Dimensionality of Visual Complexity in 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 problem is especially relevant for computer graphics, where a better understanding of visual complexity can aid in the development of more advanced rendering algorithms. In this paper, we describe a study of the dimensionality of visual complexity in computer graphics scenes. We conducted an experiment where subjects judged the relative complexity of 21 high-resolution scenes, rendered with photorealistic methods. Scenes were gathered from web archives and varied in theme, number and layout of objects, material properties, and lighting.

We analyzed the subject responses using a variety of methods. We first looked at a two-dimensional embedding of pooled subject responses using multidimensional scaling analysis. It is not completely obvious how to name the resulting axes, but we found one roughly corresponded to "numerosity" and the other roughly to "material / lighting complexity". Using a related analysis, we also derived a one-dimensional complexity ordering of the stimulus images. We then looked at the individual differences between subjects, using a combination of multidimensional scaling, individual differences scaling, and clustering. We found that each subject seemed to approach the problem of visual complexity differently and had different complexity spaces, but on average, most subjects seemed to evaluate visual complexity using 2 to 3 dimensions.

Finally, we compared our one-dimensional complexity ordering to several computable complexity metrics, such as scene polygon count and JPEG compression size, and did not find them to be very correlated. Understanding the differences between these measures can lead to the design of more efficient rendering algorithms in computer graphics.

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. A large body of literature has attempted to tackle this problem, with approaches ranging 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 explored in graphics, 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, such algorithms rely on the human visual system's inability to efficiently code certain kinds of complexity. For example, previous work has demonstrated the perceptual tradeoffs in substituting texture for complex geometry,[8] a very common and useful practice in computer graphics. By understanding visual complexity overall 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 describe some initial steps we took to explore visual complexity in computer graphics scenes. We first present a stimulus set of 21 realistic, graphics-rendered scenes collected from the academic literature and ray tracing

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

competitions. We differ from previous work on visual complexity in that we try to look at a more general variety of scenes, such as tabletop scenes of two or three objects, room-sized scenes, and outdoor scenes. For some scenes, in addition to the rendered image we also have the 3D scene description, which we can compare and contrast with other measures of visual complexity.

Next, we describe an experiment where subjects compared pairs of images in terms of their visual complexity. We analyzed the experiment results in several ways. First, we show a two-dimensional embedding of the stimulus images, based on multidimensional scaling of pooled subject responses. The dimensions of this embedding, while difficult to name, roughly correspond to "numerosity" and "material / lighting complexity". We then present a one-dimensional ranking derived from pooled responses. We also discuss additional analyses of individual subject responses, using multidimensional scaling, individual differences scaling, and clustering. While these analyses confirm that subjects did in fact behave differently, they still suggest a space with roughly the same structure as the original two-dimensional embedding, although a larger experiment is necessary to confirm this result holds in general.

Finally, using the one-dimensional ranking of the stimulus images derived from the original data, we compare how human evaluations of visual complexity correlate with image-based metrics like JPEG compression and visual clutter metrics,[9] and scene-based metrics like polygon and light counts. The disparity between these measures points to directions for future work in graphics.

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[10] 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 has been done by Patel et. al. [11] There has also been work on raw 3D shape complexity metrics in the context of architecture.[12]

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).

Our work is similar in spirit and approach to that of Rogowitz et. al.,[13] though their focus is on image similarity and ours is on image complexity. They conducted a thorough investigation and comparison of various algorithmic and perceptual measures of image similarity, and used multidimensional scaling analysis in two (and three) dimensions to identify dominant axes of similarity, such as natural vs. man-made, and more vs. less human-like.

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. Our goals are similar to those of Oliva et. al., though our methodology is different, and our interest is in synthetically rendered graphics scenes, which are more varied in type (tabletop, room-sized, outdoor) and are also associated with 3D descriptions of the geometry, materials, and illumination used to create them.

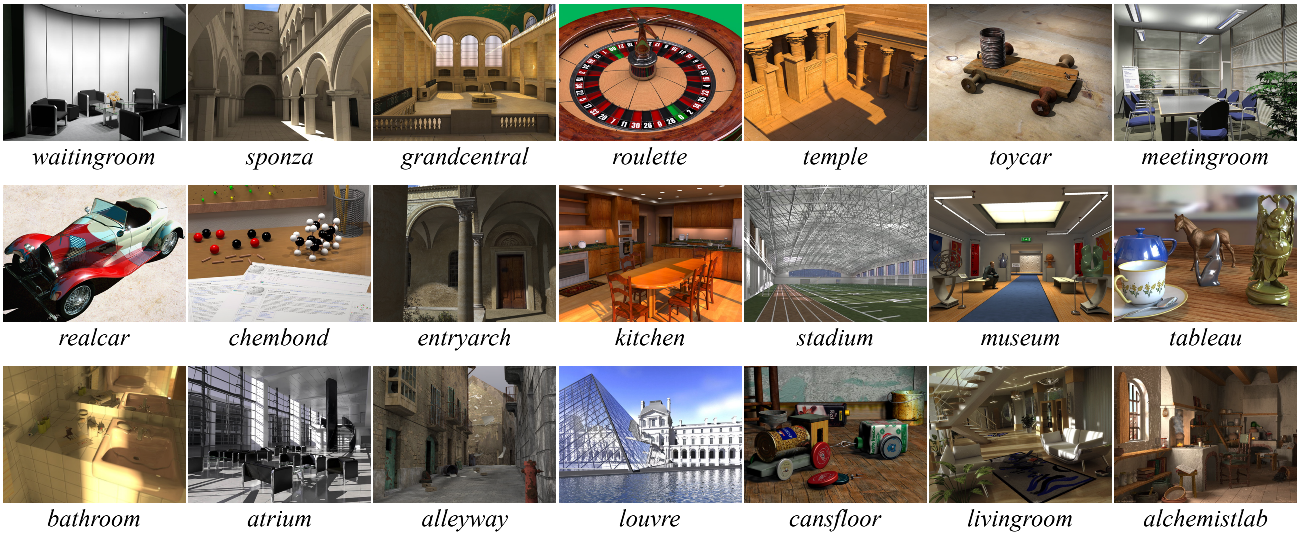


Figure 1. The final stimulus set in our study: 21 images that cover a broad range of complexity, while also representing various types of geometries, materials, and lighting. The images are ordered from top to bottom and left to right according to the one dimensional analysis described in Section 4.2.2.

3. EXPERIMENT DESIGN

We are interested in studying the visual complexity of computer graphics scenes. To this end, we picked a stimulus set of images that spans a range of geometric, material and lighting complexity, while being of manageable size. First, we describe the process of selecting our stimulus set, and next, we describe our main experiment, where we obtained dissimilarity measures of complexity for this set of images. We then describe our multidimensional scaling analysis of subject responses to discover perceptually meaningful axes in visual complexity for graphics.

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 focus on the perception of realistic computer generated 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 such as low resolution. This resulted in a set of 40 images.

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 (300 dpi) 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. These tasks were run with 21 subjects.

To keep these initial ranking tasks manageable, we split the 40 images into sets of 13—14 images based on scale: two sets for building and room-sized scenes, and one for tabletop scenes (where nearby objects occupy the majority of the image). Using the ranking results, we selected a final set of 21 images that are manageable for pair-wise comparisons in experiments, but also cover a broad range of complexity representing various types of geometries, materials, and lighting. This final stimulus set is shown in Figure 1. We also combined these images into a final pilot ranking task to ensure that subjects did not have difficulty comparing images across scales.

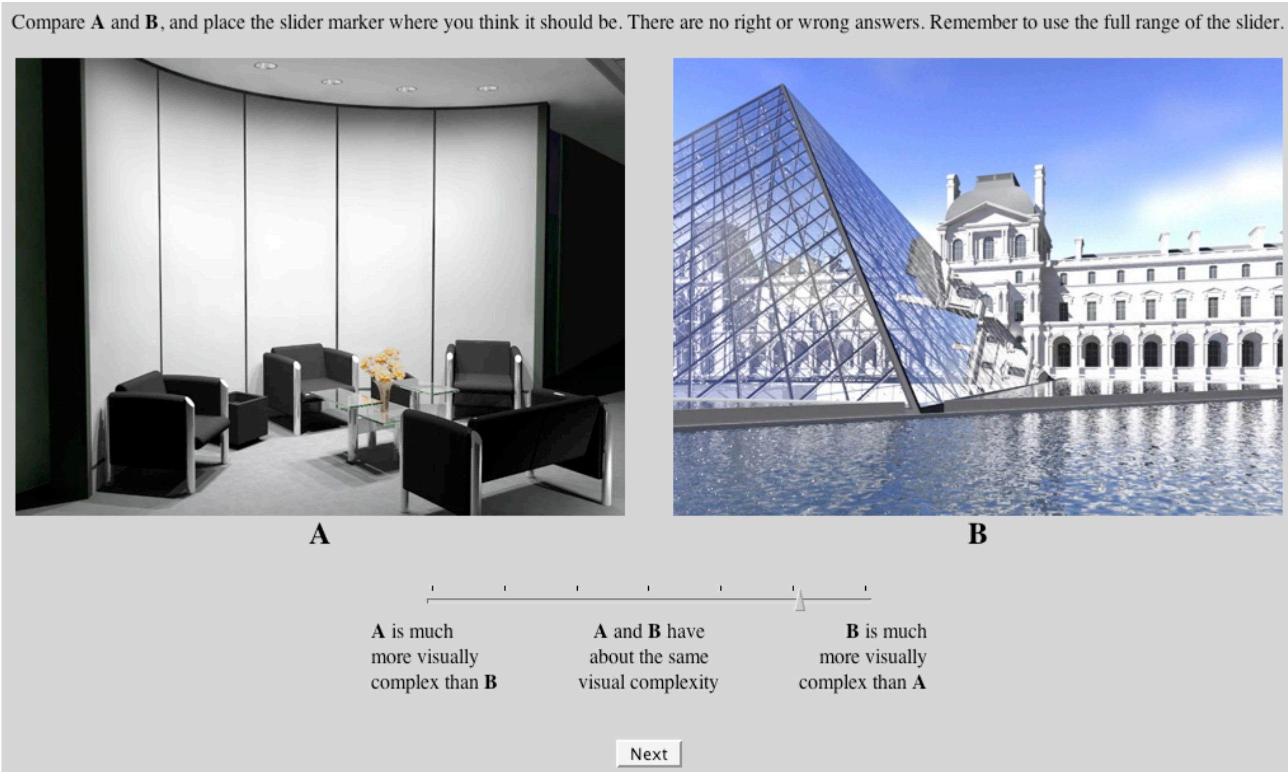


Figure 2. User interface for the experiment. Subjects compared every pair of stimulus images, using a slider to indicate which image they thought was more complex, and by how much.

3.2 Experiment Design

Our goal was 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 different subjects participated in the experiment. The experiment was conducted using a web interface on a 15" Mac Powerbook G4 (screen diagonal 15.25", 1280x854 resolution). Images were 480×360 pixels and presented on a neutral gray background. The viewing distance was 24 inches, and the visual angle subtended by each image was 11.42 degrees.

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 stimulus set we used. 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. There are different ways to ask subjects to perform this task. During pilot sessions, we found that subjects were not comfortable answering the question, "How 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 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 using 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 required 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. The pairs in the training trials were predetermined: there were 11 of them so that each of the 21 stimulus images was 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. 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

In this section, we present the analyses we performed and results we obtained from the experiment data. Section 4.1 describes some basic preliminary analyses we conducted prior to MDS. Section 4.2 then describes the results of several MDS analyses of our stimulus set, including some additional analyses that reveal differences between the individual subjects in the experiment.

4.1 Preliminary analyses

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 randomly selected from the set of 210 stimulus image pairs.

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 compared their 10 duplicate responses to their original responses. We gave subjects credit for answering consistently if their 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, $1 \leq i, j \leq k$ ($k = 21$, the number of images). The mean magnitude $mag(s)$ is given by:

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

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

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

4.2 Multidimensional Scaling Analyses

In this section we describe two MDS analyses applied to the subject responses. For both analyses, we used the PERMAP software,[14] an interactive tool supporting various forms of metric and non-metric scaling (all results in this paper are metric MDS). PERMAP has several interactive tools to explore the stability of a solution, such as jittering, parking, and the ability to manipulate individual points in proposed scaling solutions. Using these features, we were able to verify the stability of the solutions we obtained.

Our first approach ignores the sign of the dissimilarity judgments, and only considers the magnitudes of the subject responses, to compute a two dimensional visual complexity space. Our second approach also take the dissimilarity signs into consideration, which are used to compute a one dimensional ordering on visual complexity.

4.2.1 Scaling based on magnitude of dissimilarity

Our normalized subject dissimilarity judgments $nd(s)_{ij}$ are signed values, 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 then perform MDS on this matrix, finding the knee of the stress curve at two dimensions (see Figure 5). While this is not an extremely prominent knee, we found that results did not change much in 3 or more dimensions; despite the modest improvements in stress shown in Figure 5, the resulting embeddings are simply 'fatter' versions of the 2D embedding.

The 2D embedding is shown in Figure 3. While it is not clear how to definitively identify the two axes, we can observe some general trends that seem to correlate with the axes, suggesting some possible names for them. The horizontal axis appears to be related 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 is more difficult to name than the horizontal axis, but there is a rough correspondence with scene **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 is not just an indicator of the amount of 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.) are spread across the bottom of the space, because material and lighting complexity are very difficult to convey when squeezing such large scenes into a small image. It is possible 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.

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 shows some example situations. In Figure 4(a), when subject judgments are almost unanimous, the signs of responses do not provide additional information. Similarly, in Figure 4(b), when two images are judged to be fairly similar in complexity, again signs don't matter. However, consider the situation in Figure 4(c), where subjects were strongly divided on which image was more complex. Intuitively, a polarized image pair of

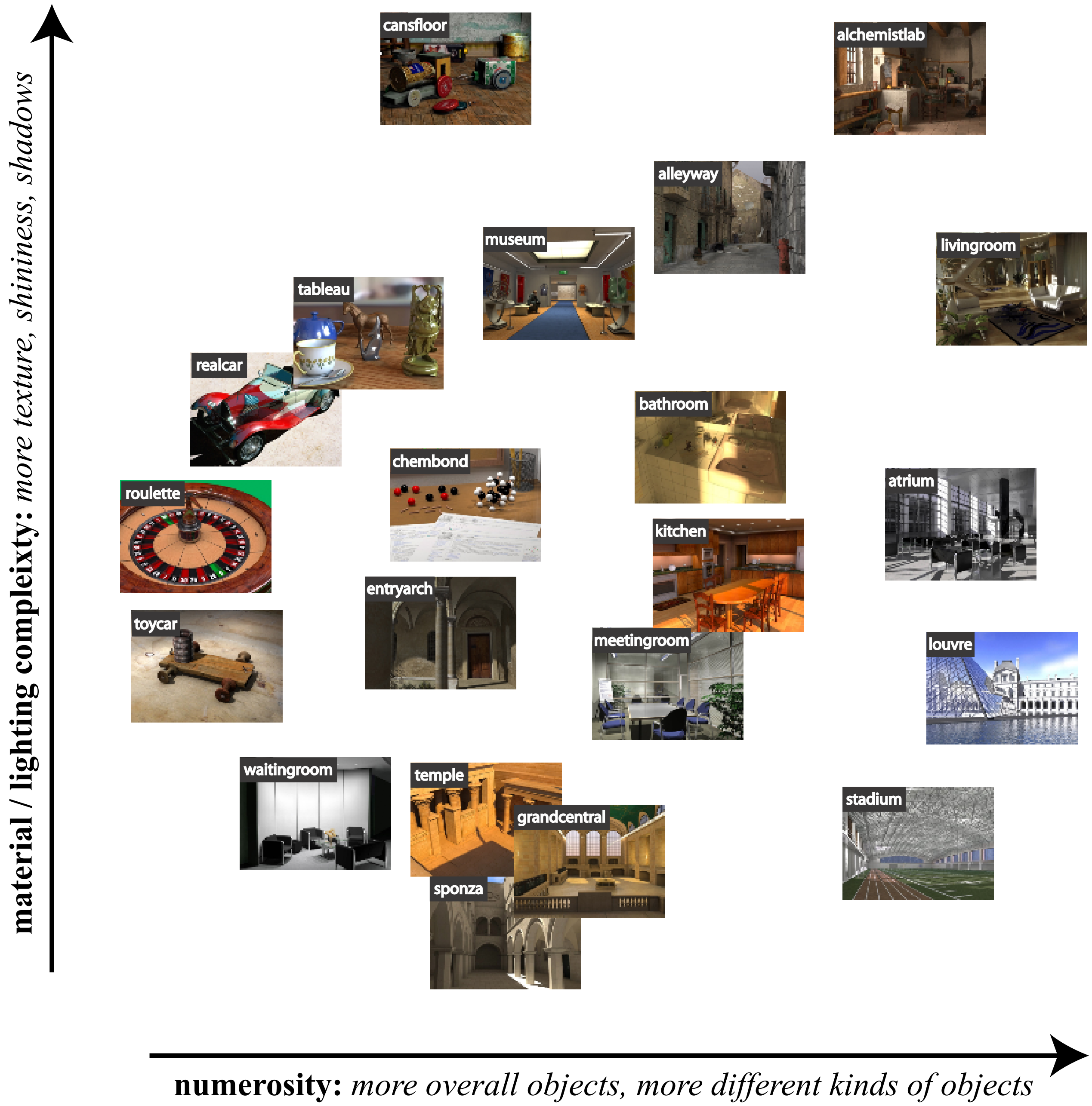


Figure 3. The 2D multidimensional scaling (MDS) analysis of the stimulus set. The horizontal axis roughly corresponds to numerosity, ranging from single objects on the left, to huge buildings and cluttered rooms on the right. The vertical axis roughly corresponds to material and lighting complexity,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

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 $ND(s)$ as-is, and computed a one dimensional scaling of this matrix.

(a) (b) (c)

Figure 4. Consensus on various image pairs. (a) Subjects uniformly judged the left image in these pairs to be more complex. A line drawn between the images in these pairs has a positive slope in the presented 2D complexity space. (b) Subjects did not strongly select either image as more complex, indicating their similarity both in complexity and content. 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. A line drawn between the images in these pairs has a negative slope in 2D complexity space.

The images in Figure 1, 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 some disagreement on the ordering of images in the middle of the scale, near the ends of the scale the rankings were identical.

4.2.3 Stress Curves

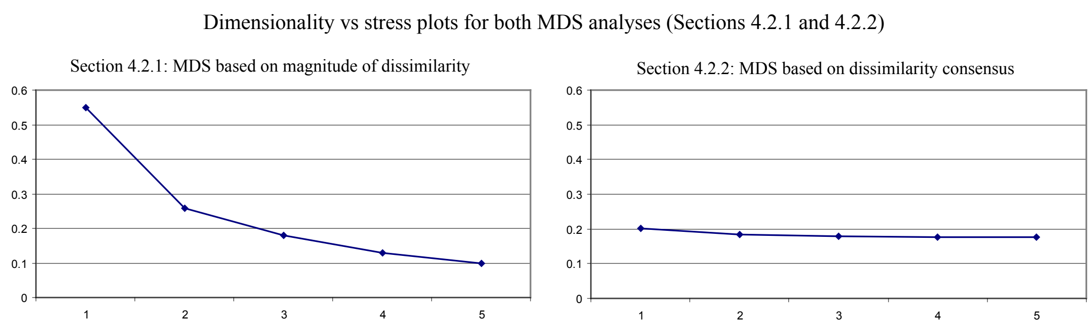


Figure 5. Stress curves for our first two multidimensional scaling analyses.

The goal of MDS is to embed a set of points in a n-dimensional space such that the distance bewteen points in this space corresponds as "closely" as possible to the input matrix of pairwise point dissimilarities. The degree of correspondence between the computed scaling result and the input matrix is measured using a stress function (we used the Kruskal Stress1 function[14]). We would like to find the dimension that has low, ideally zero, stress. Since this ideal situation never occurs in practice without considerable overfitting, a common approach is to compute the stress for each dimension, plot the stress curve, and pick the dimension at which the "knee" of the curve occurs. Figure 5 shows the stress curves for the two analyses. The analysis based on magnitudes alone (Section 4.2.1) has a knee in the curve at two dimensions; a one-dimensional fit is by comparison quite poor. On the other hand, the analysis based on the dissimilarity consensus (Section 4.2.2), despite also having a knee at two dimensions, is considerably flattened to be almost one-dimensional (the corresponding MDS plot, not shown, is a slightly bent crescent, which one can argue is a 1D manifold). This matches our intuition to use 1D scaling from the dissimilarity consensus analysis, rather than the magnitude analysis, to derive a one dimensional ordering of the stimulus images.

4.2.4 Scaling to determine individual differences across subjects

To gain more insight into the data, we decided to look at differences between subjects in the study. We first computed MDS solutions for each of the 21 subjects, and determined their dimensionalities using a combination of MDS stress and Shepard plots. We found that the dimensionalities of these solutions ranged between 2 and 5, with 17 subjects at 2 or 3 dimensions (total average: 2.66). While there are some similarities between the MDS solutions for each user, the solutions do not match exactly; in particular, there is disagreement about images in the middle of the complexity space (the images in the middle of Figure 3). This is consistent with behavior observed during our ranking pilot studies. In addition, individual MDS results all have a lower correlation measure (average $R^2 = .517$) than the all-subject MDS of Section 4.2.1 ($R^2 = .876$), presumably due to less noise in the multi-subject analysis.

In a subsequent analysis, we ran individual differences scaling using INDSCAL.[15] The goal of individual differences scaling is to compute an MDS solution over a set of stimuli, while simultaneously computing weights representing the importance each subject gave to each dimension of the solution. Using this approach to compute a 2D result, we find an embedding that is qualitatively similar to the result of Section 4.2.1. The 3D result is also similar, although the third dimension is slightly more expressive than that of Section 4.2.1; we did not find it corresponded to any meaningful property. In looking at different projections of the INDSCAL results, we discovered that images with related content are close to each other; for example, {*sponza*,*grandcentral*,*temple*}, {*kitchen*,*conferenceroom*}, and {*atrium*,*louvre*,*stadium*}. These groups can be seen in Figure 3 and Figure 4-(b) as well, but they are more pronounced in the INDSCAL result, particularly the last group.

Examination of the INDSCAL subject weights shows that all subjects are weighting dimensions quite differently. We have performed some initial clustering analyses to identify groups of subjects that behaved similarly (by looking at the norms between normalized matrices $ND(s)$). Our initial results show groups that are in line with those uncovered by INDSCAL analysis, but a larger study with more subjects is required to extract more meaningful information.

It is important to note here that one should not expect all users to behave similarly in this experiment. Analyses similar to ours used in previous studies of texture / image complexity[3–5] depend on users not behaving identically in order to compute unique positions in the MDS results for each image in the stimulus set. Furthermore, their results exhibit individual differences despite posing much more specific questions about visual complexity.

5. DISCUSSION

5.1 Comparison with other measures of complexity

In this section, we compare the one-dimensional ordering derived in Section 4.2.2 against those obtained by several computable measures of visual complexity, and also against scene-based measures of model complexity. The behavior of these other measures is not often correlated with 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. Aligning these measures should enable better rendering and scene representation algorithms.

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

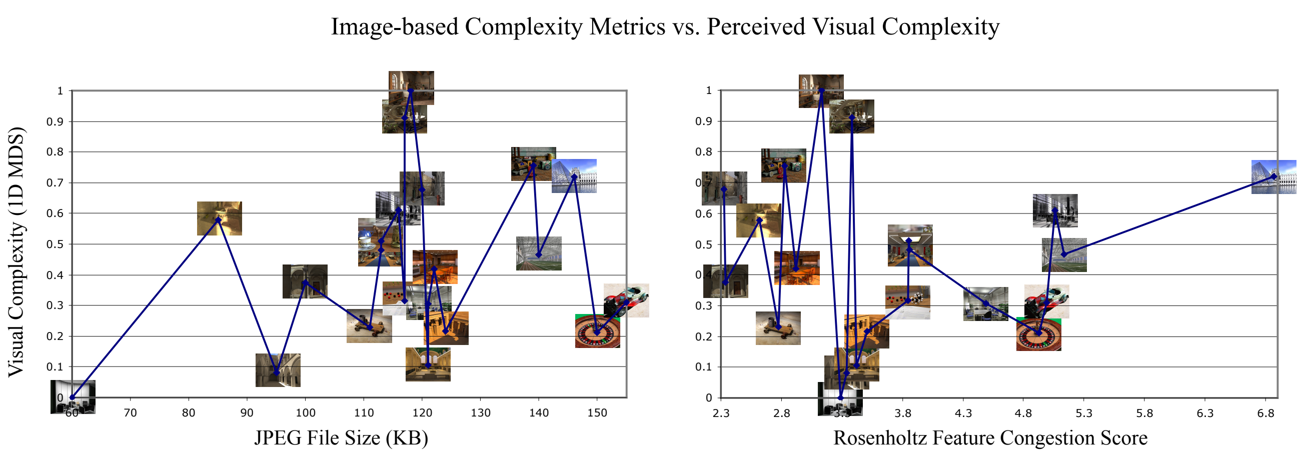
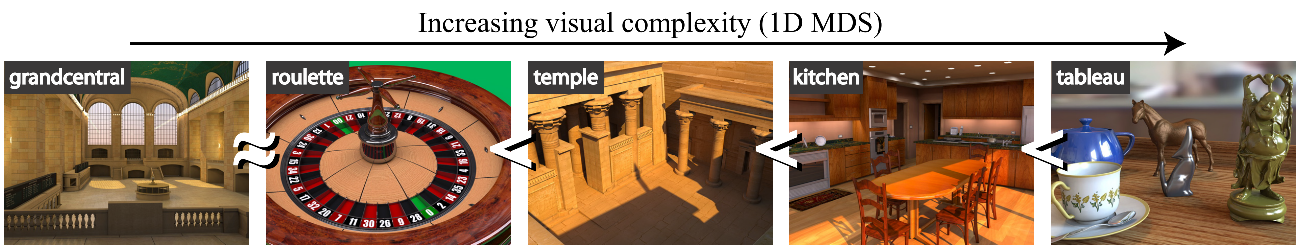


Figure 6. Comparing the experiment-derived 1D complexity ordering with orderings based on JPEG file size and the feature congestion metric.[9] Notice that neither curve is monotonic, indicating the low correlation between perceived complexity and computable measures.

Visual complexity is somewhat related to frequency content in image compression, or clutter in visual search.[9] 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,[9] respectively. The results are shown in Figure 6. As expected, neither curve is monotonic.

It is interesting to note which kinds of high frequency correlate with 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 a 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.

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



1.5M polygons	150K polygons	2.1M polygons	388K polygons	630K polygons
150K lights	23K lights	505K lights	59K lights	13K lights
409 seconds	590 seconds	225 seconds	290 seconds	79 seconds

Figure 7. 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 not much correlation between these numbers and perceived visual complexity.

In computer graphics, measures of scene complexity are often used as indicators of how difficult it is to render a final image. One common indicator is geometric complexity, the number of primitives (usually triangles) in a scene. Another common indicator is the number of lights used to render the scene. Many of these lights are actually *virtual* lights that result from dense samplings of real-life lighting conditions and effects, such as sun/sky illumination, indirect illumination (the result of multiple bounces of light), and environment maps. Typical rendering algorithms take time proportional to geometric / illumination complexity. Raw rendering time is also sometimes reported as a measure of scene complexity, as it is a direct measurement of rendering difficulty.

We obtained triangle counts, light counts, and rendering times (using the Lightcuts[16,17] 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 7. As expected, similar to image-based measures, there is little correlation between visual complexity and these algorithmic indicators of complexity from computer graphics. Looking closely at these images reveals some possible explanations. 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 considerable detail in lights, numbers of windows, and the railing in the front. However, these details are so small in screen space that they do not contribute as much to perceived visual complexity as they should. Our hope is that a better understanding of visual complexity will enable development of rendering algorithms that are more in line with how human observers perceive scenes.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a study of visual complexity for computer graphics scenes. We have discussed a 2D multidimensional scaling analysis of the data, and also several followup analyses. While the dimensions of the 2D analysis are not easy to identify, the data exhibits trends that roughly correspond to "numerosity" and "material/lighting complexity". The subsequent followup analyses suggest that there are significant individual differences between subjects. Finally, we have shown how perceived visual complexity does not necessarily correlate well with image-based and scene-based measures of complexity.

One major area of future work is to replicate these results on a larger set of images. Unlike with real photographs analyzed in other related studies,[5,13] it is difficult to find a large, broad set of computer graphics images that do not have visible artifacts which would confound perceptual studies. By focusing on some important dimensions, such as numerosity, material complexity, and lighting complexity in detail, it may be possible to generate larger image sets and perform more targeted studies. These studies can also use much more specific questions about visual complexity. For example, 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. The disadvantage is that the breadth of scenes tested is likely to be small, since each kind of scene (room, outdoor nature, etc) would need its own unique model.

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 understanding of the relationship between scale and perceived complexity.

Overall, we hope that continued investigation in this space will provide more insight into the relationship between visual complexity and graphics. This can not only serve to motivate new approaches to perceptually based rendering; it can also provide better guidelines for graphics designers and modelers aiming to convey complexity in their scenes.

ACKNOWLEDGMENTS

The authors would like to thank the subjects of the study for their time and patience, and the members of the Cornell Program of Computer Graphics for their valuable comments and feedback. This work was supported by NSF CAREER 0644175, CNS 0615240, and a grant from the Intel Corporation.

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. H. Rushmeier, B. E. Rogowitz, and C. Piatko, "Perceptual issues in substituting texture for geometry," in *Proceedings of the SPIE: Human Vision and Electronic Imaging V*, **3959**, pp. 372–383, 2000.
9. R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *J. Vis.* **7**, pp. 1–22, 8 2007.
10. 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.
11. 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.

12. 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.
13. 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.
14. 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, 1997.
15. J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika* **35**, pp. 283–319, 1970.
16. 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.
17. B. Walter, A. Arbree, K. Bala, and D. P. Greenberg, "Multidimensional lightcuts," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pp. 1081–1088, ACM Press, (New York, NY, USA), 2006.