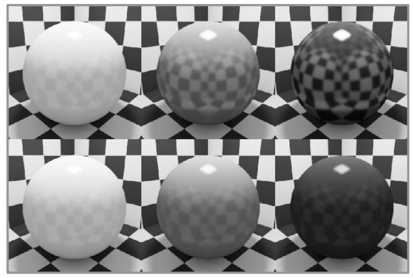

Figure 4: Gloss matching
Top row: same physical gloss parameters ($\rho_s = 0.099$, $\alpha = 0.04$)
Bottom row: same visual gloss parameters ($c = 0.057$, $d = 0.96$)

2.3 Results

With a perceptually based gloss space, working with gloss parameters becomes much easier. The main usability implications are:

- Gloss Matching - Maintain constant gloss while changing material colors (Fig. (**??**))
- Uniform Gloss Space - Change in distance equals change in appearance
- Just-Noticeable-Differences in Gloss - Identify JNDs as ρ_d, ρ_s, and α change
- Unified Material Parameters - Color and gloss specified together as five variables: (L, a, b), (c, d)

In addition to these benefits, gloss space has several other applications:

- Real Time Rendering - Perceptually based triage
- CAD/CAM Rendering - Materials change perception of shape
- Image Based Rendering - Amount of detail needed for highest visual quality
- Computer Vision - Aid in determining properties of objects (Fig. (**??**))

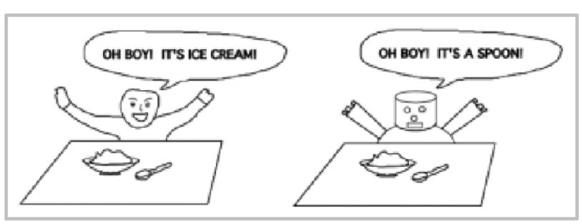

Figure 5: Computer vision needs all the help it can get

3 Rendering: Perceptually Based Rendering

Rendering is generally computationally expensive. However, perception can help speed it up by taking advantage of the limits of human vision.

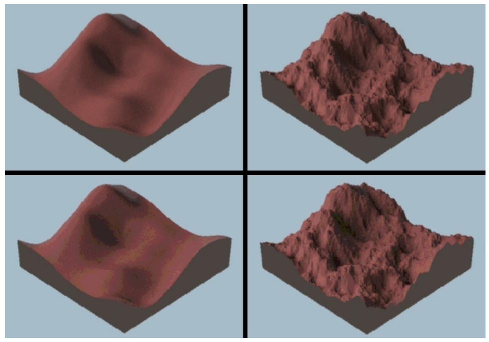

Figure 6: Masking on rough surface

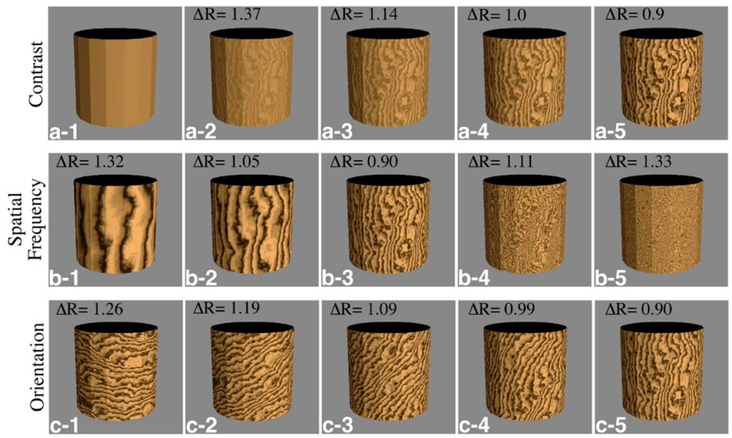

Figure 7: Contributing factors in masking

3.1 Masking

One such limit that can be used is the effect known as masking. Masking happens when rendering a rough or noisy surface. It can help disguise quantization effects that would be very noticeable on a smooth surface. The higher the frequency of the surface, the more it hides the underlying pattern. See Fig. (**??**).

Masking is also more effective under the following conditions(See Fig. (**??**) for illustration):

- High mask contrast
- Spatial frequencies of mask and artifacts are similar
- Orientations of mask and artifacts are similar

Masking can be a very big help when doing perceptually based rendering. After computing the direct lighting in a scene, the renderer can build a threshold map that indicates what areas of the scene have high levels of masking. The indirect lighting computations can then be simplified by putting more effort into areas that will not benefit from masking. This can also be used to help in interactive systems by predicting an appropriate level of detail for objects being tessellated.

3.2 Visible Difference Predictors vs. Functional Difference Predictors

Visible Difference Predictors(VDP) measure raw differences in images. This can be helpful in many cases, but one large problem with VDPs is that they don't take into account perceptual meaning. Functional Difference Predictors(FDP), on the other hand, try to measure the perceptual differences between images. In trying to define FDPs, many questions arise, such as how does one measure image quality and errors in images? There can be many different kinds of errors and artifacts in images due to the way they were rendered, but an observer may not know which one is correct. Errors may also heavily depend on the user's task.

In an attempt to categorize certain errors and work towards defining FDPs, an experiment was conducted where subjects looked at images containing errors in reflections. Two images were visible at the same time, one with a contrast error and one with a projective error. The subjects were all asked four questions: are the images different, which one is correct, do they have the same material, and are the objects in the same location. The results of the experiment showed that the viewers were very good at deciding whether or not two images are different. However, they were not so good at deciding which one was correct. They regularly decided that the two objects had different material properties in the case of contrast errors, but not in the case of projective errors. Conversely, projective errors made it seem that the objects were in different locations, but contrast errors did not have the same effect. See Fig. (**??**).

4 Imaging: Visually Predictive Imaging

4.1 Tone Reproduction

The range of luminances in the world is vastly greater than that displayable by monitors and other display devices. Tone reproduction is a way to reduce the huge dynamic range of a real-world scene down to that of a monitor, while at the same time trying to preserve the appearance of the image.

Examples of various dynamic ranges:

- Real world: 100,000,000:1
- Single scene: 10,000:1
- Display device: less than 100:1

4.1.1 Tone Reproduction Operator

There are many ways to map scene luminance values to colors on a display. A tone reproduction operator is an algorithm that implements one of these methods. The ultimate goal of all tone reproduction operators is to produce a visual match between a real-world scene and a corresponding image shown on a display device. See Fig. (**??**).

4.1.2 Visual Adaptation

In an attempt to accurately reproduce human vision, many tone reproduction operators include models of visual adaptation, the way our visual system reacts to differing levels of light. Human vision can be divided

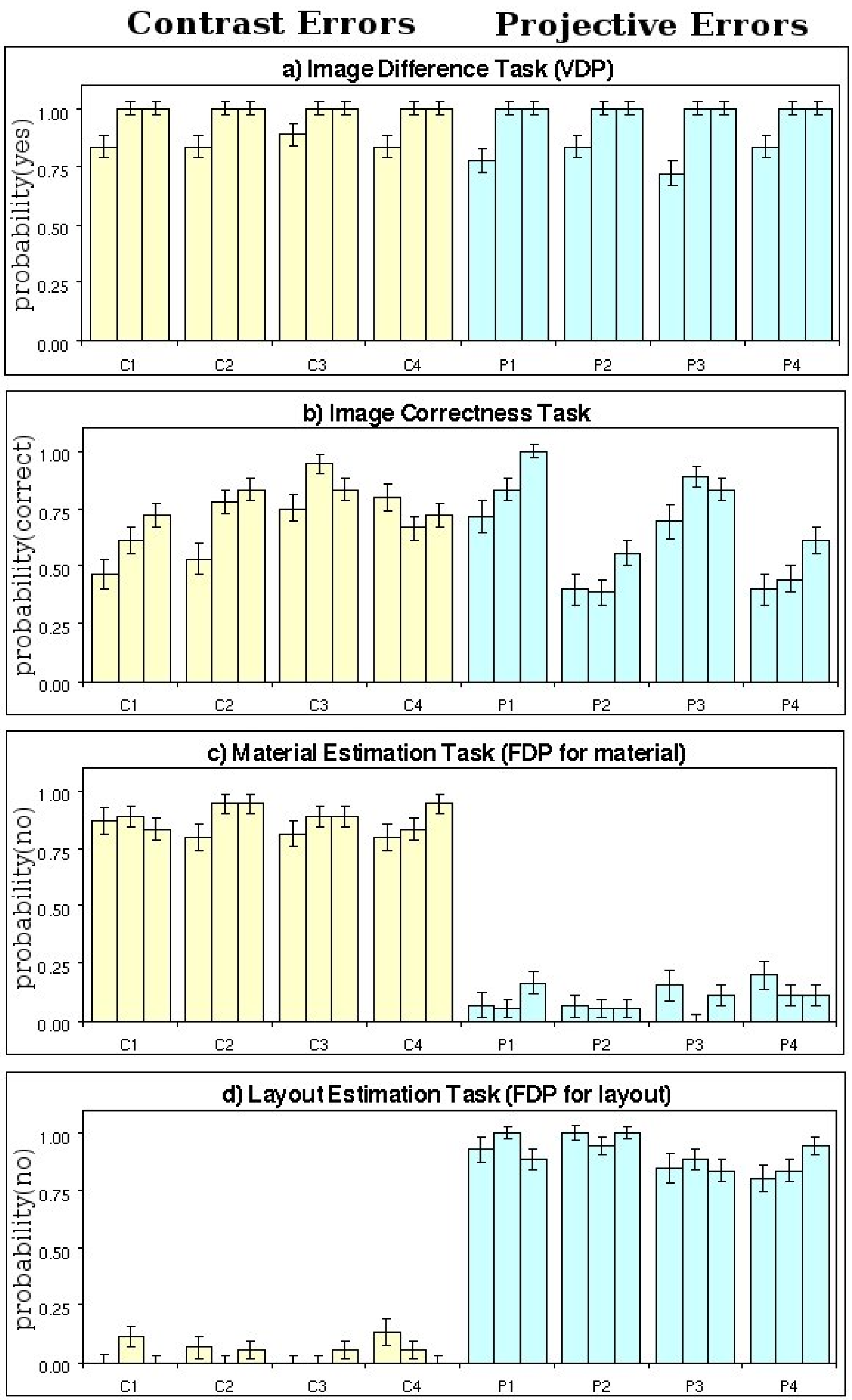

Figure 8: FDP experiment - Image error detection results

Figure 9: Tone reproduction of high dynamic range scene
Left: Pictures taken at varying exposures
Top Right: Direct linear mapping
Bottom Right: Visual mapping

into three main ranges: scotopic, mesopic, and photopic. The scotopic range is used in very low lighting conditions, such as starlight, when the rods are the dominant sensors in our eyes. It is characterized by poor visibility, no color vision, and low acuity. Photopic is in the range of indoor lighting and sunlight. The dominant sensors are the cones. The effects of this zone are just the opposite of those in the scotopic range; here we have good visibility, color vision, and acuity. Mesopic is a transition zone between scotopic and photopic and falls between moonlight and indoor lighting. Tone reproduction operators can use this information to change things like color and sharpness of images in order to convey the overall amount of light in a scene. Many tone reproduction operators also include models of glare to obscure things around very bright light sources in the image.

4.2 Low Vision

Low vision is a term that groups all kinds of uncorrectable vision impairments. These impairments can affect acuity, visual fields, contrast, adaptation, motion, searching abilities, and more. There are also many different causes such as trauma, disease, and aging. Imaging and tone reproduction can be used to assist people with many forms of low vision.

4.2.1 Simulating Effects of Aging

As people get older, their vision naturally starts to get weaker. They experience higher amounts of glare and longer adaptation times. By simulating these effects in computer visualizations, it can help people with stronger vision understand exactly what an older person sees. This can result in things like building designs that are more easily navigable by older people as well as lots of other safety implications.

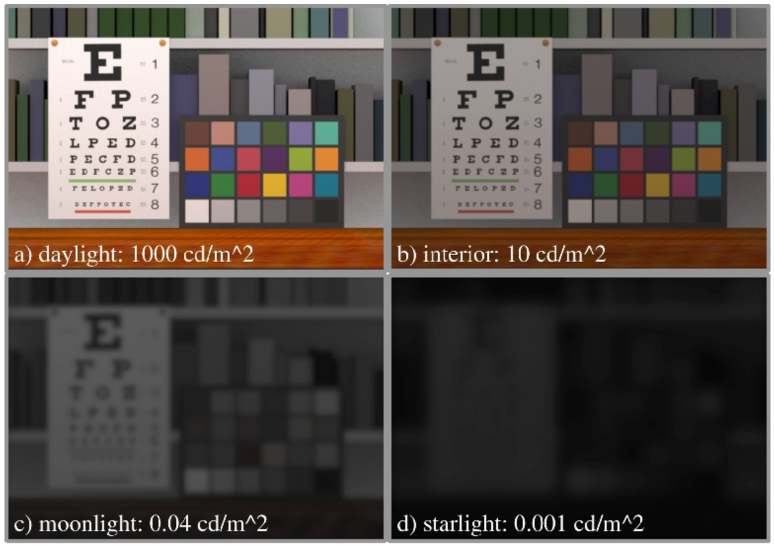

Figure 10: Appearance across range of luminances

4.2.2 Improving Accessibility

With the large amount of images now present on the world wide web, it can be difficult for people with low vision to use all the available resources. By creating models of different impairments, it's possible to enhance images in ways that can help make the image appear almost the same as the original image appears to someone with normal vision. In addition to only processing static web graphics, this technology can also be extended to text and dynamic media.

4.2.3 Vision Testing

There are many children who have vision impairments that go untreated because of a lack of vision testing. One way to help make vision testing for children easier is to integrate it into video games. As a proof of concept, a game called "Dalton's Jungle" was created. The game works by displaying a background of different circles within a certain color range. It then changes the color of some part of the image over time. The color changes are done in the shape of an animal and children are asked to find these animals as quickly as they can. Depending on whether or not the child can see the animal determines what type of color blindness he or she may have. This concept can also be applied to many other types of impairments.

4.2.4 Visualization For The Blind

Another large accessibility problem is for totally blind people needing to interact with images on a computer. This brings up the question of whether or not it is possible to render graphics in a non-visual form. One way to do this is to use audio ques, such as pitch or synthetic speech, while the user is moving a pointer over the image. The problem with this approach is that it is difficult to get an overall impression of the image because there is only a small window that can be explored at one time. A second option is to use a tactile interface to give the user a broad overview and then let them use a pointer to get more detailed information.