In Proc. NPAR 2010

Non-Photorealistic Rendering and the Science of Art

Aaron Hertzmann
University of Toronto

Abstract

I argue that Non-Photorealistic Rendering (NPR) research will play a key role in the scientific understanding of visual art and illustration. NPR can contribute to scientific understanding of two kinds of problems: how do artists create imagery, and how do observers respond to artistic imagery? I sketch out some of the open problems, how NPR can help, and what some possible theories might look like. Additionally, I discuss the thorny problem of how to evaluate NPR research and theories.

1 Introduction

Why does art work? How can a few lines or blobs drawn on a piece of paper express motion, form, mood, and emotion? How does an illustrator use lines to convey a 3D shape with greater clarity than would a photograph, even though these lines seem entirely unlike what we see in the real world? Despite centuries of interest in these questions, they remain unanswered. These questions have been raised in many disciplines, including art history, psychology, cognitive science, and neuroscience. Scientific study of these questions could yield fundamental insights into art and illustration, shed light on the workings of the human visual system, lead to new ways to create art, illustration, and design, and lead to new kinds of art.

One way to understand how a complex system works is to attempt to design and build one like it. Indeed, some insights in understanding human intelligence have directly resulted from attempts to design machine intelligence. In constrast, mainstream analyses of art—whether from art criticism or neuroscience—are highly descriptive (e.g., [Cavanagh 2005; Kemp 2009; Ramachandran and Hirstein 1999; Zeki 1998]). Art critics and neuroscientists alike might discuss general features of a painting, or trends across paintings, but without providing recipes for when these trends appear or how they are combined to create images. Image processing methods that analyze artistic style (e.g., to detect forgeries) use linear filter responses that give little insight into the style itself. Likewise, anyone who looks to books of art instruction and criticism for tips on how to build an artistic rendering algorithm will be very disappointed, because they lack a sufficiently precise language for implementation.

In this paper, **I argue that Non-Photorealistic Rendering (NPR) research will play a key role in the scientific understanding of visual art and illustration,** and sketch out some possibilities for what this understanding might look like. Attempting to create algorithms for artistic imagery forces us to think about imagery in new ways. Conversely, when an NPR algorithm is successful—even if it seems entirely ad hoc—it provides an opportunity to understand *why* it is successful, and, what, if anything, it has in common with how humans make imagery. NPR is complementary to other ways of understanding art: in the future, I believe we will see more interplay between the analysis of art and synthesis. This research must ultimately be interdisciplinary, involving NPR and other fields such as neuroscience, cognitive science, and art history. Conventional descriptions of art (and how to make it) lack a precise language for describing art, and NPR could provide that language. I believe that the development of NPR will be essential to the scientific understanding of art, and thus there exists a great opportunity for our field to contribute to human knowledge.

Historically, theory and experiment have been described as the two pillars of scientific research. In the past few decades, computer simulation has become the third pillar of modern science [Denning 2007]. For scientific study of art, NPR will provide this third pillar.

There are two types of scientific questions about art that NPR can help address. The first is:

How do artists create imagery?

For example, can we describe an artist's style as a functional mapping from a photograph or 3D scene to an image? Second,

How do observers respond to artistic imagery?

For example, can we quantify the appeal or effectiveness of a composition? It is very hard to imagine being able to answer these questions without the benefit of the computational tools of NPR.

My focus here is on 2D visual, representational art, graphic design, and illustration, which I will often just refer to as "art" for short. I mostly discuss representational painting and drawing, since these are the areas I've worked in most, as a researcher and as an art student. However, these same ideas are relevant for many kinds of art. I take a very loose definition of art; just about anything creative or expressive can be considered art. However, NPR is most relevant to understanding art for purposes such as illustration, animation, decoration, formal beauty, and so on.

Even for the reader solely interested in building tools—which is the work most of us in the field are engaged in—I think it is nonetheless useful and provocative to think about NPR as building theories of art. Many of these questions are long-term questions that could take years or even decades to address, if they are even solvable at all. Nonetheless: there is enough excellent initial research—both algorithms and experimental studies—to be optimistic that these problems are well worth tackling. Furthermore, these ideas can suggest a number of possible directions for future research, and I will suggest several open problems in NPR along the way.

This paper begins with a general framework for discussion: how can NPR build theories of art? What kinds of theories are there? I then discuss the question of experimental studies and evaluation. This problem is one of the most difficult that confronts the NPR community today, and this section is relevant to modern NPR research, as well as the long-term vision of theories of art. In the remaining sections of the paper, I sketch out several possible types of NPR theories of art, including approaches based on theories of optimization and computational neuroscience. Along the way, I describe a little about how these theories are used in other fields. This is both to provide background information, but also to provide examples of how mathematical theories are used in other sciences.

http://www.dgp.toronto.edu/~hertzman/ScienceOfArt/

(a) (b)

Figure 1: ***Some simple but effective ideas in NPR,*** *for cartoon illustration, painting, and line drawing. The algorithms involved are based on simple mathematical or algorithmic ideas, and, given a set of high-level parameters, run completely automatically. While there is much to criticize in these results, I believe it is nonetheless amazing and intriguing that they can be achieved by simple and automatic algorithms. (a) A photograph processed by the algorithm of Winnemöller et al. [2006], and the variant of Haeberli's method [1990] from Hertzmann [1998]. (b) A suggestive contour line drawing [DeCarlo 2003] of a 3D model, and stylized versions [Goodwin 2007].*

2 NPR theories of art

Before discussing specific theories, I begin by describing some ways that NPR could be used to form theories of art, and the types of questions that NPR can help answer.

2.1 How do artists create imagery?

I begin with a simple example of how an NPR algorithm can form the basis of a scientific theory. Consider an automatic version of Haeberli's image-based painting algorithm [1990] (see also Litwinowicz [1997]). Haeberli showed that, just by sampling a random set of points in a photograph and placing simple brush strokes over them, you can get something that, at first glance, resembles an Impressionist painting. Strokes can be aligned to the normal of the image gradient to better match the colors of the source image. This procedure is so simple, but so remarkably effective, that it leads one to ask: is there something there? Does this algorithm have anything common with how humans paint? In effect, Haeberli's algorithm represents a theory of how humans paint. The theory is very coarse: no one could really be convinced of the literal truth of the theory, since it cannot recreate or predict any real human paintings. But it provides a starting point: given this theory, we can evaluate it, and ask how it is successful and where it fails, and use these observations to refine this theory. There are numerous other examples in NPR that are both simple and effective (Figure 1), and lead one to wonder what insights they give about art.

The key ingredient missing from most NPR research so far is experimental study of art: we must test theories against data, not just look at the pictures. The groundbreaking paper of Cole et al. [2008] entitled "Where Do People Draw Lines?" provides an excellent demonstration of how to evaluate line drawing algorithms against human data. This work provides a bridge from the recent progress in line drawing algorithms to future work in this area. Similarly, Schmidt et al.'s [2009] recent experiments indicate that some widespread assumptions about how illustrators draw 3D shape are, in fact, incorrect.

2.2 How do observers respond to artistic imagery?

A second way that NPR research can contribute to the study of art is by modeling how human observers respond to artistic imagery. This includes both perception of shape, but also evaluating aesthetics and other properties of images. Indeed, there has been some initial progress in designing objective functions for aesthetics (e.g., [Santella et al. 2006]) and clarity of communication (e.g., [Agrawala and Stolte 2001]), as well as perceptual studies of how we view drawings [Cole et al. 2009]. We as a community have much more expertise designing such models than evaluating them. I will discuss the uses of optimization models more in Section 4.2.

There are many different types of viewer response. A work may be viewed as beautiful; it may convey the perception of world quantities such as shape, materials, and motion; it may interest or intrigue a viewer; it may instill in the viewer various emotions such as pleasure or boredom. It is important to distinguish between these responses, e.g., a work that is interesting is not necessarily beautiful.

2.3 Degrees of abstraction

NPR theories of art may operate at various degrees of abstraction. Few, if any, algorithms can be said to model the *process* by which artists operate, e.g., describing the order in which an artist paints strokes or places tiles. Instead, they model the *mapping* from inputs to outputs. One may even describe the *goals* first and foremost, e.g., via an objective function, with the choice of optimization algorithm then being, ideally, incidental. These kinds of abstractions are essential for allowing us to reason about high-level features of art without getting bogged down in details like how the artist's sensorimotor system controls the muscles in their hand.

2.4 Descriptive vs. generative theories

It is worth distinguishing between theories that are *descriptive* and those that are *generative.* Descriptive theories explain general fea-

tures of artistic imagery, without providing recipes for how to create it; generative theories describe how to create images.

Most scholarship in art history is descriptive, as are previous works in the neuroscience and psychology of art; moreover, many descriptive theories are also *qualitative*. Descriptive theories can be simple and general, and may provide useful insights, but are usually incomplete descriptions of the phenomena. For example, Ramachandran and Hirstein's "peak-shift principle" [1999] explains a possible role of exaggeration in artwork in terms of general observations that may apply broadly across many types of art works, but without making very specific predictions about any individual work.

On the other hand, NPR theories are, by their very nature, generative: they describe how to create specific types of imagery. However, they do not apply as broadly to many different styles. A major goal of applying NPR to understand art is to create generative theories: if we can explain how to generate something—or to compute the viewer's response—then we have a much better claim to saying we understand it. NPR could be a great source of possible theories about art.

Descriptive and generative theories can support and refine each other in various ways. Descriptive models give hints as to factors to take into account when creating NPR algorithms. They can even be used to define terms in an energy function for optimization methods. They provide explicit things to look for when comparing artist-made imagery to computer-generated imagery. Conversely, the trial-and-error process of designing generative models can lead to insights about what factors are necessary to describe artwork. In Section 3.6, I describe our experience with developing a model for line thickness, in which descriptive and generative models fed into each other.

Descriptive theories may also be quantitative. For example, several authors have used image statistics to quantify and classify painting style [Hughes et al. 2010; Lyu et al. 2004; Taylor et al. 1999]. One application of this work is in detecting forgeries, a task for which generative models are not strictly necessary. Taylor et al. [1999] discovered that a range of Jackson Pollock's abstract drip paintings are fractal, i.e., their statistics are self-similar at multiple scales. This observation is interesting because many natural images have the similar fractal properties, which suggests a possible explanation for the appeal of his work. However, the observation does not tell us how to create Pollock paintings: there are many classes of self-similar images that look nothing like Pollock paintings, and, so far, the theory has not been explored carefully, e.g., it has not yet been tested whether "drip" paintings without these fractal properties look less appealing or less "Pollockesque."

3 Evidence and arguments

Presenting research results requires convincing one's audience and oneself of the value of the work. When developing algorithms, we want to show that the algorithms are useful; when developing theories, we must persuade the reader that the theory is likely to have some element of truth. One must make plausible arguments, and provide evidence that support the arguments.

Broadly speaking, most current NPR algorithms are aimed at one of two types of goals: *task-oriented goals* are to communicate specific information such as shape and part relationships, and *artistic goals* are to create images that are beautiful or expressive, and/or that mimic existing artistic styles. Methods can be tested against these goals using methodology from the Human-Computer Interaction (HCI) and perceptual psychology literatures, e.g., [Agrawala and Stolte 2001; Cole et al. 2009]. Task-oriented goals are easier to define and to measure carefully, as they can easily be defined objectively and measurably. Defining the artistic goals in a useful and measurable way is much more difficult.

In the long run, developing good experimental methods for evaluation of algorithms and validation of theories will be crucial for increasing the scientific merit of our work. Our current application of experimental methods is inconsistent and varies in quality. We must develop new experimental methodologies, specific to our field, and we must expand our experimental enterprise. However, we must also become aware of the dangers of inappropriate experimental methods. An uncritical reliance on quantitative evaluation could be disastrous for the field.

3.1 Current use of studies

In most NPR papers, evaluation consists solely of demonstrating algorithmic results, possibly with comparisons to results from earlier methods. When the improvement over previous results is visually obvious, little further evaluation seems necessary. This was especially common in the early days of the field. For example, the early work of Haeberli [1990] and of Winkenbach and Salesin [1994] demonstrated very clear advances over what little prior work had come before. As the field becomes more mature, it becomes more difficult to demonstrate a major improvement over previous methods, and the visual evaluation becomes subtler. Furthermore, many of us are sensitive to the criticism that computer graphics is just a bunch of hacking and pretty pictures, and, thus, unscientific and non-rigorous. Hence, there is a push for more rigorous evaluation and testing. Consequently, many authors of NPR (and computer graphics papers in general) have turned to user studies to provide means of evaluation, even for aesthetic algorithms.

Experimental studies are extremely appealing as a way to add scientific rigor to research papers. NPR research has, so far, largely borrowed two main types of methodologies from other fields. *Perceptual studies*, from the psychology literature, aim to elicit rules of human perception by careful presentation of specific stimuli, such as perceived 3D surface orientation under conflicting cues. *User studies*, from the HCI literature, evaluate user performance on performing some task, such as the efficiency of a particular type of menu. Many types of studies are reductionist, often boiling down the effect of an image or rendering to a single number. This reductionism is necessary for performing controlled experiments. However, reductionism can also be antithetical to creativity and exploratory research.

3.2 The dangers of studies

We should learn as much as possible from the collective experience of related fields when developing our own experimental methodology. The field of HCI is more active and more mature than NPR, and may provide some lessons for the way NPR is headed. Like NPR, HCI has a significant component of designing new systems and techniques, and it has significant overlap with both design and psychology. Experimental evaluation plays a major role in most HCI papers. However, in a recent paper entitled "Usability Evaluation Considered Harmful (Some of the Time)," Greenberg and Buxton [2008] argue that HCI's over-reliance on user testing is damaging to the field. As one example, Greenberg and Buxton argue that it is very difficult to publish HCI papers without experimental results, particularly quantitative results. This biases research towards problems with established evaluation metrics. For example, there are many publications on pointing and selection, where one can compare methods by measuring time to select a desired target. More speculative but creative research is harder to publish. Greenberg and Buxton make many other relevant points; I urge anyone interested in these issues to read their paper. Kosara et al. [2003] also

discuss the benefits and dangers of using studies in the context of visualization.

An uncritical reliance on testing can be very dangerous to the field of NPR, for several reasons. First, as noted above, an uncritical demand for experimental evaluation can make it difficult to publish creative, exploratory research. (It has often been observed that institutions tend to optimize for what is easily measurable rather than what is really important.) Second, even in cases that are measurable, the requirement to publish studies puts additional burden on the authors. Performing meaningful studies requires significant effort and specialized expertise, and thus slows down progress while also erecting barriers to newcomers to the field. We should only require this effort in cases where it is really warranted.

Within an individual paper, poorly-designed studies may be uninformative or even severely misleading. The use of significance statistics (e.g., p-values) can give the false appearance of objectivity; we must remember that every study is limited by the sample size, the quality of the experimental protocol, and the assumptions of the significance test itself. Designing experimental protocols that yield meaningful results—and not just confirm the experimenter's preconceptions, even unintentionally—requires expertise and is subject to numerous non-obvious pitfalls. At our discussion at NPAR 2009, Doug DeCarlo's advice to researchers wishing to begin performing studies was: "Get help." That is, you should seek out colleagues with proper training who can collaborate in designing meaningful experiments.

At present, I do not believe that we as a community do have a good sense for how to make effective use of studies in papers. Anecdotally, as a papers committee member and reviewer, I have witnessed cases where both authors and reviewers seem to treat studies as perfunctory but necessary "check box." Authors may include a study in an attempt to inoculate against reviewer criticism, without taking care to ensure that the study is meaningful. Worse, I have seen reviewers write statements to the effect "I don't think the results look very good; your paper might be more convincing with a user study," as if experimental evaluation would be persuasive when the images are not. To paraphrase the Marx Brothers: who are you going to believe, a study, or your own lying eyes?

3.3 Building a convincing case

Presenting research requires building a coherent argument for the methods or theories. An author must attempt to persuade the reader of their theory, marshaling whatever arguments and evidence are necessary. The key point about studies is that

Experimental studies provide *evidence*, not *proof*

In other words, a study can provide evidence for the existence of an effect, or the superiority of one method over another on a particular metric, but it cannot prove these things. Some studies are more persuasive than others: in medical sciences, studies are judged according to their methodological rigor (e.g., whether the experiment is randomized and double-blind) as well as the size of the sample set and how it was selected.

In NPR research, a convincing case can often be made without the use of studies. To date, our evaluations have largely been based on providing imagery and relying on the aesthetic judgement of the reader/reviewer. This has served us well so far — we would not have made nearly so much progress had we insisted on quantitative user studies for every paper. But relying solely on arguments and pictures is not sustainable. Careful experimentation will be essential if we wish to produce results that yield fundamental insights and stand the test of time.

As in fields such as physics and HCI, we should allow some separation between exploration and evaluation. We need to explore and develop many new ideas, and later sort out which are most successful through testing and evaluation. These tasks need not be performed by the same researchers, and some may reasonably be more skilled or interested in one of these tasks or the other[1].

3.4 How do we evaluate aesthetics?

This raises the difficult question of how we judge methods designed for artistic goals, such as beauty or expressiveness, since these qualities are not easily quantified. What studies do we perform? How do we convince readers, reviewers, and ourselves that one work achieves an artistic goal better than another? How do we explore and test theories about how artists create images? Measuring subjective qualities of art — beauty, appeal, emotional expression — is a tricky business. Responses may vary from culture to culture, and from viewer to viewer. The question of how to perform these evaluations is itself arguably one of the most important research problems that we face today. To date, there have been few attempts to directly compare viewers' subjective aesthetic assessments, including [Isenberg et al. 2006; Liu et al. 2010; Santella et al. 2006].

Proxy metrics. Due to the difficulty of directly evaluating aesthetics, some researchers have assessed related measures. Santella and DeCarlo [2004] were among the first to argue for the importance of evaluation. They performed experiments to determine how their image stylization algorithm [2002] affects the viewer's eye gaze. Similarly, Gooch et al. [2004] and Winnemöller et al. [2006] measured the effect of image stylization algorithms on memory and learning tasks. In each case, these tests were interesting and informative, and supported the stated goals of the paper. But they did not directly measure the "artistic" qualities of the imagery, nor did they claim to.

Here, I call these tests *proxy metrics*: they do not directly test aesthetics, but they do test measurable properties that are relevant to evaluating art. In order for a proxy metric to be meaningful, we must gather baseline measurements from humans on real artworks and whether or not it is discriminative of types of images, or predictive of some quantity of interest. How do real artworks direct gaze or improve memory performance, as compared to photographs or photorealistic renderings? Is there any correlation between a human's assessment of the beauty of a work with fractal dimension or memory performance? Does a good score on one or several of these metrics usual imply a good score on another? Functional MRI [Kawabata and Zeki 2004] or other biophysical measurements may also provide ways to link proxies to the desired quantities. If an algorithm scores highly with a proxy metric, that provides at best weak evidence for the artistic success of the algorithm. Proxy metrics may actually be more useful when scores are bad, because they can distinguish among competing methods and point to opportunities for improvement.

Fractal dimension statistics (mentioned in Section 2.4) are another example of a proxy metric. Lee et al. [2006] describe a method for mimicking Pollock's style. They show that their method does successfully mimic the fractal dimension of his paintings, even though it was not specifically designed to do so. This does not prove the correctness of their algorithm, but supports it: if an algorithm produces images with a different fractal dimension, this is evidence of a mismatch with Pollock's style. However, it is only evidence, not

[1] The quantum theorist Wolfgang Pauli was so bad at performing experiments that it was said that merely his presence in the vicinity of an experiment would cause it to fail. This did not prevent him from winning a Nobel Prize for his theoretical contributions.

proof, because we do not yet know if fractal dimension is an informative measurement for artistic imagery; testing this with human observers and appropriate controls remains to be done.

Asking artists. One often hears the suggestion that one evaluate NPR results by "asking artists" for feedback. These suggestions completely underestimate the difficulty of doing so in a meaningful way. In my experience, the range of opinions and preferences of professional artists is so broad as to be useless for evaluation. The term "artist" itself is nearly meaningless, encompassing, for example, individuals who exhibit art in avant garde galleries, highly-trained commercial designers and animators, "outsider" artists, calligraphers, and people who paint graffiti on buildings. Even within one of these groups, tastes and preferences may vary widely. One must be wary of the usual difficulties of non-controlled experiments, e.g., the well-known "good subject effect," where people generally like to tell their questioners what they want to hear. Artists working with computer tools sometimes even prefer the visual artifacts and "computery-ness" that we researchers are working so hard to eliminate.

Informal feedback from artists can be extremely valuable when developing ideas and techniques. However, it should not be mistaken for evaluation, and can be extremely misleading if presented as such. Ideally, artist evaluations should be obtained using the same careful methodologies as any other experimental studies.

3.5 Evaluating theories

How do we test theories of how artists create images? The work of Cole et al. [2008] provides one intriguing example along this direction. In some cases, we can attempt to predict or regenerate existing works, being careful to avoid over-fitting. Proxy metrics provide another tool for comparing NPR results to traditional artworks. In the next section, I describe a complementary approach.

3.6 A case study and a proposal

In our NPAR 2007 paper [Goodwin et al. 2007], we described a formula for stroke thickness for line drawings of 3D shapes. We conjectured that this approach reveals something about how artists use line thickness. In this section, I will describe the approach we used to evaluate our hypothesis and to make an argument to support it. In doing this work, we developed a new methodology for explaining art works in terms of our theory. This methodology or variants could prove very useful for other NPR algorithms.

Our line thickness formula is based on a quantity we called Isophote Distance; see [Goodwin et al. 2007] for details. We hypothesized that

> Stroke thickness in many real-world line drawings and paintings is proportional to Isophote Distance, clamped to a fixed range of values.

While developing the work, we were met with lukewarm responses when we showed our initial results to colleagues (for the most part, computer science students in our lab). They could not tell whether we were on the right track or not. In part, this was because our preliminary renderings were not very polished. But, more importantly, line thickness is a subtle property of drawings, and there is a lot of inconsistency in line drawing style. Hence, we needed some way to "validate" or justify the method.

Our first approach was to try to recreate human-made line drawings. We created a 3D model of a character from the "Bone" graphic novels, and a 3D model of a dung beetle from a technical illustration manual. We then rendered both models with our algorithm, and, visually, we found both to match the original drawings (Figure 2). One could argue that we may have "overfitted" by designing the 3D model to match the 3D rendering, though the 3D models look visually plausible. A more rigorous test would be to attempt to predict how each artist would draw these models from another viewpoint, but no such data was available for these artists. Nonetheless, it is promising that our very simple formula was able to fit two very different examples with reasonable accuracy.

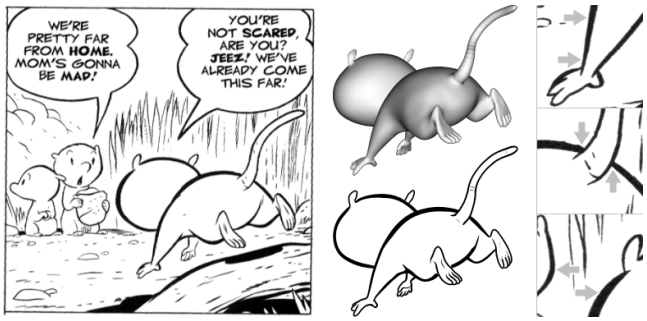

Figure 2: ***Left:*** *Illustration from "Bone" [Smith 1998] (BONE® is Copyright © 2010 Jeff Smith; image used with permission.).* ***Middle:*** *A 3D model of the possum we created, and a rendering of the possum with our algorithm described in [Goodwin et al. 2007].* ***Right:*** *The line thicknesses in the drawing exhibit a number of properties that can be derived from the Isophote Distance formula. These are rules that can be checked without knowing the 3D geometry. (top) Because the wrist has higher curvature than the upper arm, it has thinner strokes. (middle) Occluded strokes are not tapered. (right) Distant objects have thinner strokes than nearby objects of the same type.*

A broader survey. Throughout the course of this project, we looked at outlines in drawings and paintings from many sources, including from many kinds of art and animation. At first, this was just to determine whether or not these lines seemed to match or model at all, to determine if it was worth proceeding with the project. While we found many promising examples that seemed to fit, we had little hope of reproducing all these styles with our method, for two reasons. First, these examples had many different effects, such as sketchy strokes, for which we had no model. Second, it was clear that, in many cases, the artist-drawn strokes were inconsistent with our model. These artists were either following some additional rules we had not yet deduced, or else there was significant randomness or sloppiness that we could not easily model.

Nonetheless, we still believed there was some "truth" to our initial hypothesis, even if the hypothesis would need to be weakened or modified. We observed that, even if we could not *fit* the data, our method could suggest *qualitative properties* or *rules* that many of the examples did seem to follow. This observation grew out of the many hours initially spent looking at artworks, trying to determine whether it was worth conducting the project in the first place.

Without knowing the geometry in any given scene, it is not possible to directly compute Isophote Distance. However, up to a first-order approximation, Isophote Distance is inversely proportional to object depth and to radial curvature. Hence, when we could find two identical objects at different depths, we expect the more distant one to have smaller strokes (Figure 2). Furthermore, for most people and animals, the legs have lower curvature than the arms, so the legs should have thicker strokes than the arms. A more subtle consequence is that, in 3/4-view drawings of human faces, stroke

thicknesses become thinner above the cheekbone. We formulated roughly a half-dozen such rules, all of which can be tested simply by looking for specific relationships in drawings, without knowing 3D geometry.

In surveying our collected artistic imagery, we were able to find a broad range of examples that matched our rules, including examples from technical illustrations, paintings, cartoon animation, comic books, and so on[2]. Some of rules were very widely followed, including the rules about distant objects and cheekbones. Some we were unable to find much evidence for or against, particularly the rule that foreshortened objects should have thicker strokes.

These qualitative properties suggest a generalized version of our original hypothesis:

> Artists often use stroke thicknesses with the same relationships as determined by Isophote Distance.

In other words, let $\mathbf{p}$ and $\mathbf{q}$ be two different points on strokes, and let $d_I(\mathbf{p})$ and $d_I(\mathbf{q})$ be their Isophote Distances. Then, the stroke thickness at $\mathbf{p}$ should be thicker than that at $\mathbf{q}$ if and only if $d_I(\mathbf{p}) > d_I(\mathbf{q})$. This new hypothesis does not predict the actual ratio of thicknesses. Exploring this new hypothesis remains future work.

Discussion. Our survey has distinct advantages and disadvantages. We made no attempt to be systematic in our selection of examples, and we could easily be convicted of the crime of confirmation bias, as we only picked examples that support our argument. We have no quantitative measurements or statistical way to test significance. For these reasons, our study should not be taken as a rigorous scientific study that resolves the question.

Nonetheless, it is hard to imagine another way in which one could make a persuasive case about such a broad range of existing artworks. Nor would one expect to see a study like this in the typical literature about art. Ultimately, our approach is an attempt to persuade the reader that there is some truth to our hypothesis. Whether or not we were successful is for the reader to decide.

It would have also been premature to attempt a rigorous study in an area where so little is known about what artists do. Our work may be viewed as a form of "pilot" study; future work may refine or reject our theories and perform more rigorous studies.

It is worth noting that finding useful example imagery and studying was a significant effort in itself. We spent many hours poring over books, looking at art and instructional webpages, and freeze-framing animations searching for good examples and studying them.

The general methodology. This experience suggests a possible methodology for assessing NPR algorithms. For a given NPR algorithm or formula, we derive a set of properties, or rules, for how the algorithm produces images. These properties could be qualitative or quantitative, but they should be designed so that one can judge whether or not an existing artwork satisfies them. For example, it is easy to assess a rule that says that "leg strokes should be thicker than arm strokes," whereas, without knowing 3D geometry, it is impossible to judge a rule that says that stroke thickness should be inversely proportional to radial curvature. One can then assess to what degree these rules are or are not obeyed by existing artworks.

[2]The examples may be found in the slides at http://www.dgp.toronto.edu/~todd/isophote/.

4 Optimality theories

Having discussed some general topics related to NPR and science, for the remainder of this paper, I now switch to discussion of some particular possible directions. What would theories of art that make use of NPR look like?

One of the most powerful forms of abstraction in both computer science and biology is *optimization*. Expressing artistic imagery as the result of optimization allows us to abstract away details of the artistic process. More importantly, optimization provides a useful model for explaining how conflicting goals are resolved.

It might seem absurd to explain human behavior in terms of optimization of simple mathematical functions. However, researchers in many disciplines, most notably economics and biology, have demonstrated the usefulness of optimization in explaining aspects of biological systems and human behavior. It is not necessary for the optimization to be completely describe the system of interest: optimization is a useful *model*, and, like all models, it makes simplifying assumptions in order to make a problem manageable.

4.1 Optimization theories in biology

For context, it is helpful to be aware of the role that optimization has played in biology. Optimality theories of biological organisms have shown to be quite powerful; many biological structures and activities can be viewed as optimizing various quantities such as efficiency or energy consumption in search of evolutionary advantage [Alexander 1996; Alexander 2001; Parker and Smith 1990; Sutherland 2005]. Optimality theories allow us to explain what an organism (and/or evolution) is "trying to achieve" without explaining the precise mechanism by which it is achieved.

Optimality theories have been applied to a wide variety of organisms, from the structure of the body and bones, to different types of animal gait, to behavior [Alexander 1996; Parker and Smith 1990]. Modern theories of vision and neuroscience are based on optimal inferences and actions, a topic I discuss in more detail in Section 5.1.

Why would organisms be optimal? Optimization in biology can be thought of as arising from two related mechanisms. First, natural selection can be thought of as optimizing biological fitness, in order to allow the species to survive and flourish. Because species naturally compete with one another (e.g., competing for a particular niche, or in a predator-prey relationship), survival requires acquiring resources and producing healthy offspring with maximum efficiency. Second, an individual may optimize their behavior for new situations, i.e., a human can learn new actions, and, with repeated practice, can make their performance of the action better and better. The ability to learn is itself a product of evolution.

For the time being, optimality arguments for art are unlikely to directly reference evolution. Even for traits that are obviously relevant for survival such as, say, bone density or running speed, biologists have understandable difficulty making a direct mathematical link to the survival of the species; informal arguments must be used. The evolutionary rationale for art is even more unclear, and remains a topic of speculation (e.g., [Dutton 2009]). Instead, optimality models for art may appeal to the extensive training and practice that an artist conducts in order to hone their skills, and, in some cases, the many iterations required to refine a work.

In the biological literature, optimization approaches are sometimes controversial, since biological organisms are not *really* optimal. We may be near local rather than global optima, the objective function may be ill-defined and ever changing, and many evolutionary changes may be due to random events rather than adaptation. The

classic and eloquent paper of Gould and Lewontin [1979], though predating most work on optimality theory in biology, provides the template for modern attacks on optimality theory. Nonetheless, in the pages of *Nature*, Sutherland argued *"There are increasing calls for biology to be predictive. Optimization is the only approach biology has for making predictions from first principles."* [2005]. That is, if we hypothesize a possible principle (e.g., animals select the diet that maximizes energy intake), we can then optimize the animal and its behavior to see if the predictions match the animal behavior. Again, optimality is a model that we can develop and test; it should not be mistaken for the complete truth.

4.2 Optimization models of art

Modeling artwork as the result of an optimization process allows us to think about *what are we trying to compute* while largely abstracting away *the steps required to compute it*. For example, correctly modeling the order in which an artist applies strokes to a painting may be very difficult, but, ultimately, incidental to goals of the final image. In contrast, many current NPR methods are described algorithmically: given an input, a sequence of processing steps are applied to the input data to produce an image. This makes it hard to reason about the goals of the work. The idea of using optimization in NPR goes back at least to Haeberli's seminal paper [1990], but optimization methods remain very much in the minority.

One can formulate image creation as optimizing a viewer's response [Durand 2002]. The basic hypothesis can be formulated as: **an artist creates an image to maximize a desired visual response in the viewer,** subject to the constraints of the artistic medium. That is, suppose we could model the human visual system's response r to a stimulus S as a mapping

$$r = V(S) \tag{1}$$

Then, producing an artwork is a matter of finding the stimulus S that minimizes the difference between $V(S)$ and some desired response r_0:

$$\min_S d(r_0, V(S)) \tag{2}$$

subject to the constraints of a particular medium, e.g., the stimulus must be created with oil paints. The response r might be a vector that includes attributes of the perceived shape, the aesthetic response, the content of a scene, and so on. For example, the goal might to be induce in the viewer a sense of space and shape that matches the actual shape in a scene.

Of course, an accurate model of the human visual system is not yet available. Instead, we must formulate approximate models, while also justifying and validating the approximation.

Practical methods. Stroke-based optimization methods in NPR can be expressed in terms of data-fitting or compression objectives, e.g., [Haeberli 1990; Hertzmann 2001]. As a simple example, consider the problem of using a collection of paint strokes P to represent a source image I. The goal is to produce an image that conveys the same scene as the photograph, but using a limited number of brush strokes, in order to "look like a painting." The objective function then has the form:

$$E(P) = D(P, I) + \lambda C(P) \tag{3}$$

The first term $D(P, I)$ is a measure of the "perceptual distance" between the source image and the painting. The second term $C(P)$ measures the "painterliness" of the image. Some current methods use an L_2-norm as a measure of perceptual distance: $D(P, I) = ||P - I||^2$. The second term will often enforce some kind of *parsimony* on the representation. For example, we want to avoid the situation that millions of tiny strokes are placed that exactly reproduce the input photograph. Hence, a simple choice of this term is to penalize the number of strokes used, or to enforce a hard limit on the number of strokes. The parameter λ controls the trade-off between these two terms. In effect, this formulation splits our original optimization problem (Equation 2) into two terms, even though the approximation is very weak. This same formulation can be applied whether one represents the image with paint strokes, stipples, mosaic tiles, collage elements, or glued-on macaroni noodles [Hertzmann 2003].

There is a natural parallel between this formulation and information theory and compression. In a compression problem, we seek to compress a signal I with an encoding P, so that we minimize the divergence between the signal and its reconstruction from the encoding, while also minimizing the number of bits in the encoding. At a higher level, information theory seeks optimal representations that can be used to compress many signals. Perhaps the evolution of artistic style can also be expressed in these terms: painting and drawing styles have evolved to be able to best express broad classes of visual scenes.

For task-oriented rendering, one can formulate an optimization by a set of objective terms that measure specific goals that are hypothesized to be relevant to the task. For example, the LineDrive system optimizes several properties of line-drawn maps that are believed to be important for the goal of aiding navigation [Agrawala and Stolte 2001]. Automatic cropping systems optimize several terms that are hypothesized to be important for good photo composition [Liu et al. 2010; Santella et al. 2006].

Validating and improving the objectives. One benefit of an optimization formulation is that we can test and improve the individual terms of the objective function, even basing them on data from human subjects. To date, NPR research has not done nearly enough of this. Experimentation, when performed at all, is normally done on a system as a whole. Developing and testing the individual terms should allow for faster progress in combining these terms to build better and more sophisticated systems. For example, most methods use an L_2-norm for the distance term $D(P, I)$, but comparing pixel-wise differences is very poor measure of perceptual difference. Better metrics, such as SSIM [Wang et al. 2004], which have been validated for comparing images in signal reconstruction problems could be used instead.

Line drawing. Some recent line drawing algorithms are described purely in terms of mathematical sets. Contours and suggestive contours [DeCarlo et al. 2003; Judd et al. 2007] are described in terms of zero-sets on smooth surfaces, or image-space ridges [Lee et al. 2007]. These forms may also be expressed as local minima, e.g., image-space local minima of shading. This suggests a variational principle for line drawing: find a minimal set of curves that cover the darkest regions on the image. Formalizing this precisely remains an open problem.

Cole et al. [2008] note that some commonly-drawn do not correspond to shading variations. These could be explained by a more general variational principle. Suppose, given a line drawing, a viewer mentally reconstructs a particular 3D shape. The goal of line drawing is to produce a minimal set of lines so that the desired 3D shape is reconstructed by the viewer.

4.3 Random sampling

Human artists are not deterministic. If you ask me to draw one hundred pictures of a specific person, each picture will be differ-

ent. The variation will be structured, with some regularities in the images, and some of the regularity will be relatively simple (e.g., wiggles in pen strokes), but some will not (e.g., choices of overall composition and selection of which lines to draw).

One possible generalization to optimality is to replace the objective function with a probability distribution, perhaps using a Gibbs' distribution. For any objective function $E(\mathbf{x})$, the corresponding Gibbs' Distribution over variables $\mathbf{x}$ is $p(\mathbf{x}) \propto \exp(-E(\mathbf{x})/T)$, where T is a temperature parameter (assuming the function is integrable). We can then produce an image $\mathbf{x}$ by randomly sampling from the probability distribution. As long as the temperature is not too high, all of the samples will still be close to optimal, but yet all different. The use of the Gibbs distribution allows us to directly generalize the optimality formulation, since probabilities are directly related to objective function values.

In fact, many current NPR algorithms do include random steps (e.g., adding random offsets to strokes to produce wiggles), thus describing distributions over imagery. However, using this distribution for anything other than random sampling is very difficult, e.g., it is very difficult to study individual terms in the probability distribution. Furthermore, implementing interactions between elements in such an approach can be much more difficult.

4.4 Performance issues

As a practical matter, many optimization methods are extremely slow. Some specialized cases admit tractable algorithms, e.g., generalizations of k-means [Secord 2002]. In most cases of interest, however, problems are highly nonlinear, high-dimensional, mixed continuous and discrete, and thoroughly laced with local minima. For example, my work in painting by optimization [Hertzmann 2001] entailed random perturbations with a complicated proposal mechanism, parallelized across a computing cluster, and still required hours to produce an image. Nonetheless, I believe that optimization methods are worth the trouble. Once we have a good objective function, we can devise faster optimization algorithms and approximation algorithms. Understanding the underlying optimization problem we really seek to solve can often be the first step toward formulating efficient algorithms or even good heuristics.

5 The computational neuroscience of art

Much of the discussion of the neuroscience of art has employed qualitative approaches to neuroscience. These can be quite fascinating and revealing, e.g., Ramachandran and Hirsteins' theories [1999], or Sacks' [1996] descriptions of the artwork of patients with bizarre neurological conditions. However, modern computational neuroscience and vision science can provide much more quantitative, computational models of how the human vision system works, and a number of useful insights and tools that, in combination with NPR, may be crucial to understanding art.

5.1 Computational neuroscience and vision today

Computational neuroscience aims to understand the brain's functioning mathematically and algorithmically: what is the computation performed by the brain? This includes both low-level processing (what does a neuron compute? what does V1 compute? what information does a spike train store?), as well as more high-level theoretical models that are distinct from how computations are performed [Marr 1982].

As with the optimality models discussed in Section 4, many scientists have studied the hypothesis that the brain is designed to process information optimally, and, ultimately, to take optimal actions. Optimal inference and decision-making under uncertainty is given by Bayesian inference and decision theory [Jaynes 2003]. Hence, it seems plausible to model the brain as performing Bayesian computations. In fact, many experiments have been performed in which human judgements match those of an optimal Bayesian algorithm, given an appropriate model. Such experiments have been explored in visual perception [Doya et al. 2007; Kersten et al. 2004; Rao et al. 2002; Weiss et al. 2002], in cognitive science [Chater et al. 2006; Griffiths and Tenenbaum 2006], decision-making [Dayan and Daw 2008], and in motor neuroscience [Körding and Wolpert 2004; Todorov 2004]. While humans are obviously not optimal in every way and every situation, low-level vision and motor control have been modeled by Bayesian models with great initial success. The current theory is by no means complete or even widely accepted, but it is the only present theory that makes a plausible claim to future success at explaining the brain.

In Bayesian models of visual perception, there are three quantities of interest: the measurements (i.e., the image we see), the unknown variables (e.g., representing shape, reflectance, identity, and motion), and prior knowledge we obtain from experience and evolution. The rules of probability theory instruct us how to optimally determine the value (or possible values) of each unknown variable, integrating out all uncertainty and factoring out all sources of noise. This allows our visual system to extract as much information as possible from the data, and to perform well under difficult situations, e.g., perceiving shape even under scotopic (nighttime) conditions, recognizing identity just from a moving dot pattern, and so on. See [Kersten et al. 2004] for a more detailed explanation of how the visual system may be modeled by probabilistic computations.

Such models can provide the viewer model needed in Section 4.2, with response r consisting of a probability distribution over properties of the scene being depicted.

5.2 Probabilistic perception of art

So what might the neuroscience of art be: how do we explain the human visual system's interpretation of artistic imagery? For now, this is largely a matter of conjecture, and I mainly focus on the explanatory power of a probabilistic account. Given whatever cues are provided by an image, the brain infers as many unknown variables as possible. For example, a person will interpret shape—but probably not color—from a line drawing; ambiguities will remain in areas that are left sketchy or empty. We can perceive an illustration in terms of real objects and scenes (rather than just a piece of paper with ink on it), because of our visual system's ability to handle ambiguities and multiple interpretations by performing Bayesian computations. We can recognize scenes from line drawings and paintings for the same reason, even though we never see objects as pen or paint strokes in the real world. We can abstract or leave out information because the brain will "interpret around" it; this allows an artist to omit irrelevant information and thus strengthen the desired interpretation of the image. The underlying mathematical mechanisms for these processes would be described in terms of Bayesian inference. If successful, such a model of the visual perception of art could provide comprehensive explanatory power for visual imagery that is quantitative and predictive.

5.3 Measuring and creating ambiguity

In most discussions of the role of uncertainty in probabilistic models (including the discussion above), uncertainty is a tool for dealing with noisy and ambiguous measurements. Here I conjecture that uncertainty itself plays a unique and irreplaceable role in describing our experience of art. Furthermore, if we can model the

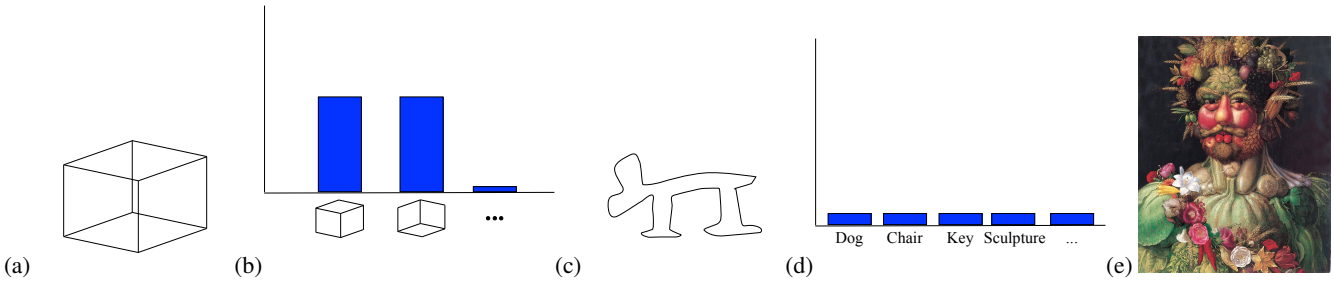

Figure 3: ***Vagueness vs. ambiguity.*** *(a) The Necker cube is ambiguous: it can be viewed in two very distinct ways. (b) Schematic illustration of a viewer's probability distribution over what the image depicts. There are two high-probability orientations of the cube. (c) A vague drawing. (d) No interpretation has high probability; probability is spread across many possible interpretations. (e) Vertumnus by Giuseppe Arcimboldo, 1591, illustrating a much more sophisticated example of ambiguity. The two interpretations of the Necker cube differ only in orientation: it's a cube either way. Here, the ambiguity is much more interesting, because the two main interpretations are so radically different in terms of shape, materials, object identity, and so on.*

uncertainties of the visual system, then we should be able to create NPR algorithms that specifically optimize for images that have a specified uncertainty — perhaps we could design an algorithm that automatically makes images that are beguilingly ambiguous and enigmatic.

Vagueness vs. ambiguity. To understand the importance of uncertainty, we can distinguish between a work that is *vague* versus one that is *ambiguous.* These terms are often applied to literary works as well as to imagery [Wikipedia 2010]. A vague image, like a blurry photograph, is one that is simply unclear: one cannot tell what is going on. In information theoretic terms, a vague image contains very few bits of information. In contrast, an ambiguous image admits two very distinct interpretations. Examples of ambiguity include the Necker cube (which side is nearer to the viewer?), and Arcimboldo paintings (is it a person or a pile of vegetables?) (Figure 3). Vagueness is often a sign of a poorly-executed work, whereas ambiguity is often a sign of a skillful work of art. Ambiguity in a book can leave you pondering and discussing the work long after you have finished reading it, whereas vagueness may have you giving up on the book without finishing it.

Ambiguous images are *interesting*, since we have a natural desire to resolve uncertainty and seek clarity. Often, we will puzzle over an image that is hard to interpret, but, once we figure it out, we lose interest. Ambiguous images defy solution: we keep searching for the unique answer, but the dilemma cannot be resolved. An ambiguous image is like a morsel stuck in your teeth that your tongue can't quite dislodge, but can't stop trying to. On the other hand, a vague image is not very interesting, because there is obviously little to be learned from studying it.

Ambiguity appears in many art works very simply in the tension between representation and the media itself. For example, a painting with distinct and clear brush strokes admits two very distinct interpretations: it is a bunch of paint, or it is an image of a 3D scene. All sorts of images, from Impressionist painting to Photomosaics [Silvers 1996], play with this ambiguity.

Probabilistic interpretation. I conjecture that we can explain these phenomena in terms of the viewer's probability distributions. When looking at a blurry picture of a face, the viewer's PDF over possible faces is vague: there are many different faces consistent with the picture. When looking at a Necker cube, there are two separate peaks in the viewer's PDF over the cube's orientation. When looking at an Arcimboldo painting, the ambiguity happens at multiple levels: the joint probability distribution between many different variables (is it a person or a vegetable? is the surface shiny or rough? etc.) has two distinct spikes.

It has been hypothesized that visual interpretations are samples from the probability distribution (e.g., [Daw and Courville 2007]). With a Necker cube, or Arcimboldo painting, at any given instant, we see one interpretation or the other. We can consciously "flip" between these two interpretations, or this flip may occur naturally if we are not concentrating on the image. Hence, our interpretation of such images is *bistable*.

If we can model the visual system (even up to coarse approximations), we can describe the degree of ambiguity and vagueness in visual interpretation. We can then measure and optimize these quantities. If our goal is clarity, we might simply optimize the image to maximize the probability of the correct interpretation. However, if our goal is to make the image *interesting*, then we might want to maximize ambiguity, e.g., maximizing the probability of the second-most likely interpretation, constrained to be far from the most-likely interpretation, with some constraint minimizing the entropy as well.

Generic vs. non-generic viewpoints. Ramachandran and Hirstein [1999] discuss related topics under the heading of generic and non-generic views. A generic view is one that, topologically, is stable to infinitesimal perturbations; a non-generic view is one in which objects line up in coincidental and misleading ways. There is a direct analogue to art. I distinctly remember an art class in which the instructor criticized a painting in which I'd unthinkingly painted a table leg to line up with the side of the window above it, thus making the composition confusing. Freeman [1996] gives a probabilistic explanation for the generic viewpoint assumption.

Ramachandran and Hirstein do not distinguish between beauty and ambiguity. Ambiguity does not create beauty — the Necker cube is not (in my opinion) beautiful — but ambiguity can contribute to the interest of an image.

5.4 Visual encodings and information theory

One of the amazing features of human vision is how we can effortlessly understand and interpret new styles of representation. A skillful artist can construct an image from a new kind of strokes or tiles or other primitive, and have it be perfectly unambiguous and clear, despite the fact that the viewer has no guide to interpret it.

One drawing might use lines to indicate outlines, another might use them for hatching; a paint stroke might illustrate a leg, a house, a tree or an effect of light reflection. In information theory terms, it is as if I can send you a message in a new encoding, and you can read the message without having to know the encoding. In machine learning terms, it is as if we perform model selection—inferring the artistic style—at the same time as interpreting the image. On the other hand, only some encodings are valid. There are severe limits to how much one can mix styles and levels of abstraction. A pointillist bather next to an line-drawn bather would be confusing. In Bayesian terms, this may be expressed as a hierarchical prior.

I believe there is an analogue in natural images: even though the space of imagery in the real world is huge, the actual range of imagery in a particular image is very small. Real-world scenes have limited ranges of materials, colors, and shapes. One does not see beach sand next to forest trees next to asphalt and so on randomly repeated throughout our field of vision. I believe our visual system exploits these regularities when interpreting the world, identifying and grouping common scene elements, and using this commonality to infer consistent shape and materials for these elements, e.g., as in [Glasner et al. 2009; Goldman et al. 2005].

A related effect may be happening with artistic imagery. If you see a single stroke in isolation, you are unlikely to be able to interpret it, or to convince me of your interpretation. It is only through seeing all strokes can you make a convincing explanation of what information each stroke conveys. Perhaps the same visual inference processes are in effect as for natural scenes, as described above.

The question for NPR is: can we develop the notion of a space of visual languages for NPR? What is the hypothesis space for interpreting the visual language used in an artistic image? The procedural language of Grabli et al. [2004] is one initial possibility, but there is clearly much more work to be done.

6 Conclusion

I have sketched some possible ways to build scientific theories of art using the tools of NPR. Though we do not know how much of art can be described scientifically, it is clear that we have only scratched the surface. In the future, working together with scholars from other fields, there is much more we will learn and discover.

Acknowledgements

I thank Maneesh Agrawala, Doug DeCarlo, David Fleet, and Ryan Schmidt for discussions and feedback on multiple paper drafts. I am also grateful for feedback from the anonymous reviewers. This work was supported by NSERC, CIFAR, and the Ontario MRI. Much of this paper was written during a sabbatical visit to the Hebrew University of Jerusalem.

References

AGRAWALA, M., AND STOLTE, C. 2001. Rendering Effective Route Maps: Improving Usability Through Generalization. In *Proc. SIGGRAPH*.

ALEXANDER, R. M. 1996. *Optima for Animals*, 2nd ed. Princeton University Press.

ALEXANDER, R. M. 2001. Design By Numbers. *Nature 412* (Aug.), 591.

CAVANAGH, P. 2005. The artist as neuroscientist. *Nature 434*, 7031, 301–307.

CHATER, N., TENENBAUM, J. B., AND YUILLE, A. 2006. Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences 10*, 7 (July).

COLE, F., GOLOVINSKIY, A., LIMPAECHER, A., BARROS, H. S., FINKELSTEIN, A., FUNKHOUSER, T., AND RUSINKIEWICZ, S. 2008. Where Do People Draw Lines? *ACM Trans. Graph. 27*, 3.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S., , AND SINGH, M. 2009. How Well Do Line Drawings Depict Shape? *ACM Trans. Graphics 28*, 3.

DAW, N. C., AND COURVILLE, A. C. 2007. The pigeon as particle filter. In *Proc. NIPS*.

DAYAN, P., AND DAW, N. D. 2008. Decision theory, reinforcement learning, and the brain. *Cognitive, Affective & Behavioral Neurosc. 8*, 4, 429–453.

DECARLO, D., AND SANTELLA, A. 2002. Stylization and Abstraction of Photographs. *ACM Trans. Graphics 21*, 3 (July).

DECARLO, D., FINKELSTEIN, A., RUSINKIEWICZ, S., AND SANTELLA, A. 2003. Suggestive Contours for Conveying Shape. *ACM Trans. Graphics 22*, 3 (July), 848–855.

DENNING, P. J. 2007. Computing is a Natural Science. *Comm. ACM 10*, 7.

DOYA, K., ISHII, S., POUGET, A., AND RAO, R. P. N., Eds. 2007. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press.

DURAND, F. 2002. An Invitation to Discuss Computer Depiction. In *Proc. NPAR*.

DUTTON, D. 2009. *The Art Instinct: Beauty, Pleasure, and Human Evolution*. Bloomsbury Press.

FREEMAN, W. T. 1996. The generic viewpoint assumption in a Bayesian framework. In *Perception as Bayesian Inference*, D. Knill and W. Richards, Eds. Cambridge University Press, 365–390.

GLASNER, D., BAGON, S., AND IRANI, M. 2009. Super-Resolution From a Single Image. In *Proc. ICCV*.

GOLDMAN, D. B., CURLESS, B., HERTZMANN, A., AND SEITZ, S. 2005. Shape and Spatially Varying BRDFs from Photometric Stereo. In *Proc. ICCV*.

GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human Facial Illustration: Creation and Psychophysical Evaluation. *ACM Trans. Graphics 23*, 1, 27–44.

GOODWIN, T., VOLLICK, I., AND HERTZMANN, A. 2007. Isophote Distance: A Shading Approach to Artistic Stroke Thickness. In *Proc. NPAR*.

GOULD, S. J., AND LEWONTIN, R. C. 1979. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proc. R. Soc. Lond. B. 205*, 581–598.

GRABLI, S., TURQUIN, E., DURAND, F., AND SILLION, F. 2004. Programmable Style for NPR Line Drawing. In *EGSR*.

GREENBERG, S., AND BUXTON, B. 2008. Usability Evaluation Considered Harmful (Some of the Time). In *Proc. CHI*.

GRIFFITHS, T. L., AND TENENBAUM, J. B. 2006. Optimal predictions in everyday cognition. *Psychological Science 17*, 767–773.

HAEBERLI, P. E. 1990. Paint By Numbers: Abstract Image Representations. In *Proc. SIGGRAPH*, vol. 24, 207–214.

HERTZMANN, A. 1998. Painterly Rendering with Curved Brush Strokes of Multiple Sizes. In *SIGGRAPH 98 Conference Proceedings*, 453–460.

HERTZMANN, A. 2001. Paint By Relaxation. *Proc. CGI*.

HERTZMANN, A. 2003. A Survey of Stroke-Based Rendering. *IEEE Computer Graphics & Applications 23*, 4 (July/August).

HUGHES, J. M., GRAHAM, D. J., AND ROCKMORE, D. N. 2010. Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. *Proc. Nat. Acad. Sci. 107*, 4, 1279–1283.

ISENBERG, T., NEUMANN, P., CARPENDALE, S., SOUSA, M. C., AND JORGE, J. A. 2006. Non-Photorealistic Rendering in Context: An Observational Study. In *Proc. NPAR*, 115–126.

JAYNES, E. T. 2003. *Probability Theory: The Logic of Science*.

JUDD, T., DURAND, F., AND ADELSON, E. 2007. Apparent Ridges for Line Drawing. *ACM Trans. Graphics*.

KAWABATA, H., AND ZEKI, S. 2004. Neural Correlates of Beauty. *J. Neurophysiol. 91*, 1699–1705.

KEMP, M. 2009. Art history's window onto the mind. *Nature 461*, 882–883.

KERSTEN, D., MAMASSIAN, P., AND YUILLE, A. 2004. Object Perception as Bayesian Inference. *Annu. Rev. Psychol. 55*, 271–304.

KÖRDING, K. P., AND WOLPERT, D. 2004. Bayesian Integration in Sensorimotor Learning. *Nature 427*, 244–247.

KOSARA, R., HEALEY, C. G., INTERRANTE, V., LAIDLAW, D. H., AND WARE, C. 2003. Thoughts on User Studies: Why, How, and When. *IEEE Computer Graphics & Applications 23*, 4, 20–25.

LEE, S., OLSEN, S., AND GOOCH, B. 2006. Interactive 3D Fluid Jet Painting. In *Proc. NPAR*.

LEE, Y., MARKOSIAN, L., LEE, S., AND HUGHES, J. F. 2007. Line drawings via abstracted shading. *ACM Trans. Graphics*.

LITWINOWICZ, P. 1997. Processing Images and Video for an Impressionist Effect. In *Proc. SIGGRAPH*, 407–414.

LIU, L., CHEN, R., WOLF, L., AND COHEN-OR, D. 2010. Optimizing Photo Composition. In *Proc. EG*.

LYU, S., ROCKMORE, D., AND FARID, H. 2004. A Digital Technique for Art Authentication. *Proc. Nat. Acad. Sci. 101*, 49, 17006–17010.

MARR, D. 1982. *Vision*. W. H. Freeman.

PARKER, G. A., AND SMITH, J. M. 1990. Optimality theory in evolutionary biology. *Nature 348* (Nov.), 27–33.

RAMACHANDRAN, V., AND HIRSTEIN, W. 1999. The Science of Art: A Neurological Theory of Aesthetic Experience. *J. of Consciousness Studies 6*, 6–7, 15–51.

RAO, R. P. N., OLSHAUSEN, B. A., AND LEWICKI, M. S., Eds. 2002. *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press.

SACKS, O. 1996. *An Anthropologist on Mars: Seven Paradoxical Tales*. Vintage.

SANTELLA, A., AND DECARLO, D. 2004. Visual Interest and NPR: an Evaluation and Manifesto. In *Proc. NPAR*.

SANTELLA, A., AGRAWALA, M., DECARLO, D., SALESIN, D., AND COHEN, M. 2006. Gaze-Based Interaction for Semi-Automatic Photo Cropping. In *Proc. CHI*, 771–780.

SCHMIDT, R., KHAN, A., KURTENBACH, G., AND SINGH, K. 2009. On Expert Performance in 3D Curve-Drawing Tasks. In *Proc. SBIM*.

SECORD, A. 2002. Weighted Voronoi Stippling. In *Proc. NPAR*, 27–43.

SILVERS, R. S. 1996. *Photomosaics: Putting Pictures in their Place*. Master's thesis, MIT.

SMITH, J. 1998. *Bone, Volume 5: Rock Jaw: Master of the Eastern Border*. Cartoon Books.

SUTHERLAND, W. J. 2005. The best solution. *Nature 435* (June), 569.

TAYLOR, R. P., MICOLICH, A. P., AND JONAS, D. 1999. Fractal analysis of Pollock's drip paintings. *Nature 399*, 422–423.

TODOROV, E. 2004. Optimality principles in sensorimotor control. *Nature Neuroscience 7*, 9 (Sept.).

WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Proc. 13*, 4, 600–612.

WEISS, Y., SIMONCELLI, E., AND ADELSON, E. H. 2002. Motion illusions as optimal percepts. *Nature Neuroscience 5*, 6 (June).

WIKIPEDIA, 2010. Ambiguity. http://en.wikipedia.org/wiki/Ambiguity.

WINKENBACH, G., AND SALESIN, D. H. 1994. Computer–Generated Pen–And–Ink Illustration. In *Proc. SIGGRAPH*.

WINNEMÖLLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-Time Video Abstraction. *ACM Trans. Graphics 25*, 3 (July).

ZEKI, S. 1998. *Inner Vision: an exploration of art and the brain*. Oxford University Press.