

Abstracted Painterly Renderings Using Eye-Tracking Data

Anthony Santella

Doug DeCarlo

Rutgers University
Department of Computer Science & Center for Cognitive Science
{asantell, decarlo}@cs.rutgers.edu

Abstract

When used by artists, manual interfaces for painterly rendering can yield very satisfying abstract transformations of images. Automatic techniques produce interesting paintings as well, but can only recast pictures in a different style without performing any abstraction. At best, information is removed uniformly across the image, without emphasizing the important content. We describe a new approach for the creation of painterly renderings that draws on a model of human perception and is driven by eye-tracking data. This approach can perform meaningful abstraction using this data, with the minimum interaction possible: the user need only *look* at the image for several seconds. We demonstrate the effectiveness of this interactive technique and compare it with a fully automatic approach.

Keywords: automatic painting, image simplification, perceptual model, eye-tracking

1 Introduction

The technical excellence found in all great art can be seen as a vehicle for the subjective interpretation afforded by it. It is this power that makes the work profound and deeply satisfying. In this respect, we can single out the work of artists such as Vermeer. See Figure 1. This picture looks realistic and does actually contain some precise spatial information, but it diverges substantially from what you would see in an ordinary photo. The artist adapts texture, lighting and detail. As observed by vision scientists (e.g., Zeki [1999]), the result is an image that directs your attention to its most meaningful places and allows you to understand the structure there without conscious effort. As the immediate perceptual features of the painting fade into the famous impression of realism, you can instead focus on deeper questions about the scene—who is that woman, and what does that letter say?

This interpretation can be traced largely to the *abstraction* performed by the artist. The organization of the image is designed to direct the viewer's attention to the relevant content and to influence their perception of it. Artists develop individual styles to achieve this, which can arise in part due to the versatility of their chosen media. Nowadays, artists find computers among the possibilities.

Haerberli [1990] described an interactive approach for transforming ordinary images into stylized variations. The transformation,



Figure 1: Johannes Vermeer's "Girl Reading a Letter at an Open Window" (1657). The realistic look belies the effort of the artist, who has organized the image to highlight meaningful information. For example, the artist creates high contrast in distinguishing the figure from the background and delineating her face, hair, dress and hands. In comparison, unimportant objects in the scene—the chair, the panes of glass, the tabletop, the back wall—lack the same sharp detail and fine contrasts.

often referred to as painterly rendering, bears similarity to traditional painting in that it results in an aggregate collection of brush strokes. Haerberli's goal was the creation of artistic, stylized and abstract images by giving the artist precise control over the result. The user places brush strokes manually, with colors taken from the source image. The amount of interaction required to produce a painting could be reduced, for instance, by determining the stroke direction using the image gradient.

Since then, there has been considerable work on painting systems working from images along two rather different fronts. First, comes the introduction of digital analogues of traditional media such as pen-and-ink [Salisbury et al. 1994] or watercolor [Curtis et al. 1997]. In these systems, as in [Haerberli 1990], the user retains control. Second, are the automatic methods for placing strokes [Hertzmann 1998; Litwinowicz 1997; Shiraishi and Yamaguchi 2000] in both static and moving imagery. Here, the user can only select global effects, such as the brush size or shape. Not surprisingly, hand-crafted images are typically more appealing than automatically generated renderings, as abstraction takes place in addition to more superficial stylization. To make these images particularly effective, however, artistic skill is required in addition to a fair amount of time. With all of this progress, however, there is no work that bridges between these two methodologies. The reason

for this is clear: abstraction of image content is not something computers can do as of yet—artistic and organizational decisions must be made by a person.

This paper introduces a new approach for making artistic renderings which takes the first step towards bridging these two techniques. The result will be a system that produces abstract imagery with several seconds of user interaction (and perhaps no artistic skill). It starts with a *model of human visual perception* that characterizes the organization of an image that the visual system imposes on it. In other words, a perceptual model describes what regions or visual objects we can look at and understand in an image. Meanwhile, a user supplies *intentional* content with the interface; it describes what the user wants the output to convey. In this case, it specifies the importance of particular visual elements in the image. The intentional structure of an image is described using *eye-tracking data*, which shows where people had to look in order to extract meaningful information from the image. To gather this data, the user is simply instructed to “look at the image” while an eye-tracking sensor monitors the user’s gaze for several seconds.

Provided with the perceptual and intentional structure of an image, the system proceeds by placing individual brush strokes, deciding which to include and how they can be transformed as dictated by the perceptual model and eye-tracking data. Only those picture elements are included that both attracted the user’s gaze and were perceivable given the limits of human vision. The resulting image encourages viewers to concentrate on the content that this particular user found important without being distracted by irrelevant detail and background content.

Looking to the future, such a system will improve as human vision research allows for the formulation of more detailed and accurate models, and computational vision techniques allow for richer automated analysis of images. For now, such a system might serve as a valuable tool in eye movement research, by providing a controlled means for adapting imagery based on a perceptual model.

This paper is structured as follows. After a review of the relevant research from the human vision community and on painterly rendering in Section 2, we describe the use of the perceptual model and eye-tracking data for painterly rendering in Section 3. Section 4 contains results of using this system on a variety of images.

2 Background

2.1 Visual Acuity

The resolution of the eye is limited by numerous factors, such as the density of photoreceptors in the retina, the nature of the neural connections between them, and the contrast of the stimulus. This resolution also varies across the retina, with the highest density of photoreceptors being located centrally in the fovea.

Psychophysical experiments to determine the limits of vision, such as those by Koenderink et al. [1978] or Kelly [1984] measure detection and discrimination of simple repeating patterns—sinusoidal gratings. These experiments produced models which describe the degree of fineness in detail that is perceptible by the human visual system (as a function of contrast and spatial frequency). Pattanaik, et al. [1998] uses such a model (along with a model of visual adaptation) for producing perceptually realistic images of scenes.

Eccentricity is an angular measurement of distance from the center of the visual field. Models describing the decrease in sensitivity as a function of eccentricity [Koenderink et al. 1978] allow us to incorporate a notion of spatially varying resolution into our model. Reddy [2001] uses a model of visual acuity (which describes the highest spatial frequency that can be resolved at maximum contrast) for performance-based visualization of 3-D environments. Reddy’s model also included terms for moving imagery. Given our use of

static imagery, this model reduces to the following:

$$\begin{aligned} G &= 60.0 \text{ cycles/degree} \\ M(e) &= \frac{1}{1 + 0.29e + 0.000012e^3} \\ H(e) &= G \times M(e) \end{aligned} \quad (1)$$

In this model, G represents the highest perceivable frequency at maximum contrast. As a function of the eccentricity e (in degrees), the sensitivity falloff factor $M(e)$ (which is 1 inside the fovea, and decreases towards 0 with increasing e) scales the frequency G to produce the visual acuity $H(e)$, which specifies the highest detectable frequency at a particular eccentricity in cycles per degree.

None of the psychophysical experiments that informed these models involved the viewing of actual scenes. Very simplified stimuli were presented instead (typically, repeating patterns). While this questions the accuracy of these models for such situations, they have been used successfully in computational settings [Pattanaik et al. 1998; Reddy 2001]. Further psychophysical studies are required to develop better models for these situations.

These limits on sensitivity within the visual field are not seen as a disadvantage of the human visual system. On the contrary, it fits hand-in-hand with its ability to integrate information across changes in viewpoint and movements of the eyes.

2.2 Eye Movements

Several basic facts about how people look at images factor into our use of recordings of eye movements. For our application, we are only concerned with eye movements on unchanging imagery, and the discussion here is limited to this.

- People can examine only a small part of an image at one time, and so understand images by scanning them using discrete, rapid movements of their eyes, called *saccades*. While saccades can be initiated voluntarily, they typically proceed in a goal-directed fashion. The motions are performed with remarkable precision and efficiency—the eyes seldom perform wasted motions, and typically land near the best place to gather the desired visual information. Decisions on where to look next seem to draw heavily upon extra-foveal information (lower resolution information from outside the fovea).
- Saccades are punctuated by stabilizing motions called *fixations*, which allow the eye to dwell on a particular stationary object. The overwhelming majority of visual processing takes place during fixations. Under normal circumstances, the attention of the viewer is at the fixation location, for at least the bulk of its duration.
- In each individual glance, people look at *something*—the eyes do not wander randomly. Mackworth and Morandi [1967] showed there is good agreement between informative locations of a display and fixation locations. More recent studies by Henderson and Hollingworth [1998] have shown that *semantic informativeness* (i.e., the meaning of an image region) strongly influences fixation positions compared to visual informativeness (i.e., structural information, such as an intensity discontinuity). Of course, locations that are deemed informative depends on the task and gaze positions change accordingly [Yarbus 1967].
- In most tasks, the time spent fixating on a particular location or object indicates that processing on that object is taking place [Just and Carpenter 1976]. More specifically, fixation duration provides a rough estimate on how much processing is expended in understanding that portion of the image. Exploratory fixations tend to be brief.

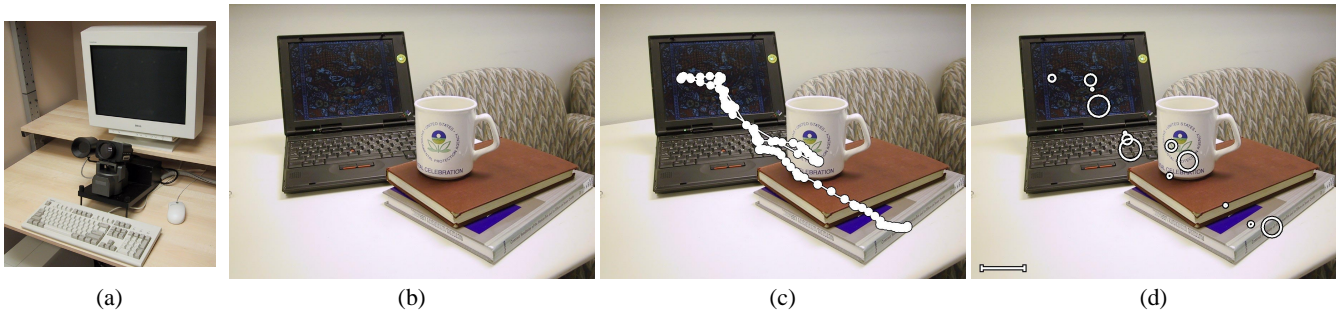


Figure 2: (a) Commercial eye-trackers can be placed just underneath a computer display to monitor where the user looks on the screen. (b) A still-life photograph which is presented to a user. (c) Eye movement data for this image. Each circle represents the point of regard at a particular moment in time; eye gaze position is marked 60 times per second for a total duration of 5 seconds. (d) The fixations positions are extracted from this data, where the diameter of the circle indicates fixation duration (the scale in the lower left indicates 1 second).

Many other types of movements are possible, such as those involved in smooth pursuit; but it is the saccades and fixations that play the largest role in gathering information from across a static image.

Much of what is known about eye movements comes from the use of *eye-trackers*, which are sensors designed to record the movements of the eyes. Although eye-trackers have used similar principles for the last thirty years, recent advancements have made them smaller, less invasive and inexpensive. When used for human-computer interaction [Duchowski 2000; Duchowski and Vertegaal 2000; Jacob 1993; Sibert and Jacob 2000; Vertegaal 1999], eye-trackers can be mounted directly on the user's head to allow for unencumbered viewing of a scene, or can be placed less invasively in the work environment. Figure 2(a) shows our ISCAN ETL-500 eye-tracker (with an RK-464 pan/tilt camera) as we have placed it, just below the computer monitor. In current applications, eye-trackers are used as a "cursor" to select objects on the screen or in a virtual environment [Sibert and Jacob 2000], to provide attentional information to enable image or scene simplification [Duchowski 2000], or to provide indirect evidence about the attention of a user [Vertegaal 1999] in teleconferencing. While using eye-trackers as a voluntary selection mechanism is not a typical manner of looking, evaluation of such systems indicates they are still quite beneficial [Sibert and Jacob 2000].

After a brief calibration period (built into commercial systems), eye-trackers are ready to gather observations of a viewer's point of regard. Upon viewing the image of a still-life in Figure 2(b) for five seconds, a subject's eye movements were tracked (at 60 Hz). The resulting eye movement pattern shown is displayed in (c). Through analysis of this raw data (which is relatively noisy), fixation locations (and their corresponding durations) are detected using a simple velocity threshold [Duchowski and Vertegaal 2000]. Fixations from viewing this image are shown in Figure 2(d) as circles centered at the fixation location; the diameter of the circles is proportional to the duration.

In gathering this data, the viewer is simply instructed to "look at the image". This encourages spontaneous looking [Kahneman 1973], whereby the viewer's gaze tends to fall upon the most informative or important parts of the image [Yarbus 1967]. The alternative of asking the user to "look at the parts of this image you find important" would set up a task for the user, and task-relevant looking [Kahneman 1973] is more complex and difficult.

2.3 Painterly Rendering

The interactive painting system presented by Haerberli [1990] enabled the production of images formed from collections of brush strokes. The placement of these brush strokes was entirely specified by the user. Automatic methods for orienting brush strokes,

including using the image gradient or a secondary image were provided.

Fully automatic methods for producing painterly images have also been presented [Hertzmann 1998; Litwinowicz 1997; Shiraishi and Yamaguchi 2000]. Automation was achieved in [Litwinowicz 1997] by covering the canvas with a jittered grid of strokes, painted in random order. While strokes were still oriented using the image gradient, they are also cut short to prevent them from crossing edges in the image, resulting in a more coherent image with crisper edges. Other edge-preserving approaches have been investigated, including the use of curved strokes [Hertzmann 1998] or the approximation of the local structure of the image [Shiraishi and Yamaguchi 2000]. A locally orderless image representation [Koenicker and van Doorn 1999], which preserves histograms of pixel values within small regions of interest, provides some explanation for why painterly approaches are effective, in addition to being another way of producing a stylized display.

In some systems, brushes over a range of sizes are combined to capture image structures at the various scales [Hertzmann 1998; Shiraishi and Yamaguchi 2000]. This allows for the omission of fine detail in the image that does not noticeably change the painting. This approach uniformly removes detail of a particular scale across the entire image if it is not very different from its surroundings (in the same way lossy image compression works). Of course, when used at coarser scales and higher tolerances, this same approach can be used to simplify an entire painting. A similar effect was achieved using relaxation in [Haerberli 1990]. Note, however, that this approach omits strokes based on frequency and intensity difference alone, and not on content, and only produces reasonable results in images with a single subject, and no distracting content. The next section describes our painterly rendering approach which uses eye-tracking data to select and emphasize structures in the image that the user found important.

3 Painting using a Perceptual Model

Our implementation of a painterly rendering system combines aspects from several existing approaches [Haerberli 1990; Litwinowicz 1997; Hertzmann 1998]. The paintings consist of a collection of curved strokes, defined on a jittered grid for a chosen set of image scales, and painted in random order. Strokes are painted using a single color and drawn from a palette with varying thickness.

The significant contribution of our work comes from its novel use of a perceptual model that is modulated with eye-tracking data. In a data-driven fashion, we transform the image by *pruning* in order to select just those perceptual elements that people looked at extensively, using an existing model of people's visual sensitivity

that was described in Section 2.1.

3.1 Eye-tracking Data

The selected image is presented for viewing by the user—this is our user interface. During this viewing period, an eye-tracker is passively taking measurements which indicate where on the screen the user is looking at any particular time. Section 2.2 describes how this is parsed into a set of n fixations:

$$\{\mathbf{f}_i = (x_i, y_i, t_i) \mid i \in [1..n]\} \quad (2)$$

where (x_i, y_i) are the image coordinates (in pixel units) describing the location of the fixation, and t_i is the duration of this fixation (in seconds). (Fixations that are outside the image are simply discarded.) Again, Figure 2(d) contains an example of such a data set.

3.2 The Perceptual Model

How can we use eye-tracking data to infer which parts of the image the user found interesting? Research in psychology suggests this data is a fairly direct means to determine this [Henderson and Hollingworth 1998; Just and Carpenter 1976; Mackworth and Morandi 1967; Yarbus 1967]. Still, care must be taken in making inferences about the user’s interest from fixations. A model of visual acuity alone does not reflect the fact that longer fixations indicate more interest, or that very brief fixations could be due to the presence of distinctive image features with irrelevant content [Just and Carpenter 1976]. Finally, even though it was possible within the limits of vision for the viewer to perceive a particular image feature, this does not mean it was consciously noticed or attended.

This section describes how this model of visual sensitivity is used for converting the eye-tracking data into a form that can be related to the image. Beyond this, simple adjustments are made to account for fixation duration and limits of attention. More research in human vision is needed here, and this model can be improved accordingly when more is known.

Visual acuity, as described by the model from Section 2.1, describes the highest perceptible frequency $H(e)$ as a function of the eccentricity angle e , which is measured from the center of the eye. To relate this to a particular point \mathbf{p} in the image, the Euclidean distance $d_i(\mathbf{p})$ (in pixel units) is measured from this point to a particular fixation location (x_i, y_i) , as in Figure 3(a). From here, simple trigonometry and calibration of the eye-tracking setup (to determine the screen size, resolution, and distance to the viewer) is used to convert distances $d_i(\mathbf{p})$ into angular eccentricities $e_i(\mathbf{p})$ [Duchowski and Vertegaal 2000].

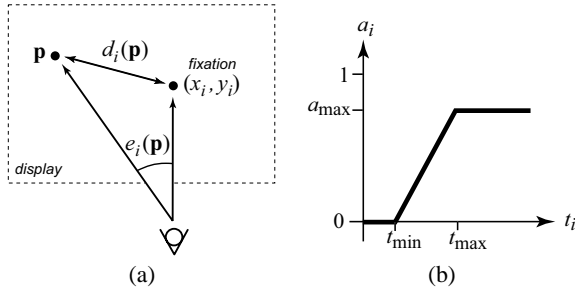


Figure 3: (a) Distance computations for computing eccentricities with respect to a particular fixation when placing a brush at \mathbf{p} . (b) A simple attention model defined as a piecewise-linear function for determining the scaling factor a_i for fixation \mathbf{f}_i based on its duration t_i . Very brief fixations (below t_{\min}) are ignored, with a ramping up (at t_{\max}) to a maximum level of a_{\max} .

The perceived frequency at \mathbf{p} is computed as the maximum of the frequencies $H(e_i(\mathbf{p}))$ considered over all fixations \mathbf{f}_i . A simple model of attention and processing is included to account for fixation duration. A frequency scaling factor $a_i \in [0, a_{\max}]$ (which depends on the duration t_i) discounts brief fixations as shown in Figure 3(b). The figure also shows how durations below t_{\min} are ignored, with the fixation’s influence increasing linearly until t_{\max} (we use $t_{\min} = \frac{1}{6}$ s and $t_{\max} = \frac{1}{3}$ s). This discounts brief fixations that are likely to be caused by visually distinctive features that are not semantically relevant.

Capping the maximum value of a_i for all fixations to $a_{\max} \in (0, 1]$ is used to scale back the acuity model to match the size of the smallest brush used, keeping the *relative* detail constant. Section 3.4 describes our method of choosing a_{\max} in terms of the brush size. Given this, the highest perceivable frequency at the location \mathbf{p} within the image is given by:

$$f_{\max}(\mathbf{p}) = \max_i [a_i \times H(e_i(\mathbf{p}))] \quad (3)$$

This model can be used directly to produce an abstract image by blurring the input with a Gaussian filter whose radius varies over the image (and corresponds to (3)), or by rendering a locally orderless image [Koenderink and van Doorn 1999] with a similarly varying aperture size. In this respect, this model effectively provides an “importance map” for the input image. The next section describes how this model is used to make decisions that affect the organization of the painting.

3.3 Applying the Model

Painterly renderings are simply collections of brush strokes. Applying the perceptual model to a painting simply prunes away those strokes that the user probably did not see. The previous section describes how this model defines a limit on the spatial frequency of information visible to the user. The decision to include a stroke in the painting is made by associating a frequency with each stroke, and determining if this frequency is less than the limit imposed by the model and data.

We follow [Hertzmann 1998], where curved strokes are applied in a coarse-to-fine fashion; coarser strokes draw their color from blurrier versions of the original image, and are correspondingly thicker. Given a brush diameter D , we can determine the relative spatial frequency of a stroke using this brush in a direction perpendicular to the stroke. A stroke of width D corresponds to one half of a cycle of the sine gratings used in contrast sensitivity experiments, which corresponds to a frequency of $f = 1/2D$ [Reddy 1997].

Consider the placement of a stroke with frequency f at the image location \mathbf{p} . The stroke is included in the painting if the visual information carried by it would have been perceptible to the user: if $f < f_{\max}(\mathbf{p})$ (measured at the start of the stroke). The resulting image contains only those strokes that convey the information as specified by the model.

3.4 Stroke Styles

The color used to paint the stroke can be adjusted to achieve various effects by altering the contrast or saturation of the stroke color [Haeberli 1990]. The perceptual model provides a means of making these adjustments based on image content. For instance, lowering the contrast in unviewed regions makes them less noticeable; raising it emphasizes objects. Color contrast is still not well understood [Regan 2000]; adjusting the contrast and saturation in this way is quite reasonable.

Define the function $u(\mathbf{p})$ ranging from 0 (where the user did not look at \mathbf{p}) to 1 (where the user fixated \mathbf{p} for a sufficiently long period

of time):

$$u(\mathbf{p}) = \frac{f_{\max}(\mathbf{p})}{a_{\max}G} \quad (4)$$

Style transformations such as those described in [Haeberli 1990] are extended to incorporate fixations using u as follows:

- *Contrast enhancement*: Important edges are emphasized by high-pass filtering the image based on u , and extrapolating from a blurred version of the image at \mathbf{p} out beyond the original pixel value. The amount of extrapolation changes linearly with $u(\mathbf{p})$, being c_{\min} when $u = 0$ and c_{\max} when $u = 1$ (an extrapolation amount of 1 produces no change). c_{\min} and c_{\max} are global style parameters for controlling the type of contrast change. For example, choosing $[c_{\min}, c_{\max}]$ to be $[0, 2]$ raises contrast where the user looked, and lowers contrast where they didn't. (Default: $[c_{\min}, c_{\max}] = [0, 2]$.)
- *Saturation enhancement*: Colors can also be enhanced; colors are intensified in important regions and de-saturated in background areas. The transformation proceeds the same as with contrast, now specified using $[s_{\min}, s_{\max}]$, and extrapolating between the original pixel value and its corresponding luminance value. As an example, choosing $[s_{\min}, s_{\max}]$ to be $[0, 1]$ makes the unattended portions of the image more grey. (Default: $[s_{\min}, s_{\max}] = [0, 1.2]$.)

There are also a number of parameters used in forming the curved strokes (not all of which use the perceptual model):

- *Brush size*: Brush strokes of varying widths are used to cover the image starting at size b_{\min} , incrementing by b_{diff} , with b_n evenly spaced brushes in total. (Default: $b_{\min} = 5$, $b_{\text{diff}} = 3$, $b_n = 8$.)
The brush size is used to determine a_{\max} , the scaling factor for the acuity model. We can set this value to place the frequency $a_{\max}G$ just above the frequency that corresponds to the smallest brush, so that the central quarter of the fovea is covered by these strokes (this yields $a_{\max} = \frac{1}{12}$ for $b_{\min} = 5$).
- *Stroke length*: l_{\min} and l_{\max} are the minimum and maximum stroke length. (Default: $l_{\min} = 1$, $l_{\max} = 15$.)
- *Early stroke termination*: The value of s_{color} is a threshold value used to determine if a stroke should be terminated based on the Euclidean distance between the color at each end of the stroke (as it is constructed). This is a simple alternative to source image comparison [Hertzmann 1998]. (Default: $s_{\text{color}} = 0.05$.)
- *Stroke path determination*: Stroke construction in [Hertzmann 1998] proceeded by following gradient contours. Noise in uniform regions that can produce spaghetti-like strokes can be ignored using a gradient magnitude threshold s_{grad} . If the gradient magnitude at a particular image location is below this threshold, the stroke will simply proceed straight ahead, instead of following the gradient contour. Should a stroke commence in a uniform region of the image, it is assigned a default orientation s_{orient} . (Default: $s_{\text{grad}} = 0.012$, $s_{\text{orient}} = 10$.)

4 Results

An interaction with our system proceeds as follows. An image is selected for transformation, and is displayed on the screen in the presence of an eye-tracker. We have experimented with two different tasks where the user is instructed to:

1. “Look at the image for five seconds.”
2. “Look at the image; press a button when you are finished.”

The image is then displayed. The user can also adjust parameters to achieve different effects. In almost all cases, the only parameter that is worth changing is the minimum brush size b_{\min} (which limits the overall detail in the painting). For the following examples, we have primarily selected “non-artistic” images to demonstrate the abstraction performed by using eye-tracking data¹. Artistic photographs (or stock photographs) often contain a single subject, with much of the content transformation already performed by the photographer. Unless specified, parameter values are the defaults.

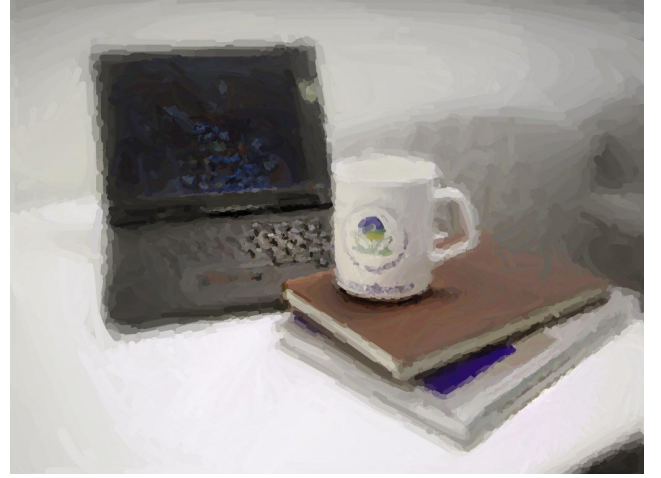


Figure 4: A painterly rendering using eye-tracking data (Task 1).

A typical abstracted painting is displayed in Figure 4 which uses the image from Figure 2(b) and fixations in Figure 2(d)—notice how the detail present in the rendering corresponds to the locations the user viewed. Figure 5 shows another typical transformation.

Two sets of results are presented in Figure 6; the observed fixations for each are displayed to the left of the resulting painting. Notice how the detail present in the image is reflected by the fixations. The first set of fixations in (a) were gathered with task 1 described above. The second set in (c) used task 2; the user viewed the image for 22 seconds.

The examples in Figure 7 compare use of the perceptual model and eye-tracking data to a fully automatic approach. A fully automatic approach can also be modulated by changing the brush size, as seen in (c) and (d). However, the detail is affected uniformly over the entire image, making the resulting paintings much less organized with regard to content. Using the data in (b), the abstracted transformation in (e) shows a nice balance of content—it has a clear subject. The automated paintings in (c) and (d) do not—they either include too much or too little detail in different parts of the painting. The painting produced with a coarser brush in (f) contains less detail, but still has the meaningful abstraction that is present in (e).

5 Discussion

We have presented a new technique for producing abstracted painterly renderings from eye-tracking data. Very brief interactions are required to gather this data, yet they can produce quite pleasing results. The use of a perceptual model, coupled with the goal-directed nature of eye movements makes this possible.

¹All images and eye-tracking data are available at:
<http://www.cs.rutgers.edu/~decarlo/abstract.html>

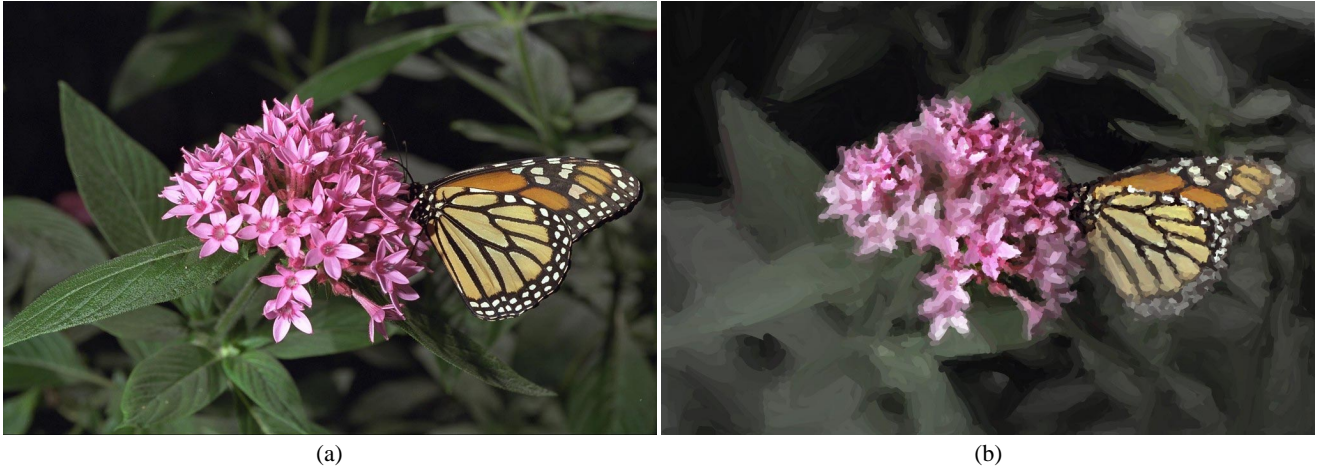


Figure 5: An example transformation of photograph in (a) to painting in (b). (Task 2). (Photo courtesy <http://philip.greenspun.com>)



Figure 6: Fixations and results depend on the user's task. Transforming (a) to (b) uses task 1 data, while (c) to (d) uses task 2 data.



Figure 7: Comparison using the photograph in (a) with and without the eye-tracking data in (b). Without data, (c) too much detail is left in using a finer brush $b_{\min} = 5$, while (d) too much is taken out using a coarse brush $b_{\min} = 10$. Using fixation data, (e) background subjects have much less detail, making the desired subject of the rendering obvious; uses finer brush $b_{\min} = 5$. Using a coarser brush $b_{\min} = 7.5$ (f) maintains the relative level of detail between foreground and background. (Task 1).



Figure 8: Adjacent detail can be inappropriately emphasized in the background, such as in the shutters located behind the main subject.

Further research for this framework is required, as right now, the results do not exhibit the level of artistry as seen in hand crafted painterly renderings—this gap can certainly be narrowed. Research in computational vision can be used to analyze the images, perhaps to separate out objects from background. Unwanted background detail can be included that is nearby important foreground objects, such as the shutters behind the subject's head in Figure 8—such an approach might alleviate this problem. Better models of human perception are required as well; our system will accommodate them as they become available. Certain artifacts are visible in the output as well. Perhaps most noticeable is the tendency for the images to appear blurry where large strokes are placed; this is caused by sampling the colors from a blurry image. While this can be alleviated using anisotropic smoothing [Perona and Malik 1990], it then requires a more sophisticated algorithm for cleanly placing strokes along sharp edges.

Our technique is also applicable to other domains where abstraction is desired, such as the non-photorealistic rendering of 3D models with a managed level of detail. The painterly processing of video streams represents a particular challenge, as it's not clear how the interaction should proceed: only one fixation location for each image frame is possible, given only a single viewing. Tracking of previously viewed objects over time, coupled with fixation data gathered from multiple viewings of the video might produce reasonable and temporally consistent abstracted video.

Acknowledgments

Thanks to Eileen Kowler, Manish Singh, Jan Koenderink, John Henderson, Spike Hughes, Cassidy Curtis, and Matthew Stone. Partially supported by NSF Instrumentation 9818322.

References

- 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.
- DUCHOWSKI, A., AND VERTEGAAL, R. 2000. Eye-based interaction in graphical systems: Theory and practice. In *SIGGRAPH 2000 Course Notes #5*. ACM SIGGRAPH.
- DUCHOWSKI, A. 2000. Acuity-matching resolution degradation through wavelet coefficient scaling. *IEEE Transactions on Image Processing* 9, 8 (Aug.), 1437–1440.
- 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.
- HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of ACM SIGGRAPH 98*, 453–460.
- JACOB, R. J. 1993. Eye-movement-based human-computer interaction techniques: Toward non-command interfaces. In *Advances in Human-Computer Interaction, Volume 4*, H. Hartson and D. Hix, Eds. Ablex Publishing, 151–190.
- JUST, M. A., AND CARPENTER, P. A. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 441–480.
- KAHNEMAN, D. 1973. *Attention and Effort*. Prentice-Hall.
- KELLY, D. 1984. Retinal inhomogeneity: I. spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A* 74, 1, 107–113.
- KOENDERINK, J. J., AND VAN DOORN, A. 1999. The structure of locally orderless images. *International Journal of Computer Vision* 31, 2/3, 159–168.
- KOENDERINK, J. J., M. A. BOUMAN, A. B. D. M., AND SLAPPEDEL, S. 1978. Perimetry of contrast detection thresholds of moving spatial sine wave patterns. II. the far peripheral visual field (eccentricity 0-50). *Journal of the Optical Society of America* 68, 6, 850–854.
- LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. In *Proceedings of ACM SIGGRAPH 97*, 407–414.
- MACKWORTH, N., AND MORANDI, A. 1967. The gaze selects informative details within pictures. *Perception and Psychophysics* 2, 547–552.
- PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of ACM SIGGRAPH 98*, 287–298.
- PERONA, P., AND MALIK, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Pattern Analysis and Machine Intelligence* 12, 7, 629–639.
- REDDY, M. 1997. *Perceptually Modulated Level of Detail for Virtual Environments*. PhD thesis, University of Edinburgh.
- REDDY, M. 2001. Perceptually optimized 3D graphics. *IEEE Computer Graphics and Applications* 21, 5 (September/October), 68–75.
- REGAN, D. 2000. *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion and Binocular Disparity*. Sinauer.
- SALISBURY, M. P., ANDERSON, S. E., BARZEL, R., AND SALESIN, D. H. 1994. Interactive pen-and-ink illustration. In *Proceedings of ACM SIGGRAPH 94*, 101–108.
- SHIRAIISHI, M., AND YAMAGUCHI, Y. 2000. An algorithm for automatic painterly rendering based on local source image approximation. In *Proceedings of the First International Symposium on Non-photorealistic Animation and Rendering*, 53–58.
- SIBERT, L. E., AND JACOB, R. J. K. 2000. Evaluation of eye gaze interaction. In *Proceedings CHI '00*, 281–288.
- VERTEGAAL, R. 1999. The gaze groupware system: Mediating joint attention in mutiparty communication and collaboration. In *Proceedings CHI '99*, 294–301.
- 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.

Abstracted Painterly Renderings Using Eye-Tracking Data: Anthony Santella, Doug DeCarlo

