Appears in the ACM SIGGRAPH 2008 conference proceedings

Perception of Complex Aggregates

Ganesh Ramanarayanan* Cornell University
Kavita Bala* Cornell University
James A. Ferwerda† Rochester Institute of Technology

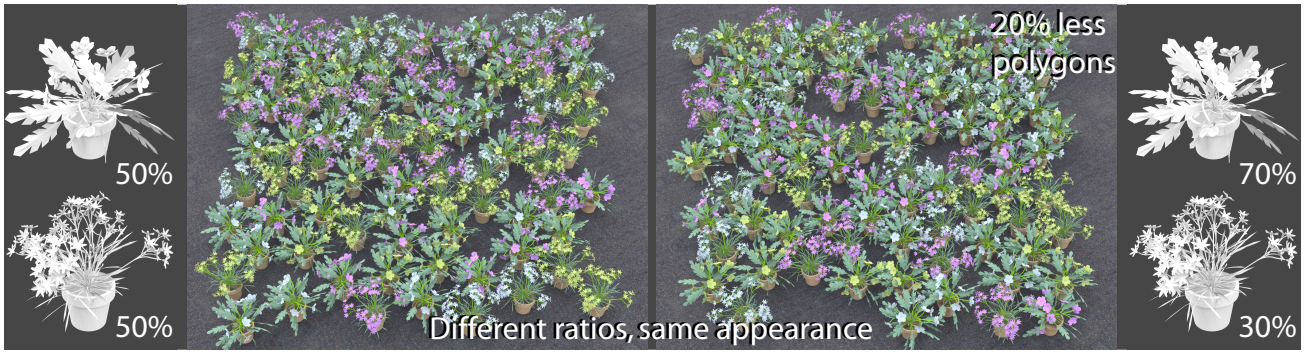

Figure 1: *Using our experiments, the aggregates of flowers in the left and right images were predicted to have the same appearance (and verified in a confirmatory study) despite a 20% difference in polygon count. This difference is achieved not by performing level-of-detail simplification, but by changing the ratios of objects in the aggregate.*

Abstract

Aggregates of individual objects, such as forests, crowds, and piles of fruit, are a common source of complexity in computer graphics scenes. When viewing an aggregate, observers attend less to individual objects and focus more on overall properties such as numerosity, variety, and arrangement. Paradoxically, rendering and modeling costs increase with aggregate complexity, exactly when observers are attending less to individual objects.

In this paper we take some first steps to characterize the limits of visual coding of aggregates to efficiently represent their appearance in scenes. We describe psychophysical experiments that explore the roles played by the geometric and material properties of individual objects in observers' abilities to discriminate different aggregate collections. Based on these experiments we derive metrics to predict when two aggregates have the same appearance, even when composed of different objects. In a follow-up experiment we confirm that these metrics can be used to predict the appearance of a range of realistic aggregates. Finally, as a proof-of-concept we show how these new aggregate perception metrics can be applied to simplify scenes by allowing substitution of geometrically simpler aggregates for more complex ones without changing appearance.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism;

Keywords: complexity, perceptually-based modeling, aggregates

*{graman,kb}@cs.cornell.edu
†jaf@cis.rit.edu

1 Introduction

Capturing the visual richness of the real world requires modeling and rendering complex scenes. One common source of complexity arises in the form of *aggregates*: large collections of objects where the properties of the group are more salient than those of the individuals. Examples of aggregates include forests, crowds, herds of animals, cars in a parking lot, bins of candy and piles of fruit. Paradoxically, the cost of manipulating and rendering scenes increases with aggregate complexity, exactly when observers are attending less to individual objects and more to the overall properties of the aggregates. Thus, exploiting our inability to perceive all the features of complex scenes should enable significant optimizations. But relatively little is known about the perception of aggregates and the factors that govern their appearance.

In this paper we present some first steps toward understanding the perception of complex aggregates of realistic objects for graphics. Our goal is to characterize the limits of visual coding of aggregates and to use these limits to efficiently represent the appearance of aggregates in computer graphics scenes.

To achieve this goal, we have conducted a series of psychophysical experiments exploring the roles played by the geometric and material properties of individual objects in observers' abilities to discriminate different aggregates. On the basis of these experiments we have derived metrics that can predict when two aggregates have the same appearance, even if they have different distributions of objects. We have confirmed that these metrics can be used to predict the discriminability of a range of realistic aggregates.

We believe that understanding aggregate perception is important in many areas of graphics, such as modeling, scene design, and scene simplification. As a proof-of-concept application, we show how these metrics can be used to substitute a geometrically simpler aggregate for a more complex one without a change in appearance. Figure 1 shows an example with an aggregate of flowers. Other possible applications include optimizing appearance for a given scene graph and polygon budget, or appearance-based procedural geometry generation.

The paper is organized as follows. Section 2 presents an overview of related work. In Section 3 we characterize the dimensions of

aggregate appearance. In Section 4 we describe the experiments we conducted to quantify the psychophysics of aggregate perception. In Section 5 we discuss experimental results and the resulting thresholds, and in Section 6 we validate our perceptual metrics. Finally, in Section 7 we explore some applications of these metrics for realistic image synthesis. In Section 8 we conclude with a discussion of limitations and future work.

2 Related Work

Due to the correlation between visual complexity and realism in image synthesis, the modeling and rendering of complex scenes has been a topic of significant interest in computer graphics. An important focus in this area has been on finding ways to reduce the memory and/or processing required to model and render scenes containing geometric or material aggregates.

2.1 Aggregates in computer graphics

Instancing. The paradigm of instancing, originally developed by Sutherland [1963], is ubiquitous in graphics applications today as a means to add complexity to scenes while simplifying model specification and reducing memory storage. There has been particular interest in converting fractal and other procedural models into instancing hierarchies [Hart 1992]. Work in plant generation has recognized that in complex plant scenes, parametrically similar objects (such as a collection of trees derived from a common L-system parse tree) can be approximated as a single instance without affecting quality [Smith 1984; Deussen et al. 1998; Brownbill 1996].

Level-of-detail. Most of the traditional work on level-of-detail (see [Luebke et al. 2002] for a survey) focuses on reducing the polygon count of a single object. Cohen et. al. [1998] and Luebke et. al. [2001] have studied the perception of LOD by considering contrast and texture masking. Rushmeier et. al. [2000] studied how shape and texture interact with each other, where LOD representations of objects are combined with better textures to maintain object fidelity. Watson et. al. [2004] have looked at supra-threshold perception for LODs. Lee et. al. [2005] have looked at mesh saliency for simplification of individual meshes.

Our focus in this paper is on aggregates as a whole, where we must reason about the numbers and kinds of objects in the aggregate. Recent work by Cook et. al. [2007] has shown how some aggregates can be stochastically simplified based on camera position by using single large elements as proxies for multiple small, often subpixel, elements; however, their focus is on sub-threshold simplification, while we propose simplifications that are supra-threshold but still appearance-perserving.

Capturing appearance. Recently there has been promising work in how images convey scene appearance, and how this appearance can be characterized and manipulated [Ramanarayanan et al. 2007; Vangorp et al. 2007; Khan et al. 2006]. In this work we hope to develop a similar understanding of appearance for aggregates.

2.2 Perception of aggregates

While there is a large literature on the perception of individual objects (see [Palmer 1999] for a review), there has been relatively little work done on the perception of aggregates of real-world objects. However, useful insights can be drawn from the literatures on texture perception, visual attention, and scene understanding.

Perceptual grouping and texture perception. Our ability to recognize objects in a complex scene and distinguish them from one another is closely related to perceptual grouping [Koffka 1935] and texture perception [Julesz 1962]. Early work on pre-attentive texture segregation hypothesized the fundamental importance of texture elements [Beck 1972; Marr 1976], most famously Julesz' textons [Julesz 1981], which are based on features such as size, crossings, and terminators. There has been followup work on the importance of arrangement [Beck 1982], structural edges between textures [Nothdurft 1985], and asymmetries in texture segregation [Gurnsey and Browse 1989]. Work on the perceptual dimensions of texture [Rao and Lohse 1993; Heaps and Handel 1999] suggests the importance of properties such as numerosity, orientedness, and so on. There are also models of pre-attentive segmentation based purely on spatial frequency and orientation responses, such as those found in early vision processing [Bergen and Adelson 1988; Malik and Perona 1990].

Visual attention, search, and popout. Given the sheer volume of visual information conveyed by aggregates, it is natural to ask how the human visual system is able to quickly sort through this information and identify important local properties, such as finding individual objects. The field of visual search and attention is concerned with exactly this question (see [Wolfe 1998] for a review). Influential models such as feature integration theory [Treisman and Gelade 1980] and saliency maps [Koch and Ullman 1985] formulate attention in terms of feature maps for primary features such as color, orientation, and luminance, that are then used for pre-attentive and attentive visual processing. Significant work has focused on understanding the relative strengths of these cues [Found and Müller 1995; Desimone and Duncan 1995], the effects of density and set size on them [Sagi and Julesz 1987; Bacon and Egeth 1991], and computational models based on them [Itti and Koch 2001]. Metrics of clutter have been also developed [Rosenholtz et al. 2007] based on the idea of multiple competing feature maps.

Scene understanding. The most general kind of aggregate is a complex image or scene, made of multiple varied objects, such as a bustling street corner. Despite the immense complexity of such scenes, the human visual system is quite adept at processing the whole and extracting meaningful information quickly [Potter 1975]. In this sense, basic aggregate perception is fundamentally tied to scene understanding. Research on scene 'gists' [Oliva and Torralba 2001] suggests that people estimate the global properties of a scene robustly, without paying specific attention to individual objects. Another interesting approach to understanding the perception of scenes is try to embed them in a multidimensional perceptual space, seeing which dimensions people identify as globally important properties (similar to texture dimensionality studies mentioned above). Oliva et. al. [2004] suggest the importance of dimensions such as numerosity and structure; these cues are similar to those used in visual discrimination and search tasks.

2.3 The need for a psychophysics of aggregate perception

Although a variety of techniques have been developed in computer graphics for efficiently modeling and rendering complex aggregates, such as instancing or level-of detail, manipulations are based largely on computational considerations and do not formally take the characteristics of visual perception into account. On the other hand, while research in visual perception has provided insights into the visual coding of complex patterns, research on the perception of aggregates of realistic objects in 3D scenes that might provide a foundation for graphics techniques has been scant. Therefore, in the following sections we will first outline a conceptual framework to describe the perceptual dimensions of complex realistic aggregates, and then we will conduct a series of psychophysical experiments to relate the physical and mathematical dimensions used to describe aggregates to their perceptual appearance.

3 Appearance of Aggregates

Here we describe our characterization of the appearance of aggregates, based on their fundamental properties.

3.1 Properties of Aggregates

When viewing an aggregate, there are many properties that affect how we perceive it. Due in part to the difficulty of generating and controlling stimulus sets, there have been relatively few systematic studies of the perception of aggregates of realistic objects in 3D scenes. As described above, research in texture perception, visual search, and scene understanding are all relevant. While there are substantial differences in the studies and the models proposed in these areas, many of the important visual properties of real-world aggregates are captured effectively by a few common dimensions. These include: *numerosity*—the number of objects visible in a scene; *variety*—the range of variation in the shapes, sizes, and material properties of the objects that make up the aggregate; and *arrangement*—the layout of objects in the scene. We briefly describe each of these dimensions:

Numerosity. Numerosity is the fundamental defining property of an aggregate. Intuitively, as the number of objects in a collection increases, an observer perceives the collection more as a whole, rather than focusing on individual objects. The exact point at which this perceptual transition takes place is likely to depend on the properties of the objects in the aggregate.

Variety. The variety of the objects in an aggregate also affects how it is perceived. Variety can arise from differences in the shapes and materials of objects and encompasses variations in geometry, color, gloss, pattern, and texture.

Arrangement. Human observers use the structure and arrangement of objects to help them understand real-world scenes. Arrangement itself is multidimensional, including regularity, density, and spatial dimensions. For example, the *regularity* of trees arranged in patterns in an orchard is perceived differently from the disordered arrangement of trees in a natural forest. Similarly, the *density* of flowers in a garden is perceived differently from a uniform scattering of flowers over a large meadow. The *spatial dimensions* of an aggregate can also vary, e.g. one dimension (a string of beads), two dimensions (plants on terrain), or three dimensions (leaves on a tree).

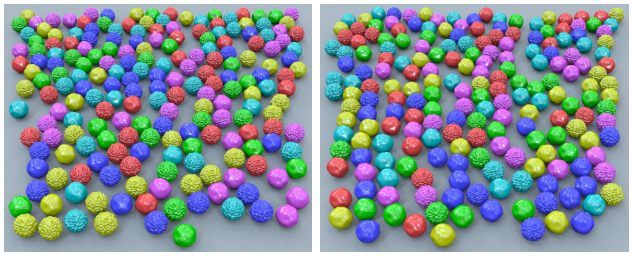

Figure 2: *Two aggregates with the same appearance but different distributions of objects - the left image has 40 more bumpy objects and 40 fewer smooth objects.*

3.2 Understanding the Appearance of Aggregates

We wanted to study transformations that can be applied to aggregates that continue to preserve their appearance as a whole, while potentially reducing their complexity. Various transformations were considered. For example, how many individuals can be replaced by lower level-of-detail representations to gain performance? The concern with this strategy is that while performance can be improved using lower level-of-detail representations, if an observer pays attention to a particular object, the lower resolution of that object will be apparent, reducing the perceived quality of the aggregate. A level-of-detail approach is effective, but only to the point where all differences are sub-threshold.

We took a different approach and decided to study how sensitive observers are to the numbers of each kind of object in an aggregate. We hypothesized that within limits, observers will be relatively insensitive to variations in the ratios of objects that make up a heterogeneous aggregate; Figure 2 shows an example of this. Therefore, when modeling aggregates, we would be able to change their composition to favor the objects with lower geometric complexity. Our aim is to to develop predictive perceptual metrics that can be used to guide the modeling of complex aggregates and go beyond the ad-hoc tools currently used.

4 Experiments

To achieve the goals above, we have conducted a series of psychophysical experiments to understand the insensitivity of human observers to the exact composition of individual objects constituting an aggregate. First, we describe the objects we studied, and the numerosities, varieties, and arrangements we used to create stimulus images. Then, we describe our experimental design.

4.1 Scope of our experiments

The aggregate properties of numerosity, variety, and arrangement defined in the previous section are each multidimensional in their own right. It is not feasible to study all these dimensions within the scope of a single research paper, therefore, in this section we describe the set of choices we made to define a reasonable subspace that is (a) small enough to be manageable, and (b) general enough to be practically useful for graphics applications. Where possible, we have made conservative choices to broaden the range of cases in which our study could apply.

Numerosity. Ideally we would like to study complex models where the numerosity of objects is on the order of real-world scenes. Prior research in search tasks [Sagi and Julesz 1987] has shown that humans are more accurate in estimating properties of numerous, dense distributions; thus, we conservatively chose to study as high a numerosity as possible. However, if too high, the limits of image size and resolution make objects so small as to be indistinguishable from one another. For our stimulus set, an aggregate size of $n = 200$ provided maximum numerosity and density while still keeping the individual objects distinguishable (see Section 4 for more details).

Variety. Variety in an aggregate arises from differences in both the shapes and material properties of objects. Again, this is a very large space to study. To keep our study conservative and tractable, we focused on *binary aggregates*: aggregates consisting of two types of shapes. Our exploratory studies showed that ternary and higher aggregates are likely to have even larger perceptual thresholds for discriminability. We studied a range of different pairs of shapes and materials, as described in the next section.

Arrangement. Because of the powerful interactions between arrangement and perceptual grouping, aggregate appearance is af-

fected greatly by arrangement. While studying the effects of regularity and disorder on the perception of aggregates would be valuable, for our experiments we have chosen to study aggregates of disordered objects as these are representative of the aggregates encountered in many natural scenes. Even with disordered arrangement, we had to decide how to distribute the objects spatially. Distributions in 3D raise the problem of occlusion which from a single viewpoint might distort an observer's perception of both the number and variety of objects in an aggregate. To avoid these problems, we restrict our studies to random arrangements of objects on a flat surface (no stacking), viewed at an angle from above (though not directly overhead), to maintain 3D scene understanding while minimizing occlusion.

4.2 Aggregate objects and materials

The space of all possible shapes and materials is very large. The two objects in a binary aggregate can be almost the same or vastly different. To develop a sense of how differences between individual objects can affect aggregate perception, we studied a range of pairs of shapes, from similar to dissimilar. For materials we focused on color variations, both correlated and uncorrelated with shape variation.

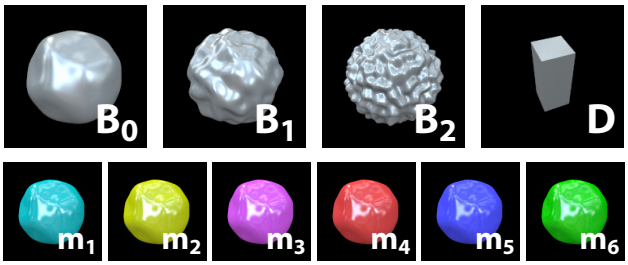

Figure 3: *The geometries and materials used in our experiments. The* B *objects are similar to each other, and the* D *object is completely different. The colors used are, in order from left to right, cyan, yellow, magenta, red, blue, and green.*

Shape. We selected the set of geometries shown in Figure 3. Our canonical reference object is B_0, a smooth, sphere-like object. B_0 is a 'blobby' created in the spirit of shape studies [Todd et al. 1997]. We selected this object because it is recognizable, while not being a perfect shape. To generate B_0, we used the technique of [Ramanarayanan et al. 2007], where sphere vertices are perturbed along their normal direction as a function of a 3D Perlin noise cube [Perlin 2002]. We used a sphere of radius 0.81 and scaled the noise cube by a factor of 1.68.

To understand how geometric similarity affects perception, we created two object shapes, B_1 and B_2, that are bumpier versions of B_0. [1] B_1 and B_2 were selected to be equally spaced perceptually and clearly distinguishable. To confirm this we passed image pairs through the visible differences predictor (VDP) of Mantiuk et. al. [2005]. At a nominal viewing size of 30×30 pixels (.72 degrees visual angle), the percentage of pixels above VDP threshold comparing diffuse renderings of B_0 and B_1 is 45.89%, and the percentage for B_0 and B_2 is 69.82%. To understand the effect of object dissimilarity, we also looked at D, a small rectangular prism, with a completely different shape, volume, and features from B_0.

Material and lighting. Previous research has shown that specular reflections of realistic lighting are a strong cue for shape [Fleming et al. 2004]; so we picked glossy Ward materials and lit the objects with the UC Berkeley Eucalyptus Grove environment map [Debevec and Malik 1997]. Gloss properties were defined using the perceptually uniform c, d gloss space of Ferwerda et. al. [2001]. All materials had a c value of 0.09 which shows specular cues, falls within a range of realistic materials [Ward 1992], and is not unrealistically shiny. Instead of focusing on changing glossy material properties, which are hard to perceive in large aggregates, we varied the objects' diffuse color to create variety in the material properties of the aggregate. Furthermore, color can vary correlated with shape (red balls and blue cubes), or uncorrelated with shape (red and blue balls, or red and blue cubes). We studied both possibilities.

4.2.1 Rendering aggregates with random arrangement

This section describes the scene setup we used to display our aggregate stimuli, in particular our specification of random arrangements on the plane, and the camera / rendering setup.

Object arrangement. To model the aggregates we used in the experiments, we randomly placed individual objects on a gray, diffuse plane. We use fast Poisson disk sampling [Dunbar and Humphreys 2006] to cover the plane with n candidate locations in the view frustum. We achieve a dense distribution [Sagi and Julesz 1987] by matching the radius of the Poisson disks with the radius of the test objects. To stay maximally different, D's size remains much smaller than the Poisson disk size.

Camera view. We set the camera view at a viewing angle of 45 degrees above the plane to balance two effects; one, that the aggregate is viewed in perspective, and two, that occlusion is low. The camera view distance was calibrated to maximize the number of objects on screen, while keeping the stimulus images a reasonable size so that objects are distinguishable from each other even when they are viewed in perspective. Stimulus images were 590×475 pixels (14 degrees visual angle) with 200 objects in the aggregate, each with a minimum object size of 30×30 pixels (.72 degrees visual angle). Figures 4 shows some examples.

Rendering. Based on the experiment design described in the next section, each image was rendered by placing the appropriate objects in the locations computed by the Poisson disk sampling, and assigned suitable material properties. The final images were rendered with environment map lighting using the Lightcuts algorithm [Walter et al. 2005; Walter et al. 2006].

4.2.2 Specifying aggregate distributions: numerosity and variety

For our experiment we must create different object distributions. In a binary aggregate, two shapes are being tested. We specify r, the ratio between the two shapes. Since all our aggregates had 200 objects, an aggregate of B_0 and B_1 with $r = 50{:}50$ would have 100 B_0 and 100 B_1, and an aggregate of of B_0 and B_2 with $r = 20{:}80$ would have 40 B_0 and 160 B_2. These examples are shown in Figure 4-(i) and (ii), respectively, where all objects are of one color.

The materials of an aggregate are varied by changing the color of the objects. We have two choices: to correlate the color change with shape, or not. When color is correlated, then each shape is always associated with the same color. Figure 4-(iii) shows such an aggregate where all B_0 are cyan, and all D are yellow, and $r = 50{:}50$ (100 of each). The colors can also be uncorrelated, in which case each object can be either cyan or yellow with equal probability, independent of its shape (see Figure 4-(iv), and (v-vi) for examples of more colors).

[1] B_1 and B_2 were generated using Perlin noise scaling factors of 3.2 and 5.66, respectively.

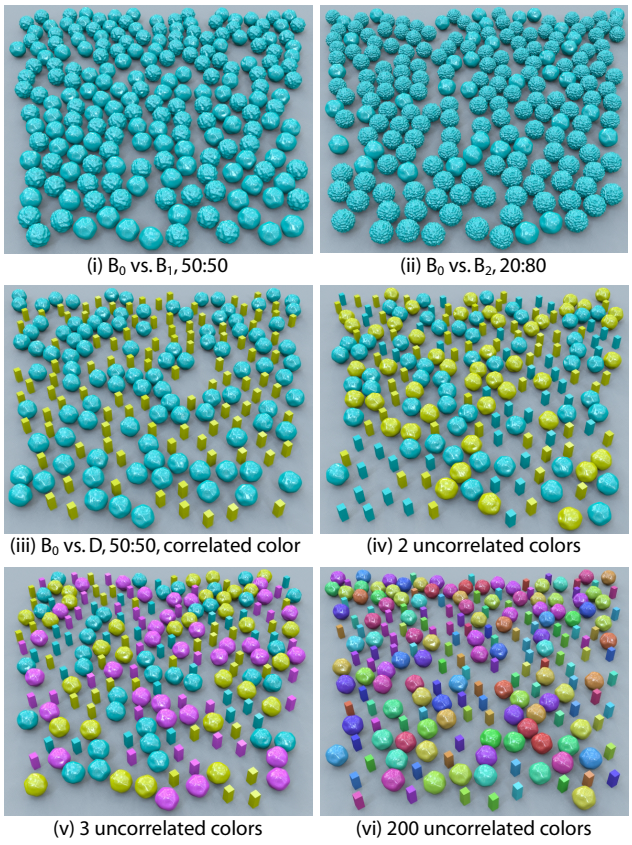

Figure 4: *Example stimuli from our experiments. (i) and (ii) show aggregates composed with* B_0 *and* B_1 / B_2*, respectively. (iii)-(vi) show different examples of color variation for* B_0 *and* D*: correlated color in (iii) and uncorrelated color in (iv)-(vi).*

4.3 Experiments

Our goal is to study the sensitivity of human observers to the composition of objects in binary aggregates. We now describe the psychophysical experiments we conducted with our stimulus set to study this. We ran two experiments: in Experiment 1, we focused on varying shapes in the aggregate, and in Experiment 2 we focused on varying material (color).

4.3.1 Study Question

We begin with a binary aggregate of two shapes (α, β). We looked at many conditions, with different pairs of shapes, different colors, and ratios. For each condition, the shape variation of any stimulus image is characterized by the ratio r, e.g. 50:50 (see Figure 4-(i)). We would like to find how much r can be perturbed without affecting the appearance of the aggregate.

The specific question we asked is shown in Figure 5. On the top is the *base ratio* $br = br_\alpha : br_\beta$, shown as the reference image. Below that are two stimulus images, with ratios r' and r''. One of these two images (either the left or the right) is always the SAME aggregate as the reference; i.e., $r' = br$, except the objects are shuffled around / arranged differently. The other image has a different ratio $r'' = br_\alpha - \Delta : br_\beta + \Delta$. The question asked is: **Which image shows the same collection of objects as the reference?** Our goal is to find Δ at which the observer cannot pick which aggregate is the same as the reference, which indicates he believes r' and r'' have the same appearance.

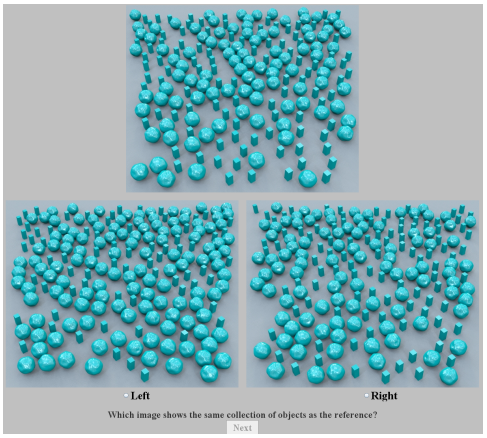

Figure 5: *Screenshot of the experiment GUI. Shown:* B_0 *vs.* D.

4.3.2 Experiment 1: Single-color ratio sensitivity

In Experiment 1 we explored sensitivity to ratios purely in terms of shape variation. All objects were made of the cyan material. We looked at three specific kinds of binary aggregates (see Figure 4):

- B_0 **vs.** B_1**:** two geometrically similar objects.
- B_0 **vs.** B_2**:** two geometrically similar objects, but less similar than B_0 vs. B_1.
- B_0 **vs.** D**:** two geometrically dissimilar objects.

Our goal was to measure when a change in ratio resulted in a noticeable change in aggregate appearance. We expected that thresholds could be different depending on the initial composition of the aggregate, so we considered 5 base ratios for the reference image: 20:80, 35:65, 50:50, 65:35, and 80:20; i.e., these had $(40, 70, 100, 130, 160)$ B_0 objects. For each of these ratios, our goal was to find the smallest Δ such that the aggregate with $\Delta * 200$ more or fewer of B_0 (and vice-versa for the other object) was judged to have the same appearance as the reference aggregate. There is a further variation. It is not clear if the threshold should be the same going up and down in the count of an object. While it may seem that Δ_{up} and Δ_{down} are close to each other, pilot studies showed that they behave quite differently, so we chose to measure them separately. Thus, our final goal was to find Δ_{up} and Δ_{down}, such that the aggregate with $200 * \Delta_{up}$ more, or $200 * \Delta_{down}$ fewer, B_0 objects, had the same appearance as the reference.

We expected that B_0 vs. B_1 would have the highest thresholds, because B_1 looks the most like B_0 and is most easily confused with it. Likewise, we expected that B_0 vs. D would have the lowest thresholds, since D is completely different from B_0.

4.3.3 Experiment 2: Multi-color ratio sensitivity

In our second experiment, we looked at material variety (by changing color), while holding the ratio of the objects in the aggregate fixed. We tested the following two kinds of binary aggregates:

- B_0 **vs.** B_1**:** two geometrically similar objects.
- B_0 **vs.** D**:** two geometrically dissimilar objects.

For each of these, we tested a wide variety of color conditions, all at the ratio 50:50. Each object had a color associated with it at random

from the following 6 color distributions: only one color {cyan}; two colors {cyan, yellow}, three colors {cyan, yellow, magenta}, and so on, in the order shown in Figure 3. We also considered two extreme cases: correlated color, and 'lots' of color. In the correlated case, each kind of object was associated with a single color; i.e. B_0 was cyan, and B_1 or D were yellow. In the 'lots' case, each object in the aggregate had assigned to it a unique color from the HSV cone (H sampled in 200 divisions, $S = 0.91$, $V = 0.338$).

Color is a very strong grouping cue that often cannot be ignored even in unrelated search tasks [Theeuwes 2004]. For this reason, we expected that color would raise thresholds for aggregate appearance. The one exception to this is the correlated color case. When both color *and* geometry work together to distinguish two objects, we expected subjects to be quite accurate.

4.4 Experimental Methodology

4.4.1 Measuring thresholds

To measure thresholds we built on the QUEST [Watson and Pelli 1983] method, as implemented in the MATLAB PsychToolbox [2]. QUEST adaptively presents trials to a subject based on their past performance, which is useful when a good a priori estimate of threshold is not available. It maintains a likelihood function, defined over all potential threshold values, that is updated with each subject response, and used to determine the final measured threshold.

In our study, we were interested in locating many (around 60) thresholds, spanning a number of aggregate test conditions. We also wanted to keep the study size reasonable (under 1 hour worst case). Thus, we limited ourselves to 7 QUEST-suggested trials per subject per threshold value. To ensure each subject saw a full range of stimuli and to keep subject performance independent, state was not shared between subjects; that is, subject i's responses did not affect the trials provided to subject $i + 1$.

To combine all subject data into a single threshold value, rather than computing a threshold per subject and averaging, which is noisy with only 7 trials per subject per threshold value, we pooled all subject responses into a single likelihood function and used it to compute threshold and error bars (given by standard deviations). Computationally, this is equivalent to treating all responses as belonging to a single super-subject, and estimating the threshold for that super-subject. To validate that this technique estimates reasonable thresholds, we tried bootstrapping methods [Wichmann and Hill 2001] on our data and found they produced similar results.

4.4.2 Experimental procedure

Prior to taking the experiment, the study question (Section 4.3.1) was explained to each subject, using simple examples with a few objects. Then the subject began the actual experiment on a computer using the GUI mentioned earlier (Figure 5). The GUI indicated the trial number and also the number of correct responses the subject had given thus far - we found this improved subject performance and morale in what was otherwise a monotonous task.

For each aggregate type (i.e. B_0 vs. B_1), trials were presented in random order, one at a time. They were also interleaved across all base ratios in order to ensure statistical independence [Watson and Pelli 1983]. Experiments 1 and 2 were run on 17 subjects each, all in their twenties to fifties and with normal or corrected-to-normal vision. Experiment 1 took about 45 minutes per subject and Experiment 2 took 30 minutes per subject. The experiment was conducted in a dark room using an LCD display (Dell 2000FP, 20" diagonal, 1600x1200 resolution, sRGB, max luminance 200 cd/m^2, 60:1 dynamic range, gamma 2.2).

[2] http://www.psychtoolbox.org

5 Results

We now present the results of our experiments on binary aggregate perception. We will first discuss the single-color case, which investigates the impact of shape similarity alone on the appearance of aggregates with different distributions. We then discuss the multi-color case, where we look at how color variety modulates the thresholds observed in the single-color case, and finally we summarize the results.

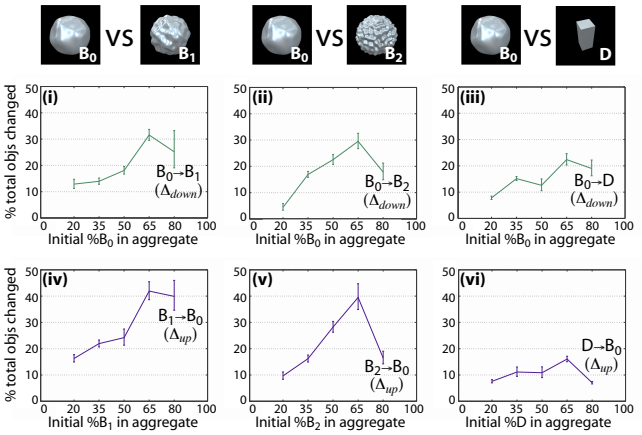

Figure 6: *Thresholds for shape variations. The top row shows insensitivity to decreasing B_0 in the ratio. The bottom row shows insensitivity to decreasing the other object in the ratio. In both cases, the more of an object you start with, the more you can decrease it before subjects noticed. The final points in the curves dip because the total number of objects in each image is constant, so when one kind of object is too numerous, the other one starts to pop out more.*

5.1 Experiment 1 - Effects of Shape on Aggregate Appearance

The results of Experiment 1 are shown in Figure 6, indicating how sensitive observers are to changes in shape ratios. The top row corresponds to Δ_{down}, and the bottom row corresponds to Δ_{up}. For example, looking at Figure 6-(i), 50 on the x-axis is associated with the value 18. This indicates that if you start with 50% B_0, you can reduce this by 18%, replacing B_0 with B_1, and the new aggregate (which only has 32% B_0) will have the same appearance. In other words, for base ratio 50:50, $\Delta_{down} = 18\%$.

Effect of shape. The first thing to notice in Figure 6 is that overall, the most similar object to B_0, B_1, has the highest thresholds, and the least similar object, D, has the lowest thresholds. This supports one of our hypotheses, which is that objects more similar to B_0 will have higher thresholds (i.e., more replacements are possible). B_2 lies in between, but is not very different from B_1.

Threshold curve. The second thing to notice is that for the first four data points in each graph, thresholds are trending up (linear regression R^2 between .76 and .95). That is, the more α you start with, the less sensitive you are to replacing α with something else. Intuitively, this makes sense; with a large number of α, estimates of numerosity are likely to be less accurate than with a small number

of α. This is an interesting trend that indicates that aggregate perception may have Weber's law like characteristics [Palmer 1999], but confirming this would require further studies.

The fifth data point in every graph reveals some interesting behavior - the upward trend stops, and either stays flat or changes direction. Recall that each aggregate has a total of 200 objects, so when we have 160 of object α (80% point on the x-axis), we only have 40 of object β. If β is salient in the field of αs, then we can use estimates of β to better understand how many objects there are. For example, it is easier to notice 20 objects becoming 40 than to notice 180 objects becoming 160.

By this reasoning, the behavior of the fifth data point in these curves gives a sense of how much effort is required to discriminate α and β. For B_1 (Figure 6-(i,iv)), the effect is slight; this reflects the high similarity of B_0 and B_1. Many subjects reported that they found the shape difference subtle, despite the fact that the VDP indicates that the differences are well above threshold. For B_2 (Figure 6-(ii,v)), the effect is dramatic; this shows that B_0 and B_2 are easily distinguishable. For D (Figure 6-(iii,vi)), the effect exists, but it is much larger when D is getting replaced with B_0 (Figure 6-(vi)). This is because B_0, being bigger and more densely packed, groups together better than D, and therefore provides better information for judging aggregate properties.

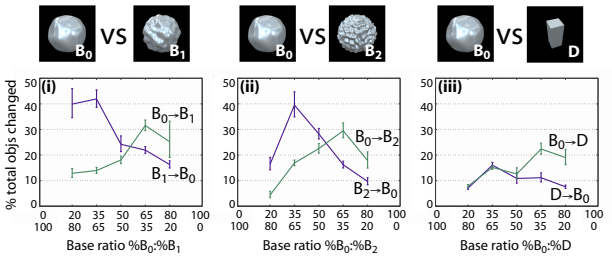

Figure 7: *Threshold asymmetry. When plotting all result curves together, notice that the curves can differ for skewed distributions, but they meet near the 50:50 point. The lack of perfect symmetry between the curves possibly indicates a subject preference for focusing on one object over another.*

Threshold asymmetry. There is an inherent symmetry in these graphs - since the binary aggregates are of fixed total numerosity, when B_0 is decreased, B_1 is increased, and vice-versa. Thus, the curves Figure 6(i-iii) are naturally related to the curves (iv-vi). To understand this further, in Figure 7, we combine each set of two curves into a single plot. The blue curves in the bottom row of Figure 6, which show the effect of increasing B_0 (Δ_{up}), have been flipped and overlaid on the green curves, which show the effect of decreasing B_0 (Δ_{down}). The x-axes have also been combined to show the aggregate base ratio for each point.

The first thing to notice here is that the curves on each plot cross. For example, looking at Figure 7-(i), for 20:80 B_0 vs. B_1, observers are far more sensitive to decreasing B_0 than decreasing B_1. However, for 80:20, the thresholds are reversed. This is to be expected, since as mentioned earlier, driving an object close to extinction is very noticeable. As we near the 50:50 point, the two curves get closer together and the thresholds are more symmetric. The curves in Figure 7-(ii) show similar behavior. In Figure 7-(iii), for high numerosities of B_0 ($50-80\%$) the curves in the plot have the same form as the other two and the same explanation can be applied. However, for low numerosities ($20-30\%$) the crossing has collapsed and the curves are essentially the same. This can again be attributed to differences in how well the dense B_0 and sparse D objects group with each other.

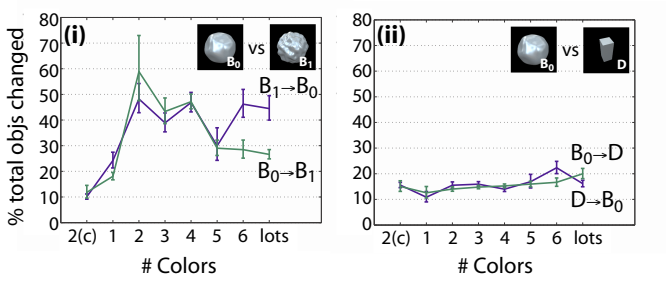

Figure 8: *Binary aggregates under color variation. (i) For similar objects B_0 and B_1, correlated color (2(c)) enhances shape distinctions, but uncorrelated color does the opposite because color is a strong cue, especially when there are only two. As colors increase, thresholds decrease but never beyond single color performance (Figure 6). (ii) For completely dissimilar objects B_0 and D, color has virtually no effect because the shape difference dominates the color difference.*

5.2 Experiment 2 - Effects of Color on Aggregate Appearance

The purpose of Experiment 2 was to gauge the impact of color on the perception of aggregates. Recall that we tested two geometrically similar objects, B_0 vs. B_1, and two dissimilar objects, B_0 vs. D, in a variety of color conditions. The results are summarized in Figure 8, where the x-axis indicates the number of uncorrelated colors tested ('2(c)' and 'lots' indicate the extreme cases of correlated color and one unique color per object, as described in Section 4.2. As shown in the figure, while color did not have a large effect on B_0 vs. D, it has a large effect on B_0 vs. B_1, often making subjects completely ignore the geometric differences between the objects. Furthermore, when correlating color with geometry, thresholds for B_0 vs. B_1 became much lower, whereas no similar effect was noticed for B_0 vs. D.

We will now examine the results in more detail. Looking at Figure 8-(i), thresholds increase dramatically from 1 color to 2 colors, and then taper off as even more colors are added. At the rightmost data point, which uses a unique color for each object on screen, thresholds in one direction are comparable to thresholds for a single color, but always greater. One possible explanation is that as the number of different colors on screen becomes large, color is less usable as a cue for aggregate properties, so shape difference information dominates, as is the case with a single color. Thus, while uncorrelated color can result in significant threshold elevation, its worst-case behavior is bounded by the single color case.

The correlated color case (2(c), the leftmost data point) shows another extreme. When geometry and color are correlated, differences in either cue can be used to discriminate objects, and for B_0 vs. B_1 or B_2 geometries, color is the stronger cue, so thresholds are reduced.

In contrast, B_0 vs. D tells a different story. Regardless of the amount of color variety or its correlation, thresholds are almost the same across all conditions, with perhaps a small amount of elevation for extreme color cases. This insensitivity to color variation arises because the geometric differences between D and B_0 are so great that they provide a much stronger cue for aggregate appearance than color differences.

5.3 Summary of results

We summarize the results from our experiments:

- The more numerous an object, the less noticeable when it is replaced. Objects of the same type behave similarly (B_0 vs. B_1, B_2). Very different objects (B_0 vs. D) have much lower thresholds.
- If objects are similar, correlated color decreased thresholds, and uncorrelated color increased thresholds, although with too many colors this effects weakens. If objects are very different, color has virtually no effect on threshold.

6 Confirmatory Study and Validation

To understand if these results generalize to other scenes, we created aggregates of more realistic kinds of objects and materials, and tried to predict when these new aggregates are perceived as equivalent even when the ratios of objects in them are different. We did this by associating each aggregate with the appropriate result curve measured in our experiment, and testing whether our measured thresholds predicted the performance of subjects on these new aggregates. We used the exact same study setup of Experiments 1 and 2, except that we used the method of constant stimuli, testing specific ratio modifications for each aggregate, rather than using an adaptive mechanism like QUEST. 11 subjects participated in the confirmatory study, and it took about 15 minutes per subject.

Predictive Metric. To predict the results of this study, we used the results of Section 5 as a metric by associating each aggregate with the most similar object pairings we tested, as follows. When color was correlated between similar objects (e.g. flower1 vs. flower2), we used the B_0 vs. B_1 correlated color threshold of Experiment 2. When color was uncorrelated across similar objects (e.g. toyota vs. bmw), we used the B_0 vs. B_2 curve of Experiment 1. In all other cases, where shapes are significantly different, we used the B_0 vs. D curve, which, by Experiment 2, can apply to all color situations. The geometrically simpler object was always associated with B_0.

Figure 9 summarizes our results for the 94 data points tested. In each case, we picked several base ratios (x-axis of Figure 7), and tested various modifications of these ratios, both below and above threshold. We were able to conservatively predict the results for 96% of cases; 82% were exact predictions, and 14% were conservative in that subjects judged the appearance to be the same at a higher threshold than predicted by the curve. In the 4% cases where our predictions were not conservative, predicted thresholds were within 5% of the number of total objects n.

Aggregate	Setup	Result
Using B0 vs. B1 2(c) (correlated color) threshold (Figure 8-(i))		
	n=100; 2(c) colors; 10 trials Base ratio 50:50 Tested +/– {5, 10, 15, 20, 40}	**90% predicted** (70% exact)
Using B0 vs. D curve (Figure 7-(iii))		
	n=200; 2(c) colors; 30 trials Base ratios 30:70,50:50,70:30 Tested +/– {5, 10, 15, 20, 30}	**93% predicted** (87% exact)
	n=200; 2(c) colors; 10 trials Base ratio 50:50 Tested +/– {5, 10, 15, 20, 40}	**100% predicted** (90% exact)
Using B0 vs. B2 curve (Figure 7-(ii))		
	n=100; 3 colors; 10 trials Base ratio 50:50 Tested +/– {5, 10, 15, 20, 40}	**100% predicted** (90% exact)
	n=160; 2 colors; 18 trials Base ratios 30:70,50:50,70:30 Tested +/– {10, 20, 30}	**94% predicted** (78% exact)
	n=160; 3 colors; 8 trials Base ratio 50:50 Tested +/– {10, 20, 30, 40}	**100% predicted** (88% exact)
	*Ternary aggregate* n=160; 2 colors; 8 trials Base ratio 50:50 Tested +/– {10, 20, 30, 40}	**100% predicted** (63% exact)
Total: 96% predicted (82% exact)		

Figure 9: *The results of our confirmatory study on aggregates of more realistic objects. The metrics presented in Section 5 predicted results 96% of the time, being exactly accurate 82% of the time, even including difficult untested cases such as the ternary (3 shape) car aggregate (bottom).*

7 Possible Applications

As demonstrated in Figure 1 and validated in Section 6, it is possible to vary the composition of an aggregate without affecting how observers perceive its appearance; furthermore, this can be predicted with the thresholds we have presented. We now describe how this can help in the modeling and rendering of complex scenes.

Aggregate Simplification. One approach is to use our thresholds on existing aggregates. In the simplest case, consider a *unary* aggregate of one object. In a pilot study, we found that subjects' ability to count objects reflected performance very similar to our B_0 vs. D curve, with slightly lower thresholds. Stated simply, people appear to be insensitive to 5% changes in numerosity without affecting appearance. This argument is similar to Weber's law [Palmer 1999] for light intensity, where it is possible to eliminate some fraction of light energy without it being perceptually salient.

For aggregates of multiple objects, there are several ways to proceed. For binary aggregates, we can use object similarity to pick a threshold metric from Figure 6, which can be used to replace some objects with others. In addition to the flowers example in Figure 1, we show some more proof-of-concept examples in Figure 10. In all of these scenes, the two objects don't have the same polygon count, so we can skew the distribution in favor of the geometrically simpler object. For instance, the jacks are 9K polygons, and the balls are 2.5K; in the flower case, the polygon counts are 30K vs. 10K. Polygon savings are $7-20\%$, depending on the relative model complexities in the aggregate. For ternary and higher aggregates, one can treat them as a binary aggregate of 1 object and $k-1$ objects, as demonstrated in the confirmatory study (Figure 9 bottom).

Aggregate Design. Our metrics also speak fundamentally to tradeoffs that can be made by designers of aggregates, procedural models, and other high-complexity geometric graphics content. Taking into account object similarity and color variation, designers will have a good idea of how much they can play with the distribution of an aggregate before its appearance fundamentally changes. Additionally, the threshold elevation that occurs with uncorrelated

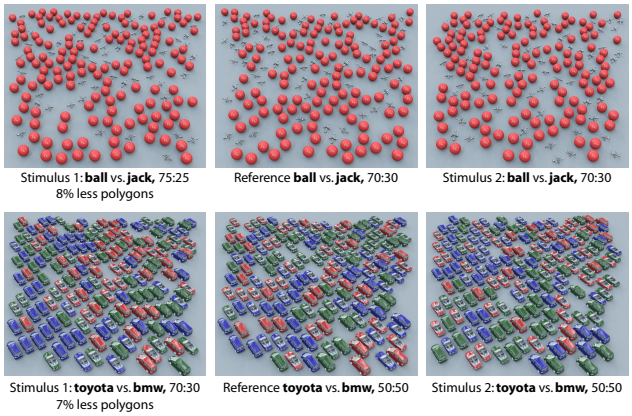

Figure 10: *Validation and application. Like Figure 1, we show a few examples where we predicted and experimentally validated a change in object ratios, preserving appearance while reducing polygon count. Top row: balls and jacks, 5% ratio change, Bottom row: cars, 20% ratio change.*

color is a useful tool for a designer. Note that uncorrelated color does occur somewhat frequently in the real world (autumn leaves, plants, fruits, toys, stationery, etc).

8 Conclusions

In this paper we have presented some first steps toward understanding and taking advantage of the perception of complex aggregates in computer graphics. We first described a set of psychophysical studies that explored the roles that object geometry and material properties play in observers' abilities to discriminate the ratios of objects in heterogeneous aggregates. The main results of the experiments showed that depending on the properties of the objects, observers can be insensitive to changes in object ratios and judge aggregates to have the same appearance with respect to the dimensions of numerosity, variety, and arrangement. We then used these results to develop metrics for predicting when different aggregates will have the same appearance, and we validated that these metrics are predictive for aggregates of novel objects. In addition we have discussed some proof-of-concept applications for how these results may be applied to simplify aggregate geometry in complex scenes.

Limitations and Future Work. While these initial findings are both interesting and useful, there is still much more to do. First, in terms of psychophysical studies, the scope of our experiments can be expanded to include aggregates of more than two or three varieties of objects. We also need to further explore the effects of numerosity (low and high) and arrangement (especially regularity and 3D layout) on the perception of aggregates. With respect to perceptual metrics for aggregate modeling and rendering, it would be useful to develop more detailed models of the effects we have observed so that we could use more aggressive simplification where thresholds are higher. It would also be valuable to have better metrics for the interplay of geometry and material properties in aggregate appearance. In addition, while we have only characterized aggregate appearance in terms of the entire aggregate being in view, it would be interesting to look at different formalizations depending on camera zoom and occlusion, to develop more advanced metrics that can be applied to dynamic aggregates and moving viewpoints with changes in scale. Finally, it would be valuable to explore the relationships between the statistical foundations of texture / material perception and aggregate perception, and develop theoretical and computational frameworks that can explain and take advantage of the limits of visual coding of these different kinds of "stuff".

Acknowledgements

We would like to thank the study subjects for their time and patience, and the members of the Cornell Program of Computer Graphics for valuable discussions and feedback. This work was supported by NSF CAREER 0644175, 0615240, 0403340, and grants from Intel Corporation and Microsoft Corporation.

References

BACON, W. F., AND EGETH, H. E. 1991. Local processes in preattentive feature detection. *J. Exp. Psych.: Human Percep. and Perf. 17*, 77–90.

BECK, J. 1972. Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology 85*, 1–19.

BECK, J. 1982. *Textural segmentation*. Erlbaum, Hillsdale, NJ, 285–317.

BERGEN, J. R., AND ADELSON, E. H. 1988. Early vision and texture perception. *Nature 333*, 6171, 363–365.

BROWNBILL, A. 1996. *Reducing the storage required to render L-system based models*. Ph.D. thesis, University of Calgary.

COHEN, J., OLANO, M., AND MANOCHA, D. 1998. Appearance-preserving simplification. In *SIGGRAPH '98*, ACM, New York, NY, USA, 115–122.

COOK, R. L., HALSTEAD, J., PLANCK, M., AND RYU, D. 2007. Stochastic simplification of aggregate detail. *ACM Trans. Graph. 26*, 3, 79.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97*, 369–378.

DESIMONE, R., AND DUNCAN, J. 1995. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience 18*, 193–222.

DEUSSEN, O., HANRAHAN, P., LINTERMANN, B., MĚCH, R., PHARR, M., AND PRUSINKIEWICZ, P. 1998. Realistic modeling and rendering of plant ecosystems. In *SIGGRAPH '98*, 275–286.

DUNBAR, D., AND HUMPHREYS, G. 2006. A spatial data structure for fast poisson-disk sample generation. *ACM Trans. Graph. 25*, 3, 503–508.

FERWERDA, J. A., PELLACINI, F., AND GREENBERG, D. P. 2001. A psychophysically-based model of surface gloss perception. In *Proceedings of the SPIE: Human Vision and Electronic Imaging VI*, vol. 4299, 291–301.

FLEMING, R. W., TORRALBA, A., AND ADELSON, E. H. 2004. Specular reflections and the perception of shape. *Journal of Vision 4*, 9, 798–820.

FOUND, A., AND MÜLLER, H. J. 1995. Searching for unknown feature targets on more than one dimension: further evidence for a 'dimension weighting' account. *Perception and Psychophysics 58*, 1, 88–101.

GURNSEY, R., AND BROWSE, R. A. 1989. Asymmetries in visual texture discrimination. *Spatial Vision 4*, 31–44.

HART, J. C. 1992. The object instancing paradigm for linear fractal modeling. Proceedings of Graphics Interfaces, 224–231.

HEAPS, C., AND HANDEL, S. 1999. Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance 25*, 2, 299–320.

ITTI, L., AND KOCH, C. 2001. Computational modeling of visual attention. *Nature Reviews Neuroscience 2*, 3 (Mar), 194–203.

JULESZ, B. 1962. Visual pattern discrimination. *IRE Transactions on Information Theory 8*, 84–92.

JULESZ, B. 1981. Textons, the elements of texture perception, and their interactions. *Nature 290*, 91–97.

KHAN, E. A., REINHARD, E., FLEMING, R. W., AND BÜLTHOFF, H. H. 2006. Image-based material editing. *ACM Trans. Graph. 25*, 3, 654–663.

KOCH, C., AND ULLMAN, S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology 4*, 219–227.

KOFFKA, K. 1935. *Principles of Gestalt Psychology*. Harcourt, Brace and World, New York, NY.

LEE, C. H., VARSHNEY, A., AND JACOBS, D. W. 2005. Mesh saliency. *ACM Trans. Graph. 24*, 3, 659–666.

LUEBKE, D. P., AND HALLEN, B. 2001. Perceptually-driven simplification for interactive rendering. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, Springer-Verlag, London, UK, 223–234.

LUEBKE, D., REDDY, M., COHEN, J., VARSHNEY, A., WATSON, B., AND HUEBNER, R. 2002. *Level of Detail for Computer Graphics*. Morgan Kaufmann.

MALIK, J., AND PERONA, P. 1990. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A 7*, 923–932.

MANTIUK, R., DALY, S., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Predicting visible differences in high dynamic range images - model and its calibration. In *Proceeindgs of the SPIE: Human Vision and Electronic Imaging X*, vol. 5666, 204–214.

MARR, D. 1976. Early processing of visual information. *Phil. Trans. of the Royal Society London B 275*, 483–519.

NOTHDURFT, H. C. 1985. Sensitivity for structure gradient in texture discrimination tasks. *Vision Research 25*, 1957–1968.

OLIVA, A., AND TORRALBA, A. 2001. Modeling the shape of a scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision 42*, 3, 145–175.

OLIVA, A., MACK, M. L., SRESTHA, M., AND PEEPER, A. 2004. Identifying the perceptual dimensions of visual complexity of scenes. 26th Annual Meeting of the Cognitive Science Society.

PALMER, S. E. 1999. *Vision science: From Photons to Phenomenology*. Bradford Books/MIT Press, Cambridge, MA.

PERLIN, K. 2002. Improving noise. In *SIGGRAPH '02*, ACM, New York, NY, USA, 681–682.

POTTER, M. C. 1975. Meaning in visual search. *Science 187*, 4180, 965–966.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual equivalence: towards a new standard for image fidelity. *ACM Trans. Graph. 26*, 3, 76.

RAO, A. R., AND LOHSE, G. L. 1993. Identifying high-level features of texture perception. *Graphical Models and Image Processing 55*, 218–233.

ROSENHOLTZ, R., LI, Y., AND NAKANO, L. 2007. Measuring visual clutter. *J. Vis. 7*, 2 (8), 1–22.

RUSHMEIER, H., ROGOWITZ, B. E., AND PIATKO, C. 2000. Perceptual issues in substituting texture for geometry. In *Proceedings of the SPIE: Human Vision and Electronic Imaging V*, vol. 3959, 372–383.

SAGI, D., AND JULESZ, B. 1987. Short-range limitation on detection of feature differences. *Spatial Vision 2*, 1, 39–49.

SMITH, A. R. 1984. Plants, fractals, and formal languages. In *SIGGRAPH '84*, 1–10.

SUTHERLAND, I. E. 1963. A man-machine graphical communication system. Proceedings of the Spring Joint Computer Conference.

THEEUWES, J. 2004. Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin and Review 11*, 1, 65–70.

TODD, J. T., NORMAN, J. F., KOENDERINK, J. J., AND KAPPERS, A. M. L. 1997. Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception 26*, 7, 807–822.

TREISMAN, A., AND GELADE, G. 1980. A feature integration theory of attention. *Cognitive Psychology 12*, 97–136.

VANGORP, P., LAURIJSSEN, J., AND DUTRÉ, P. 2007. The influence of shape on the perception of material reflectance. *ACM Trans. Graph. 26*, 3, 77.

WALTER, B., FERNANDEZ, S., ARBREE, A., BALA, K., DONIKIAN, M., AND GREENBERG, D. P. 2005. Lightcuts: a scalable approach to illumination. *ACM Trans. Graph. 24*, 3, 1098–1107.

WALTER, B., ARBREE, A., BALA, K., AND GREENBERG, D. P. 2006. Multidimensional lightcuts. *ACM Trans. Graph. 24*, 3, 1081–1088.

WARD, G. J. 1992. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92*, 265–272.

WATSON, A. B., AND PELLI, D. G. 1983. Quest: a bayesian adaptive psychometric method. *Perception and Psychophysics 33*, 2, 113–120.

WATSON, B., WALKER, N., AND HODGES, L. F. 2004. Suprathreshold control of peripheral LOD. *ACM Trans. Graph. 23*, 3, 750–759.

WICHMANN, F. A., AND HILL, N. J. 2001. The psychometric function ii: Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics 63*, 8, 1314–1329.

WOLFE, J. M. 1998. *Visual search*. University College London Press, London.