

# Visual Interest and NPR: an Evaluation and Manifesto

Anthony Santella

Doug DeCarlo

Department of Computer Science  
Center for Cognitive Science  
Rutgers University

## Abstract

Using eye tracking, we study the way viewers look at photos and image based NPR illustrations. Viewers examine the same number of locations in photos and in NPR images with uniformly high or low detail. In contrast, viewers are attracted to areas where detail is locally preserved in meaningfully abstracted images. This accords with the idea that artists carefully manipulate detail to control interest and understanding. It also validates the method of meaningful abstraction used in DeCarlo and Santella [2002]. Results also suggest eye tracking can be a useful tool for evaluation of NPR systems.

**Keywords:** abstraction, evaluation, eye tracking, visual perception

## 1 Introduction

Research in NPR has provided computer graphics with a wealth of attractive and visually interesting styles. NPR techniques have also been applied in visualization [Interrante 1996; Gooch et al. 1998; Lum and Ma 2002], and the design of effective diagrams [Agrawala and Stolte 2001; Agrawala et al. 2003]. In this context, the argument is often made that NPR can provide a more compact and easily understood presentation of information. This is plausible. Artists often carefully craft their imagery for easy understanding [Zeki 1999], and there are general rules of visual design [Tufte 1990] that can be realized by a NPR system. There has been, however, relatively little quantitative evaluation of these claims.

In a purely entertainment or artistic context, evaluation may not be necessary. When an artist has near complete control over output [Haerberli 1990; Curtis et al. 1997; Kalnins et al. 2002], poor results are simply bad art. Characterizing bad art might be a genuinely fascinating topic of research, but in practice you know it when you see it. However, in visualizations, when a computer has significant control, this is an engineering concern. Design decisions about presenting information must be made on the basis of empirical evidence of users' perceptions. General knowledge of cognition can guide the design process [Agrawala and Stolte 2001], but there is still a need to evaluate the final system.

In this work, we evaluate the effectiveness of our image based NPR system [DeCarlo and Santella 2002]. It is not aimed at visualization, but is motivated by the use of meaningful abstraction (directed removal of detail) in crafting artistic imagery, and so has



Figure 1: Image abstracted via eye tracking. Notice how detail has been removed in the background while the license plate (which the original viewer read) remains clear. Our findings show that later viewers look at these high detail areas.

similar goals. Eye movements over imagery are directed in a meaningful and economical manner, and are tightly linked to cognition (Section 2.2). Because of this, they can provide evidence for evaluation.

Our aims are to:

- Present a method of evaluation new to NPR (Section 3)—one based on tracking viewers' eye movements.
- Use this method to provide quantitative validation for our system (Section 5) as well as interesting new insights (Section 6).
- Explain why this methodology is widely applicable in NPR, even when the NPR system itself does not use eye tracking.

## 2 Background and Related Work

### 2.1 Evaluation of NPR

Prior methodologies used to evaluate NPR fall into one of two categories. The first method polls a representative number of users, collecting their opinions to find out how they respond to the system. Schumann et al. [1996] polled architects for their impressions of sketchy and traditional CAD renderings, and based on the results, argued for the suitability of sketchy renderings for conveying the impression of tentative or preliminary plans. Similarly, Agrawala and Stolte [2001] demonstrate the effectiveness of their map design system using feedback from real users.

The second approach measures users' performance at specific tasks as they use a system (or its output). When the task depends

on information gained from using the system, performance provides a measure of how effectively the system conveys information. An early study [Ryan and Schwartz 1956] looked at the time required to judge the position of features in photos and hand rendered illustrations in different styles. Faster responses suggested clearer illustrations. Interrante [1996] assessed renderings of transparent surfaces using medical imaging tasks. Performance provided a measure of how clearly the rendering method conveyed shape information. Gooch and Willemsen [2002] tested users' ability to walk blindly to a target location in order to understand spatial perception in a non-photorealistic virtual environment. Gooch et al. [2004] compared performance on learning and recognition tasks using photographs and NPR images of faces. Investigations like this draw on established research methodologies in psychology and psychophysics.

Both of these methods have their limitations. For example, the goal of imagery is not always task related. In advertising or decorative illustration, the goal is more to attract the eye than to convey information. Success is measurable, but not by a natural task. Surveys have their own limitations. The information desired may not be reliably available to subjects by introspection. In addition, both task performance and user approval ratings assess only the quality of a system as a whole. Neither directly say why a pattern in performance or experience occurs. To understand this, the system needs to be systematically changed and the experiment repeated. This process can be costly and time consuming (or impossible). Any additional information to explain performance more directly is therefore highly valuable.

The system we evaluate [DeCarlo and Santella 2002] creates artistic renderings from photographs, using flat regions of color and bold black lines. The rendering is based on a hierarchical segmentation tree, and meaningful abstraction is performed by choosing different levels of the tree in different locations. These decisions are based on a viewer's observation of the original image, recorded by eye tracking. The hope is that removing detail will enhance understanding of the image. Further, viewers may be encouraged to examine the image in a way similar to the first viewer, and take away a similar meaning or impression. This technique was inspired by the kind of abstraction artists sometimes use to make imagery visually satisfying and easy to understand. Artistic application of our method is not intended to convey information in a specific context; therefore there is no applicable task. Systematic questioning of viewers might substantiate the intuition that the images are well designed, but would not inform future work.

Here, we present an alternate methodology which draws on established psychophysical research. The approach analyzes eye movements and provides an objective measure of cognition. It can be the basis of evaluation, or provide complementary evidence when a task or other method is available. Regardless of the context in which the user is viewing an image, the common factor is the act of looking. This mediates all information that passes from the display to the user. In our original system this provided an easy and intuitive method for abstraction; for the same reason, we apply eye tracking to evaluation. These choices are independent; evaluation via eye tracking is a general methodology that can be used regardless of how the imagery is created. Our study also looks at renderings that are created without the use of eye tracking.

## 2.2 Eye Movements and Cognition

Drawing conclusions about cognitive processes by observing eye movements is reasonable due to the close connection between looking and cognition. People examine scenes in a series of fixations (periods where the eye is stable, viewing a single point) and saccades (quick eye movements between points). Viewers don't look around at random; instead, they fixate meaningful and informative

parts of images [Mackworth and Morandi 1967; Henderson and Hollingworth 1998], with a fixation duration related to processing [Just and Carpenter 1976; Henderson and Hollingworth 1998]. Viewing is highly influenced by task. The classic example of this [Yarbus 1967] showed that viewers examining the same image, but performing different tasks, showed drastically different patterns of viewing in which they focused on the features relevant to the task.

This link between looking and cognitive processing is not only investigated in the context of human vision. It is also used as an experimental methodology in psychology. Eye tracking has been used in attempts to understand task performance in a variety of contexts. For example, [Land et al. 1999] studied the everyday activity of making tea and emphasized that actions, even those thought of as automatic are usually preceded by (largely unperceived) fixations of features relevant to the upcoming action. For these reasons, eye tracking has been used to evaluate informational displays including application interfaces [Crowe and Narayanan 2000], web pages [Goldberg et al. 2002], and air traffic control systems [Mulligan 2002]. Eye movements may even reveal information that viewers are trying to report, but cannot, because it is not consciously available. For example, experiments have shown that professional radiologists examining slides look longer at locations where tumors are present, even when they fail to identify and report them [Mello-Thoms et al. 2002]. All this said, there is still a great deal that remains unknown about eye movements. The factors involved in their control and planning are highly debated [Kowler 1990]. One hindrance has been the lack of methods to systematically manipulate images in a way that could reveal what factors are important. Manan et al. [1995] compared patterns of eye movements over high and low pass filtered images. A variety of experiments have been conducted that involve cutting and pasting scene elements [Henderson and Hollingworth 1998]. Controlled manipulation of images however, remains difficult. This may be an area in which computer graphics can inform psychology, and our results may serve as a bridge for such efforts.

## 2.3 Analysis of Eye Movement Data

To understand the specifics of how eye tracking data can be used, it is necessary to understand what the data looks like and how it is analyzed. Eye tracking provides location in screen space (the point of regard or POR), evenly sampled through time. As mentioned before, people examine images in a series of fixations, and saccades. Because viewing is inhibited during saccades, all samples recorded during them can be discarded. A number of methods exist to do this [Widdel 1984; Salvucci and Anderson 2001]. The method used here enforces a limit on the distance the POR has moved within a window of time.

Even when individual fixations have been isolated, it is often useful to impose more structure on the data. In looking at an image, viewers examine many different features, some closely spaced on a single object, others more distant. A common pattern of looking is to scan a number of different features and then return back to particularly interesting ones. Multiple close fixations suggest interest and increased processing in the same location. Because of this, cumulative interest in a location is often a valuable measurement. When the location of features is known, this is often measured by counting viewing time spent within a bounding box [Salvucci and Anderson 2001]. When there are not predetermined features, *clustering* can be used to characterize regions of interest in a data driven fashion [Privitera and Stark 2000; Santella and DeCarlo 2004]. Nearby fixations are clumped together, yielding larger, spatially structured units of visual interest. The number of clusters indicates the number of regions of interest (ROI) present, and the number of points contained in them provides a measure of cumulative interest. In the experiment described in the next section this will reveal important



Figure 2: Example Stimuli

information about how viewers process images.

### 3 Experiment

#### 3.1 Stimuli

The images used in this experiment were 50 photographs, and four NPR renderings of each photo for a total of 250 images and five conditions. Most photos were taken from an on-line database<sup>1</sup>. Photos spanned a broad range of scenes. Images that could not be processed successfully were avoided, such as blurry or heavily textured scenes. Prominent human faces were also excluded, although human figures were present in a number of the images. All NPR images were generated using the method of DeCarlo and Santella [2002]. This method can make local decisions about the amount of detail to include using a model of contrast sensitivity. The four renderings differed in how these decisions were made.

The five conditions are pictured in Figure 2, they are:

**Photo:** This is the unmodified photograph.

**High Detail:** A low global threshold on contrast ensures that most detail is retained, removing primarily areas of low contrast texture and shading.

**Low Detail:** A high contrast threshold is used, removing most detail throughout the image. The resulting image is drastically simplified but still for the most part recognizable.

**Eye Tracking:** Detail is modulated as in [DeCarlo and Santella 2002], using a prior record of a viewer’s eye movements over the image. Detail is preserved in locations the original viewer examined (we call these locations *detail points*) and removed elsewhere. The eye tracking data was recorded from a single subject who viewed each image for five seconds (and was instructed to simply look at the image).

**Saliency Map:** Detail is modulated in the same manner as eye tracking, but the detail points are selected automatically by a

saliency map algorithm [Itti et al. 1998]<sup>2</sup>. This method uses filter responses and other techniques to select locations of potential interest based on local image structure. The algorithm has a model of the passage of time. So, like fixations, each point has an associated duration. Five seconds worth of detail points were created. The locations viewed by people and chosen by the saliency algorithm can be similar in some cases, but in general result in renderings with noticeably different distributions of detail.

This set of conditions represents a systematic manipulation of an image. The effects of NPR style, detail, and abstraction are separated. Local simplification is present in two forms: one based on a viewer, and the other on purely low level features. Because detail is controlled by choosing the levels of a hierarchical segmentation, simplified images consist of a subset of the features in higher detail images. The eye tracking and saliency conditions are rendered literally using a part of the tree used to render the high detail condition, while the low detail case includes the least content.

#### 3.2 Subjects

Data was collected from a total of 74 subjects including 50 undergraduates participating for course credit and 24 subjects (graduate and undergraduate) participating for pay.

#### 3.3 Physical Setup

All images were displayed on a 19 inch LCD display at 1240 x 960 resolution. The screen was viewed at a distance of approximately 33.75 inches, subtending a visual angle of approximately 25 degrees horizontally. Eye movements were monitored using an ISCAN ETL-500 table-top eye-tracker (with a RK-464 pan/tilt camera). The movement of the pan/tilt unit introduces too much noise in practice, and it was not active during the experiment. Instead, subjects placed their heads in an optometric chin rest to minimize head movements.

<sup>1</sup> <http://philip.greenspun.com>

<sup>2</sup> available at <http://iLab.usc.edu>

### 3.4 Calibration and Presentation

Eye trackers need to be calibrated in order to map a picture of a subject's eye to a position in screen space. This is accomplished by having the viewer look at a series of predetermined points. In our experiments, a five point calibration was used. The quality of this calibration was checked visually, and also recorded. Every 10 images, the calibration was checked and re-calibration was performed if necessary. Recordings were used to measure the average quality of the calibrations. Errors had a standard deviation of approximately 24 pixels (about a half degree), which agrees with the published sensitivity of the system. Note that this does not account for systematic drift from small head movements during a viewing.

After calibration, subjects were instructed to look at a target in the center of the screen and click the mouse to view the first picture when ready. On the user's click, the image was presented for 8 seconds, and eye movements were recorded. After this, the target reappeared for one second. A question then appeared. The subject clicked on a radio button to select their response, clicked again to go on, and the process repeated.

Subjects normally saw one condition of each of the 50 images. The condition and order were randomized. While viewing the images, subjects were told to pay attention so they could answer questions which came after each image. Questions were divided into two types, the order of which was randomized. Questions asked the viewer either to rate how much they liked the image on a scale of 1 to 10, or whether they had already seen the image, or a variant of it, earlier in the experiment. Occasional duplicate images were inserted randomly when this question was used; data for these repeated viewings is not included in the analysis. The questions were selected to keep the viewer's attention from drifting, while at the same time not giving them specific instructions which might bias the way they looked at the image.

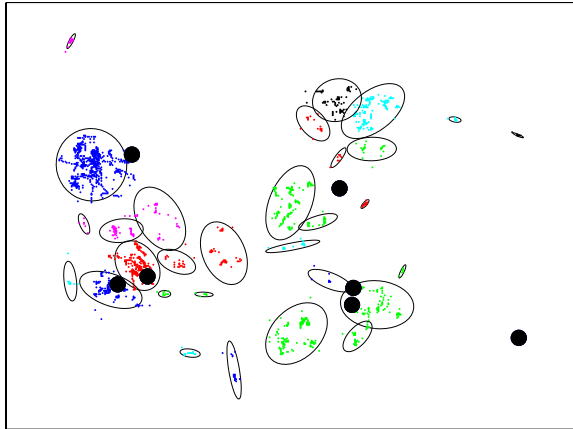


Figure 3: Illustration of data analysis, per image condition. Each colored collection of points is a cluster. Ellipses mark 99% of variance. Large black dots are detail points. We measure the number of clusters, distance between clusters and nearest detail point, and distance between detail points and nearest cluster.

## 4 Analysis

### 4.1 Data Analysis

Analysis draws upon a number of established measures and techniques tailored to our experiment, to provide complimentary evidence about how stylization and abstraction modifies viewing.

Some processing is common to all our analysis. First, all eye movement data is filtered to discard saccades. We then perform clustering on the filtered samples [Santella and DeCarlo 2004]. The clusters are not always meaningful, but on the whole they correspond well to features of interest in the image. There is some reason to believe the number of points contained in a cluster may reveal how important a feature is, this is not considered here.

Our clustering method requires a scale choice. Clusters whose *modes* are closer than this scale value will be collapsed together. We select a scale of 25 pixels (roughly half a degree) for all analysis, which is about the level of noise present. Results depend on the scale choice used in the clustering process. Clearly, at coarser scales there will be fewer clusters and a smaller difference between the condition means. We argue below that this does not affect interpretation of our results.

All clustering was conducted in two ways. In the first, which we will refer to as *per viewer analysis*, each viewer's data was clustered separately. In the second analysis, which we will refer to as *per image analysis*, data for all viewers of a particular image was combined *before* clustering. It is reasonable to think that as one adds data from individual viewers, the data will approach some hypothetical distribution of image feature interest [Wooding 2002]. This second analysis may therefore provide a better measure of aggregate effects.

Below, we describe the measurements performed using the clusters. See Figure 3 for an illustration of the data.

**Clusters:** Because clusters roughly correspond to objects examined in the image, we would expect to find fewer clusters in the eye tracking and salience cases if they succeed in focusing viewer interest. We might also expect uniform simplification to reduce the number of clusters, because it reduces detail.

**Distance (from data to detail points):** In the eye tracking and salience conditions, we wish to measure whether interest is focused on the locations where detail is preserved. The change in distance from each of the cluster centers to the closest detail point between conditions tells us how effective the manipulation is in drawing interest to these locations. If the abstraction is successful, we would expect that clusters will be closer. This tests the system as a whole. There will be no change in distance if our hypothesis is wrong, which would mean that varying detail does not attract more focused interest. It is also possible there was no detail that could be put in a particular location, because there was none in the original image, or because it cannot be represented in our system's visual style.

**Distance (from detail points to data):** Implicit in the choice of detail points is the assumption that viewers should look at *all* of the locations. This is not captured by the distance measure. A viewer could spend all the time looking at one detail point yielding a zero distance. To quantify this, it is possible to measure the distance *from* each detail point to the closest cluster. A high average value means the locations of a significant number of detail points were not closely examined. This distance will decrease in salience and eye tracking conditions if detail modulation makes people look at high detail areas that were not normally examined.

### 4.2 Statistical Analysis

Data for all subjects was clustered as discussed in Section 4.1. In total there are 10 eye tracking records for each of the 50 images in each of the 5 conditions for a total of 2500 individual recordings. More data than this was gathered; a matched number of recordings for each condition was selected randomly. As noted in Section 3.4, data was recorded in blocks where one of two questions was asked. Analysis showed no effect of the questions, and these results are based on roughly equal numbers of images presented in blocks of each question type.

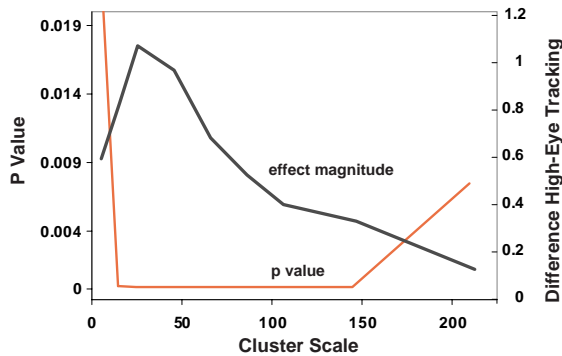


Figure 4: Statistical significance is achieved for number of clusters over a wide range of clustering scales. The magnitude of the effect decreases, but significance remains quite constantly over a wide interval. Our results do not hinge on the scale value selected.

Analysis of variance (ANOVA) are used to test whether differences between condition are significant. These tests produce a  $p$  value: the probability the measured difference could occur by chance. The per viewer case gives itself naturally to statistical testing by a two-way repeated measure ANOVA. A two-way ANOVA separately tests both the contribution that the particular image and the condition make to the results. This lets one look at the effect of a condition while factoring out the variation among the different images. A repeated measure analysis treats each viewers' eye tracking record as an independent measurement, so there are 10 data points per image and condition pair. In the per image analysis, the 10 recordings are collapsed together and data is analyzed instead by a simple two-way ANOVA. There is now only one data point per image and condition pair, so it is more difficult to show a statistically significant effect. In performing these tests a single ANOVA is used to confirm all conditions are not equivalent, ANOVA are then conducted between each pair of conditions, to identify which conditions differ.

## 5 Results

Figure 5 graphs the average results for all measures. The take-away message, quantified below, is that on the whole:

- Eye tracking and salience conditions have fewer clusters than photo and uniform detail conditions in all analyses. In the per image analysis, eye tracking has fewer clusters than salience.
- Distance between the viewed locations and the detail points decreased as a result of modulating detail.
- Distance between detail points and viewed locations showed no change; however the distances for salience points were significantly higher than those for eye tracking points.

### 5.1 Quantitative Results

**Clusters:** In the per viewer analysis, there was about one fewer cluster in the eye tracking and salience conditions, compared to the others. This means each viewer examined one fewer region on average. Analysis showed this difference was significant ( $p < .001$ ). There was no significant difference ( $p > .05$ ) between the photo or uniform detail conditions, or between eye tracking and salience.

In the per image analysis, eye tracking had about 6 fewer clusters than uniform detail and photo conditions, while salience had about

3 fewer. Eye tracking differed significantly from all other conditions including salience ( $p < .001$ ). Salience differed from original at  $p < .01$ , and from high and low at  $p < .05$ .

**Distance (from data to detail points):** Clusters in the eye tracking condition were about 20 pixels closer to the eye tracking detail points than high detail clusters, in both per viewer and per image analysis ( $p < .0001$ ). Salience clusters are about 10 pixels closer to salience detail points (per viewer:  $p < .0001$ , per image:  $p < .01$ ). This is not spatially very large, but it represents a consistent shift of cluster centers towards the detail points. The magnitudes of the two shifts (10 and 20 pixels) were not significantly different from each other. For per image analysis, distances measured to eye tracking detail points were significantly higher  $p < .01$  than corresponding distances to salience points.

**Distance (from detail points to data):** There was no significant change ( $p > .05$ ) for saliency or eye tracking renderings in either analysis. In both analyses however, the distances were significantly smaller ( $p < .001$ ) when measured from eye tracking detail points than from salience detail points (a difference of about 40 in the per viewer and 10 in the per image condition).

All of the two-way ANOVAs test the significance of both the experimental condition, and the particular image. In all tests, the effect of the image was highly significant ( $p < .001$ ). This is neither surprising nor particularly informative. It simply states that individual images have varying numbers of interesting features and they are distributed differently in relation to the detail points.

As mentioned above, all of this analysis uses clusters created with a particular choice of scale. Figure 4 shows that results do not depend on this choice. The difference between mean number of clusters in the high detail and eye tracking conditions (per viewer analysis) is plotted along with the corresponding  $p$  value. Though the magnitude of the difference varies,  $p$  values show an effect of approximately equal significance over a range of scales. The effects we have shown are therefore not due to the particular scale selected.

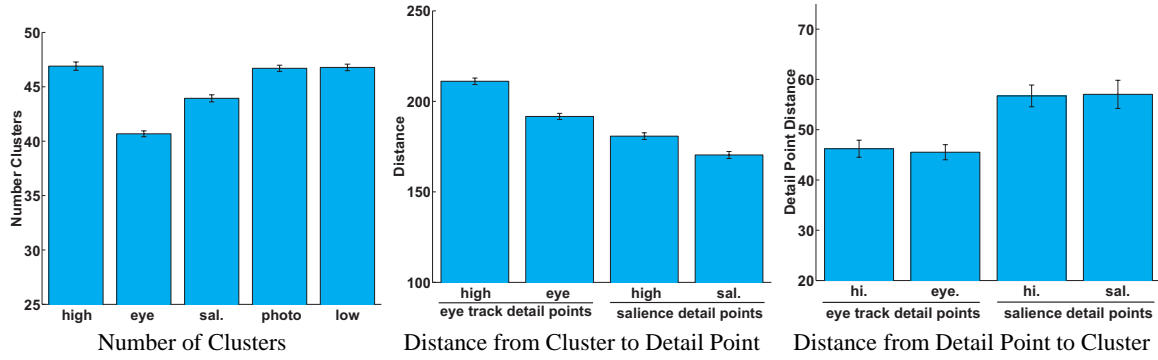
### 5.2 Discussion

These results provide evidence that local detail modulation does change the way viewers examine an image. Eye tracking and salience renderings each have significantly fewer clusters than all uniform detail images in both per image and per viewer analysis (significance is stronger in the per viewer analysis, but that is to be expected based on the number of samples). Distances to detail points also show an improvement for both salience and eye tracking renderings. This indicates that not only were fewer places examined, but the examined points were closer to the detail points. Distances from detail points to data do not show improvement. This indicates that though interest was concentrated by the manipulation, it did not bring new interest to detail points that were not already interesting in the high detail renderings. Results do not prove enhanced or facilitated understanding per se; however, this is strongly suggested by the more focused pattern of looking.

Results also indicate that although improvement can be seen with detail modulation based on both eye tracking and salience, the two behave differently. Modulation based on both produces fewer clusters of interest, and decreased distance to detail points. However, in the per image analysis, the number of clusters for the eye tracking condition was significantly lower than the salience condition. Also, the distances measured from salience points are consistently higher than those from eye tracking points; this is further evidence that eye tracking points are more closely examined. Distance to detail points shows the opposite relationship (though more weakly) and argues against this conclusion. However, we show below that this is almost certainly due to the number of detail points, and is not meaningful.

These results fit our intuition that the locations a viewer exam-

**Per Image Analysis, data for all viewers of an image is clustered together.**



**Per Viewer Analysis, data for each viewing is clustered separately.**

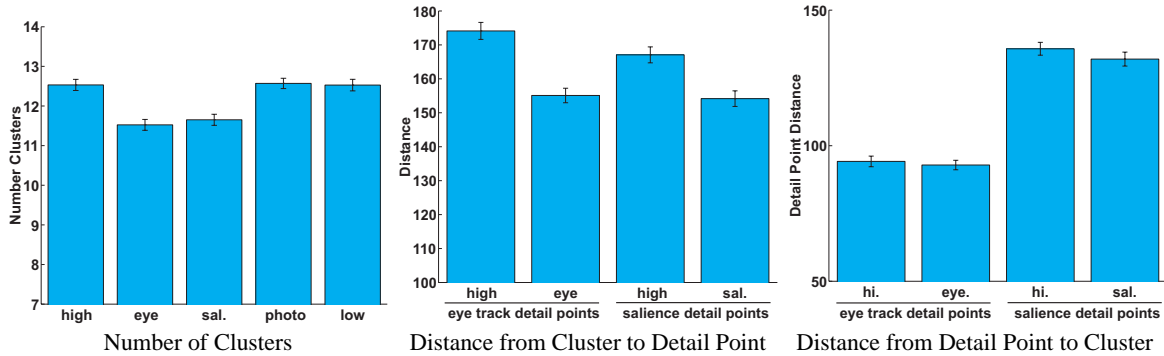


Figure 5: Average results for all analysis

ined will, in general, be a better predictor of future viewing than a saliency model, which has no sense of the meaning of image contents. There is considerable controversy in the human vision literature about how much of eye movements can be accounted for by low level saliency features. Some optimistically state that saliency predictions correlate well with real eye motions [Privitera and Stark 2000; Parkhurst et al. 2002]. Others are more doubtful and claim that when measured more carefully and in the context of a goal driven activity, the correlation is quite poor [Turano et al. 2003; Land et al. 1999]. Our results show saliency points (at least those produced by the algorithm used) are less interesting in general. Though abstraction does attract increased interest to saliency points; people look nearer to some of them, but at less of them.

It is not clear at first glance that distance values measured against the eye tracking and saliency detail points can legitimately be compared to each other in order to judge if they are functionally equivalent. Differences may be due to the number and distribution of the two kinds of detail points, rather than their locations relative to features in the images and hence the data collected. The saliency algorithm produces more fixations than real viewers, so there were typically more detail points in the saliency case (10.9 per image on average) than eye tracking (5.96 on average) This would seem to bias distances *to* saliency detail points toward lower values, and make comparison difficult.

Fortunately, some simple controls indicate that the lower distances *to* saliency points is an unimportant artifact, while the higher distances *from* saliency detail points is meaningful. Replacing recorded data with random data indicates if effects are driven by the data or by the detail points. Doing this eliminates all effects in distances *from* detail points. In contrast, distances *to* detail points are still significantly higher for eye tracking and lower for saliency points. Data drives the difference in distance *from* detail points. The difference in distances *to* detail points is at least partly driven

by the points themselves. A second control is to discard some detail points so that the number of detail points is equal across conditions. Effects driven by the number of points should change or disappear. There is no qualitative change in distances *from* detail points. Distances *to* saliency detail points however become significantly higher than those *to* eye track points: a reversal (the main effect of improvement from detail modulation in both conditions is not affected). We conclude that eye tracking detail points really are more examined than saliency points overall.

In contrast to the changes caused by abstraction, there is little evidence that the style manipulation alone produces a significant change in viewing. There is no significant difference between the photo and high detail images in number of clusters. A qualitative comparison of scatter plots of the data in these two conditions agrees that the distribution of points isn't very different. There is however, high variance between the average effects for particular images. There are images where large areas of low contrast texture are removed by the stylization itself. In these cases, viewer interest is different between the high detail and photo conditions (see Figure 6 for an example). Removing prominent but low contrast texture is abstraction, but it is abstraction over which one has no control. Rather, it is built implicitly into the system (in this case into the segmentation technique). The opposite effect can also occur: the style can attract attention to less noticeable features. Notice in Figure 6, how drawing ripples on the water in black has attracted the eye to them. A method for quantifying when and where these effects occur is a topic for future research. These appear to be primarily low-level effects, so work on saliency [Itti et al. 1998] may provide a good starting point.

Interestingly, our results also indicate the number of regions of interest is not primarily driven by detail. It is surprising how much the pattern of interest in the low detail case qualitatively and quantitatively resembles the high detail. The highly detailed and highly

simplified renderings have the same number of clusters while the mixed detail images, in which the number of regions lies between these extremes, have less. This implies it is *locally* increased detail that attracts the eye. Substantiating and quantifying this is an interesting subject for future research and has direct application in designing future NPR and visualization systems.

In summary:

- viewers look at fewer locations in images simplified using eye tracking and salience data,
- these locations tend to be near locations where detail is locally retained,
- neither the NPR style itself, nor application of uniform simplification modify number of locations examined, and
- this effect exists for both eye tracking and salience detail points, but there is less interest in salience points.

## 6 Conclusion

These results validate our original system's design to focus interest by manipulating image detail using eye tracking data. Results also have broader implications for those designing NPR systems, using salience maps in graphics, and designing future experimental evaluations of NPR systems.

Our results show *meaningful* abstraction is important for effective NPR. Abstraction that does not carry any meaning is implicit in many NPR styles; for example, there are no shading cues in a pure line drawing. Uniform control of detail is also common in NPR systems. These are important considerations. But, both were tested in this study and produced no change in the number of locations viewers examined. In contrast, meaningful abstraction clearly affected viewers in a way that supports an interpretation of enhanced understanding. Directed meaningful abstraction should be considered seriously in designing future NPR systems.

Similarly, although low level (salience map) and high level (eye track) detail points behave similarly in their ability to capture *increased* interest, they differ in their *absolute* capture of interest. The increased capture of interest seems to be a low level effect; people don't bother looking where there isn't anything informative. However, semantic factors are also active. The locations that interest another person are influenced by image meaning and are a better predictor than salience of where future viewers will look.

This has implications for the use of salience maps in graphics. It would be highly desirable to automatically locate places viewers will look in a number of applications, not the least of which would be automatic abstraction in NPR and visualization. Though salience points behave similarly to eye tracking points in part of our analysis, results indicate that on the whole salience is not suitable for this purpose. It can be successful in adaptive rendering applications Yee et al. [2001], where it is only necessary that people be somewhat more likely to look at selected locations. Salience does provide information about how likely the structural qualities of a feature are to attract interest. However, we want to encourage a later viewer to get the same content from an image that an earlier, perhaps more experienced viewer examined closely. Current salience map algorithms are hardly expert viewers. This kind of application requires better predictions, motivated by semantic information that salience is generally unlikely to provide.

In addition, eye tracking may be a useful technique for evaluations of other NPR systems. It provides an alternative to questionnaires and task performance measures. Even when a task based method is possible, eye tracking can be useful in investigating what features underlie the performance observed. Information is, after all, extracted from the imagery by looking, and the large body of

research available suggests that locations examined indeed reveal the information being used to complete a task.

The experiments performed by Gooch et al. [2004] can serve to illustrate this. If users can perform a task better using an NPR drawing of a face rather than a photograph, it is valuable to see where clusters of visual interest occur in the two conditions. This information may explain performance. It could focus future experiments and inform design choices about rendering faces, without exhaustive experimental testing. Similarly, in evaluation of assembly diagrams [Agrawala et al. 2003], eye tracking can provide very specific information about how people use such instructions. For example, eye tracking could help further explain the way users interleave actions in the world and examination of the relevant part of the instructions. Eye tracking records are directly and obviously related to the imagery that evokes them. This makes them very interpretable—a desirable quality in any measurement. This in turn guards against the danger [Kosara et al. 2003] of performing a user evaluation that ultimately doesn't yield any useful result.

This leads us to an agenda for future investigation into visual perception and NPR:

- An increasingly quantified understanding of the importance of abstraction in NPR, where abstraction is understood not as mere removal of information but as targeted control of detail in pursuit of a point.
- Further use of eye tracking, combined with tasks where possible, to understand how style and abstraction affect perception and use of visual imagery.
- Experimentation used not only to evaluate whole systems, but to build quantitative models of perception as it relates to questions of interest in NPR.
- The integration of these models into future NPR systems.

We hope this informs the design of smarter, more carefully tailored NPR systems which provide imagery that is not only appealing, but also carefully engineered for efficient understanding and use.

## Acknowledgments

Special thanks to Eileen Kowler for her advice and encouragement. Thanks also to Philip Greenspun (<http://philip.greenspun.com>) for the use of his photos, and to Jim Housell, Paul Kry, Dennis Santella and Matthew Stone. This material is based upon work supported by the National Science Foundation under the grant SGER 0227337.

## References

- AGRAWALA, M., AND STOLTE, C. 2001. Rendering effective route maps: improving usability through generalization. In *Proceedings of ACM SIGGRAPH 2001*, 241–249.
- AGRAWALA, M., PHAN, D., HEISER, J., HAYMAKER, J., KLINGNER, J., HANRAHAN, P., AND TVERSKY, B. 2003. Designing effective step-by-step assembly instructions. In *Proceedings of ACM SIGGRAPH 2003*.
- CROWE, E. C., AND NARAYANAN, N. H. 2000. Comparing interfaces based on what users watch and do. In *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2000*, 29–36.
- CURTIS, C. J., ANDERSON, S. E., SEIMS, J. E., FLEISCHER, K. W., AND SALESIN, D. H. 1997. Computer-generated watercolor. In *Proceedings of ACM SIGGRAPH 97*, 421–430.
- DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. In *Proceedings of ACM SIGGRAPH 2002*, 769–776.

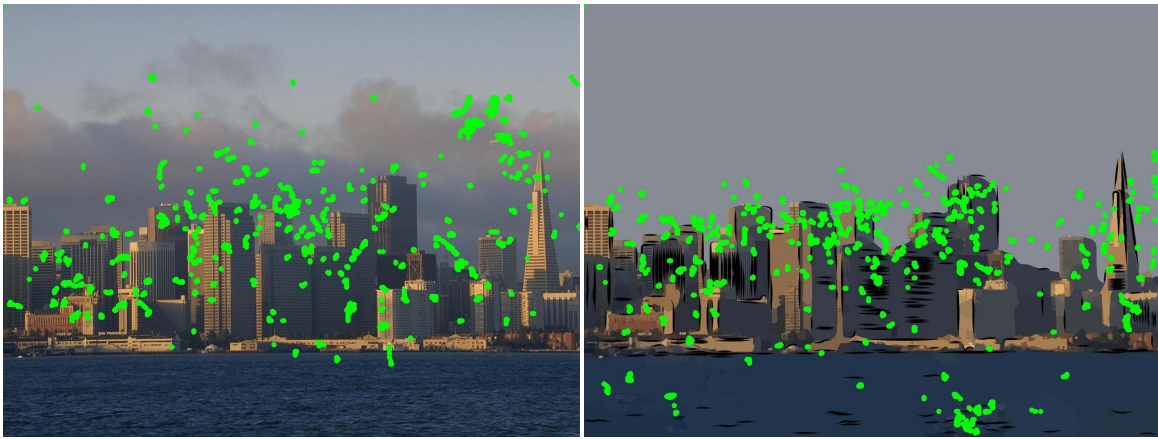


Figure 6: Original photo and high detail NPR image with viewers' filtered eye tracking data. Though we found no effect globally across these image types, there are sometimes significantly different patterns of viewing, as can be seen here.

- GOLDBERG, J. H., STIMSON, M. J., LEWENSTEIN, M., SCOTT, N., AND WICHANSKY, A. M. 2002. Eye tracking in web search tasks: design implications. In *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2002*, 51–58.
- GOOCH, A. A., AND WILLEMSSEN, P. 2002. Evaluating space perception in NPR immersive environments. In *Proceedings of the Second International Symposium on Non-photorealistic Animation and Rendering (NPAR)*, 105–110.
- GOOCH, A. A., GOOCH, B., SHIRLEY, P., AND COHEN, E. 1998. A non-photorealistic lighting model for automatic technical illustration. In *Proceedings of ACM SIGGRAPH 98*, 447–452.
- GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustration: Creation and psychophysical evaluation. *ACM Transactions on Graphics* 23 (January), 27–44.
- HAEBERLI, P. 1990. Paint by numbers: Abstract image representations. In *Proceedings of ACM SIGGRAPH 90*, 207–214.
- HENDERSON, J. M., AND HOLLINGWORTH, A. 1998. Eye movements during scene viewing: An overview. In *Eye Guidance in Reading and Scene Perception*, G. Underwood, Ed. Elsevier Science Ltd., 269–293.
- INTERRANTE, V. 1996. *Illustrating Transparency: communicating the 3D shape of layered transparent surfaces via texture*. PhD thesis, University of North Carolina.
- ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 1254–1259.
- JUST, M. A., AND CARPENTER, P. A. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 441–480.
- KALNINS, R. D., MARKOSIAN, L., MEIER, B. J., KOWALSKI, M. A., LEE, J. C., DAVIDSON, P. L., WEBB, M., HUGHES, J. F., AND FINKELSTEIN, A. 2002. WYSIWYG NPR: Drawing strokes directly on 3D models. In *Proceedings of ACM SIGGRAPH 2002*, 755–762.
- KOSARA, R., HEALEY, C., INTERRANTE, V., LAIDLAW, D., AND WARE, C. 2003. User studies: Why, how and when? *IEEE Computer Graphics and Applications* 23, 4, 20–25.
- KOWLER, E. 1990. The role of visual and cognitive processes in the control of eye movements. In *Eye Movements and Their role in Visual and Cognitive Processes*, E. Kowler, Ed. Elsevier Science Ltd., 1–70.
- LAND, M., MENNIE, N., AND RUSTED, J. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 1311–1328.
- LUM, E. B., AND MA, K.-L. 2002. Hardware-accelerated parallel non-photorealistic volume rendering. In *Proceedings of the Second International Symposium on Non-photorealistic Animation and Rendering (NPAR)*, 67–74.
- MACKWORTH, N., AND MORANDI, A. 1967. The gaze selects informative details within pictures. *Perception and Psychophysics* 2, 547–552.
- MANNAN, S., RUDDOCK, K. H., AND WOODING, D. S. 1995. Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision* 9, 363–386.
- MELLO-THOMS, C., NODINE, C. F., AND KUNDEL, H. L. 2002. What attracts the eye to the location of missed and reported breast cancers? In *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2002*, 111–117.
- MULLIGAN, J. B. 2002. A software-based eye tracking system for the study of air traffic displays. In *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2002*, 69–76.
- PARKHURST, D., LAW, K., AND NIEBUR, E. 2002. Modeling the role of saliency in the allocation of overt visual attention. *Vision Research* 42, 107–123.
- PRIVITERA, C. M., AND STARK, L. W. 2000. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 9, 970–982.
- RYAN, T. A., AND SCHWARTZ, C. B. 1956. Speed of perception as a function of mode of representation. *American Journal of Psychology*, 60–69.
- SALVUCCI, D., AND ANDERSON, J. 2001. Automated eye-movement protocol analysis. *Human-Computer Interaction* 16, 39–86.
- SANTELLA, A., AND DECARLO, D. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2004*.
- SCHUMANN, J., STROTHOTTE, T., AND LASER, S. 1996. Assessing the effect of non-photorealistic rendering images in computer-aided design. In *ACM Human Factors in Computing Systems, SIGCHI*, 35–41.
- TUFTE, E. R. 1990. *Envisioning Information*. Graphics Press.
- TURANO, K. A., GERUSCHAT, D. R., AND BAKER, F. H. 2003. Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research* 43, 333–346.
- WIDDEL, H. 1984. Operational problems in analysing eye movements. In *Theoretical and applied aspects of eye movement research*, A. G. Gale and F. Johnson, Eds. Elsevier Science Ltd., 21–29.
- WOODING, D. S. 2002. Fixation maps: quantifying eye-movement traces. In *Proceedings of the Eye Tracking Research and Applications (ETRA) Symposium 2002*, 31–36.
- YARBUS, A. L. 1967. *Eye Movements and Vision*. Plenum Press.
- ZEKI, S. 1999. *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press.