



**HAL**  
open science

## Structure-preserving manipulation of photographs

Alexandrina Orzan, Adrien Bousseau, Pascal Barla, Joëlle Thollot

► **To cite this version:**

Alexandrina Orzan, Adrien Bousseau, Pascal Barla, Joëlle Thollot. Structure-preserving manipulation of photographs. International Symposium on Non-Photorealistic Animation and Rendering (NPAR), Aug 2007, San Diego, United States. pp.103-110, 10.1145/1274871.1274888 . inria-00362868

**HAL Id: inria-00362868**

**<https://inria.hal.science/inria-00362868v1>**

Submitted on 29 Apr 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Structure-preserving manipulation of photographs

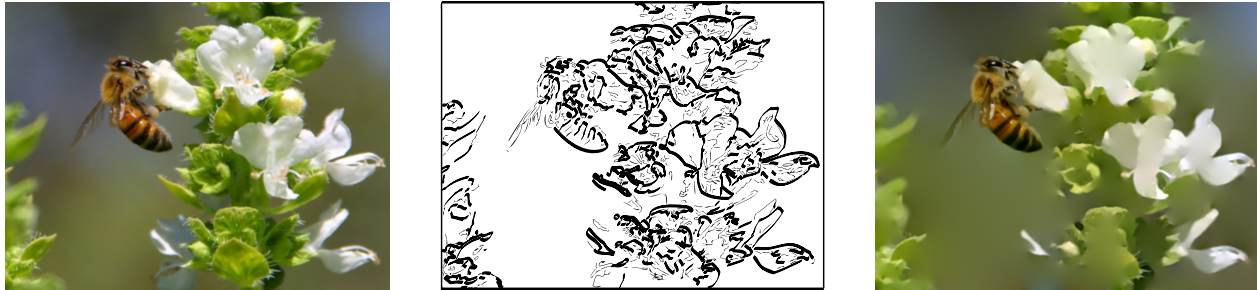
Alexandrina Orzan

Adrien Bousseau

Pascal Barla

Joëlle Thollot

ARTIS - INRIA Grenoble University



**Figure 1:** Our approach takes as input an image (left), and allows a user to manipulate its structure in order to create abstracted or enhanced output images. Here we show a line drawing with line thickness proportional to their structural importance (middle), and a reconstruction of color information that focuses on the bee and removes detail around it (right).

## Abstract

Visual content is often better communicated by simplified or exaggerated images than by the “real world like” images. In this paper, we offer a tool for creating such enhanced representations of photographs in a way consistent with the original image content. To do so, we develop a method to identify the relevant image structures and their importance. Our approach (a) uses edges as the basic structural unit in the image, (b) proposes tools to manipulate this structure in a flexible way, and (c) employs gradient domain image processing techniques to reconstruct the final image from a “cropped” gradient information. This edge-based approach to non-photorealistic image processing is made feasible by two new techniques we introduce: an addition to the Gaussian scale space theory to compute a perceptually meaningful hierarchy of structures, and a contrast estimation method necessary for faithful gradient-based reconstructions. We finally present various applications that manipulate image structure in different ways.

**CR Categories:** I.4.3 [Image Processing and Computer Vision]: Enhancement; I.3.3 [Computer Graphics]: Picture/Image Generation;

**Keywords:** Visual communication, multiscale analysis, image processing, image reconstruction.

## 1 Introduction

Effective visual communication is not always best achieved by the “real-world like” images. Simplified objects or exaggerated features can often improve perception and facilitate comprehension. This observation led researchers to investigate new non-photorealistic image processing techniques, as in the work of De-Carlo and Santella [2002] that explicitly aims at better conveying a

message visually by means of perceptual considerations. In particular, such an image manipulation approach offers a tool for transmitting effective visual information by grabbing visual attention.

However, to modify images in a way consistent with the information content of the image, one must first identify the relevant *structures* and their *importance* and then ensure they are preserved through manipulations. The main goal of this paper is precisely to give insights into “what structure means” when we have **no a priori** about semantics, and to provide the user with image manipulation tools to create enhanced or abstracted representations of photographs in accordance with their structural information.

Taking a look at Figure 2 that represents a hand-made scientific illustration, it is clear that the main subject of the image is the butterfly: it is depicted with many details, while plants around are more or less suggested. However, while abstracted, secondary elements of the image retain their look and are easily identified; in other words, their relevant structural components are preserved through the abstraction process. Similarly, our approach is to provide high-level *structural* information to guide user image manipulations. This is a significant step compared to previous approaches.

Gaussian scale space theory has been developed by the computer vision community to deal with structure identification in images with no a priori information. This theory models the first stages of human vision (front-end vision) and extracts features that are perceptually important. A relevant structure is defined as an element that is *invariant* to various viewing conditions; other elements can be considered “accidental”, and of less importance. Out of many invariants in an image (edges, corners, ridges, curvatures, etc), *edges* contain most of the visually important information, because the human visual system is very sensitive to contrast variations [Palmer 1999]. Moreover, an edge-based representation of images provides a flexible and simple way of merging multi-scale information. We therefore use edges as the structural unit in our image manipulation and derive their importance from the scale space analysis.

Edges together with their importance form a hierarchy of structures that can be easily manipulated by the user to reflect what is semantically important to her. We then build on existing gradient domain manipulation techniques to reconstruct the final image from the modified edge structure.



**Figure 2:** “Le Papillon” (*The Butterfly*), watercolor by Eric Alibert. From “Leman, mon ile”, © 2000 by Editions Slatkin. As seen in the guidebook of scientific illustration [2003].

The major contribution of this paper is to combine these two existing techniques: scale space analysis and gradient domain image editing, to provide new *structure-oriented* image manipulation methods. Moreover, the successful adaptation of these techniques to this new domain has required extending them via two specific technical contributions. For scale space analysis, we propose a new approach to extract image structure that identifies meaningful edges according to their importance in the scale space. For gradient domain image manipulation, we propose a novel reconstruction method from sparse edge fields that carefully reintroduces blur and contrast information.

Regarding our contributions, we would like to emphasize that this work is *not* about a new stylization technique. It is rather a starting point for any subsequent stylization. However, we show applications, that differ in the way they manipulate image structure. We provide three types of manipulation: levels-of-detail, shape simplification, and importance-based line drawing. We hope that our research will motivate the development of other applications that take advantage of image structure.

The paper is organized in six sections. After presenting previous work (Section 2), we give quick overviews of Gaussian scale-space image analysis and gradient domain image manipulation (Section 3). We then present our method (Section 4) and various applications along with results (Section 5) and implementation details (Section 6). We finally discuss limitations, and propose possible extensions as future work (Section 7).

## 2 Previous work

A number of previous techniques focused on creating enhanced or abstracted renderings from arbitrary photographs.

Generally the previous methods manipulate an image globally without using the image structure [Winnemöller et al. 2006], or rely on the user to define what is important [Wang et al. 2004; Kang et al. 2006; Wen et al. 2006]. As a result, the content either cannot be controlled, or its control involves tedious user interactions. We are interested in automatically extracting the relevant structural information to enrich automatic systems or assist the user in her task.

Previous work made use of Gaussian scale space [Hertzmann 1998] or saliency maps [Collomosse and Hall 2005; Collomosse and Hall 2003] in order to guide painterly stylizations. However, saliency maps identify image regions that already grab visual attention in the original image, and using them to guide stylization will only preserve these attention-grabbing regions. In contrast, our goal is to extract a structure that allows the user to *intentionally* manipulate the image, possibly modifying its attention focus (i.e. changing its subject, see Figure 1 - right), and hence conveying a particular message.

DeCarlo and Santella [2002; 2004] were the first to use a meaningful visual structure in photo abstraction. They use color regions as structural units and create their hierarchy of regions from a pyramid of down-sampled versions of the image. But for coarser-level regions the shape simplifies and the borders move slightly. There is therefore no perfect overlap between finer and coarser regions. When mixing different levels of detail in the same image, this becomes problematic because one doesn’t know how to unify information at different scales.

Bangham et al. [2003] extend DeCarlo and Santella’s work by improving the region segmentation. Their region hierarchy is based on a morphological scale-space and has the advantage of preserving region shapes. But since only the region size is considered, and not its contrast, they tend to eliminate visually important cues that have a high contrast but small size.

In general, multi-scale region approaches have the inconvenience of associating a solid color to each region. This fact introduces visible color discontinuities between regions and demands to explicitly treat color mixing when two regions merge together. The end result is a poster-like effect in the final rendering. In contrast, our structural hierarchy is an edge-based one that allows us to avoid the problems generated by region-based methods. In particular, our edges are not required to be closed contours, as opposed to region boundaries, and hence they do not create erroneous color discontinuities.

Edge representation of images has been used in previous work of course, although not with the same purpose. Elder et al. [2001] use the edge domain to ease image editing operations (crop, delete, paste), but have no concept of edge importance. Perez et al. [2003] suggest using gradient information only at edge locations as input for a Poisson solver, in order to obtain a texture flattening effect. We improve on this method with the aim of manipulating an image for abstraction and/or enhancement purposes by (a) giving insights into how image structure can be manipulated, and (b) by providing a new reconstruction method that extends [Pérez et al. 2003].

## 3 Background

In order to manipulate images in a structure-preserving way, our method relies on two image processing tools: Gaussian scale space and gradient domain image manipulation. In the following, we give a quick overview of both tools and we provide the reasons for choosing them for our purpose. The bottom line is that Gaussian scale space will be responsible for extracting the structure of edges, while gradient domain processing will be used for reconstruction.

### 3.1 Gaussian scale space

Scale space methods base their approach on representing the image at multiple scales, ensuring that fine-scale structures are successively suppressed and no new elements are added (the so-called “causality property” [Koenderink 1984]).

The motivation for constructing scale-space representations originates from the basic fact that real-world objects are composed of different structures at different scales of observation. Hence, if no prior information is available about the image content, the state-of-the-art approach for deriving the image structure is to use the successive disappearance of scale features to create a hierarchy of structures [Romeny 2003].

Gaussian scale space is the result of two different research directions: one looking for a scale-space that would fit the axiomatic basis stating that “we know nothing about the image” and the other searching for a model for the front-end human vision [Fischler and Firschein 1987; Wandell 1995; Romeny 2003]. Since our purpose is to define a human-vision-like importance measure for an image content we have no a priori on, this scale-space fits our needs.

A scale-space is a stack of images of increasing scales. The basic Gaussian scale space is thus a stack of images convolved by Gaussian kernels of increasing variance<sup>1</sup>. In the general case, Gaussian derivatives of any order can be used to build the stack, allowing one to create scale-spaces of edges, ridges, corners, laplacians, curvatures, etc.

Edge representations, as discontinuities in image brightness, retain important data about objects in the image (shape, surface orientation, reflectance) [Lindeberg 1998]. We thus settle on studying the image structures represented by a hierarchy of edges in the Gaussian scale space. As edges are defined by gradient information, we only need to convolve the original image with Gaussian derivatives of order 1, one for each image dimension. These Gaussian derivatives  $G_x$  and  $G_y$  are computed as follows:

$$\begin{aligned} G_x(x, y; \sigma) &= g(y) \cdot g'(x) \text{ and} \\ G_y(x, y; \sigma) &= g(x) \cdot g'(y) \end{aligned}$$

with

$$g(i) = \frac{e^{-\frac{i^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \quad \text{and} \quad g'(i) = -\frac{e^{-\frac{i^2}{2\sigma^2}} i}{\sqrt{2\pi}\sigma^3}$$

where the width  $\sigma$  of the kernel corresponds to scale and  $i \in \{x, y\}$ . Given an input image  $I$ , we thus build two different scale spaces: an horizontal gradient  $I_x = I \otimes G_x$  and a vertical gradient  $I_y = I \otimes G_y$ .

### 3.2 Gradient domain image manipulation

Many recent works introduced gradient manipulations as an efficient tool for image processing. The main reason is that gradient represents the image variations independently of the original colors, allowing more flexibility in image manipulations. Handling directly the image variations makes possible applications such as seamless image editing [Pérez et al. 2003] and image fusion [Agarwala et al. 2004; Raskar et al. 2004]. Gradient domain is also an intuitive representation for image contrast [Fattal et al. 2002].

We propose to combine the flexibility of gradient domain manipulations to the high level control provided by the gradient scale space. This allows us to seamlessly mix information from multiple scales.

Working in the gradient domain implies one can reconstruct an image  $I$  from its gradient field  $\mathbf{w} = (\mathbf{w}_x, \mathbf{w}_y)$ . As a manipulated gradient is unlikely to be conservative and integrable, a common approach is to compute an estimation of the image whose gradient field best fits  $\mathbf{w}$  in a least-square minimization sense:

$$\arg \min_I \int_{\Omega} (\nabla I - \mathbf{w})^2 d\Omega$$

<sup>1</sup>For numerical stability, one usually starts with a variance  $\sigma_0 = 1$

This estimation corresponds to the unique solution of the Poisson equation  $\Delta I = \text{div} \mathbf{w}$ , where  $\Delta$  and  $\text{div}$  are the Laplace and divergence operators [Pérez et al. 2003; Fattal et al. 2002].

## 4 Our approach

Figure 3 illustrates our approach: we first apply Gaussian scale-space analysis to the input image  $I$  to get gradient values at multiple scales  $(I_x, I_y)_\sigma$ ; then we manipulate this rich information in a way that preserves the structure of the image, giving rise to a gradient field  $\mathbf{w} = (\mathbf{w}_x, \mathbf{w}_y)$ ; finally, the output image  $O$  is built from  $\mathbf{w}$  using Poisson reconstruction. When using a color image as input, we use only its luminance during the multi-scale analysis.

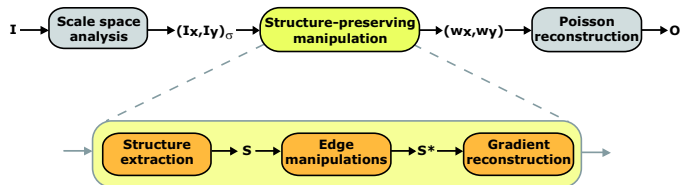


Figure 3: Overview of our method.

Our structure-preserving manipulation represents the heart of the approach and is composed of three steps:

1. **Structure extraction:** starting from the raw multi-scale gradient values, we extract the image structure  $S$  corresponding to the edges, their importance and profile.
2. **Edge manipulations:** We then use the structure  $S$  as a high-level control for user-defined image manipulations, and output a manipulated structure  $S^*$ .
3. **Gradient reconstruction:** We finally reconstruct a gradient field from the set of manipulated edges with their profile.

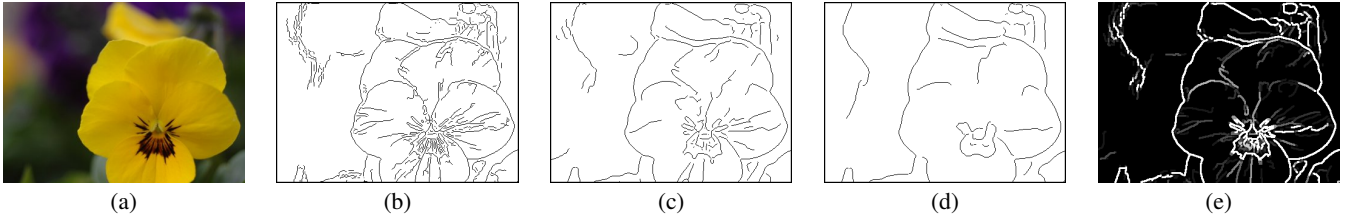
In the following section we mainly present the two technical steps of the method: structure extraction and gradient reconstruction. Several edge manipulation techniques are presented in Section 5.

### 4.1 Structure extraction

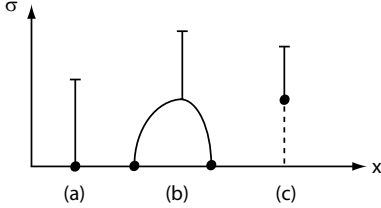
#### 4.1.1 Edge extraction

From the first-order Gaussian derivative scale spaces, we want to build a hierarchy of edges holding structural importance. Before defining what we mean by “structural importance”, we first extract edges at all the available scales in order to get the richest possible information. For this task we use a Canny edge detector [Canny 1986]: it is a state-of-the-art edge detection method that processes the Gaussian derivative information at each scale to give thin, binary edges. Its main quality resides in using hysteresis thresholding that results in long connected paths and avoids small noisy edges (see Figure 4).

After applying the Canny detector, we are left with a multi-scale binary mask  $C_\sigma$  that indicates at each scale the edges locations. Figure 5 illustrates such a typical edge scale-space for a simple 1D example. Due to the nature of Gaussian scale-space, three different cases can occur: (a) an edge exists and suddenly drops off at a higher scale; (b) two edges are coming toward each other and collapse at a higher scale; (c) some “blurry” edges only appear at a higher scale. To simplify further computations, we “drag” edges corresponding to case (c) down to the minimum scale and note  $C_\sigma^*$  the resulting multiscale edge mask.



**Figure 4: Edge importance.** (a) The input image. (b-d) Canny edges at increasing scales. (e) The lifetime measure reflects the importance of edges: “older” edges correspond to more stable and important structures.



**Figure 5: Three different events in a 1D Gaussian scale-space:** (a) an edge drops off at a high scale; (b) two edges collapse; (c) a blurry edge is created. In the last case, we drag the edge down to the finest scale for convenience.

#### 4.1.2 Edge importance

As shown in Figure 5, there is a great deal of coherence along the scale dimension in the multi-scale edge representation. The main idea behind scale-space techniques is to try to extract this coherent *deep structure*, by linking edges at different scales. In particular, because of the causality property of Gaussian scale-space, an edge that disappears at a given scale will not reappear at a higher scale; hence an important measure of structure along scale is *lifetime*, as edges that live longer will correspond to more stable structures.

Unfortunately, extracting an edge lifetime is not trivial, since edges move in Gaussian scale-space (this corresponds to Figure 5 case (b)). This motivated edge focusing techniques, that track edges at increasing scales. In this paper, we take an alternative approach which revealed simpler to implement: instead of considering each pixel  $p$  belonging to an edge, we consider its projected point  $\mathcal{P}_\sigma(p)$  onto the closest edge at scale  $\sigma$  (we use a distance field for this purpose). We can then define the membership of any pixel  $m_\sigma(p)$  as the binary function that indicates whether  $p$  can be considered to belong to an edge at scale  $\sigma$ :

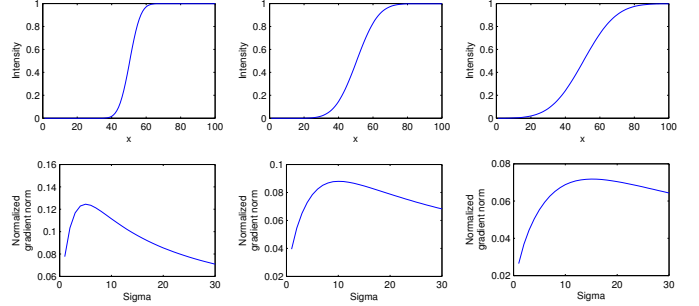
$$m_\sigma(p) = \begin{cases} 1 & \text{if } \|\mathcal{P}_\sigma(p) - p\| < T_\sigma \\ 0 & \text{otherwise} \end{cases}$$

The choice of the threshold distance  $T_\sigma$  is essential to get a good approximation for our membership function. Bergholm [1987] proved that the edge shifting is less than a pixel when the scale  $\sigma$  varies by less than 0.5. Therefore, we increase our  $\sigma$  values by  $\Delta\sigma = 0.4$  at each scale and use  $T_\sigma = \sigma/\Delta\sigma$ . This approach is similar in spirit to the morphological linking method of Papari et al. [2007].

Finally, using membership for linking purpose, we compute the lifetime  $L(p)$  at each edge pixel  $p$  in the finest scale by summing up membership values. Considering the successive scale values  $\sigma_i, i \in 1..N$ , where  $N$  is the size of our scale-space stack, we write lifetime as:

$$L(p) = \arg \min_i \{\sigma_i | m_{\sigma_i}(p) = 0\}$$

This can be seen as a simpler, easier-to-manipulate version of Lindeberg edge strength measure [Lindeberg 1998]. We can now use



**Figure 6: Best scale estimation.** Top: 1D edges blurred with  $\{\sigma_i\} = \{5, 10, 15\}$ . Bottom: normalized gradient magnitude scale space proposed by Lindeberg. The best-scale measures (the local maxima) are at the  $\{\sigma_i\}$  used for blurring, hence they represent well each edge profile.

lifetime as a measure of structural importance to manipulate edges in a structure-preserving way, as shown in Section 5.

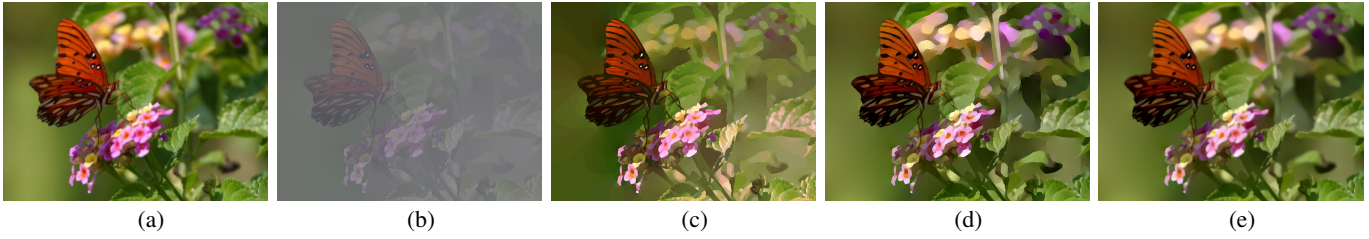
#### 4.1.3 Edge profile

In the previous section, we mainly relied on edge locations and their persistence along scale. Another concern is to deal with their *profile* (contrast value and degree of blur). In this paper, as in previous work [Lindeberg 1998; Elder and Goldberg 2001], we rely on a simple assumption: the profile of an edge gradient is modeled as the convolution of a Dirac (its location and contrast) with a spatially varying Gaussian kernel (its blur). For instance, in a photograph with depth-of-field, out-of-focus edges are blurry (with a wide profile) while in-focus edges are sharp (with a thin profile).

Our second measure of structure then consists, for each edge, in finding the *best scale* that locally corresponds to its blur. This is also the scale where we measure the contrast.

The best scale search is another form of *deep structure* that has been studied by Lindeberg [1998]. Following his approach, we first compute a normalized gradient magnitude scale-space by  $\|\nabla I\| = \sqrt{\sigma(I_x^2 + I_y^2)}$ . The best scale  $B(p)$  at an edge pixel  $p$  is then identified as the one which gives the first local maxima along the scale axis in this normalized gradient magnitude stack. But as with lifetime computation, we need to link “moving edges” at different scales using the projection operator  $\mathcal{P}_\sigma$  again:  $\|\nabla I(p)\| = \|\nabla I(\mathcal{P}_{\sigma_i}(p))\|$ . Figure 6 shows how best scales can be well estimated for edges of increasing blur.

We are now able to “re-blur” the edges using the best scale. Moreover, we will also make use of this measure to find a correct contrast in order to get edge profiles back into the output image.



**Figure 7: Gradient reconstruction.** (a) Input image. (b) Reconstructed image using only the original gradient values at edge positions. (c) Reconstructed image with histogram equalization. Note the quantization artefacts. (d) Reconstructed image using contrast correction. Note that blurry edges become sharp if the profile is not taken into account. (e) Full reconstruction using contrast correction and re-blurring.

## 4.2 Edge manipulations

The multi-scale Canny edges, together with their lifetime and best scale finally constitute the structure  $S = \{C_\sigma^*, L, B\}$  we extracted from the input image. This structure can be manipulated in various ways depending on the application (see Section 5). The main idea is to select a subset  $E$  of the multi-scale Canny edges  $C_\sigma^*$  according to lifetime  $L$ . After manipulation, we are thus left with a new, simpler structure  $S^* = \{E, B\}$ .

## 4.3 Gradient reconstruction

We wish to reconstruct the corresponding image by solving a Poisson equation, i.e. we want to build a vector field  $\mathbf{w}$  that corresponds to our new edges.

We propose to use the scale space information to estimate the original gradient profiles and correctly reproduce the contrast and blur of the input image. However, taking the original gradient values at edge locations as suggested by Perez et al. [2003] results in a gradient field that does not capture the whole original contrast, nor the original blur (Figure 7, (a) and (b)). This is because we only consider the central value of the profile, loosing all its surrounding informations.

A simple solution to the contrast problem would be to apply a histogram equalization on the reconstructed image to match the original contrast. However the very low dynamic range of the reconstructed image leads to strong quantization artifacts (Figure 7(c)).

We thus need to take into account our knowledge of edge profiles to compute the correct contrast. Our model of an edge represents blurry edges that appear in the input image  $I$  as the convolution of a step function  $H$  by a 2D Gaussian kernel  $G_B$ , where  $B$  is the local best scale. When we measure  $I_x$  (resp.  $I_y$ ) at scale  $B$  on edge locations, we get the following contrast values:

$$I_x = H \otimes G_B \otimes \frac{\partial G_B}{\partial x} = H \otimes \frac{\partial G_{B_2}}{\partial x} = \frac{\partial H}{\partial x} \otimes G_{B_2}$$

with  $B_2 = \sqrt{2}B$ . However, to recover the original contrast value of the profile, we are precisely interested in the value of  $\frac{\partial H}{\partial x}$ . This corresponds to the deconvolution of  $I_x$  (resp.  $I_y$ ) by  $G_{B_2}$ . Unfortunately, deconvolution is known as an ill-posed problem, particularly sensitive to noise and quantization [Romeny 2003]. To avoid this problem, we propose to simplify our model for the sake of contrast correction: we replace the 2D Gaussian derivative by a 1D Gaussian derivative  $\tilde{G}_x = g'(x)$ . This way, we can derive an analytical solution for the correction problem (see appendix): for each edge pixel  $p$ , we only need to multiply the gradient value found in  $I_x$  (resp.  $I_y$ ) by  $2B(p)\sqrt{\pi}$ . This correction gives a final contrast close to the original one, and we find that our approximation works well in practice, with no visible artefacts (see Figure 7(d)).



**Figure 8: Detail removal:** (a) original image, and (b-d) several levels-of-detail automatically generated by our method.

Finally, even if using edge locations and correcting their contrast does give a convincing result, blurry edges become sharp in the reconstructed image. Therefore, we also re-blur the edges, as seen in Figure 7(e). This process remains optional as the sharp result provides an interesting cartoon style.

## 5 Applications

Most of the image manipulations presented in this paper can be seen as variations of recently proposed methods that take advantage of the flexibility of the gradient domain. Our contribution is to use the high-level structural information provided by our approach to guide these gradient manipulations.

### 5.1 Detail removal

We use the lifetime information as a threshold value to seamlessly remove details while keeping important structures. Such image editing operations are similar to the seamless cut and paste operations proposed by Perez et al. [2003] and Elder et al. [2001], except that we provide a high level control to the user, who has only to select the desired level of detail (Figure 8).

### 5.2 Multi-scale shape abstraction

We propose a shape abstraction method that adapts the level of abstraction to the scale of the features in order to preserve the informative content of the picture. In practice, we select for each edge its last available version in the scale space using lifetime. As



**Figure 9:** Shape abstraction: (a) original image, and (b) our shape abstraction result. Notice how the thin details are kept, while shapes of bigger objects are abstracted (e.g. the poles).

shapes become more and more smoothed along scales due to the Gaussian filter, relevant structures will have increasingly rounded shapes while details will keep their original silhouettes.

In opposition to previous approaches [DeCarlo and Santella 2002] that remove texture details and abstract shapes at the same time, our approach selects for each edge (including edges belonging to texture details or other small elements) the shape of its last scale. Hence, our approach still keeps most of the meaningful structural information, while simplifying its shape, as seen in Figure 9.

This application can be seen as a fusion of multi-scale images, similar in spirit to other image fusion methods like the ones of Agarwala et al. [2004] and Raskar et al. [2004].

### 5.3 Line drawing

The edge lifetime information offers a powerful high-level parameter for any line drawing algorithm. Figure 10 presents the rendering of vectorized edges with a different width to enhance important structures from details. Figure 1 (middle) also shows an example of this application.

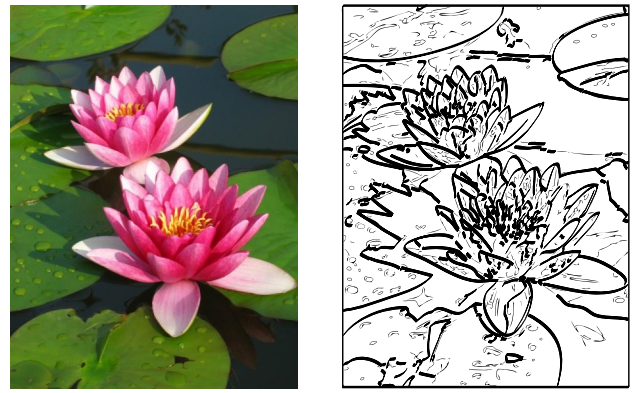
### 5.4 Local control

In order to offer a local control to the user, each image manipulation can be weighted by a gray-level map indicating the desired amount of abstraction (Figure 12). This mechanism is essential to be able to focus on a given zone of the input image, and efficiently grabs visual attention. We take advantage of the Poisson reconstruction to obtain seamless transitions between regions of different weight.

## 6 Implementation

In our approach, we clearly did not focus on performance, but rather on how to extract and use image structure: our current implementation<sup>2</sup> is in Matlab, with performance times of approximately 10 minutes for the whole process of our approach, considering an  $800 \times 600$  input image and a scale-space depth of  $N = 30$ . However, most of this time is spent in the structure extraction, and the Poisson reconstruction takes only about 2 seconds; once structure has been computed, it can be manipulated rather efficiently.

To solve the Poisson equation on the manipulated gradient field, we use the sine transform based Poisson solver of Simchony et al.



**Figure 10:** Vectorized edges, with a larger width for relevant structures (i.e. those having greater lifetime).

[1990] with Dirichlet conditions. We use the Matlab implementation provided by Agrawal et al.<sup>3</sup>

## 7 Discussion and future work

### 7.1 Discussion

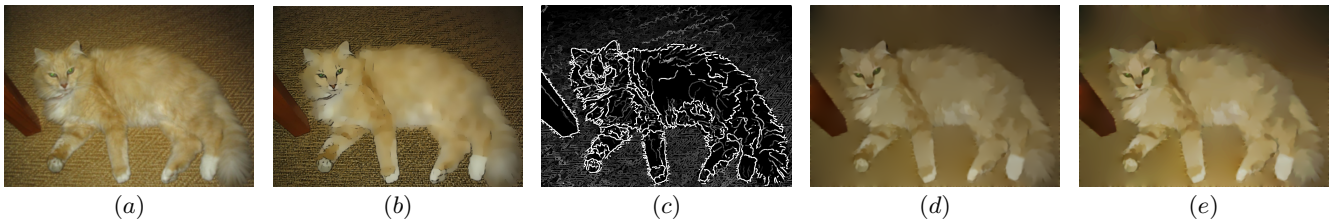
In our exploration of deep structure, we mainly took inspiration from Lindeberg [1998]; indeed, he has a notion similar to lifetime, and the best scale measure is directly borrowed from his approach. One alternative for measuring the best scale is the method introduced by Elder et al. [1998; 2001], based on local signal-to-noise ratios. But this approach is not easy to combine with our importance measure, making Lindeberg’s method better suited to our purposes. However, there is a main difference between Lindeberg’s work and ours: we separate the importance of edges from their contrast and profile, while he deals with all this information at once. Our approach has the advantage of being easier to manipulate: one can modify any property without affecting the others.

This is well illustrated in Figure 11: here we show a failure case of Winnemöller et al.’s abstraction approach [2006]. Although their method gives convincing results in many cases, this specific example shows how they cannot get rid of high-contrast texture lines without abstracting the cat too far; in contrast, our approach allows us to simply remove detail edges regardless of their contrast. Another interesting example comes up when we compare our results to DeCarlo et al. [2002] as in Figure 13: while their method couple simplification of shape with detail suppression, ours allows to remove details *without* necessarily simplifying shapes.

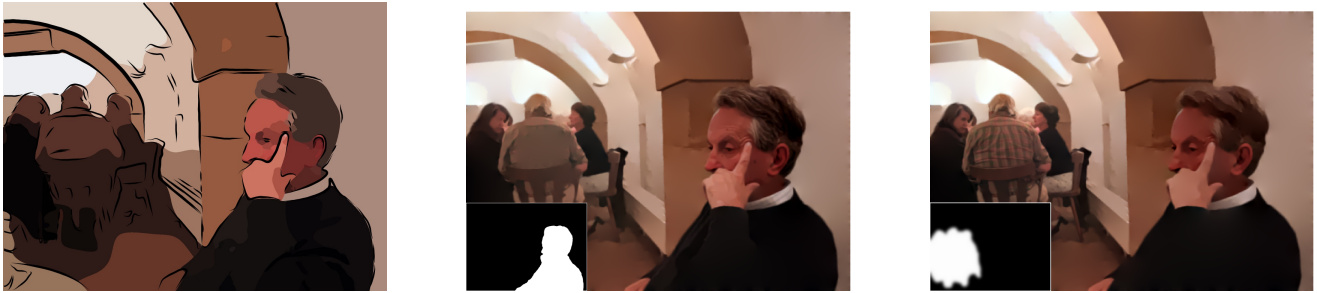
Another choice we made is to use Poisson reconstruction methods. This body of techniques have an advantage over other diffusion approaches: it is independent from the input image. For instance, while a diffusion method will try to blur an unwanted detail, a Poisson approach will simply ignore it in the reconstruction. This is again well illustrated by the example in Figure 11, since the texture lines simply do not appear in our image. Another advantage is that it gives smooth results: when compared to the stylized image of DeCarlo et al. in Figure 13, we can clearly see that our method better represents color variations and avoids the introduction of arbitrary color discontinuities. However, these advantages come at a cost: it is hard to reconstruct an image with a correct contrast. This is the reason why we introduced our contrast correction method. We can also perform histogram equalization as a post-process, as shown in Figure 11 (d) and (e).

<sup>2</sup> <http://artis.imag.fr/Publications/2007/OBBT07/>

<sup>3</sup> <http://www.umiacs.umd.edu/users/aagrawal/software.html>



**Figure 11:** Comparison with the failure case of Winnemöller et al. [2006]. (a) Original picture. (b) Winnemöller et al. abstraction failure: note how the carpet details are preserved while the fur is abstracted away. (c) Our lifetime map. (d) Our detail removal abstraction preserves the cat structure and abstract the carpet. (e) We apply histogram equalization as a post-process to fine tune contrast.



**Figure 12:** Local control: original image of DeCarlo et al. [2002] and our results for two different user-specified control maps.

Finally, one may wonder why we have not used Elder et al.’s contour domain approach [Elder and Goldberg 2001] instead of the Poisson reconstruction. Although their approach could be used for most of the stages of our method, the fact that they need to handle colors on both side of edges makes the manipulation stage less flexible.

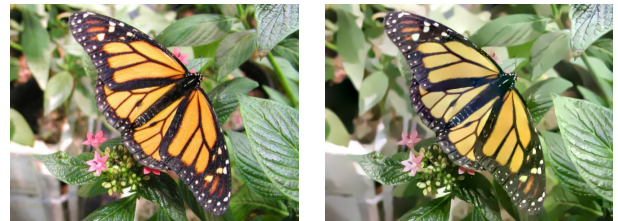
## 7.2 Future work

Our approach extracts structure from a luminance image, but uses this structure to reconstruct color images. This approach works in most cases, since luminance carries a lot of the structural information. However, when iso-luminant color regions “touch” each other in the input image, our method tends to fail to reconstruct correct colors, as shown in Figure 14. We envision two ways to solve this problem in future work: one would be to use color-to-gray methods [Gooch et al. 2005] that would introduce color discontinuities in our original luminance image; the other one would require to extract edges for different channels, but it raises the problem of combining this disparate information in the end.

Another avenue for future work resides in the design of an intuitive and efficient user interface. For instance, we plan to develop free-hand tools to manipulate directly edges at different scales: when selecting an edge at a coarse scale, the user could manipulate many details at a time, while selecting an edge at a fine scale would allow her to edit details in subtle ways. Applying our approach to process videos is also an exciting area of future work: Gaussian scale-space and Poisson reconstruction approaches have been already applied successfully to videos, and the flexibility of our structural measures would make them tractable for manipulating image sequences.

Finally, as we see our approach as a starting point for any subsequent stylization, we are also interested in developing such stylized renditions that take advantage of structural information. As an example, we created two preliminary results, shown in Figure 15: a drawing, and a watercolor. There are many connections to establish between style parameters and structure information, and we hope

this work will motivate future research along this direction.



**Figure 14:** Failure in the image reconstruction due to the iso-luminance of the pink flowers and green leaves, which leads to a greenish butterfly.



**Figure 15:** Different stylizations obtained from our abstracted images, in a drawing and watercolor style.

## Acknowledgments

Alexandrina Orzan is supported by a grant from the European Community under the Marie-Curie project VISITOR MEST-CT-2004-008270.





**Figure 13:** Comparison with the DeCarlo et al. [2002]. (a) Original picture. (b) DeCarlo et al. results exhibit flat color regions with shape simplification (c) Our result simplifies the image while keeping smooth color variations and original shapes.

## Appendix

We model a directional edge gradient  $I_{\{x,y\}}$  as the 1D convolution of a step function  $H$  of amplitude  $A$  by a Gaussian kernel  $g_\sigma$  and a Gaussian derivative  $g'_\sigma$ , resulting in:

$$\begin{aligned} I_x(0) &= (H \otimes g_\sigma \otimes g'_\sigma)(0) = (H \otimes g'_{\sqrt{2}\sigma^2})(0) \\ &= \int_{-\infty}^{+\infty} H(t) \cdot g'_{\sqrt{2}\sigma^2}(-t) dt = \int_0^{+\infty} A \cdot g'_{\sqrt{2}\sigma^2}(-t) dt \\ &= A \cdot g_{\sqrt{2}\sigma^2}(0) = \frac{A}{2\sigma\sqrt{\pi}} \end{aligned}$$

## References

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM TOG (Proceedings of SIGGRAPH 2004)*, 294–302.
- BANGHAM, J., GIBSON, S., AND HARVEY, R. 2003. The art of scale-space. In *British Machine Vision Conference*, 569–578.
- BERGHOLM, F. 1987. Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9, 6, 726–741.
- CANNY, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, 6, 679–698.
- COLLOMOSSE, J. P., AND HALL, P. M. 2003. Cubist style rendering from photographs. *IEEE Transactions on Visualization and Computer Graphics* 9, 4, 443–453.
- COLLOMOSSE, J. P., AND HALL, P. M. 2005. Genetic paint: A search for salient paintings. In *Proc. of EvoMUSART*, 437–447.
- DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. *ACM TOG (Proceedings of SIGGRAPH 2002)* 21, 3, 769–776.
- ELDER, J. H., AND GOLDBERG, R. M. 2001. Image editing in the contour domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 3, 291–296.
- ELDER, J. H., AND ZUCKER, S. W. 1998. Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 7, 699–716.
- FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM TOG (Proceedings of SIGGRAPH 2002)*, 249–256.
- FISCHLER, M. A., AND FIRSCHEIN, O. 1987. *Intelligence: The Eye, the Brain and the Computer*. Addison-Wesley.
- GOOCH, A. A., OLSEN, S. C., TUMBLIN, J., AND GOOCH, B. 2005. Color2gray: salience-preserving color removal. *ACM TOG (Proceedings of SIGGRAPH 2005)* 24, 3.
- HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *SIGGRAPH '98*, 453–460.
- HODGES, E. R. S. 2003. *Guild Handbook of Scientific Illustration*. John Wiley & Sons, Inc.
- KANG, H. W., CHUI, C. K., AND CHAKRABORTY, U. K. 2006. A unified scheme for adaptive stroke-based rendering. *The Visual Computer (Proceedings of Pacific Graphics 2006)* 22, 9, 814–824.
- KOENDERINK, J. J. 1984. The structure of images. *Biological Cybernetics* 50, 5, 363–370.
- LINDBERG, T. 1998. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision* 30, 2, 117–154.
- PALMER, S. 1999. *Vision Science : Photons to Phenomenology*. MIT Press.
- PAPARI, G., CAMPISI, P., PETKOV, N., AND NERI, A. 2007. A biologically motivated multiresolution approach to contour detection. *EURASIP Journal on Advances in Signal Processing* 2007.
- PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM TOG (Proceedings of SIGGRAPH 2003)*.
- RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. In *Proceedings of NPAR 2004*, 85–152.
- ROMENY, B. T. H. 2003. *Front-End Vision and Multi-Scale Image Analysis*. Kluwer Academic Publishers.
- SANTELLA, A., AND DECARLO, D. 2004. Visual interest and npr: an evaluation and manifesto. In *Proceedings of NPAR 2004*, 71–150.
- SIMCHONY, T., CHELLAPPA, R., AND SHAO, M. 1990. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 12, 5, 435–446.
- WANDELL, B. A. 1995. *Foundations of Vision*. Sinauer Associates.
- WANG, J., XU, Y., SHUM, H.-Y., AND COHEN, M. F. 2004. Video tooning. *ACM TOG (Proceedings of SIGGRAPH 2004)* 23, 3, 574–583.
- WEN, F., LUAN, Q., LIANG, L., XU, Y.-Q., AND SHUM, H.-Y. 2006. Color sketch generation. In *Proceedings of NPAR 2006*.
- WINNEMÖLLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-time video abstraction. *ACM TOG (Proceedings of SIGGRAPH 2006)* 25, 3, 1221–1226.