

Abstract Painting with Interactive Control of Perceptual Entropy

MINGTIAN ZHAO and SONG-CHUN ZHU, University of California, Los Angeles and Lotus Hill Institute

This article presents a framework for generating abstract art from photographs. The aesthetics of abstract art is largely attributed to its greater perceptual ambiguity than photographs. According to psychological theories [Berlyne 1971], the ambiguity tends to invoke moderate mental effort in the viewer for interpreting the underlying contents, and this process is usually accompanied by subtle aesthetic pleasure. We study this phenomenon through human experiments comparing the subjects' interpretations of abstract art and photographs, and quantitatively verify, the increased perceptual ambiguities in terms of recognition accuracy and response time. Based on the studies, we measure the level of perceptual ambiguity using entropy, as it measures uncertainty levels in information theory, and propose a painterly rendering method with interactive control of the ambiguity levels. Given an input photograph, we first segment it into regions corresponding to different objects and parts in an interactive manner and organize them into a hierarchical parse tree representation. Then we execute a painterly rendering process with image obscuring operators to transfer the photograph into an abstract painting style with increased perceptual ambiguities in both the scene and individual objects. Finally, using kernel density estimation and message-passing algorithms, we compute and control the ambiguity levels numerically to the desired levels, during which we may predict and control the viewer's perceptual path among the image contents by assigning different ambiguity levels to different objects. We have evaluated the rendering results using a second set of human experiments, and verified that they achieve similar abstract effects to original abstract paintings.

Categories and Subject Descriptors: I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—*Perceptual reasoning*; I.3.4 [Computer Graphics]: Graphics Utilities—*Paint systems*; I.4.10 [Image Processing and Computer Vision]: Image Representation—*Hierarchical*; J.5. [Computer Applications]: Arts and Humanities—*Fine arts*

General Terms: Algorithms, Experimentation, Human Factors

Additional Key Words and Phrases: Abstract art, entropy, image parsing, painterly rendering, perceptual ambiguity, semantics

ACM Reference Format:

Zhao, M. and Zhu, S.-C. 2013. Abstract painting with interactive control of perceptual entropy. *ACM Trans. Appl. Percept.* 10, 1, Article 5 (February 2013), 21 pages.

DOI = 10.1145/2422105.2422110 <http://doi.acm.org/10.1145/2422105.2422110>

1. INTRODUCTION

Abstract artworks like Monet's famous *Wheatstack* (Figure 1(a)) have a characteristic charm beyond photographic and representational arts. In particular, observing and interpreting an abstract artwork,

The work at UCLA was partially supported by NSF grant IIS-1018751 and ONR MURI grant N000141010933; the work at LHI was supported by NSFC grant 60970156.

Authors' addresses: M. Zhao and S.-C. Zhu, UCLA Dept. of Statistics, 8125 Mathematical Sciences Building, Box 951554, Los Angeles, CA 90095-1554; emails: mtzhao@ucla.edu, sczhu@stat.ucla.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2013 ACM 1544-3558/2013/02-ART5 \$15.00

DOI 10.1145/2422105.2422110 <http://doi.acm.org/10.1145/2422105.2422110>

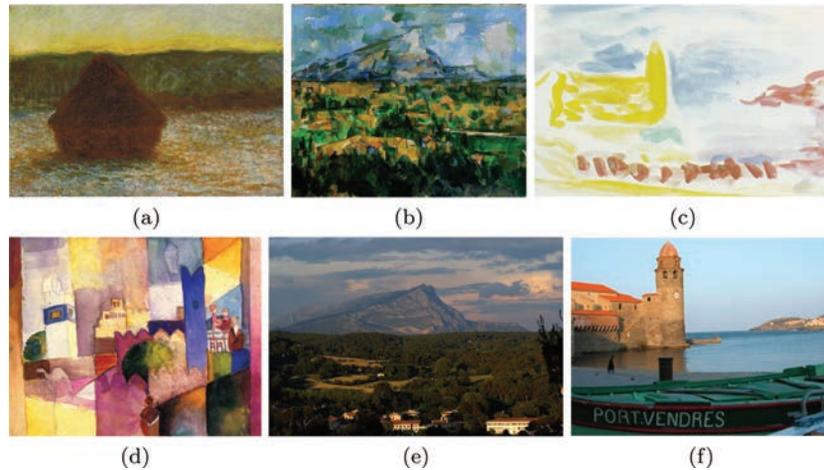


Fig. 1. Some abstract artworks. (a) *Wheatstack (Thaw, Sunset)*, 1890–1891 by Claude Monet. (b) *Le Mont Sainte-Victoire*, 1902–1904 by Paul Cézanne. (c) *View of Collioure (The Bell Tower)*, 1905 by Henri Matisse. (d) *Kairouan (III)*, 1914 by August Macke. (e) A photograph with a similar scene to (b). (f) A photograph with a similar scene to (c).

in some sense, is like playing a guessing game with the artist, which typically has ambiguities and causes confusion, but this experience is usually also interesting and rewarding.

This subtle beauty of abstract art has long been noted by both artists and psychologists. Wassily Kandinsky (1866–1944), a Russian abstract painter, attributed the “fairy-tale power and splendor” of Monet’s haystacks to the surprise and confusion caused by their indistinct painting style [Lindsay and Vergo 1994, p.363]. Daniel Berlyne (1924–1976), a pioneer in theoretical and experimental psychology, further explained this phenomenon with his theory of the motivational aspects of perception [Berlyne 1971, pp.61–114; Konečni 1978; Funch 1997, pp.26–33]. According to Berlyne, the process of observing and interpreting aesthetic patterns in abstract art involves certain levels of *perceptual ambiguity*. To resolve the ambiguity, the observer subconsciously puts in mental effort (e.g., continuous guesses until reaching the correct answer [Kersten 1987]) that can lead to moderate changes in the arousal levels in his/her nervous system, which in turn reward him/her with emotional pleasure.

The confusion and ambiguity of abstract art may exist in various forms, styles, and levels, as shown by the examples in Figure 1. In many abstract artworks, such ambiguities are often achieved by

- *preserving* visual features in certain semantic dimensions (e.g., scene configuration, identity of object/part, color/shape/texture characteristics), and
- *freeing* (e.g., spatially disarranging, obscuring, randomizing) the other dimensions.

While the former leaves clues for our visual perception, the latter usually poses challenges, for example:

- In Monet’s wheatstack, Cézanne’s mountain and Matisse’s bell tower shown in Figures 1(a) through 1(c), global structures of the scenes are mostly preserved in the sense that they are recognizable, while the appearance and shapes of individual objects are obscured. In particular, the objects in Figure 1(b) are obscured in different degrees, so the viewer usually recognizes the mountain first, which further helps to recognize the trees and huts in the context. We call this sequential recognition effect (i.e., the viewer recognizes less obscured objects first, then understands the scene and other objects with the help of contextual information) the *perceptual path*.

— In Macke’s *Kairouan (III)* shown in Figure 1(d), as well as in Picasso’s famous *Guernica* and *Violin and Guitar*, the identifiability of individual objects/parts are well preserved, while the spatial configurations of the scenes are disarranged.

— In some modern paintings, such as Pollock’s drip paintings, only some low-level colors and shapes are preserved, while high-level semantic and geometric structures are randomized.

In this article, we focus on the style of Monet’s wheatstack and Cézanne’s mountain, which preserves the scene structures while obscuring individual objects. We conducted human experiments to study the different ambiguity levels between such abstract paintings and photographs. Based on the experiments, we define and measure the level of perceptual ambiguity, and propose an abstract rendering method using image obscuring operators such as color shift and shape deformation, in which we can compute and interactively control the ambiguity levels and perceptual paths using kernel density estimation and message-passing algorithms. Through a second set of human experiments, we verified that our rendering results achieve similar abstract effects as original abstract paintings by artists.

The rest of this article is organized as follows. Section 1.1 summarizes related work on abstract art in computer graphics, image analysis, and perception; Section 1.2 lists our contributions, and improvements over our previous work [Zhao and Zhu 2010]. We report on our human experiments and analyze the experimental results in Section 2. In Section 3, we introduce a numerical measure for perceptual ambiguity, called perceptual entropy, which is defined on a hierarchical parse tree representation for image content. We also explain how a parse tree is constructed using interactive image segmentation and labeling methods. Then, in Section 4, we present the image-obscuring and painterly rendering techniques to manipulate the perceptual entropy. To complete the system pipeline, in Section 5, we show how the perceptual entropy is computed and adjusted, and how the perceptual path is predicted. Section 6 illustrates our rendering results. In Sections 7 and 8 we present the second set of human experiments, which verifies the rendering results. Finally, we conclude in Section 9 with discussion.

1.1 Related Work

Recently, in the computer graphics and image analysis communities, especially in the nonphotorealistic rendering (NPR) area [Gooch and Gooch 2001; Strothotte and Schlechtweg 2002], there have been continuing efforts to understand and render abstract artworks in different styles.

In computer graphics, Haerberli [1990] first proposed abstract image representations using brush strokes. Image representation with brush strokes essentially abstracts images by modifying many high-frequency details and only preserves relatively low-frequency surfaces and gentle gradients. Later, the study on stroke-based rendering was further extended by many painterly rendering methods [Meier 1996; Litwinowicz 1997; Hertzmann 1998; Zeng et al. 2009] for better visual effects. To achieve nonuniform abstraction across an image which is naturally performed by artists, DeCarlo and Santella [2002] developed an approach for stylization and abstraction of photographs, which identifies visually attractive elements utilizing eye-tracking data, and preserves more details of such areas during rendering. Recently, automatic methods for image and video simplification or abstraction have been developed [Orzan et al. 2007; Kyprianidis 2011; Olsen and Gooch 2011; Mould 2012]. The main idea of these methods is to filter images to remove weak edges/ridges in relatively flat areas, which human vision is not very sensitive to. For vector graphics, Mi et al. [2009] proposed a method for 2D shape abstraction using part-based representations by identifying and preserving important parts. Moreover, many specific styles of abstract art have also been widely studied and simulated, including image mosaics [Finkelstein and Range 1998; Orchard and Kaplan 2008], drip-painting [Lee et al. 2006], cubism [Collomosse and Hall 2003], abstract texture synthesis by sampling [Morel et al. 2006], emerging images [Mittra et al. 2009], and pixelation [Gerstner et al. 2012].

On the image analysis aspect, Pollock's famous drip paintings were analyzed using fractal mathematics [Mureika et al. 2005; Jones-Smith and Mathur 2006; Taylor et al. 2007]. Statistical and computer vision methods have also been applied in analyzing and classifying paintings of various styles [Wallraven et al. 2009; Hughes et al. 2010]. Recently, Rigau et al. [2008] proposed aesthetic informational measures to evaluate artistic images based on information-theoretic principles. There is also growing interests in the subtle effects of perceptual ambiguity in abstract art [Arnheim 1971; Yevin 2006; Hertzmann 2010].

In the literature of perception and psychophysics, there are also studies on abstract images generated using artistic rendering techniques. Gooch et al. [2004] presented a study on human facial illustrations, and showed that the rendered facial illustrations and caricatures are as effective in communicating complex facial information as photographs. Wallraven et al. [2007] also studied the effects of artistic stylization in stylized facial expressions, using both real-world and computer-generated images. Redmond and Dingliana [2009] compared different NPR styles in the perception of abstracted scenes, and observed that salient target objects can be effectively emphasized using NPR, given appropriate scene context and level of stylization. Mandryk et al. [2011] evaluated and compared the emotional responses to NPR images of different styles.

1.2 Our Contributions

Most of the above studies on rendering focused on relatively low-level image features (e.g., color, gradient). A few methods also work in the perceptual space by dealing with visual saliency and attention [DeCarlo and Santella 2002]. In contrast, the creation and appreciation of abstract art entail the manipulation of categorical recognition for scenes, objects, and parts where ambiguity and confusion may occur. Our method for rendering abstract paintings is based on the hypothesis that they usually have greater ambiguities for understanding than photographs, which is fundamentally different from previous image abstraction methods. To achieve this, this article makes the following contributions:

- We introduce the image parsing method [Tu et al. 2005] to provide a hierarchical descriptor of image contents for studying the mechanism of abstract art at the semantic level.
- We compare abstract art and photographs containing different categories of objects using human experiments, and quantitatively measure the differences in recognition accuracy and response time between them, which reflect their differences in perceptual ambiguity.
- In the frameworks of Bayesian statistics and information theory, we define a numerical measure of the levels of perceptual ambiguities, called perceptual entropy, and develop algorithms to compute the entropy for images and predict their most likely perceptual paths.
- We propose a painterly rendering method for generating abstract painting images from photographs, in which we have interactive control of the ambiguity levels and perceptual paths.

This article extends our previous work on abstract painting [Zhao and Zhu 2010]. Compared with the previous study, this article presents additional or improved methods and results in two main aspects:

- Improved models and algorithms to compute the entropies over hierarchical image structures, for simulating human visual perception better. These include a logistic-regression-based distance metric between image regions using color, shape, and texture features (in Section 5.1); a more accurate approximate of the joint perceptual entropy according to the most probable parse tree configurations (in Section 5.2); and a sequential algorithm for predicting the most likely perceptual path among image regions (in Section 5.4).

Table I. List of Scene and Object Categories In this Article (which distribute widely over common categories usually appearing in paintings)

7 Scene Categories	42 Object Categories			
close-up	abstract background	big mammal	bike	bird
indoor	bridge	building	bus/car/train	butterfly/insect
landscape	chimney	clothing/fabrics	door/window	face/skin
portrait	fish	flag/sign	flower	fruit
seascape	furniture/bench	glass/porcelain	grass/straw/reed	ground/earth/pavement
skyline	hair	house/pavilion	human	kite/balloon
streetscape	lamp/light	leaf	mountain	pillar/pole
	road/street/alley	rock/stone/reef	sand/shore	ship/boat
	sky/cloud/glow	small mammal	snow/frost	statue
	sun/moon/star	tower/lighthouse	tree/trunk/twig	umbrella
	wall/roof	water/spindrift		



Fig. 2. Example image patch pairs of abstract paintings of different object categories and their corresponding photographs used in our human experiments. The dog belongs to the small mammal category.

— More comprehensive human experiments. These include more extensive experiments and analyses on the effects of perceptual ambiguity reflected in recognition accuracy and response time (in Sections 2 and 7); and an additional experiment on the effect of perceptual paths (in Section 8).

2. HUMAN EXPERIMENTS ON THE LEVELS OF PERCEPTUAL AMBIGUITY: PART ONE

We use human experiments to compare the mental effort in interpreting abstract art images and photographs, so as to verify our hypothesis that abstract art images should generally have greater ambiguity levels than photographs, which may be reflected by lower recognition accuracy and longer response times.

We collected 123 abstract art images from works of well-known artists and divided them into different scene and object categories. Table I shows the list of categories we use in this article, some of which do not exist in the 123 images but frequently appear in other paintings. We manually pair these images up with 123 photographs collected using Web search engines, which match the abstract art images in both categories and contents well. Figure 2 includes some example image patches from the matched pairs.

These images are then presented to 20 human subjects (voluntary college and graduate students of art, science, and engineering majors) within a limited time span (one minute per image) on a 17-inch color monitor. During the experiment, these images are depaired and presented in random order, and each image is seen by a subject only once. Following our pre-experiment instructions, as soon as the subject feels that he/she recognizes the foreground object (highlighted with a bounding box) in the

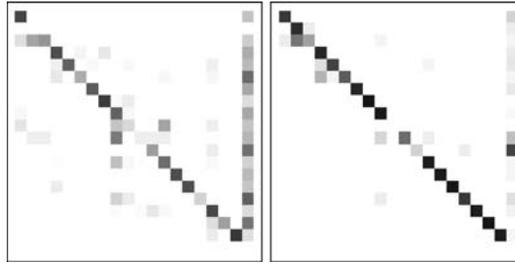


Fig. 3. Object category confusion matrices obtained in our experiments for abstract paintings (left) and photographs (right). The horizontal axis for reported categories and the vertical axis for true categories. The darkness of the each grid is proportional to its corresponding frequency of the subjects' reports. The rightmost column of each matrix stands for either the "none of these categories" report or failed attempts at recognition within the limited time span (one minute per image).

center of the image, he/she hits the keyboard. Then the image disappears, and the response time is recorded. The subject is immediately asked to choose one of the categories in Table I provided on the screen or report "none of these categories."

Recognition Accuracy. The recognition accuracy can be visually reflected by the confusion matrices summarizing reported vs. true interpretations, as shown in Figure 3, in which the horizontal axes of the matrices stand for reported categories and the vertical axes stand for true categories, so the diagonal elements correspond to correct recognition results. The rightmost column of each matrix stands for either the "none of these categories" report, or failed attempts at recognition within the limited time span (one minute per image). The matrix for abstract art is more scattered with weaker diagonal elements, and has a darker rightmost column. This means the subjects generally have lower recognition accuracy for abstract art than for photographs. Meanwhile, even for abstract art, the diagonal elements are still darker than other grids in each row, which means that the images are usually still recognizable with effort, otherwise they may become meaningless and inaeesthetic like purely flat or noise images.

Response Time. For abstract art images, we expect greater mental effort by the human subjects, reflected by their longer response times for recognition. Figure 4 displays a few box plots of the recorded response times for object recognition in abstract paintings and photographs, in which six categories are included as examples (corresponding to the image pairs displayed in Figure 2). These plots show greater average response times for abstract paintings than that for photographs. But we also notice that not all significance levels are high, as confirmed by paired one-sided t -tests shown in Table II. The negative t -score for the bird is due to the extreme outlier of 28 seconds in the photograph sample (not shown in Figure 4). If we remove that pair, we get t -score = 1.4648 and p -value = 0.08012. This confirms the significance of the difference in mental effort.

3. IMAGE UNDERSTANDING AND PERCEPTUAL AMBIGUITY IN A COMPUTATIONAL PERSPECTIVE

From a computational perspective, vision is an ill-posed problem. It is widely acknowledged in the human and computer vision communities that the imaging process loses lots of information about the 3D world, and thus we cannot restore the contents uniquely from an image. Instead, visual perception is achieved by computing the most probable interpretations of the observed images. When a dominant interpretation with significantly larger probability than all the other interpretations does not exist, the image causes perceptual ambiguity. Thanks to the artists' exquisite skills, good abstract artworks usually have carefully calculated (but implicitly calculated in terms of numerical computing) probabilities for competing interpretations in order for the viewers to enjoy the guessing game with the artists.

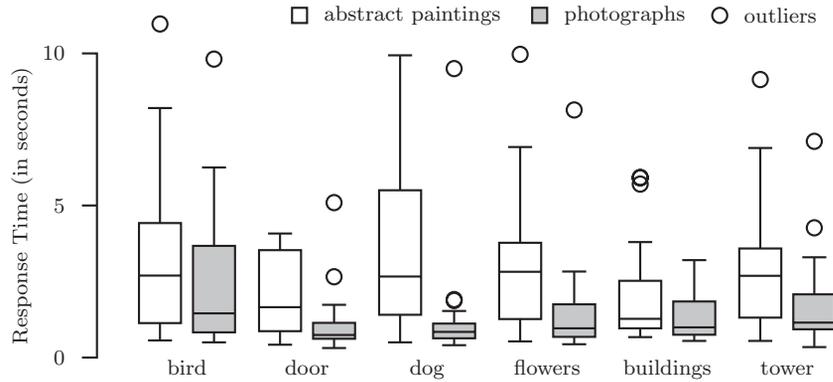


Fig. 4. Box plots of response time for object recognition in abstract paintings and photographs. Three outliers greater than 11s are not shown (photograph of bird: 28s, abstract painting of door: 14.8s, and abstract painting of dog: 21.1s).

Table II. Paired One-Sided *t*-Tests Comparing the Response Time for Object Recognition in Abstract Paintings and Photographs, Corresponding to Figure 4

	bird	door	dog	flowers	buildings	tower
<i>t</i> -score	-0.3775	2.0045	2.524	4.2201	2.4147	1.995
<i>p</i> -value	0.645	0.02974	0.01033	0.0002318	0.013	0.03029

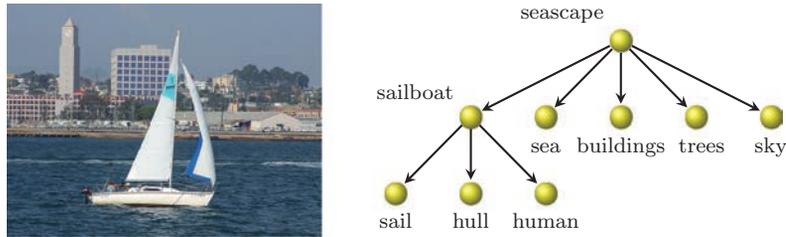


Fig. 5. A seascape image (left), Courtesy of pdphoto.org. and its example parse tree (right).

To quantitatively measure the level of ambiguity, we can compute the information (Shannon) entropy of the probabilities of all interpretations [Cover and Thomas 2006], and define

$$\text{ambiguity level} = \text{entropy}(\text{probabilities of interpretations}).$$

Therefore, to proceed, we need a representation for the interpretation of image content.

3.1 Parse Tree and Entropy

We adopt the *parse tree* introduced by Tu et al. [2005] for computer vision from computational linguistics. Similar to parse trees for English sentences, a parse tree for image representation is a hierarchical decomposition. It has a root node corresponding to the entire scene of the image, which has a few children/descendent nodes corresponding to the constituent objects and parts. As shown in Figure 5, the photograph is a seascape scene (i.e., the label of the root node of the parse tree), which is then decomposed into five objects/regions: sailboat, sea, buildings, trees, and sky. The sailboat node is further decomposed into three children: sail, hull, and human on board. In general, we view a parse tree as a directed acyclic graph (DAG) $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, whose vertices \mathcal{V} represent the nodes, and directed edges

\mathcal{E} represent the parent→child links in the parse tree. Each node $i \in \mathcal{V}$ is associated with its category label ℓ_i (e.g., a category in Table I) and visual features \mathcal{A}_i (e.g., shape, color, and texture).

To model perceptual ambiguity with the parse tree representation, we make two assumptions.

Assumption 1. The main cause of perceptual ambiguity is due to the obscured objects, rather than unclear parse tree structures (i.e., we do not study the abstract style, as shown in Figure 1(d)).

Assumption 2. For understanding abstract art in the sense of recognizing the contents, we only care about computing the category labels, ignoring visual features specific to object instances (e.g., we do not have to describe whether a human is tall or short).

Hence, we simplify the parse tree to a vector representation of the category labels of its nodes, $\mathcal{L} = (\ell_1, \ell_2, \dots, \ell_K)$, where the labels for a correct interpretation should (i) correspond well to the image; and (ii) be compatible to each other, for example, a boat rather than a bus is compatible with the sea surface. Under the Bayesian framework, in computer vision and pattern recognition it is a standard practice to compute the *maximum a posteriori* (MAP) estimate, $\hat{\mathcal{L}}_{\text{MAP}} = \arg \max_{\mathcal{L}} p(\mathcal{L}|\mathbf{I})$, as the best interpretation of image \mathbf{I} . But the MAP estimate only captures the major mode or peak of the posterior probabilities $p(\mathcal{L}|\mathbf{I})$, and cannot tell how much better this best interpretation is than the other interpretations, which influences the ambiguity, and thus our mental efforts in visual perception.

Compared with MAP, the perceptual entropy defined by

$$\mathcal{H}(\mathcal{L})|\mathbf{I} = \sum_{\mathcal{L}} -p(\mathcal{L}|\mathbf{I}) \log p(\mathcal{L}|\mathbf{I})$$

describes the uncertainty/ambiguity associated with the posterior possibilities. It is worth mentioning that $\mathcal{H}(\mathcal{L})|\mathbf{I}$ differs from the conditional entropy [Cover and Thomas 2006]

$$\mathcal{H}(\mathcal{L}|\mathbf{I}) = \sum_{\mathbf{I}} p(\mathbf{I}) \mathcal{H}(\mathcal{L})|\mathbf{I} = \sum_{\mathcal{L}, \mathbf{I}} -p(\mathcal{L}, \mathbf{I}) \log p(\mathcal{L}|\mathbf{I}),$$

in the way that $\mathcal{H}(\mathcal{L})|\mathbf{I}$ only deals with a specific image and thus does not sum over \mathbf{I} . For the abstract paintings we study here, with $p(\mathcal{L}|\mathbf{I})$ often having more than one local maxima (i.e., multimodality) corresponding to multiple competing interpretations [Yevin 2006], we expect $\mathcal{H}(\mathcal{L})|\mathbf{I}$ to be significantly greater than the close-to-zero ambiguities of photographs, but still much lower than the upper bound $\log |\Omega_{\mathcal{L}}|$, where $|\Omega_{\mathcal{L}}|$ is the volume of the space of \mathcal{L} (i.e., number of all possible category label combinations).

3.2 Constructing the Parse Tree

Given an input photograph, as automatic image parsing is not a solved problem in general, we use an interactive program to construct the parse tree in three steps.

Step 1. Image Segmentation. We first segment the given image into regions corresponding to different objects. To achieve this, we adopt a scribble-based interactive segmentation algorithm [Lombaert et al. 2005]. Using this method, each time we draw foreground and background scribbles, we can segment the image into two parts. We continue with this procedure to further segment each part until every object is separated from its neighboring regions, or a resolution limit is reached (i.e., we are not interested in even smaller objects or parts). With the number of nodes $K < 15$ for most images, the segmentation is usually completed within several minutes.

Step 2. Hierarchical Organization. Using the recursive foreground-background segmentation scheme above, we obtain a binary tree in which each non-leaf node corresponds to a region we have already segmented into two parts. However, some nodes might not correspond to individual semantic objects (e.g., a node containing parts of two different objects), and sometimes an object is mistakenly divided

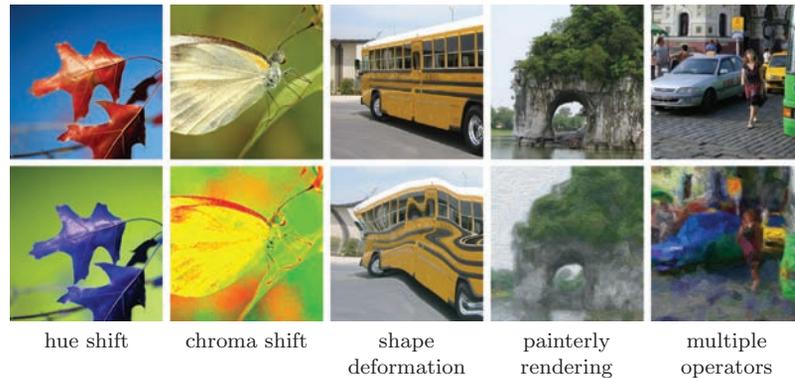


Fig. 6. An illustration of the image obscuration and painterly rendering operators.

into multiple branches in the binary tree. In order to obtain a meaningful hierarchy conforming to the image semantics, we delete and merge nodes to form a multiway tree in an interactive manner on the software interface.

Step 3. Category Labeling. We manually label the categories of all nodes in the parse tree (scene category for the root node and object category for the other nodes). This ground-truth parse tree with all category labels is helpful for later computing the ambiguity level. But the category labeling step is optional. Even without the manually selected category labels, our method can still compute the ambiguity level, possibly with slightly lower accuracy. The usage of ground-truth labels will be explained in Section 5.1.

4. OBSCURATION AND RENDERING

During rendering, our method allows interactive control of perceptual entropy by sliding a bar on the software interface, and the system obscures and abstracts the image accordingly. Different objects are allowed to have different entropy levels, which makes some areas of the image easier to understand than the others, leading to the perceptual path effect further as mentioned in Section 1. We will discuss this effect further in Section 5.4. In the rendering process, the parse tree, including the segmentation map, is the central representation. It preserves the configuration of the scene and allows us to propagate the contextual information between nodes in order to estimate the ambiguity levels.

The main task of the rendering engine is to transfer the visual appearance of an input photograph into an abstract painting style. According to vision research [Marr 1982], color, shape, and texture are the key features of an image for visual perception. Therefore, we transfer the visual appearance using two groups of image processing operators: (i) image obscuration, which processes the color and shape of the input image; and (ii) painterly rendering, which processes the texture. Figure 6 illustrates these operators.

Image Obscuration. We first transfer the input image into the CIELCH color space, whose three channels are lightness, chroma, and hue, respectively. It is a cylindrical form of the perceptually uniform CIELAB color space. To obscure the color information of the image, random noises are added to the hue and the color tone. The noise follows a truncated Gaussian distribution whose standard deviation is positively related to the desired ambiguity level. Since paintings are usually more saturated than photographs, a positive shift also related to the ambiguity level (e.g., following a Gamma distribution with its location parameter proportional to the ambiguity level) is added to the chroma channel to increase the saturation. To obscure the shape information, an image region is warped using a thin

plate spline (TPS) transformation [Barrodale et al. 1993] with its boundary pixels as control points. The offsets of these boundary points are randomly sampled (2D Gaussian, truncated) whose average distance is related to the specified ambiguity level. To ensure smoothness for the warped image, a diagonal regularization term is added to the kernel matrix of the TPS transformation.

Painterly Rendering. For the texture appearance of paintings, we adopt our earlier work on stroke-based painterly rendering [Zeng et al. 2009; Zhao and Zhu 2010]. The layout and attributes of the brush strokes are controlled by stochastic stroke processes [Zhao and Zhu 2011], whose parameters are related to the desired ambiguity level. The entire rendering scheme can be viewed as a top-down, hierarchical data-generating process. In a stochastic way, the rendering parameters are generated according to the desired ambiguity levels, and they further generate the painting image. Using the above stochastic operations on color, shape, and texture, we expect that the final ambiguity level of the rendered abstract painting should be significantly larger than the original photograph, and we will verify this through computation and human experiment.

5. COMPUTATION AND INTERACTIVE CONTROL OF PERCEPTUAL ENTROPY

We compute the actual ambiguity level of the rendered image and compare it with the desired value, in order to ensure that we have achieved the expected effects. Otherwise, the image should be re-rendered with (automatically) adjusted parameters.

Since visual perception involves both direct object recognition using visual features and indirect recognition using contextual information [Oliva and Torralba 2007], we compute $p(\mathcal{L}|\mathbf{I})$ using a method that accounts for both aspects. The probability of the category labels in the parse tree can be factorized according to

$$p(\mathcal{L}|\mathbf{I}) = \frac{1}{Z} \prod_{i \in \mathcal{V}} \phi_i(\ell_i) \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(\ell_i, \ell_j) = \frac{1}{Z} \prod_{i \in \mathcal{V}} p(\ell_i|\mathbf{I}_i) \prod_{(i,j) \in \mathcal{E}} \tilde{f}(\ell_i, \ell_j)$$

in which we assume each node is only correlated with its parent and children (i.e., the Markov property).

— The unary term $\phi_i(\ell_i) = p(\ell_i|\mathbf{I}_i)$ is the posterior probability of object recognition for image region \mathbf{I}_i without the context (we call it *local evidence*). We compute this using a nonparametric model with a large human annotated dataset from LHI [Yao et al. 2007]. Details are explained in the next section.

— The binary term $\psi_{ij}(\ell_i, \ell_j) = \tilde{f}(\ell_i, \ell_j)$ models the contextual relations in terms of prior/empirical pairwise frequencies between parent-child nodes. We approximate these frequencies by counting numbers of parent-child pairs with corresponding categories on the human labeled parse trees in the LHI dataset.

5.1 Local Evidence

The computation of $p(\ell_i|\mathbf{I}_i)$ is achieved using kernel density estimation [Duda et al. 2000], with a sample of approximately $N = 25,000$ kernels (probabilistic voters) from the LHI image dataset. Each voter (\mathbf{J}_n, ℓ_n) has its image region \mathbf{J}_n and category label ℓ_n . Figure 7 displays a few example voters.

With the voters, the local evidence is computed with

$$p(\ell_i|\mathbf{I}_i) \propto \sum_n \exp\{-\lambda D(\mathbf{I}_i, \mathbf{J}_n)\} \mathbf{1}(\ell_i = \ell_n),$$



Fig. 7. Example probabilistic voters from the LHI image dataset for computing the local evidence for object recognition.

in which $\mathbf{1}(\cdot)$ is the indicator function, and λ is a rate parameter controlling the overall entropy level. The logistic distance function

$$D(\mathbf{I}_i, \mathbf{J}_n) = \frac{1}{1 + \exp\{-\beta_0 - \sum_j \beta_j \|h_j(\mathbf{I}_i) - h_j(\mathbf{J}_n)\|_2\}}$$

measures the dissimilarity between two image regions, in which each h_j extracts a feature statistic in color, shape, or texture channels of the image regions:

- For color, h_j is a normalized 2D hue-chroma histogram with 32 blocks (8 sectors for hue and 4 levels for chroma).
- For shape, h_j is a normalized 2D spatial histogram of image region boundary pixels with $8^2 = 64$ blocks, assuming rough alignment according to the bounding box of each image region.
- For texture, h_j are normalized histograms of sine and cosine Gabor filter responses over the image region (2 types, 4 directions, 3 scales, and 16 bins each).

The β s are precomputed using logistic regression by setting $D(\mathbf{J}_{n_1}, \mathbf{J}_{n_2}) = \mathbf{1}(\ell_{n_1} \neq \ell_{n_2})$ for pairs of voters. Since the number of pairs $\binom{N}{2}$ is huge, we randomly take a small sample of 50,000 pairs for regression.

The motivation of using a logistic distance is that it is difficult to define a reasonable metric distance function between categories, for example, it is unclear whether “flower” is closer to “building” or “furniture”. Instead, we usually only care about the two states of the distance function: (i) zero for the same category and (ii) nonzero otherwise.

As we mentioned before, if the ground-truth categories are manually labeled for the parse tree, we can compute $p(\ell_i|\mathbf{I}_i)$ more accurately by including the original image region $\mathbf{I}_i^{\text{ori}}$ from the input photograph in the group of voters. Usually, when the desired ambiguity level is not too high, the rendered image region should still be quite similar to the original one in terms of $D(\mathbf{I}_i, \mathbf{I}_i^{\text{ori}})$, then the original image region will have a heavy voting weight and bring significant information gain [Cover and Thomas 2006] to $p(\ell_i|\mathbf{I}_i)$.

Figure 8 illustrates the idea of this voting scheme for computing the local evidence performed in the space spanned by color, shape, and texture features.

5.2 Contextual Information Propagation and Entropy Approximation

Even with all unary terms $p(\ell_i|\mathbf{I}_i)$ and binary terms $\tilde{f}(\ell_i, \ell_j)$ available, it is still infeasible to compute $p(\mathcal{L}|\mathbf{I})$ for $\mathcal{H}(\mathcal{L}|\mathbf{I})$, since the space of \mathcal{L} is usually too huge to explore. For example, if there are $K = 10$ nodes of 40 possible categories, the space volume of \mathcal{L} is $|\Omega_{\mathcal{L}}| = 40^K \approx 10^{16}$.

Fortunately, the parse tree is a singly connected network, and we can compute the most probable joint configurations of its category labels efficiently using max-product message passing [Sy 1993].

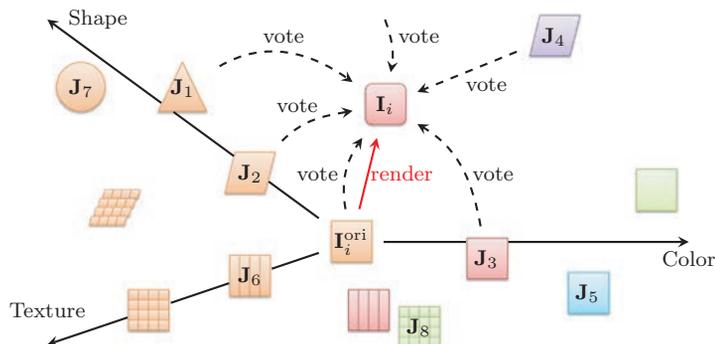


Fig. 8. Illustration of our voting algorithm for the local evidence in the space spanned by color, shape, and texture features.

Using Sy's algorithm, a user-specified number of the most probable configurations can be derived in the descendent order of their probabilities. During this process, both the local evidence and compatibility terms are considered, and messages are propagated between connected nodes in an iterative way to update local beliefs.

With the probabilities $p(\mathcal{L}|\mathbf{I})$ of the top- M ($M \ll |\Omega_{\mathcal{L}}|$) most probable joint configurations, we can approximate the entropy using

$$\hat{\mathcal{H}}(\mathcal{L})|\mathbf{I} = \sum_{\text{top-}M} -q(\mathcal{L}|\mathbf{I}) \log q(\mathcal{L}|\mathbf{I}),$$

in which $q(\mathcal{L}|\mathbf{I})$ are renormalized from $p(\mathcal{L}|\mathbf{I})$ over the top- M configurations according to

$$q(\mathcal{L}|\mathbf{I}) = \frac{p(\mathcal{L}|\mathbf{I})}{\sum_{\text{top-}M} p(\mathcal{L}|\mathbf{I})}.$$

This approximation essentially drops all the rest configurations after the top- M , which is reasonable, since usually only very few configurations are possible due to the compatibility terms (e.g., a bus cannot be part of a tree), and when we try to understand an abstract artwork, we can quickly eliminate configurations with too low probabilities at early stages, and the ambiguity is only caused by those with relatively high probabilities among the top- M . In practice, the choice of M is a balance between computational precision and cost. According to our experiments, for $K = 10$ nodes, $M = 100 \sim 1000$ should work fairly well.

5.3 Normalized Perceptual Entropy

For different images, the number of nodes and the space volume $|\Omega_{\mathcal{L}}|$ of \mathcal{L} may vary. In order to have a common measure of the ambiguity level, we use a normalized version of perceptual entropy defined as

$$\tilde{\mathcal{H}}(\mathcal{L})|\mathbf{I} = \frac{\hat{\mathcal{H}}(\mathcal{L})|\mathbf{I}}{\log M} \in [0, 1].$$

This number is then compared to the desired ambiguity level which the user specified before rendering.

According to the comparison, we determine whether the rendered painting image has the desired ambiguity level (e.g., the difference between the computed and desired ambiguities is within $\epsilon = 10\%$). If it does not, the painting is re-rendered with adjusted parameters according to a negative feedback mechanism. Suppose the desired ambiguity level is $\tilde{\mathcal{H}}_0^*$, and after the first rendering, the computed ambiguity is $\tilde{\mathcal{H}}_1$ (normalized perceptual entropy), we will adjust the rendering parameters according

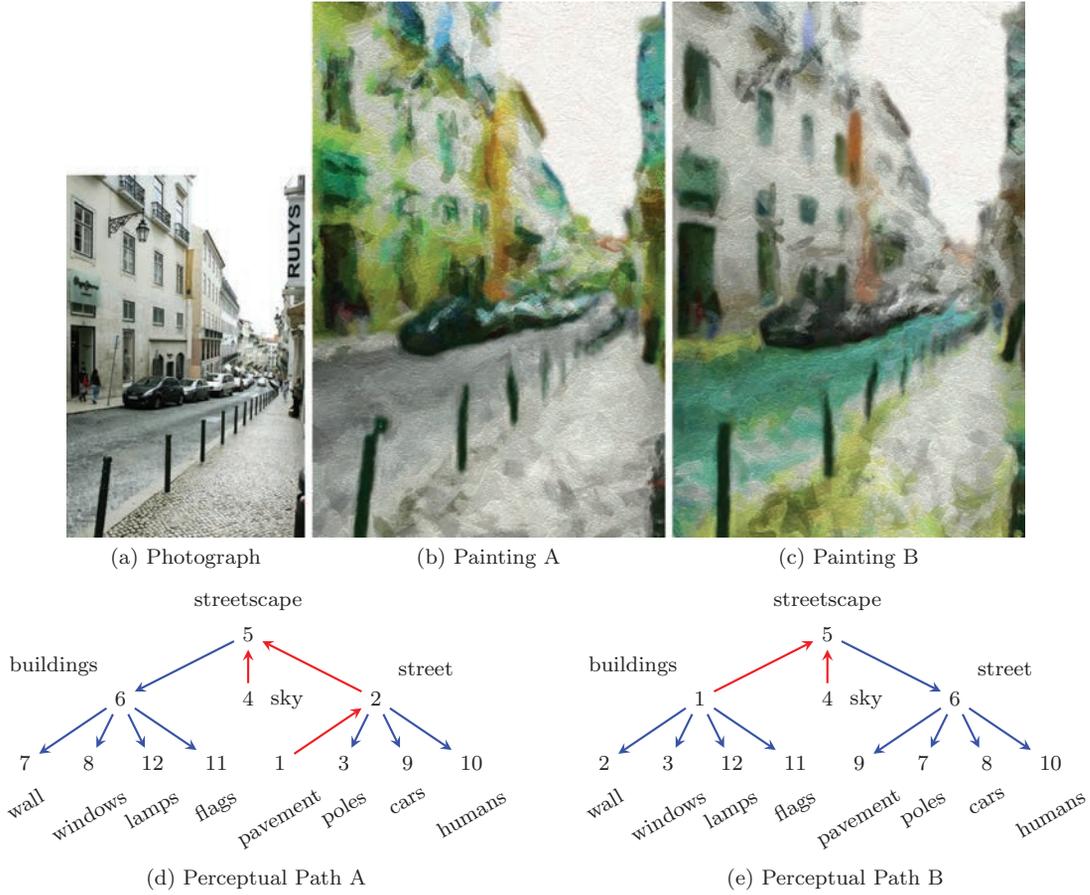


Fig. 9. A streetscape photograph (courtesy of public-domain-image.com) and its two abstract paintings rendered by setting different ambiguity levels for different objects, to simulate the perceptual path effect. Zoom to 400% to view details. Their predicted different perceptual paths using the algorithm described in Section 5.4 are displayed in (d) and (e), respectively. The numbers indicate the sequences of the nodes in the paths. The arrows indicate the propagation of contextual information (red: bottom-up; blue: top-down).

to a *virtual* desired ambiguity level $\tilde{\mathcal{H}}_1^*$ and redo the rendering, in which

$$\tilde{\mathcal{H}}_1^* = \begin{cases} (\tilde{\mathcal{H}}_0^*)^2 / \tilde{\mathcal{H}}_1, & \text{if } \tilde{\mathcal{H}}_1 > \tilde{\mathcal{H}}_0^* + \epsilon, \\ 1 - (1 - \tilde{\mathcal{H}}_0^*)^2 / (1 - \tilde{\mathcal{H}}_1), & \text{if } \tilde{\mathcal{H}}_1 < \tilde{\mathcal{H}}_0^* - \epsilon. \end{cases}$$

If necessary, we continue to compute $\tilde{\mathcal{H}}_2, \tilde{\mathcal{H}}_2^*, \tilde{\mathcal{H}}_3, \tilde{\mathcal{H}}_3^*, \dots, \tilde{\mathcal{H}}_t, \tilde{\mathcal{H}}_t^*, \dots$ and repeat the rendering until $\tilde{\mathcal{H}}_t$ is close to $\tilde{\mathcal{H}}_0^*$. Due to the randomness involved in the process, the convergence of $\tilde{\mathcal{H}}$ is not guaranteed. But in practice, with a relatively generous difference threshold ϵ (e.g., 10% to 20%), we can usually get close to the desired level within a few iterations.

5.4 Perceptual Paths

Image understanding can be achieved through various top-down and bottom-up computing processes [Han and Zhu 2009], during which the viewer recognizes the image contents in an order with the propagation of contextual information. For example, in Figure 1(b), the mount is usually recognized in

the first place, which further helps the recognition of the highly abstracted trees and huts in the front. In this article, we call it the perceptual path effect, and control the path/order by assigning different ambiguity levels to different objects, letting the viewer recognize less obscured objects before more heavily obscured ones, as illustrated in Figure 9. In Figure 9(b) we set lower ambiguity level to the street than to the buildings, while for Figure 9(c) we set the levels in the opposite way. We can predict the perceptual paths as shown in Figures 9(d) and 9(e) by simulating a greedy information propagation process. We first identify the node with the lowest ambiguity level and take its most probable label as our interpretation. Then we propagate this information to the other nodes and identify the one with the lowest ambiguity among them. This process continues until we have reached all the nodes. Here is the detailed algorithm:

- (1) We represent a perceptual path as a sequence of the nodes in the parse tree, denoted by \mathbf{S} .
- (2) According to our previous work [Zhao and Zhu 2010], we are able to compute the marginal probabilities $p(\ell_i|\mathbf{I})$ and entropies $\mathcal{H}(\ell_i|\mathbf{I})$ of the nodes given the whole image (with contextual information), using sum-product belief propagation [Yedidia et al. 2001]. The first recognized node is

$$s_1 = \arg \min_{i \in \mathcal{V}} \mathcal{H}(\ell_i|\mathbf{I})$$

with the lowest entropy, and we push s_1 into \mathbf{S} .

- (3) We fix the category label of node s_1 to

$$\ell_{s_1}^* = \arg \max_{\ell_{s_1}} p(\ell_{s_1}|\mathbf{I}),$$

which essentially sets the entropy of ℓ_{s_1} to zero. Then with this new information, we redo the belief propagation for the rest nodes to compute their probabilities $p_1(\ell_i|\mathbf{I})$ and entropies $\mathcal{H}_1(\ell_i|\mathbf{I})$.

- (4) We compute the second recognized node s_2 and its label $\ell_{s_2}^*$ in similar ways to steps (2) and (3),

$$s_2 = \arg \min_{i \in \mathcal{V} \setminus s_1} \mathcal{H}_1(\ell_i|\mathbf{I}) \quad \text{and} \quad \ell_{s_2}^* = \arg \max_{\ell_{s_2}} p_1(\ell_{s_2}|\mathbf{I}),$$

and push s_2 into \mathbf{S} .

- (5) We continue the steps above to sequentially figure out the labels of the other nodes,

$$s_2, \ell_{s_2}^* \rightarrow p_2, \mathcal{H}_2 \rightarrow s_3, \ell_{s_3}^* \rightarrow p_3, \mathcal{H}_3 \rightarrow \dots \rightarrow s_K, \ell_{s_K}^*,$$

until we have reached all the nodes. Now we have obtained the whole sequence \mathbf{S} .

During the process of fixing the nodes' category labels ℓ_i in a sequence, the number of unknown labels decreases, and thus the perceptual entropy of the parse tree decreases. The entropy reaches zero when all labels are fixed. Actually, the above algorithm starting from the lowest marginal entropies is a greedy method to minimize the overall ambiguity eliminated by fixing labels

$$E(\mathbf{S}) = \sum_{i=1}^K \mathcal{H}_{i-1}(\ell_{s_i}|\mathbf{I})$$

with $\mathcal{H}_0(\ell_i|\mathbf{I}) = \mathcal{H}(\ell_i|\mathbf{I})$. In contrast, if the sequence \mathbf{S} starts from nodes with high entropies, $E(\mathbf{S})$ tends to be higher. Considering that the elimination of ambiguity is associated with the mental effort to make decisions, the greedy process described above minimizes the effort in interpreting the image.

6. RENDERING RESULTS

Using the pipeline introduced above, we have rendered many abstract painting images from photographs collected from Web search engines.



Fig. 10. Three abstract paintings of different ambiguity levels rendered using our method, corresponding to the example in Figure 5. Zoom to 600% to view details.



Fig. 11. A photograph of UCLA Royce Hall, and two corresponding abstract paintings of different ambiguity levels rendered using our method. Zoom to 600% to view details.

Figure 10 displays three abstract paintings corresponding to the example in Figure 5, with $\tilde{\mathcal{H}}$ at approximately 0.25, 0.5, and 0.75, respectively. We can see that as the ambiguity level increases, the color and shape in the images become more heavily obscured, making both the sailboat and the background objects more difficult to recognize. Note that your perception of the three paintings should have already been affected by seeing the source photograph beforehand. In fact, knowing the image contents may also make it difficult for an artist to assess the ambiguity level during the creation of abstract art. Being able to numerically compute the perceptual entropy in object recognition, our program is helpful.

Figure 11 displays a photograph of UCLA Royce Hall, and its two corresponding abstract paintings rendered using our method, with $\tilde{\mathcal{H}}$ at approximately 0.25 and 0.75, respectively. For this example, we have segmented the image into five objects: sky, building, trees, grass, and road. Note that in both paintings, the road is almost impossible to recognize without the context.

The example displayed in Figure 12 includes many object categories: sky, mountain, sand, water surface, human, bench, and so on. Most of these objects are heavily obscured, but the rendered paintings for this landscape are still clear enough for appreciation.

Figure 13 shows a few more abstract paintings rendered using our method, whose ambiguity levels are between 0.25 and 0.75 for all examples according to our computation.

7. HUMAN EXPERIMENTS ON THE LEVELS OF PERCEPTUAL AMBIGUITY: PART TWO

In addition to the rendering pipeline and entropy computing method introduced above, we would like to further verify that the rendered abstract paintings do have our expected ambiguity effects, similar



Fig. 12. *Promenade Morecambe* (photograph courtesy of Tom Curtis / FreeDigitalPhotos.net). Zoom to 600% to view details.



Fig. 13. More abstract paintings rendered using our method. Zoom to 600% to view details.

to original paintings. We do this with another set of human experiments, comparing our rendered paintings with their source photographs. The computed ambiguity levels of these rendered abstract paintings are between 0.25 and 0.75.

Most experimental settings remain the same as in Section 2. We selected approximately 100 photographs, and have them segmented and their parse trees constructed manually. Then we render them into abstract paintings at different, desired ambiguity levels, and ask 15 human subjects (from the 20 in Section 2) to recognize the objects (highlighted using bounding boxes) in these images. The recognition accuracy and speed of the subjects are recorded.

Recognition Accuracy. Figure 14 displays the confusion matrices for object recognition in our rendered abstract paintings and corresponding source photographs. Still, the horizontal and vertical axes stand for reported and true categories, respectively. The two matrices show that subjects generally have lower recognition accuracy for our rendered paintings than photographs. Comparing these matrices to those in Figure 3, we can see our rendered images have similar ambiguity effects against photographs, to those of original paintings by artists, with diagonal elements still having the highest frequencies in most cases.

Response Time. Figure 15 displays a few box plots of the recorded response time for object recognition in our rendered abstract paintings and their corresponding photographs, in which five categories are included as examples. Its corresponding t -test results are shown in Table III. We can see similar statistics in response time to those in Figure 4 and Table II.

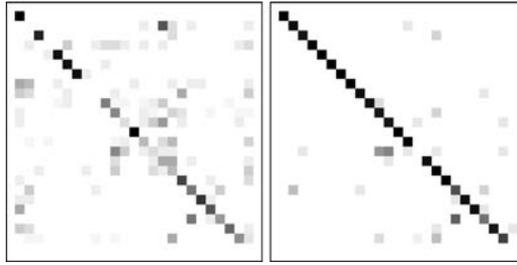


Fig. 14. Object category confusion matrices for our rendered abstract paintings (left) and their original photographs (right). The horizontal axis stands for reported categories and the vertical axis stands for true categories. The darkness of each grid is proportional to its corresponding frequency in the subjects’ reports. The rightmost column of each matrix stands for either the “none of these categories” report, or failed attempts at recognition within the limited time span (one minute per image).

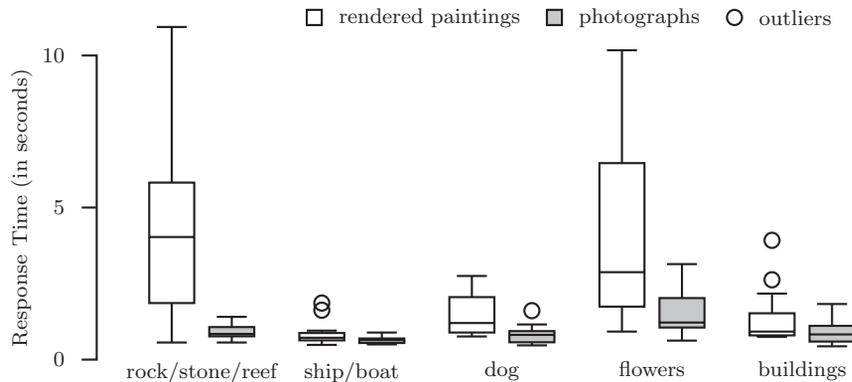


Fig. 15. Box plots of response time for object recognition in rendered abstract paintings and corresponding source photographs. Two outliers greater than 11s are not shown (rendered painting of rock/stone/reef: 11.8s, and photograph of flowers: 15.3s).

Table III. Paired One-Sided t -Tests (comparing the response time for object recognition in rendered paintings and their source photographs, corresponding to Figure 15)

	rock/stone/reef	ship/boat	dog	flowers	buildings
t -score	3.91	2.532	4.2238	2.0427	2.5585
p -value	0.0007852	0.01197	0.0004252	0.03019	0.01137

8. HUMAN EXPERIMENTS ON THE EFFECT OF PERCEPTUAL PATHS

Besides the global ambiguities of images, we also use human experiments to verify the effect of perceptual paths introduced in Section 5.4. The perceptual paths have higher dimensions than perceptual entropies and are more difficult to observe. In the literature of perception, researchers have used eye-tracking techniques to study the paths of viewers’ attentions across images [DeCarlo and Santella 2002], but this is not suitable for our case, since attention and semantic understanding are very different phases in vision [Marr 1982].

As a simplified investigation, we set up a verbal experiment to extract the rough order in which objects in an image are recognized. We select 12 human subjects, and randomly divide them into two groups of 6 people. Figures 9(b) and 9(c) are presented to the two groups, respectively, on a 23-inch color monitor. Each subject views the presented image 10 times in limited time spans with increasing

and their pairwise order is weak. The latter objects are in different branches of the parse tree. It is possible that after recognizing “buildings,” the bottom-up process prevails over the top-down processes for exploring other regions of the image, instead of looking into details within the region. The pursuit criterion in Section 5.4 needs improvement to address this issue.

Although this experiment cannot directly capture the propagation/flow of contextual information shown in Figure 9 from inside the minds of the subjects, the orders of object recognition reflected by Figure 16 mostly agree with the predicted paths, which partially supports our explanation of the perceptual path effect.

9. DISCUSSIONS AND CONCLUSION

In this article we have presented both human and computerized studies on a type of abstract art which obscures the shapes and/or appearances of objects in images, and preserves global scene structures. Our studies are based on the hypothesis that abstract arts usually have higher ambiguity levels than photographs and representational arts, which differs fundamentally from most previous work on abstract art in NPR, image analysis, and perception. After verifying this hypothesis using human experiments, we defined the perceptual entropy as a numerical measure for the level of perceptual ambiguity, and proposed a method for the rendering of abstract paintings that is capable of controlling the entropy to the user’s desired levels. By assigning different ambiguity levels to different image regions, we may predict and thus roughly control the perceptual paths in whose orders viewers are most likely to understand the image content. We have also examined the ambiguity levels and perceptual paths of our rendered abstract paintings using human experiments, and showed that they have achieved our expected effects.

The methods described in this article still have several limitations.

- The cognitive processes involved in our visual perception are certainly much more complex than the parse tree model and bottom-up/top-down computing that we have studied in this article. Semantic understanding and the perceptual path should be determined not only by the categories and ambiguities of image contents, but also other factors such as object saliency which affects visual attention in early vision.

- Simplifying the parse tree into a vector of object category labels only works for limited abstract art styles. In many images the segmentation and/or the semantic grouping are also perceptually ambiguous. To address this problem, more general image parsing methods could be utilized to consider different parse tree structures.

For future work, there are a few directions in which we may further explore the proposed framework.

- To make our method more general, it is necessary to conduct further human and computerized studies on more abstract art styles, including those freeing image information not only in the object level, but also in the scene level to the high end (e.g., Picasso’s paintings), or in the local statistics level to the low end (e.g., Pollock’s paintings).

- On the rendering aspect, our current method can be improved by integrating with better semantics-based painterly rendering algorithms for frequently-depicted object categories in paintings, and more artistic ways of image obscuration (e.g., feature exaggeration [Gooch et al. 2004]).

- We look forward to better algorithms for predicting the perceptual paths, for example, by considering bottom-up and top-down processes differently with certain priorities in belief propagation, or having different weights for image regions of different sizes and saliency levels.

- We may use the rendered abstract art as testing images to study human perception and attention mechanisms [Gooch et al. 2004; Wallraven et al. 2007; Redmond and Dingliana 2009], for example, by

extending the human experiments presented in this article, or by other techniques such as recording the eye saccades and fixations.

ACKNOWLEDGMENTS

We would like to thank our colleagues and friends at UCLA and LHI for their participation in the experiments, and the anonymous reviewers for their suggestions on improving the presentation of this article.

REFERENCES

- ARNHEIM, R. 1971. *Entropy and Art: An Essay on Disorder and Order*. University of California Press.
- BARRODALE, I., SKEA, D., BERKLEY, M., KUWAHARA, R., AND POECKERT, R. 1993. Warping digital images using thin plate splines. *Pattern Recogn.* 26, 2, 375–376.
- BERLYNE, D. E. 1971. *Aesthetics and Psychobiology*. Appleton-Century-Crofts, Inc.
- COLLOMOSSE, J. P. AND HALL, P. M. 2003. Cubist style rendering of photographs. *IEEE Trans. Vision. Comput. Graph.* 9, 4, 443–453.
- COVER, T. M. AND THOMAS, J. A. 2006. *Elements of Information Theory* 2nd Ed. Wiley-Interscience.
- DECARLO, D. AND SANTELLA, A. 2002. Stylization and abstraction of photographs. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '02)*. 769–776.
- DUDA, R. O., HART, P. E., AND STORK, D. G. 2000. *Pattern Classification* 2nd Ed. Wiley-Interscience.
- FINKELSTEIN, A. AND RANGE, M. 1998. Image mosaics. In *Proceedings of the 7th International Conference on Electronic Publishing (EP/RIDT '98)*. 11–22.
- FUNCH, B. S. 1997. *The Psychology of Art Appreciation*. Museum Tusculanum Press.
- GERSTNER, T., DECARLO, D., ALEXA, M., FINKELSTEIN, A., GINGOLD, Y., AND NEALEN, A. 2012. Pixelated image abstraction. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering (NPAR '12)*. 29–36.
- GOOCH, B. AND GOOCH, A. A. 2001. *Non-Photorealistic Rendering*. A. K Peters, Ltd.
- GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph.* 23, 1, 27–44.
- HAEBERLI, P. 1990. Paint by numbers: Abstract image representations. In *Proceedings of SIGGRAPH Computer Graphics (SIGGRAPH '90)*. 207–214.
- HAN, F. AND ZHU, S.-C. 2009. Bottom-up/top-down image parsing with attribute grammar. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 1, 59–73.
- Hertzmann, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)*. 453–460.
- HERTZMANN, A. 2010. Non-photorealistic rendering and the science of art. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '10)*. 147–157.
- HUGHES, J. M., GRAHAM, D. J., AND ROCKMORE, D. N. 2010. Quantification of artistic style through sparse coding analysis in the drawings of Pieter Brueghel the Elder. *Proc. Nat. Acad. Sci.* 107, 4, 1279–1283.
- JONES-SMITH, K. AND MATHUR, H. 2006. Fractal analysis: Revisiting Pollock's drip paintings. *Nature* 444, E9-E10.
- KERSTEN, D. 1987. Predictability and redundancy of natural images. *J. Opt. Soc. Amer. A* 4, 12, 2395–2400.
- KONEČNÍ, V. J. 1978. Daniel E. Berlyne: 1924–1976. *Amer. J. Psychol.* 91, 1, 133–137.
- KYPRIANIDIS, J. E. 2011. Image and video abstraction by multi-scale anisotropic kuwahara filtering. In *Proceedings of the 9th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '11)*. 55–64.
- LEE, S., OLSEN, S. C., AND GOOCH, B. 2006. Interactive 3D fluid jet painting. In *Proceedings of the 4th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '06)*. 97–104.
- LINDSAY, K. C. AND VERGO, P. 1994. *Kandinsky: The Complete Writings on Art*. Da Capo Press.
- LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. 407–414.
- LOMBAERT, H., SUN, Y., GRADY, L., AND XU, C. 2005. A multilevel banded graph cuts method for fast image segmentation. In *Proceedings of the International Conference on Computer Vision (ICCV '05)*. Vol. 1, 259–265.
- MANDRYK, R. L., MOULD, D., AND LI, H. 2011. Evaluation of emotional response to non-photorealistic images. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering (NPAR '11)*. ACM, New York, 7–16.

- MARR, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman.
- MEIER, B. J. 1996. Painterly rendering for animation. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*. 477–484.
- MI, X., DECARLO, D., AND STONE, M. 2009. Abstraction of 2D shapes in terms of parts. In *Proceedings of the 7th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '09)*. 15–24.
- MITRA, N. J., CHU, H.-K., LEE, T.-Y., WOLF, L., YESHURUN, H., AND COHEN-OR, D. 2009. Emerging images. *ACM Trans. Graph.* 28, 5, 163:1–163:8.
- MOREL, J.-M., ALVAREZ, L., GALERNE, B., AND GOUSSEAU, Y. 2006. Texture synthesis by abstract painting technique. <http://www.cmla.ens-cachan.fr/membres/morel.html>.
- MOULD, D. 2012. Texture-preserving abstraction. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering (NPAR '12)*. 75–82.
- MUREIKA, J. R., DYER, C. C., AND CUPCHIK, G. C. 2005. On multifractal structure in non-representational art. *Phys. Rev. E* 72, 046101.
- OLIVA, A. AND TORRALBA, A. 2007. The role of context in object recognition. *Trends Cogn. Sci.* 11, 12, 520–527.
- OLSEN, S. AND GOOCH, B. 2011. Image simplification and vectorization. In *Proceedings of the 9th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '11)*. 65–74.
- ORCHARD, J. AND KAPLAN, C. S. 2008. Cut-out image mosaics. In *Proceedings of the 6th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '08)*. 79–87.
- ORZAN, A., BOUSSEAU, A., BARLA, P., AND THOLLOT, J. 2007. Structure-preserving manipulation of photographs. In *Proceedings of the 5th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '07)*. 103–110.
- REDMOND, N. AND DINGLIANA, J. 2009. Investigating the effect of real-time stylisation techniques on user task performance. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization (APGV '09)*. 121–124.
- RIGAU, J., FEIXAS, M., AND SBERT, M. 2008. Informational aesthetics measures. *IEEE Comput. Graph. Appl.* 28, 2, 24–34.
- STROTHOTTE, T. AND SCHLECHTWEG, S. 2002. *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*. Morgan Kaufmann, San Francisco, CA.
- SY, B. K. 1993. A recurrence local computation approach towards ordering composite beliefs in Bayesian belief networks. *Int. J. Approx. Reason.* 8, 17–50.
- TAYLOR, R. P., GUZMAN, R., MARTIN, T. P., HALL, G. D. R., MICOLICH, A. P., JONAS, D., SCANNELL, B. C., FAIRBANKS, M. S., AND MARLOW, C. A. 2007. Authenticating Pollock paintings using fractal geometry. *Pattern Recogn. Lett.* 28, 6, 695–702.
- TU, Z., CHEN, X., YUILLE, A. L., AND ZHU, S.-C. 2005. Image parsing: Unifying segmentation, detection, and recognition. *Int. J. Comput. Vision* 63, 2, 113–140.
- WALLRAVEN, C., BÜLTHOFF, H. H., CUNNINGHAM, D. W., FISCHER, J., AND BARTZ, D. 2007. Evaluation of real-world and computer-generated stylized facial expressions. *ACM Trans. Appl. Percept.* 4, 3, 16:1–16:24.
- WALLRAVEN, C., FLEMING, R., CUNNINGHAM, D., RIGAU, J., FEIXAS, M., AND SBERT, M. 2009. Categorizing art: Comparing humans and computers. *Comput. Graph.* 33, 4, 484–495.
- YAO, B., YANG, X., AND ZHU, S.-C. 2007. Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks. In *Proceedings of the International Conferences on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR '07)*. 169–183.
- YEDIDIA, J. S., FREEMAN, W. T., AND WEISS, Y. 2001. Understanding belief propagation and its generalizations. IJCAI 2001 Distinguished Lecture Track.
- YEVIN, I. 2006. Ambiguity in art. *Complexus*, 3, 74–83.
- ZENG, K., ZHAO, M., XIONG, C., AND ZHU, S.-C. 2009. From image parsing to painterly rendering. *ACM Trans. Graph.* 29, 1, 2:1–2:11.
- ZHAO, M. AND ZHU, S.-C. 2010. Sisley the abstract painter. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '10)*. 99–107.
- ZHAO, M. AND ZHU, S.-C. 2011. Customizing painterly rendering styles using stroke processes. In *Proceedings of the 9th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '11)*. 137–146.

Received January 2012; revised October 2012; accepted October 2012