Dimensionality of Visual Complexity for Computer Graphics Scenes

Ganesh Ramanarayanan[a], Kavita Bala[a], James A. Ferwerda[b], and Bruce Walter[a]

[a]Cornell University, Ithaca, NY, USA
[b]Rochester Institute of Technology, Rochester, NY, USA

ABSTRACT

How do human observers perceive visual complexity in images? This long-studied problem is especially relevant for computer graphics, where a better understanding of complexity can aid in the development of more advanced perceptually based rendering algorithms. In this project, we describe a study of the dimensionality of visual complexity in computer graphics scenes. We present an experiment where subjects judged the relative complexity of 21 high-resolution building, room, and tabletop scenes, rendered with photorealistic methods. Scenes were gathered from web archives and varied in theme, number and layout of objects, material properties, and lighting. Multidimensional scaling of pooled subject responses embeds the stimulus images in a two-dimensional space, with axes of "numerosity" and "material / lighting complexity". In a follow-up analysis, we derive a one-dimensional complexity ordering of the stimulus images and show discrepancies between this ordering and several computable complexity metrics, such as scene polygon count and JPEG compression size.

Keywords: Visual complexity, computer graphics, perceptually based rendering, multidimensional scaling

1. INTRODUCTION

How does one evaluate the visual complexity of an image? What are the main components of visual complexity? While these questions are very easy to pose, they are notoriously difficult to answer because the space of possible images is so vast, and the higher levels of coding in our visual system are not well understood. Previous work in this space has ranged from theoretical frameworks of structured information[1] to understanding the perceptual building blocks of images[2] to studies of specific image classes, where researchers have proposed many important factors in visual complexity, such as numerosity, symmetry, and so on.[3–5]

Aside from deepening our understanding of human vision and cognition, visual complexity is also useful to study in the context of computer graphics. Unlike textures and natural images, which are simply photographed from the real world, graphics renderings are produced entirely from 3D scene descriptions, or models, stored on a computer. For very complex scenes, rendering is an expensive process that can often take days on clusters of machines. Many avenues of optimization are being aggressively explored, one of which is human perception. Researchers have made progress in exploiting the limits of the human visual system to speed up rendering, both in low-level[6] and high-level[7] contexts. In either case, algorithms rely on the human visual system's inability to efficiently code certain kinds of complexity. By understanding visual complexity and how it relates to 3D scene descriptions, we can take steps towards more powerful and efficient perceptually based rendering algorithms.

In this paper, we take some steps towards deriving perceptually meaningful axes of visual complexity for computer graphics scenes. We first present a stimulus set of of 21 realistic graphics-rendered scenes collected from the academic literature and ray tracing competitions. We then discuss a simple experiment where users compared pairs of images in terms of their complexity, resulting in an MDS (multi-dimensional scaling) analysis of complexity space with two main axes: **numerosity** and **material / lighting complexity**. Finally, we show how human evaluations of complexity on this data set correlate with image-based metrics like JPEG compression and the Rosenholtz visual clutter metric,[8] and scene-based metrics like polygon and light counts, concluding with some possible directions for future work.

Author e-mails: graman@cs.cornell.edu, kb@cs.cornell.edu, jaf@cis.rit.edu, bjw@graphics.cornell.edu

2. RELATED WORK

At a basic level, complexity occurs when something is difficult to describe, explain, or record; in other words, it is tied to information theory and coding. Early work by Leeuwenberg in structural information theory[1] laid a rigorous foundation for the perceptual coding of patterns, based on notions of structural regularity such as alternation and symmetry. This work has been extended in various directions[9] to derive predictive complexity metrics for various kinds of pattern recognition.

While the insight provided by coding-based approaches is invaluable, there is a large gap between the patterns and strings tested in structural information theoretic papers, and the richness of realistic imagery. Biederman's theory of recognition-by-components[2] took a step towards conceptualizing image pattern recognition and understanding by focusing on the objects in these images. He proposed a set of generalized shapes, called *geons*, that humans use to parse arbitrary scenes. Complexity can then be modeled and measured in terms of these shapes, as is done by Patel et. al. [10] There has also been work on raw 3D shape complexity metrics in the context of architecture.[11]

Shape complexity, despite these advances, is still a difficult problem in its own right, and challenging to extend fully to visual complexity of images. Thus, another natural avenue of attack is to directly conduct psychophysical studies on image perception. In particular, significant work has been done on the perceptual dimensions of image understanding. Rao and Lohse[3] looked at a subset of the Brodatz texture album and tried to understand the important perceptual dimensions for these textures. Subjects were asked to group textures in terms of their perceived similarity, and the authors used these groupings to derive a similarity matrix for multidimensional dimensional scaling (MDS) analysis. This analysis placed the textures in a 3D space with axes they identified as repetitiveness, orientedness, and complexity. In later work, Heaps and Handel[4] found they could duplicate the Rao and Lohse results on the Brodatz textures, but not on the MIT VisTex texture album. They also found that subjects were not consistent at proposing axis names for MDS solutions, nor could they consistently rank images according to proposed dimensions (such as repetitiveness).

Recently, Oliva et. al.[5] looked specifically at visual complexity in the context of indoor photographs, using a hierarchical grouping task to derive a similarity matrix and perform MDS analysis. The results placed the images in a 2D space, mostly along an axis corresponding to number / variety of objects. We adopt similar goals as Oliva et. al., in the context of graphics scenes. Our work is very similar in spirit and approach to that of Rogowitz et. al.,[12] who focused on image similarity. They conducted a thorough investigation and comparison of various algorithmic and perceptual measures of similarity, including multidimensional scaling analyses.

3. EXPERIMENT DESIGN

We are interested in studying complexity for computer graphics scenes, and we would like to use MDS analysis to discover perceptually meaningful axes in complexity. To this end, we designed an experiment to obtain dissimilarity measures for a representative yet manageable set of graphics renderings. First we describe the process of selecting our stimulus set, and next we describe the main experiment.

3.1 Stimulus set

A large body of work in the human vision and perception literature focuses on perception of the real world, either directly or through photographs. Computer graphics images, by comparison, are synthetic in nature and can range from completely unrealistic to almost photorealistic. In this experiment we decided to focus on the perception of realistic imagery. We sought out a set of high-quality graphics scenes, rendered with realistic, accurate rendering algorithms, to use for our experiment. We gathered a set of images from the Web, drawing primarily from academia and raytracing competitions, to build an initial set of 100 images. We then examined each image closely and eliminated those with rendering artifacts or low resolution. This resulted in a set of 40 images.

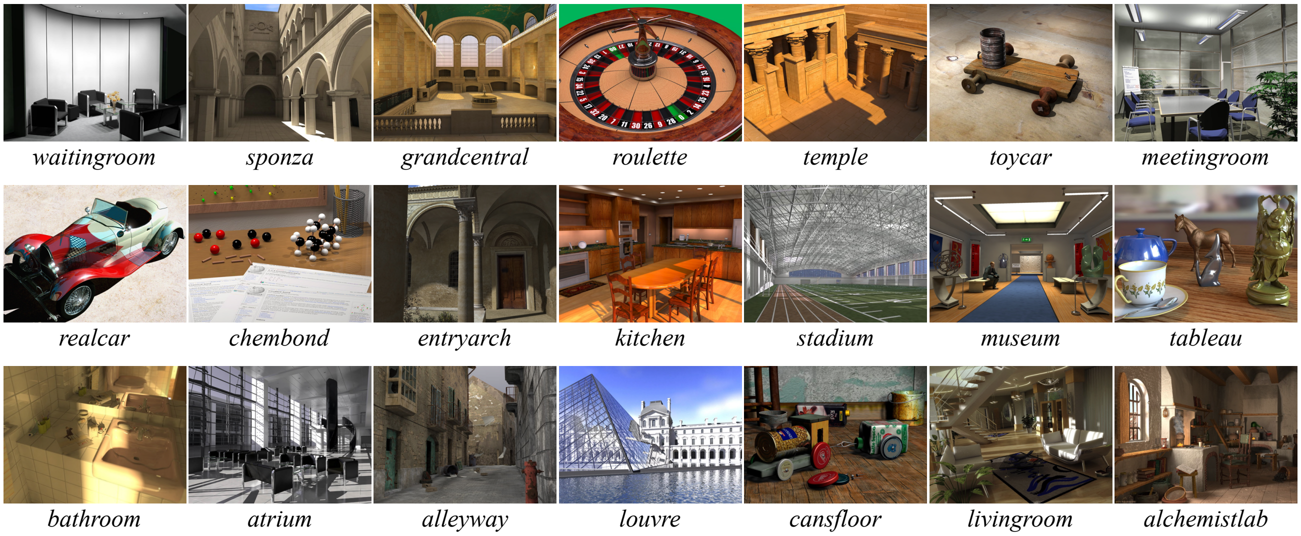

Figure 1. The final stimulus set in our study, ordered with respect to complexity as determined by the final experiment analysis. See Section 4.2.2 for details.

3.1.1 Ranking task for additional pruning

To get some initial feedback about our 40 image dataset and understand how it spanned the space of complexity, we ran several pilot ranking tasks. In these tasks, subjects were given $6'' \times 4\frac{1}{2}''$ high-quality dye-sub printouts of the images, mounted on foam board for easy handling, and asked to order them in terms of visual complexity. The term 'visual complexity' was not defined; subjects were instructed to use whatever definition of visual complexity they believe applied.

To keep the initial ranking tasks manageable, we split the 40 images into sets of 13-14 images based on scale: 2 sets for building and room-sized scenes, and 1 for tabletop scenes (nearby object occupies majority of the image). Using the ranking results, we selected 21 images to cover a broad range of complexity, while also representing various types of geometries, materials, and lighting. This is our final stimulus set, showin in figure 3. We also combined these images into a final pilot ranking task to ensure that subjects did not have difficulty comparing images across scales.

3.2 Experiment design

Our goal is to determine the perceptual distance, in complexity space, between each pair of images in the stimulus set. We designed a simple experiment to measure these values for all image pairs. As with the ranking task, subjects were instructed to evaluate 'visual complexity' using whatever definition they believe applied. 21 subjects participated in the experiment, which was conducted using a web interface on a 15" Mac Powerbook G4. Images were 480×360 pixels and presented on a neutral gray background.

3.2.1 Familiarity with stimulus set

The initial part of the experiment consisted of showing the subject all 21 images present in the study. The purpose of this part was to give users some familiarity with the stimuli so they could develop a model of what visual complexity meant for these images. Subjects saw the images in slideshow fashion, one at a time, for as much time as needed (subjects could click a "Next" button when they were ready to see the next image). This phase typically took a few minutes per subject.

3.2.2 Pairwise Comparisons

The main part of the experiment consisted of a series of trials where subjects were shown a pair of images and asked to provide dissimilarity measures in complexity space. What is the correct way to ask subjects to perform this task? During pilot sessions, we found that subjects were not comfortable answering the question, "How

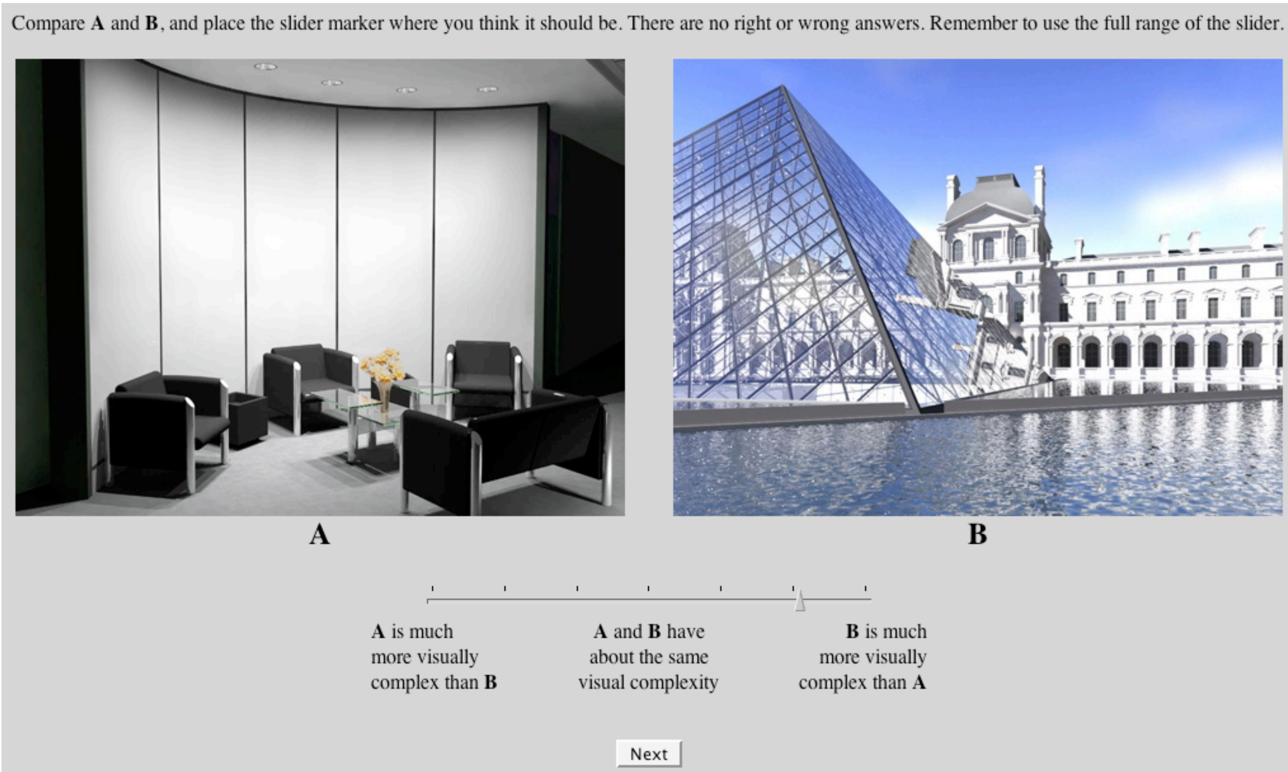

Figure 2. User interface for the experiment. Subjects compared every pair of stimulus images, using a slider to indicate which image they thought more complex, and by how much.

similar are these images in terms of complexity?" Thus, we presented subjects with the more natural task of picking which image they felt was more complex. The user interface is shown in Figure 3.2. For each pair, one of the two images was randomly designated as image A, and the other was designated as image B. Subjects recorded their responses a slider underneath the two images. The slider was marked as follows:

- left: "A is much more visually complex than B"
- center: "A and B have about the same visual complexity"
- right: "B is much more visually complex than A"

The slider was fully continuous; any placement of the bar on the slider resulted in a valid response. The slider was also marked with 7 evenly spaced ticks, one at the far left, one at the center, and one at the right, with two between each. This was to provide users with an easier-to-use discrete scale if desired. After interacting with the slider, subjects clicked a "Next" button to advance to the next trial. Subject responses were recorded by noting the position of the slider for each trial pair, with the leftmost part of the slider corresponding to -10, and the rightmost part to +10.

To obtain a data point for each image pair requires 210 trials. In addition to this, subjects completed extra trials in the beginning and end of the experiment. The first 11 trials were training trials for subjects to get accustomed to the task. These trials were clearly identified as training, and subjects were informed that their responses to these trials were not being used. The pairs in the training trials were predetermined (there are 11 of them so that each of the 21 stimulus images is represented at least once). Following this, each of the 210 pairs was presented in random order. Finally, at the end, 10 additional trials were performed with 10 pairs randomly selected from the full set of 210. These 10 "duplicate" tests were used to check the consistency of subject responses for the experiment. Subjects were not informed of the duplicate trials and did not ever notice them. In total, there were 231 trials .

The experiment was run on 21 subjects, graduate students in various fields unrelated to computer graphics or human vision. The entire experiment took half an hour on average for each subject.

4. ANALYSIS AND RESULTS

At the conclusion of the experiment, we obtained a complexity distance matrix for each of the 21 subjects. In addition, for each subject, we had 10 duplicate trial measurements we could use to check subject consistency. In this section, we detail the analyses applied and the results obtained.

4.1 Preliminary analyses

4.1.1 Subject consistency

First we looked at the subjects' ability to answer in a consistent fashion when presented with the same image pair twice. For each subject, we recorded duplicate responses for 10 random image pairs, and compared the duplicate responses to their original responses. We gave subjects credit for consistently answering for a given pair if the responses were within one standard deviation of each other (standard deviations were computed per subject). Over all subjects and all duplicate trials, we found that responses were 77.6% consistent with respect to this measure.

4.1.2 Data normalization

Next, we normalized the data across all subjects to prepare for MDS analysis. For every pair of images, each subject has a recorded response between -10 and 10, with negative numbers indicating A is more complex, and positive numbers indicating B is more complex. Since all subjects used different ranges of the scale (some used the extremes, while some never went beyond the halfway point in either direction), each subject's responses were normalized with respect to their mean magnitude. Specifically, let $D(s)$ be the matrix of responses for subject s, with entry $d(s)_{ij}$ indicating the result of his comparison of the images with index i and j. The mean magnitude $mag(s)$ is given by

$$mag(s) = \sum_{i}^{n} \sum_{j>i}^{n} \frac{2d(s)_{ij}}{(n(n+1))}$$

and the normalized response matrix $ND(s)$ is given by

$$nd(s)_{ij} = \frac{d(s)_{ij}}{mag(s)}$$

4.2 MDS analyses

In this section we describe two multidimensional scaling analyses applied to the subject responses. For both analyses, we used the PERMAP software,[13] an interactive tool supporting various forms of metric and non-metric scaling (all results in this paper are metric MDS). Using PERMAP's features for jittering, parking, and otherwise manipulating individual points in proposed scaling solutions, we were able to verify the stability of the solutions we obtained.

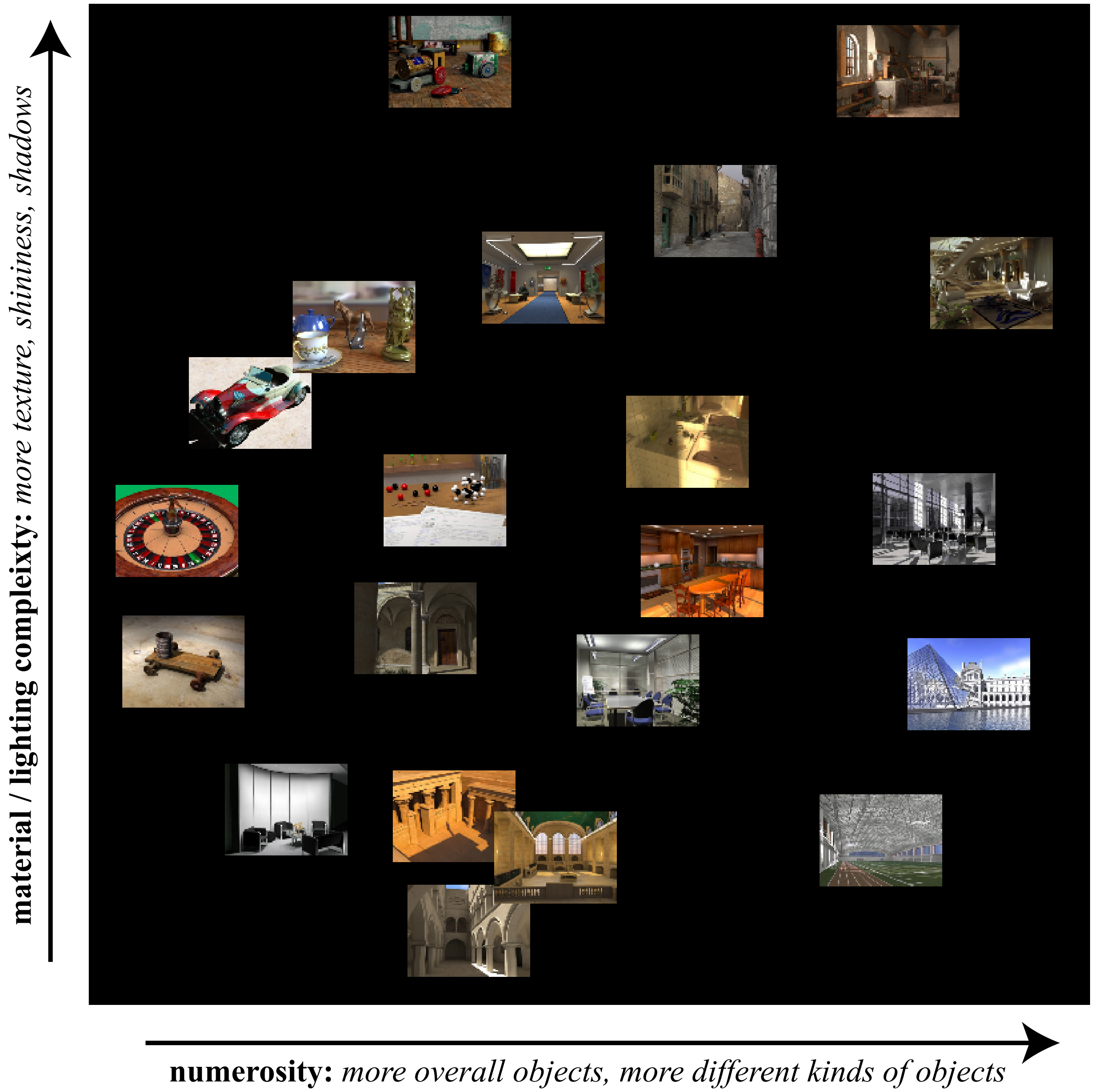

Figure 3. The 2D multidimensional scaling of the stimulus set. The horizontal axis corresponds to numerosity, ranging from single objects on the left, to huge buildings and cluttered rooms to the right. The vertical axis corresponds to material and lighting complexity, ranging from scenes with a few uninteresting materials, to complex scenes with rich textures and interesting lighting.

4.2.1 Scaling based on magnitude of dissimilarity

Recall that our normalized subject dissimilarity judgments $nd(s)_{ij}$ are signed, additionally encoding information about which image the subject perceived to be more complex. In our first analysis, we will treat these values purely as absolute distances. We first sum $|nd(s)_{ij}|$ across all subjects, obtaining a dissimilarity matrix $MagD$, and we find the knee of the stress curve at two dimensions.

The resulting 2D embedding is shown in Figure 4.2. The horizontal axis corresponds to **numerosity** in the stimulus images; in moving from left to right, the scenes depicted have more and more objects. For instance, on the left side of the space we have scenes with very few objects, such as *toycar* and *roulette*. In the middle, we have scenes such as *museum* and *meetingroom* which have many objects in view. Finally, on the right we scenes suggesting uncountably numerous objects, such as metal bars in *louvre*, and multiple vials / containers in *alchemistlab*.

The vertical axis we find corresponds to **material and lighting complexity**. Moving from bottom to top, the scenes depicted show richer materials, either in the form of glossiness or texture, and also more varied lighting with interesting shadows. For instance, *sponza* and *stadium* find themselves in the bottom half of the space, with their relatively plain materials and lighting (*sponza* is shadowed, but the effect is not as striking as in *temple*, which is just above it in the MDS result). In the middle of the space we have scenes such as *realcar* and *atrium*, marked by striking shininess and shadows, respectively. Finally, at the top of the space we have images such as *cansfloor* and *alleyway*, dominated by rich textures.

It is interesting to note that the vertical axis does not simply arise from color variation in the stimulus images. One indicator of this is the *bathroom* image, which, while largely yellow, has interesting lighting, a mirror, and glossy tiles / sink, and accordingly resides in the middle of the space. It is also interesting to note the effect of scale; most of the large scenes (*temple*, *grandcentral*, *stadium* etc.) find themselves at the bottom of the space, because material and lighting complexity are very difficult to convey when squeezing such large scenes into a small image. We fully expect that closeups of these scenes would be evaluated differently, and this is an interesting direction for future work.

From a graphics standpoint, the identified axes have a direct relationship to how models are specified: graphics scenes are described in terms of geometry, materials, and illumination. Geometry is naturally related to the numerosity axis, and the material / lighting variations make up the other axis. However, it is difficult to draw any stronger conclusions of dimensionality from this experiment; for instance, it is not clear which of shininess, texture, lighting, or shadows is the greatest indicator of complexity in the vertical direction in our MDS result, and this is another interesting direction for future work.

(a) (b) (c)

Figure 4. Consensus on various image pairs. (a) Subjects uniformly judged the left image in these pairs to be more complex; accordingly, a line drawn between the images in these pairs has a positive slope in 2D complexity space. (b) Subjects did not strongly select either image as more complex, indicating their similarity both in complexity and content; thus, these images pairs are very close in 2D complexity space. (c) Subjects were divided on these pairs because each image has a different kind of complexity; accordingly, a line drawn between the images in these pairs has a negative slope in 2D complexity space.

4.2.2 Scaling based on dissimilarity consensus

In our first analysis, recall that we ignored the sign of subject responses when pooling to obtain an all-subject dissimilarity matrix. What happens if we reintroduce this sign? Figure 4.2.1 shows some example situations. In Figure 4.2.1(a), when subject judgments are almost unanimous, the signs of responses do not provide additional information. Similarly, in Figure 4.2.1(b), when two images are judged to be fairly similar in complexity, again signs don't matter. However, consider the situation in Figure 4.2.1(c), where subjects were strongly divided on which image was more complex. Intuitively, a polarized image pair of this kind is showing how different subjects are reacting to different scene properties to make their complexity judgments, which is evidence of a multidimensional complexity space. If we were forced to pick one image as more complex, we can sum the signed responses of subjects to come up with an answer.

We can use this idea to derive a consensus scaling of the stimulus images. We form the consensus dissimilarity matrix $ConsD$ by summing all subject response matrices $D(s)$ as-is, and computed a one dimensional scaling of this matrix. The images in 3, from left to right and top to bottom, are ordered according to this scaling result. Informally, this result is similar to rankings obtained in our pilot studies; while there was disagreement on the ordering of images in the middle of the scales, the ends of the scale were fairly consistent across most subjects.

5. DISCUSSION

5.1 Comparison with other measures of complexity

We now show how the ordering we just derived in Section 4.2.2 does not match orderings obtained by several computable measures of visual complexity. This is expected, since current algorithmic means of characterizing visual complexity do not take scene understanding into account; instead, they interpret images and scenes in terms of their raw data.

5.1.1 Image-based measures: JPEG file size, feature congestion

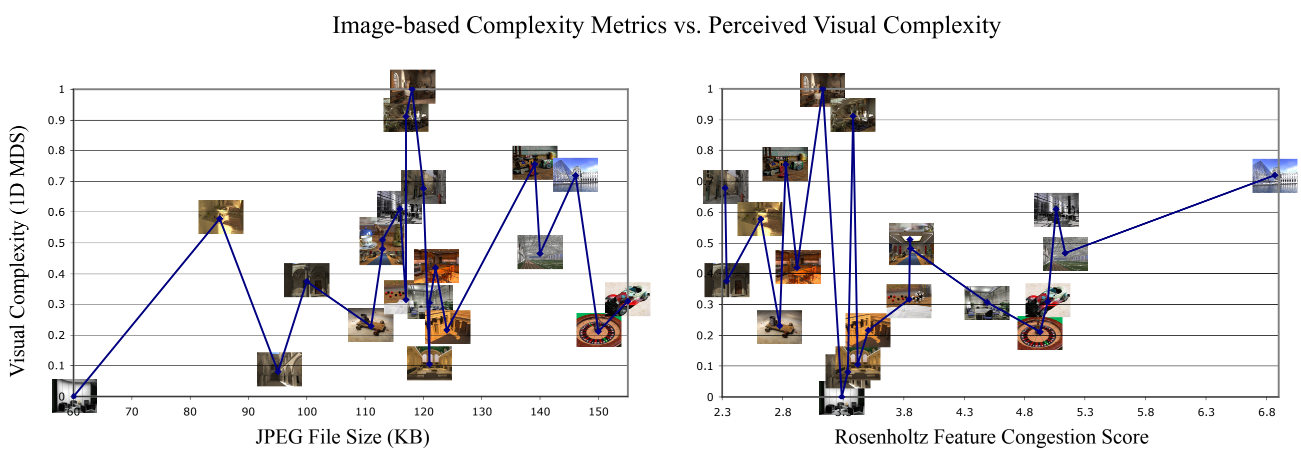

Figure 5. Comparing the experiment-derived 1D complexity ordering with orderings based on JPEG file size and the feature congestion metric.[8] Notice that neither curve is monotonic, indicating the poor correlation between perceived complexity and computable measures.

Visual complexity is somewhat related to frequency content in image compression, or clutter in visual search. To that end, we compared our 1D measure of visual complexity for this data set with two computable measures: JPEG file size, and the Rosenholtz feature congestion metric,[8] respectively. The results are shown in Figure 5.1.1. As expected, neither curve is monotonic.

It is interesting to note which kinds of high frequency correlate and feature congestion. When frequency results from material variation on one or two objects, such as in *roulette* and *realcar*, the large file sizes do not correlate well with visual complexity. However, when the high frequencies result from raw numerosity, there is is correlation, as in *louvre*. Also, when there is material variation through an entire scene, such as in *cansfloor*, file size is a good predictor of complexity. The results for the feature congestion metric are similar, although there appears to be more of a response to numerosity.

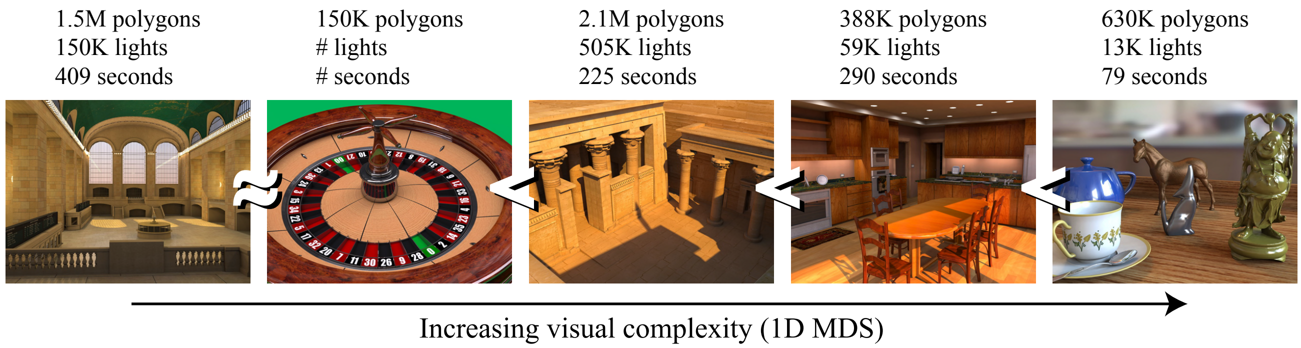

Figure 6. Comparing the experiment-derived 1D complexity ordering with orderings based on graphics scene complexity measures: triangle count, light count, and rendering time. Again, there is little to no correlation between these numbers and perceived visual complexity.

5.1.2 Scene-based measures: triangle count, number of lights, rendering time

In computer graphics, measures of scene complexity are often used as indicators of how difficult it is to render a final image. Common indicators include geometric complexity (the number of triangles in the scene), illumination complexity (the number of lights in the scene), and even raw rendering time. We obtained triangle counts, light counts, and rendering times (using the Lightcuts[14] algorithm) for a subset of our scenes, and compared these with the visual complexity orderings obtained in our experiment. The results are shown in Figure 5.1.2. As with the image-based measures, there is little correlation between visual complexity and these algorithmic indicators of complexity. Looking closely at these images reveals why. For example, *temple* (center), a massive model of 2.1 million polygons, is given the same pixel real estate as *tableau* (right), a much smaller scene of 630 thousand polygons. This makes it much harder to see subtle geometry and texture variations in *temple*. Furthermore, the high polygon count in this scene contributes to several large, countable pillars, not raw numbers of objects as in *louvre* for example. Accordingly, *temple* has understated numerosity and material/lighting complexity despite its graphics scene description. *grandcentral*, another very complex scene, has tremendous detail in lights, numbers of windows, and the railing in the front; however, this detail is so small in screen space that it does not contribute as much to perceived visual complexity as they should.

5.2 Conclusions and future work

In this paper we have presented a simple study of visual complexity for computer graphics scenes, resulting in a two dimensional space characterized by axes of numerosity and material / lighting complexity. Having identified some general areas to focus on, a natural next step is to look more deeply at numerosity, material complexity, and lighting complexity in detail. Does numerosity mean raw number of objects, or unique objects? Which materials are perceived to be more complex? Are reflections or shadows more complex? The advantage of studying these problems in the context of computer graphics is that custom scenes can be modeled and rendered for all of these particular conditions much more easily and consistently than they can be photographed in the real world.

Another interesting area for future work is the issue of scale and moving viewpoint. In the real world, and in interactive graphics applications, we are not restricted to a single view of an environment; if there is complexity in a region, we can focus our attention on it and view it in greater detail. It would be useful to understand how to incorporate this behavior and derive some understading of the relationship between scale and perceived complexity.

REFERENCES

1. E. L. J. Leeuwenberg, “A perceptual coding language for visual and auditory patterns,” *American Journal of Psychology* **84**, pp. 307–349, 1971.
2. I. Biederman, “Recognition–by–components: A theory of human image understanding,” *Psychological Review* **94**(2), pp. 115–147, 1987.

3. A. R. Rao and G. L. Lohse, "Identifying high-level features of texture perception," *Graphical Models and Image Processing* **55**, pp. 218–233, 1993.
4. C. Heaps and S. Handel, "Similarity and features of natural textures," *Journal of Experimental Psychology: Human Perception and Performance* **25**(2), pp. 299–320, 1999.
5. A. Oliva, M. L. Mack, M. Srestha, and A. Peeper, "Identifying the perceptual dimensions of visual complexity of scenes," *The 26th Annual Meeting of the Cognitive Science Society Meeting*, 2004.
6. K. Myszkowski, "Perception-based global illumination, rendering, and animation techniques," in *SCCG '02: Proceedings of the 18th spring conference on Computer graphics*, pp. 13–24, 2002.
7. G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala, "Visual equivalence: towards a new standard for image fidelity," *ACM Trans. Graph.* **26**(3), p. 76, 2007.
8. R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *J. Vis.* **7**, pp. 1–22, 8 2007.
9. E. L. J. L. Peter A. van der Helm, Rob J. van Lier, "Serial pattern complexity: irregularity and hierarchy," *Perception* **21**, pp. 517–544, 1992.
10. L. N. Patel and P. O. Holt, "Modelling visual complexity using geometric primitives: Implications for visual control tasks," in *Proceedings of the 19th European Annual Conference on Human Decision Making and Manual Control*, pp. 3–8, 2000.
11. J. S. Gero and V. Kazakov, "On measuring the visual complexity of 3d objects," *Journal of Design Sciences and Technology* **12**(1), pp. 35–44, 2004.
12. B. E. Rogowitz, T. Frese, J. R. Smith, C. A. Bouman, and E. B. Kalin, "Perceptual image similarity experiments," in *Proceedings of the SPIE: Human Vision and electronic Imaging III*, pp. 576–590, 1998.
13. R. B. Heady and J. L. Lucas, "Permap: An interactive program for making perceptual maps," *Behavioral Research Methods, Instruments, and Computers* **29**, pp. 450–455.
14. B. Walter, S. Fernandez, A. Arbree, K. Bala, M. Donikian, and D. P. Greenberg, "Lightcuts: a scalable approach to illumination," *ACM Trans. Graph.* **24**(3), pp. 1098–1107, 2005.