

Sisley the Abstract Painter

Mingtian Zhao* Song-Chun Zhu*
University of California, Los Angeles & Lotus Hill Institute

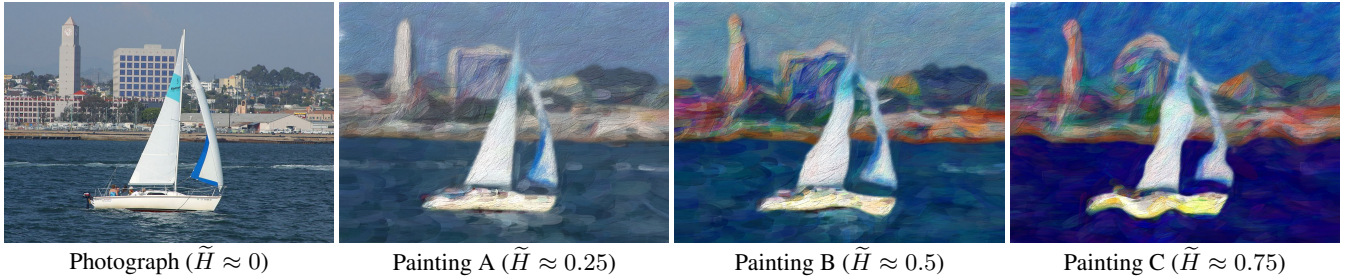


Figure 1: *Sisley renders a photograph (courtesy of pdphoto.org) into three abstract paintings at different perceptual ambiguity levels.*

Abstract

We present an interactive abstract painting system named *Sisley*. *Sisley* works upon the psychological principle [Berlyne 1971] that abstract arts are often characterized by their greater perceptual ambiguities than photographs, which tend to invoke moderate mental efforts of the audience for interpretation, accompanied with subtle aesthetic pleasures. Given an input photograph, *Sisley* decomposes it into a hierarchy/tree of its constituent image components (e.g., regions, objects of different categories) with interactive guidance from the user, then automatically generates corresponding abstract painting images, with increased ambiguities of both the scene and individual objects at desired levels. *Sisley* consists of three major working parts: (1) an interactive image parser executing the tasks of segmentation, labeling, and hierarchical organization, (2) a painterly rendering engine with abstract operators for transferring the image appearance, and (3) a numerical ambiguity computation and control module of servomechanism. With the help of *Sisley*, even an amateur user can create abstract paintings from photographs easily in minutes. We have evaluated the rendering results of *Sisley* using human experiments, and verified that they have similar abstract effects to original abstract paintings by artists.

CR Categories: 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

Keywords: abstract art, hierarchical image parsing, painterly rendering, perceptual ambiguity

1 Introduction

*e-mails: {mtzhao|sczhu}@stat.ucla.edu; mailing address: Department of Statistics, University of California, Los Angeles, CA 90095-1554, USA

©ACM, 2010. This is the author's version of the work posted here by permission of ACM for your personal use. Not for redistribution. The definitive version will be published in Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering (NPAR 2010), Annecy, France.



"I considered that the painter had no right to paint indistinctly ... and I noticed with surprise and confusion that the picture not only gripped me, but impressed itself in-eradicably on my memory."

— Kandinsky's comments on Monet's famous "Haystack"

1.1 Motivation

Wassily Kandinsky, one of the most credited abstract artists ever, considered that Claude Monet's "Haystack" painting caused some confusion which, as a consequence, makes it impressive. More precisely, because of the great *mental efforts* devoted to perception, interpreting this painting, like solving hard puzzles, becomes an interesting exploratory experience, which makes the picture unforgettable. This confusion, usually named *perceptual ambiguity*, and the mental efforts it invokes, are extraordinary for abstract arts, distinguishing them from photographs and representational arts.

This subtle phenomenon has been explained by Berlyne [1971, 61–114] using his theory of the motivational aspects of perception, that is, the perception of aesthetic patterns involves certain levels of perceptual ambiguities and mental efforts that lead to arousal level changes, which in turn cause emotional rewards and pleasures (also see [Funch 1997, 26–33]). Following this vein of thought, we think it is possible to simulate abstract paintings like Monet's "Haystack" by controlling the level of perceptual ambiguity. This can be achieved by considering vision as a process of statistical inference—a perspective that dates back to the 19th century [von Helmholtz 1866]:

"What we see is the solution to a computational problem, our brains compute the most likely causes for the photon absorptions within our eyes."

As commonly accepted in vision research, visual perception is achieved by computing the most probable interpretation of the observed image, during which perceptual ambiguity is often caused by the absence of a dominant interpretation with significantly larger probability than all the other interpretations. For the numerical measure of perceptual ambiguity, it is a common practice to adopt the information (Shannon) entropy of the probabilities of different interpretations. By carefully (but also subconsciously in virtually all cases) constructing these probabilities and the entropy using their specialized techniques [Cooke 1978], abstract painters manage to play duets with their activated audience.



Figure 2: Five abstract painting masterpieces. From left to right: “No. 5” by Jackson Pollock, “Violin and Guitar” and “Guernica” by Pablo Picasso, “Le Mont Sainte-Victoire” by Paul Cézanne, and “The Red Vineyard” by Vincent van Gogh. These abstract paintings all preserve some visual features in certain semantic dimensions/levels and free the others.

This paper aims to simulate the techniques of abstract painters on computers, in order to render abstract paintings from photographs. We try to achieve this simulation in our abstract painting system named Sisley (after Alfred Sisley, the English impressionist landscape painter), by addressing the primal problem: *how does the computer assess and control the perceptual ambiguities of images?* Fig.1 demonstrates several representative results of Sisley.

1.2 A Further Look into Abstract Arts

From ancient Chinese and Islamic calligraphy, to impressionism and expressionism, then to modern minimalism in the 21st century, abstract arts have existed in various forms and styles for centuries. Despite their great variety, these different categories actually share a key characteristic, namely, they try to preserve visual features in some semantic dimensions/levels (e.g., scene configuration, object category, color scheme), and free (i.e., randomize) the others. Take abstract paintings for example (see Fig.2),

- (1) Pollock’s famous abstract drip paintings preserve some low-level geometric statistics (e.g., fractal dimension [Taylor et al. 1999]) and totally free any other figurative features such as identifiable scenes or objects;
- (2) Picasso’s “Violin and Guitar” and “Guernica” preserve the identifiability of some individual parts, while the former frees the spatial configuration and integrity of the objects and the latter frees configurations of both the objects and the scene;
- (3) In most impressionist artworks (e.g., Monet’s Haystacks, Cézanne’s Mountains, van Gogh’s landscapes), the scene hierarchies are preserved, while appearances of objects/regions are freed. We focus on this type of techniques of impressionists (and probably some expressionists) in this paper.

In an information-theoretic language, this characteristic of preserving some features and freeing others corresponds to an increase of uncertainty or degree-of-freedom, which is commonly measured by the Shannon entropy [Jaynes 1957; Cover and Thomas 2006]. And psychological studies [Kersten 1987] have proved that the increase of uncertainty further leads to an increase of perceptual ambiguity, which is usually reflected by the demanded amount of mental efforts in recognition (e.g., number of guesses needed until the correct answer). Relying on the above connections from fidelity of semantics, to uncertainty and entropy, then to perceptual ambiguity and mental efforts, we consider the level of perceptual ambiguity as a numerical measure of how much semantic information is preserved and how much is freed. Although the ambiguity is anisotropic in different semantic dimensions (e.g., blurring an object may have different abstract effect from perturbing the spatial configuration), since this general numerical measure is not restricted to any specific techniques of artistic creation, the above idea, up to concrete analysis in future research, is capable of explaining other abstract art forms and styles (e.g., paper-cut [Xu and Kaplan 2008]).

1.3 Previous Work

In recent years, there have been plenty of investigations towards understanding abstract artworks in mathematical or statistical ways. For example, Pollock’s drip paintings have been analyzed using fractal analysis techniques [Taylor et al. 1999; Mureika et al. 2005]. Rigau et al. [2008] studied aesthetic measures in the information-theoretic framework. Wallraven et al. [2009] tried to categorize different artistic styles using image statistics. Readers may refer to [Stork 2009] for a more comprehensive survey of the literature.

For rendering abstract art images (not limited to specific styles or genres) on computers, many interesting studies have been contributed to the computer graphics community, especially the non-photorealistic rendering area [Gooch and Gooch 2001; Strothotte and Schlechtweg 2002; Durand 2002]. Haeberli [1990] proposed abstract image representations by brush strokes, which was extended by further painterly rendering studies on brush modeling, stroke layout and animation [Meier 1996; Litwinowicz 1997; Hertzmann 1998; Hertzmann and Perlin 2000; Zeng et al. 2009]. DeCarlo and Santella [2002] developed an approach for stylization and abstraction of photographs by identifying visually attended elements using eye tracking data. Winnemöller et al. [2006] built an automatic and real-time framework for image and video abstraction by image filtering preserving visual salience. Orzan et al. [2007] developed a method for structure-preserving image manipulation and enhancement. Mi et al. [2009] proposed a part-based computational method for the abstraction of 2D shapes. Specific abstract art styles have also been widely studied, including mosaics and collages [Finkelstein and Range 1998; Klein et al. 2002; Smith et al. 2005; Gal et al. 2007; Orchard and Kaplan 2008], Escherization [Kaplan and Salesin 2004], cubism [Klein et al. 2001; Colloso and Hall 2003], surrealism by image fusion [Raskar et al. 2004], futurism by motion emphasis cues [Collomosse and Hall 2006] and abstract texture synthesis [Morel et al. 2006]. Besides, Lee et al. [2006] developed a fluid dynamics based system for creating drip paintings of Pollock’s style.

Most past research focused on low-level image features (e.g., color, gradient), except a few trying to work in the perceptual space by modeling visual salience or attention [DeCarlo and Santella 2002; Winnemöller et al. 2006]. In contrast, the appreciation of art usually requires an exploratory behavior of image understanding, in the sense of recognizing the scene and objects through certain *paths of perception* [Berlyne 1971; Funch 1997], and finally grasping the event semantics. For example, one tends to recognize the mountain then the trees and huts while looking at Cézanne’s “Le Mont Sainte-Victoire” (see Fig.2), and finally get a global impression as if being personally on the scene. And for van Gogh’s “The Red Vineyard”, the path of perception usually starts from the sun in the sky, then continues to the field and the working people. For such pictures, the balance between obviousness and obscurity of image semantics is particularly important. In other words, it is crucial

to control the level of perceptual ambiguity for appropriate arousal level shifts which are just right to cause aesthetic pleasure [Berlyne 1971; Funch 1997].

1.4 Our Approach

Motivated by the above points, we deploy a three-step method for rendering abstract paintings in Sisley.

Interactive Image Parsing An input image (usually photograph) is first decomposed into a hierarchical tree representation, with each node in the tree corresponding to an image component (e.g., region, object). This is achieved by (1) interactive image segmentation for the regions, (2) hierarchical organization of the nodes, and (3) manual annotation of the nodes’ category labels. These three tasks are all performed on Sisley’s software interface. With image parsing, Sisley “understands” the image before painting it.

Customization and Rendering Users can specify desired levels of perceptual ambiguity (or *abstract levels* as named in Sisley) for the result painting and/or its individual components. According to these levels, Sisley automatically changes the image appearance, and synthesizes an abstract painting image. With different image components having different abstract levels, Sisley simulates the effect of paths of perception, to increase the subtlety of the painting.

Computation and Control Sisley assesses the abstract levels of the scene and objects for the result painting image, by referencing a large dataset of human annotated images [Yao et al. 2007], propagating information in the context [Yedidia et al. 2001], and computing the Shannon entropies. If the difference between the computed abstract levels and the user’s desired values is too large, Sisley will adjust the parameters and repaint an image. This process iterates in a servomechanism until the difference is within a threshold, or an allowed maximum number of iterations is reached. This enforces appropriate abstract levels for the painting to be enjoyable.

Sections 2 through 4 will explain the above steps in detail. To emphasize the importance of the third step for the system, it is worth pointing out that abstract levels are quite subjective measures in the sense that they vary a lot among people who have or haven’t seen a painting’s corresponding photograph before. In contrast, Sisley does the numerical computation in a relatively objective way, to match the perspectives of the majority of people instead of only the software user whose sense has been greatly affected by the knowledge of the photograph.

The main contribution of this paper is the execution of a computational idea towards the explanation and simulation of abstract art. More specifically, we find a quantitative way to define and compute the abstract level as the degree of perceptual ambiguity in scene/object recognition, using which we manage to synthesize abstract art images at abstract levels desired by the user. This paper aims to introduce this novel numerical measure and the corresponding computational process, rather than merely a rendering system focusing on good painterly effects.

2 Parse Tree and Abstract Levels

We adopt a hierarchical image representation named *parse tree* introduced by Tu et al. [2005]. A parse tree has a root node corresponding to the scene, and a few other nodes corresponding to the objects in the image. As shown in Fig.3, the photograph in Fig.1 is labeled as a seascape scene, which is then decomposed into five objects/regions: sailboat, sea, buildings, trees, and sky. The sailboat node is further decomposed into sail and hull. In general, a parse tree is a directed acyclic graph (DAG) $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, whose vertices \mathcal{V} represent the nodes and edges \mathcal{E} represent the parent→child links

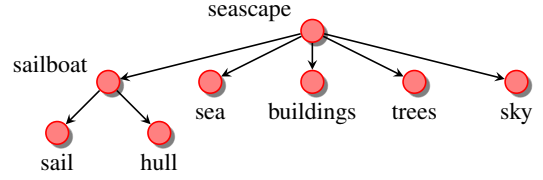


Figure 3: An example parse tree of the photograph in Fig.1.

in the tree, and each node $i \in \mathcal{V}$ is associated with its category label ℓ_i (see Table 1 for the categories covered by Sisley) and visual features \mathcal{A}_i (e.g., color and shape statistics).

Suppose a parse tree of image \mathbf{I} has K nodes, for interpreting \mathbf{I} in the sense of recognizing the scene and objects, we are interested in finding the most probable combination of the nodes’ category labels

$$\mathcal{L} = (\ell_1, \ell_2, \dots, \ell_K) \quad (1)$$

for the image, namely, the combination that maximizes the conditional probability $p(\mathcal{L}|\mathbf{I})$. However, this most probable interpretation only captures the major mode/peak of $p(\mathcal{L}|\mathbf{I})$, but fails to describe the uncertainty or ambiguity associated with multimodal probabilities. To address this problem, in an information-theoretic way, we adopt the Shannon entropy

$$\mathcal{H}(\mathcal{L})|\mathbf{I} = \sum_{\mathcal{L}} -p(\mathcal{L}|\mathbf{I}) \log p(\mathcal{L}|\mathbf{I}) \quad (2)$$

to measure the abstract level of image \mathbf{I} . For the abstract paintings we study in this paper, we expect $\mathcal{H}(\mathcal{L})|\mathbf{I}$ to be significantly greater than those of photographs, with $p(\mathcal{L}|\mathbf{I})$ usually having more than one local maxima, corresponding to multiple competing understandings [Yevin 2006].

In order to obtain the parse tree for abstract level assessment, Sisley provides three functions on its graphical user interface.

Image Segmentation We adopt the banded version [Lombaert et al. 2005] of the graph-cut algorithm [Boykov and Kolmogorov 2004] for interactively segmenting the image into two parts using foreground and background scribbles. Then each part is further binarily segmented using the same method. This continues until the user considers that every object is separated from its neighboring regions.

Hierarchical Organization The previous step generates a binary tree, in which some middle-level nodes might not correspond to meaningful individual objects (e.g., a node containing object A and part of object B), and some objects might be mistakenly separated and placed in two or more branches. In order to obtain the desired multiway tree as shown in Fig.3, Sisley provides an interactive tool for tree editing (e.g., node deleting, merging).

Category Labeling Users can label the category of each scene/object node in the tree resulted from the previous step. This step is optional, but proper labeling can much improve the accuracy of the assessment of abstract levels (we will explain this in Section 4.1). For any category not directly covered by Table 1, users may keep the node unlabeled, or simply choose a similar item, for example, choose “cloth” for “sail”, and “boat” for “hull”, for the image in Fig.1, by assuming they are equivalent.

3 Customization and Rendering

Before rendering the painting, Sisley lets the user slide a bar to indicate the desired abstract level of the result image. It also allows

Table 1: Scene and object categories adopted in Sisley, which distribute widely over those usually appearing in paintings.

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

the user to specify different abstract levels for different nodes in the parse tree, to simulate the phenomenon of paths of perception. For example, if the user expects that most people should recognize node i earlier than node j , he/she can assign a much lower desired abstract level to i than that of j . In fact, the specified abstract level of each node is treated as an extra factor to be multiplied to the global abstract level during rendering.

The main task of Sisley’s rendering engine is to transfer the input image’s visual appearance from photograph style to abstract painting styles. Since colors, shapes and textures are the key features of an image that affect its perception [Marr 1982], Sisley tries to operate on the statistics of these three aspects.

Color Sisley transfers the image (or image region) into the HSV color space, then adds a random shift to the hue channel, which follows a truncated normal distribution $\mathcal{N}(0, \sigma^2, a = -3\sigma, b = 3\sigma)$. The standard deviation σ is proportional to the expected abstract level, with $\sigma_{\max} = 15^\circ$. This lets hue move in a 90° interval. Meanwhile, a positive shift also proportional to the expected abstract level is added to the saturation channel to make the image more bright-colored.

Shape Sisley captures the boundary pixels of an image region, and shifts each of them by a 2D truncated normal offset, whose standard deviation and truncation for each dimension are also proportional to the desired abstract level, with a factor corresponding to the size of the image region. For boundary points shared by two regions with different desired abstract levels, Sisley takes the value of the region closer to the viewer (i.e., marked by foreground scribbles in segmentation).

Texture Sisley applies painterly rendering to simulate a textured painting appearance. We adopt the painterly rendering algorithm introduced by Zeng et al. [2009] with adaptations for fast processing. For example, we use a smoothed *proposal map* from the first phase of *primal sketch* [Guo et al. 2007] as the *orientation field* which determines the directions of strokes, thus avoid the relatively slow *sketch pursuit* and *orientation diffusion* phases (see the two references above for detailed explanations of the italic terms). This field also contains the magnitude (i.e., salience) of each pixel, which determines the corresponding stroke size. While laying out the brush strokes, we run inhomogeneous Poisson disk sampling [Deussen et al. 2000; Bridson 2007] instead of the original greedy algorithm, to determine the positions and sizes of strokes, and the radius of each disk is inversely proportional to its central pixel’s magnitude. This method can cover the canvas with near-minimal overlap among the strokes.

During the rendering process, Sisley also perturbs the color and geometry of each brush stroke according to desired abstract levels. This adds local randomness to the painting in addition to the global randomness obtained above. All these stochastic operations can usually increase the perceptual ambiguity of the image. This is because in the very sparse natural image space [Ruderman 1994],

the photograph is among the very few meaningful images located in areas with local minima of $\mathcal{H}(\mathcal{L})|\mathbf{I}$, and the above operations will move the painting image slightly away from the photograph, thus increase the abstract level.

The entire rendering scheme can be viewed as a hierarchical data generating process [Gelman et al. 2004]. The rendering parameters are generated according to the desired abstract levels (i.e., hyper-parameters), and they further generate the painting image in the next level, with the original photograph as a constant condition.

4 Computation and Control

Once Sisley renders a painting from the input photograph, it is necessary to assess the actual abstract level of this output and compare it with the desired level, in order to ensure that the expected results are obtained. Otherwise, Sisley needs to repaint the image with adjusted parameters according to the feedback from the output.

Since visual perception involves direct object recognition from visual features and indirect recognition using contextual information, in order to compute $p(\mathcal{L}|\mathbf{I})$, we treat the parse tree as a Markov random field (MRF) composed of pair cliques, which covers both of these two aspects. In this way, the probability of labels can be factorized as

$$p(\mathcal{L}|\mathbf{I}) = \frac{1}{Z} \prod_{i \in \mathcal{V}} \phi_i(\ell_i) \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(\ell_i, \ell_j) \quad (3)$$

in which $\phi_i(\ell_i) = p(\ell_i|\mathbf{I}_i)$ measures the local evidence of node i , corresponding to direct recognition of the image region \mathbf{I}_i , and $\psi_{ij}(\ell_i, \ell_j)$ measures the compatibility between two neighboring nodes i and j in the parse tree, affecting the propagation of contextual information in indirect recognition. In Sisley, $p(\ell_i|\mathbf{I}_i)$ is computed using a non-parametric method, and $\psi_{ij}(\ell_i, \ell_j)$ is approximated by counting the joint frequencies $\tilde{f}(\ell_i, \ell_j)$ in the LHI image dataset [Yao et al. 2007], which includes over 10,000 natural/artificial scene images with human annotated parse trees.

4.1 Local Evidence

Sisley computes the local evidence $p(\ell_i|\mathbf{I}_i)$ using non-parametric (kernel) density estimation (a.k.a. probabilistic voting) [Duda et al. 2000; Torralba et al. 2008]. We use images from the LHI image dataset as voters. For fast voting, current version of Sisley includes a subset of 101 scene voters and 470 object voters.

For each query node i corresponding to image region \mathbf{I}_i , $p(\ell_i|\mathbf{I}_i)$ is computed by accumulating the weighted votes:

$$p(\ell_i|\mathbf{I}_i) \propto \sum_n \exp\{-\lambda D(\mathbf{I}_i, \mathbf{J}_n)\} \mathbb{1}(\ell_i = \ell_n) \quad (4)$$

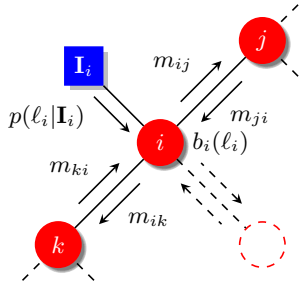


Figure 4: In belief propagation, the parse tree is treated as a MRF composed of pair cliques, in which each node's belief is computed with its local evidence and messages from its parent and children.

where $\{J_n\}$ are the voters' images, $\mathbb{1}$ is the indicator function, and λ is a rate parameter controlling the overall entropy level. The distance function D measures the difference between two images or image regions. As suggested by van de Sande et al. [2010], we adopt the Opponent-SIFT descriptor [Lowe 2004; van de Sande et al. 2010] of the images/regions which covers color, shape, and texture features, and compute D as the squared Euclidean distance between the two Opponent-SIFT feature vectors.

As we mentioned in Section 2, if the user has correctly labeled the nodes in the parse tree, this computation can be made more accurate. This is achieved by including the original node (from the parse tree of the input photograph) as an additional voter, which is reliable thus heavily weighted. In most cases, if the output painting image is not very different from the original photograph (i.e., within or near the photograph's neighborhood area in the image space as mentioned in the last paragraph of Section 3), the abstract level of the node should not be too high, thus adding this powerful voter will improve the accuracy of estimation by pushing up a strong mode in the probability distribution $p(\ell_i|\mathbf{I}_i)$.

4.2 Belief Propagation

With the local evidences $p(\ell_i|\mathbf{I}_i)$, probabilities $p(\ell_i|\mathbf{I})$ (i.e., beliefs with contextual knowledge) are computed using belief propagation [Yedidia et al. 2001] over the parse tree, as shown in Fig.4. Using uniform initialization, Sisley visits the nodes in sequential order to update their messages and beliefs, and iterates the process to continue the propagation. Each time when node i is visited,

- (1) Its outgoing message to neighbor node j is updated using

$$m_{ij}(\ell_j) = \sum_{\ell_i} p(\ell_i|\mathbf{I}_i) \tilde{f}(\ell_i, \ell_j) \prod_{k \in \partial i \setminus j} m_{ki}(\ell_i) \quad (5)$$

where $\partial i \setminus j$ denotes the neighborhood of node i excluding node j , and

- (2) Its belief is updated using

$$b_i(\ell_i) \propto p(\ell_i|\mathbf{I}_i) \prod_{j \in \partial i} m_{ji}(\ell_i) \quad (6)$$

where local evidence and incoming messages are combined.

Sometimes it is necessary to swap ℓ_i and ℓ_j in $\tilde{f}(\ell_i, \ell_j)$ since we must make sure the first label corresponds to the parent node. We finally set $p(\ell_i|\mathbf{I}) = b_i(\ell_i)$ after convergence, which is guaranteed for our tree structure [Yedidia et al. 2001].

4.3 Assessment and Control of Abstract Level

Even if all local evidences and pairwise terms are available, it is still impractical to compute $p(\mathcal{L}|\mathbf{I})$ since the space of \mathcal{L} is usually too large as the amount of nodes or categories grows. Instead, Sisley looks at the marginal probabilities $p(\ell_i|\mathbf{I})$, and gives an approximate estimate of $p(\mathcal{L}|\mathbf{I})$ as the weighted average of their entropies

$$\mathcal{H}(\mathcal{L})|\mathbf{I} \sim \sum_i w_i \mathcal{H}(\ell_i)|\mathbf{I} = \sum_i w_i \sum_{\ell_i} -p(\ell_i|\mathbf{I}) \log p(\ell_i|\mathbf{I}) \quad (7)$$

where the normalized weight w_i of node i is proportional to its lattice size on image. This approximation is reasonable for our case because the correlations between ℓ_i and its neighboring nodes' labels are already greatly decreased by the propagation.

Based on this approximation, Sisley further computes the *relative abstract level* of image \mathbf{I} as

$$\tilde{\mathcal{H}} = \frac{\sum_i w_i \mathcal{H}(\ell_i)|\mathbf{I}}{\sum_i w_i \log |\Omega_{\ell_i}|} \in [0, 1] \quad (8)$$

which is the ratio of the approximated $\mathcal{H}(\mathcal{L})|\mathbf{I}$ over its upper bound (here $|\Omega_{\ell_i}|$ denotes the volume of ℓ_i 's space Ω_{ℓ_i}). This relative number is actually the one to be compared with user's desired abstract level. If the computed (output) level $\tilde{\mathcal{H}}_0$ is close to the user's desired (input) level $\tilde{\mathcal{H}}_1$ (e.g., within $\pm 10\%$), then Sisley achieves success and finishes the job. Otherwise, it is necessary to adjust the parameters and repaint an image from the photograph. For the adjustment, after the t -th iteration, Sisley compares $\tilde{\mathcal{H}}_1^{(t)}$ and $\tilde{\mathcal{H}}_0^{(t)}$ for assigning $\tilde{\mathcal{H}}_1^{(t+1)}$ to the next iteration according to

$$\tilde{\mathcal{H}}_1^{(t+1)} = \begin{cases} \frac{[\tilde{\mathcal{H}}_1^{(t)}]^2}{\tilde{\mathcal{H}}_0^{(t)}}, & \text{if } \tilde{\mathcal{H}}_0^{(t)} > \tilde{\mathcal{H}}_1^{(t)}, \\ 1 - \frac{[1 - \tilde{\mathcal{H}}_1^{(t)}]^2}{1 - \tilde{\mathcal{H}}_0^{(t)}}, & \text{if } \tilde{\mathcal{H}}_0^{(t)} < \tilde{\mathcal{H}}_1^{(t)}. \end{cases} \quad (9)$$

with the desired $\tilde{\mathcal{H}}$ as the initial input. The rendering-computation-adjustment loop ends when the difference between $\tilde{\mathcal{H}}_1$ and $\tilde{\mathcal{H}}_0$ drops below the predefined threshold, or an allowed maximum number of iterations is reached. This idea of servomechanism is similar in concept to the bisection method in root-finding, and gradient descent in optimization, but convergence is not guaranteed for our case due to the existence of randomness (e.g., random color shifts), especially if the nodes have not been properly labeled.

5 Experimental Results

Fig.1 demonstrates three abstract paintings with $\tilde{\mathcal{H}}$ at approximately 0.25, 0.5 and 0.75 respectively. In these images, both the sailboat and the background become harder to recognize as $\tilde{\mathcal{H}}$ increases. While different viewers may feel different abstract levels due to various personal backgrounds (e.g., some may have seen the photograph before), Sisley gives a more objective estimate of the abstract level. In this way, the user can create abstract painting images at proper levels of perceptual ambiguity for others to enjoy, by diminishing the affection of his/her knowledge of the photograph and subjective feelings during the interactive process.

Fig.5 includes more results generated by Sisley. These images were all generated at medium abstract levels ($\tilde{\mathcal{H}} \approx 0.5$), but the results may seem somewhat different because of the different complexities of their parse trees (thus also different absolute abstract levels in spite of similar relative abstract levels). For example, with

Abstract Paintings



Photographs



Figure 5: More results produced by Sisley. Photographs (bottom row from left to right): “Palazzo Vecchio” (by thephotoholic), “Neptune Fountain” (by thephotoholic), “Old Wrecks At West Mersea” (by Tom Curtis), “Sunday Walk” (by Simon Howden), “Old Man And Gull” (by Federico Stevanin), “Ullswater” (by Susie B), “Promenade Morecambe” (by Tom Curtis), and “Conwy Bay” (by Matt Banks). Photographs courtesy of FreeDigitalPhotos.net.

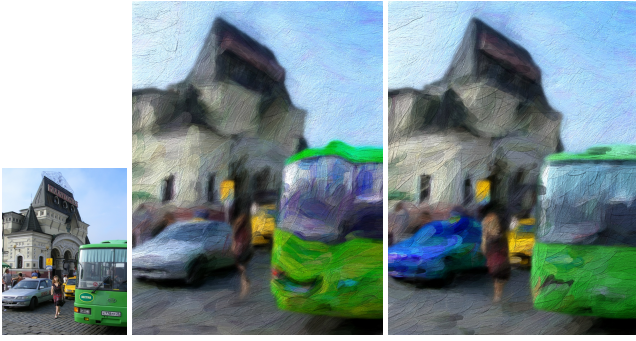


Figure 6: “Vladivostok Transport” (photograph courtesy of Matt Banks / FreeDigitalPhotos.net). The two paintings have similar abstract levels according to computation, but their ambiguities distribute over the objects in different ways, which may lead to different paths of perception for the audience.

more objects and greater depth of field in the scene, the last painting “Conwy Bay” often appears more ambiguous than the others.

Fig.6 displays two paintings with similar global abstract levels according to Sisley’s computation, but their abstract patterns differ in the way that the ambiguities distribute over the image components differently, which may lead to different paths of perception. In the first painting, the car in the front is more figurative than the bus, and the second painting was created in an opposite way.

6 Evaluations

In addition to the above computation which might not be intuitive enough, to further verify that Sisley really achieves satisfactory abstract effects, we performed comparative human experiments over three groups of images:

- (1) Photographs,
- (2) Original abstract paintings by artists, and
- (3) Our synthesized images (at various relative abstract levels between 0.25 and 0.75).

Our studies focused on two potential hints of abstract effects: *recognition accuracy* and *response speed* by human subjects. These two statistics can be objectively measured and they reflect the perceptual ambiguity and mental efforts, respectively, which are of our main interests (see Section 1.1).

We randomly selected 40 images from each of the three groups above. These images cover approximately half of the categories in Table 1. We labeled several objects in each image as *query objects* whose associated recognition accuracies were observed. Fig.7 shows a few example query objects cropped from the images.

The 40×3 images were then displayed in random order on a color monitor to 20 human test subjects (graduate and college students of different majors), with the query objects highlighted. As soon as a subject felt he/she understood an image, he/she hit a key to record the response time, then reported the category labels for the query objects by choosing from Table 1.

Recognition Accuracy As shown in Fig.8, both original paintings by artists and our synthesized images have slightly lower recognition accuracy than that of photographs. It is also noticed that the diagonal elements still dominate for most categories, suggesting that usually test subjects could still recognize objects in abstract



Figure 7: Example query objects cropped from images used in our human experiments described in Section 6.

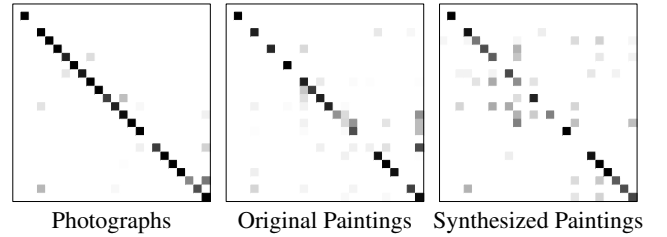


Figure 8: Confusion matrices visualizing the recognition accuracies for the three groups of images studied in Section 6. Horizontal and vertical axes represent the reported and true categories, respectively. The darkness of each grid is proportional to the frequency. Uncovered categories in the experiments are not displayed.

paintings (either original or synthesized) correctly through certain amount of efforts of thinking.

Response Speed We have recorded the data of response time used for recognition during the experiments, and analyzed them using standard statistical hypothesis testing techniques [Montgomery 2000]. Observing that analysis of variance (ANOVA) F -test on the effect of group difference gave an extremely small p -value at 2.955×10^{-8} , we further computed Tukey’s Honest Significant Differences (HSD) for testing the significance of pairwise difference in response speed, and the adjusted p -values for multiple comparisons are shown in Table 2. At level $\alpha = 0.05$, we have not observed significant difference between original and synthesized paintings in response speed, but they both differ significantly from photographs with longer response time.

As observed in the above experiments, our synthesized paintings reproduce both of the two examined statistics of the original paintings, especially, at levels where identifiability is mostly available through certain amount of mental efforts. In contrast, readers may see that some of the individual objects in Fig.7 are indeed very difficult to recognize without the their contexts. This suggests the importance of scene configuration information to visual perception, which was realized and greatly utilized by the impressionists, as we mentioned in Section 1.2.

Table 2: Summary of Tukey’s HSD test on response speed.

Group Pair	$\Delta \bar{t}$ (ms)	p -value
Photographs vs. Original Paintings	2165	< 0.01
Photographs vs. Synthesized Paintings	1183	0.03
Original vs. Synthesized Paintings	−982	0.11

7 Conclusion and Future Work

We have presented a system which augments painterly rendering with perceptual ambiguity computation and control for the simulation of abstract paintings. The system relies on the psychological findings that abstract arts are usually more ambiguous than photographs and representational arts, and accordingly tries to render abstract paintings by increasing the perceptual ambiguities of images with numerical control. Compared with past research [DeCarlo and Santella 2002; Winnemöller et al. 2006; Orzan et al. 2007; Rigau et al. 2008; Mi et al. 2009], Sisley works in the way corresponding to a different level in our biological visual system [Marr 1982], by dealing with scene/object recognition using probability density estimation and belief propagation.

One potential future work for improving the performance of the current system is to generalize the ambiguity measure. In Sisley, we assume the parse tree structure of a painting is apparent to the audience thus unchanged during image manipulation, although the nodes’ categories are unknown. But this is not always true for real abstract artworks. A probability model covering both image structure and component attributes with a feasible computational process is necessary to solve the problem.

Another important aspect to study is how does each stochastic operation (including its associated parameters) on color, shape or texture in the rendering process affect the final result. For example, it is interesting to see whether the changes of color and shape lead to different abstract effects, and whether painterly rendering can generate better effect than simply blurring or adding noise¹. Starting from this paper’s viewpoint on the common characteristic of abstract arts in perceptual ambiguity, the detailed study on the behavior of different stochastic operations actually takes one step further to discover the intrinsic differences among various abstract art styles and techniques, which will contribute to a more comprehensive understanding of the subject. In the meanwhile, however, necessary systematic strategies to choose among and apply different operations properly will greatly increase the complexity of the rendering engine.

¹In the sense of increasing the perceptual ambiguity, image degrading operations such as blurring and adding noise are also capable of producing abstract effects, although usually they are not considered as common artistic techniques. Fig.9 displays a few examples of such *nonstandard* operations.

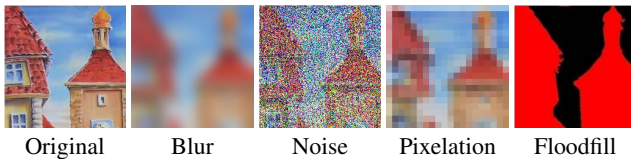


Figure 9: A few nonstandard operations to process images for abstract effects. “Nonstandard” means these operations are less frequently used in artistic depiction than common techniques such as brush strokes, color enhancement, etc.

Project Website

The latest paintings/demos/executables of Sisley are available at <http://www.stat.ucla.edu/~mtzhao/research/sisley/>.

Acknowledgements

We thank Amy Morrow, Brandon Rothrock, Zhangzhang Si, Benjamin Yao, Yibiao Zhao, and the anonymous reviewers for nice suggestions on improving the presentation of this paper. The work at UCLA was supported by an NSF grant IIS-0713652, and the work at LHI was supported by a Chinese National 863 grant 2009AA01Z331 and two National Natural Science Foundation of China (NSFC) grants 60728203 and 60970156.

References

- BERLYNE, D. E. 1971. *Aesthetics and Psychobiology*. Appleton-Century-Crofts, Inc.
- BOYKOV, Y., AND KOLMOGOROV, V. 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 9, 1124–1137.
- BRIDSON, R. 2007. Fast poisson disk sampling in arbitrary dimensions. In *SIGGRAPH 2007 Sketches*, 22.
- COLLOMOSSE, J. P., AND HALL, P. M. 2003. Cubist style rendering of photographs. *IEEE Trans. Vis. Comput. Graph.* 9, 4, 443–453.
- COLLOMOSSE, J. P., AND HALL, P. M. 2006. Video motion analysis for the synthesis of dynamic cues and futurist art. *Graphical Models*, 5, 402–414.
- COOKE, H. L. 1978. *Painting Techniques of the Masters*. Watson Guptill / Pitman Publishing.
- COVER, T. M., AND THOMAS, J. A. 2006. *Elements of Information Theory, 2nd Edition*. Wiley-Interscience.
- DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. In *Proceedings of SIGGRAPH 2002*, 769–776.
- DEUSSEN, O., HILLER, S., VAN OVERVELD, C., AND STROTHOTTE, T. 2000. Floating points: A method for computing stipple drawings. *Comput. Graph. Forum* 19, 3, 40–51.
- DUDA, R. O., HART, P. E., AND STORK, D. G. 2000. *Pattern Classification, 2nd Edition*. Wiley-Interscience.
- DURAND, F. 2002. An invitation to discuss computer depiction. In *Proceedings of NPAR 2002*, 111–124.
- FINKELSTEIN, A., AND RANGE, M. 1998. Image mosaics. In *Proceedings of the 7th International Conference on Electronic Publishing, held jointly with the 4th International Conference on Raster Imaging and Digital Typography (EP/RIDT ’98)*, 11–22.
- FUNCH, B. S. 1997. *The Psychology of Art Appreciation*. Museum Tusculanum Press.
- GAL, R., SORKINE, O., POPA, T., SHEFFER, A., AND COHEN-OR, D. 2007. 3D collage: Expressive non-realistic modeling. In *Proceedings of NPAR 2007*, 7–14.
- GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. 2004. *Bayesian Data Analysis, 2nd Edition*. Chapman & Hall.

- GOOCH, B., AND GOOCH, A. A. 2001. *Non-Photorealistic Rendering*. A K Peters, Ltd.
- GUO, C.-E., ZHU, S.-C., AND WU, Y. N. 2007. Primal sketch: Integrating structure and texture. *Comput. Vis. Image Understand.* 106, 1, 5–19.
- HAEBERLI, P. 1990. Paint by numbers: Abstract image representations. In *Computer Graphics (Proceedings of SIGGRAPH '90)*, 207–214.
- HERTZMANN, A., AND PERLIN, K. 2000. Painterly rendering for video and interaction. In *Proceedings of NPAR 2000*, 7–12.
- HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of SIGGRAPH '98*, 453–460.
- JAYNES, E. T. 1957. Information theory and statistical mechanics. *Phys. Rev.* 106, 4, 620–630.
- KAPLAN, C. S., AND SALESIN, D. H. 2004. Dihedral Escherization. In *Proceedings of Graphics Interface 2004*, 255–262.
- KERSTEN, D. 1987. Predictability and redundancy of natural images. *J. Opt. Soc. Am. A* 4, 12, 2395–2400.
- KLEIN, A. W., SLOAN, P.-P. J., COLBURN, R. A., FINKELSTEIN, A., AND COHEN, M. F. 2001. Video cubism. Tech. Rep. MSR-TR-2001-45, Microsoft Research.
- KLEIN, A. W., GRANT, T., FINKELSTEIN, A., AND COHEN, M. F. 2002. Video mosaics. In *Proceedings of NPAR 2002*, 21–28.
- LEE, S., OLSEN, S. C., AND GOOCH, B. 2006. Interactive 3D fluid jet painting. In *Proceedings of NPAR 2006*, 97–104.
- LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. In *Proceedings of 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 ICCV 2005, Volume 1*, 259–265.
- LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 2, 91–110.
- 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 SIGGRAPH '96*, 477–484.
- MI, X., DECARLO, D., AND STONE, M. 2009. Abstraction of 2D shapes in terms of parts. In *Proceedings of NPAR 2009*, 15–24.
- MONTGOMERY, D. C. 2000. *Design and Analysis of Experiments, 5th Edition*. Wiley.
- 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>.
- MUREIKA, J. R., DYER, C. C., AND CUPCHIK, G. C. 2005. On multifractal structure in non-representational art. *Phys. Rev. E* 72, 046101.
- ORCHARD, J., AND KAPLAN, C. S. 2008. Cut-out image mosaics. In *Proceedings of NPAR 2008*, 79–87.
- ORZAN, A., BOUSSEAU, A., BARLA, P., AND THOLLOT, J. 2007. Structure-preserving manipulation of photographs. In *Proceedings of NPAR 2007*, 103–110.
- RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. In *Proceedings of NPAR 2004*, 85–152.
- RIGAU, J., FEIXAS, M., AND SBERT, M. 2008. Informational aesthetics measures. *IEEE Comput. Graph. and Appl.* 28, 2, 24–34.
- RUDERMAN, D. L. 1994. The statistics of natural images. *Network: Computation in Neural Systems* 5, 4, 517–548.
- SMITH, K., LIU, Y., AND KLEIN, A. 2005. Animosaics. In *Proceedings of SCA 2005*, 201–208.
- STORK, D. G. 2009. Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. In *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns (CAIP '09)*, 9–24.
- STROTHOTTE, T., AND SCHLECHTWEIG, S. 2002. *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*. Morgan Kaufmann.
- TAYLOR, R. P., MICOLICH, A. P., AND JONAS, D. 1999. Fractal analysis of Pollock's drip paintings. *Nature* 399, 422.
- TORRALBA, A., FERGUS, R., AND WEISS, Y. 2008. Small codes and large databases for recognition. In *Proceedings of CVPR 2008*, 1–8.
- TU, Z., CHEN, X., YUILLE, A. L., AND ZHU, S.-C. 2005. Image parsing: Unifying segmentation, detection, and recognition. *Int. J. Comput. Vis.* 63, 2, 113–140.
- VAN DE SANDE, K., GEVERS, T., AND SNOEK, C. 2010. Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, in press.
- VON HELMHOLTZ, H. 1866. *Physiological Optics*. Transl. JP Southhall, 1911. Washington, DC: *Opt. Soc. Am.* (From German).
- WALLRAVEN, C., FLEMING, R., CUNNINGHAM, D., RIGAU, J., FEIXAS, M., AND SBERT, M. 2009. Categorizing art: Comparing humans and computers. *Computers & Graphics* 33, 4, 484–495.
- WINNEMÖLLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-time video abstraction. *ACM Trans. Graph.* 25, 3, 1221–1226.
- XU, J., AND KAPLAN, C. S. 2008. Artistic thresholding. In *Proceedings of NPAR 2008*, 39–47.
- 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* 2006, 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.