Image and Video Abstraction using Cumulative Range Geodesic Filtering

David Mould[a]

[a] *Carleton University*

Abstract

Image abstraction traditionally eliminates texture, flattening gradients and removing small-scale details. However, abstracting while preserving irregular silhouettes and medium-scale details can produce a richer abstracted image. We propose a variant of geodesic image filtering which preserves the locally strongest edges, leading to preservation of both strong edges and weak edges depending on the surrounding context.

Our contribution is to introduce *cumulative range* geodesic filtering, where the distance in the image plane is lengthened proportional to the color distance. We apply the new filtering scheme to abstraction applications in images and video, and demonstrate that it has powerful structure-preserving capabilities, especially regarding preservation and indication of irregular details. The basic technique, where every pixel is equally abstracted, is further extended with explorations of variable mask size based on spatial location, salience, intensity, and location combined with intensity.

Keywords: non-photorealistic rendering, image processing, abstraction, texture, texture indication, video processing

Email address: mould@scs.carleton.ca (David Mould)

1. Introduction

Since the beginnings of NPR, the field has been populated with algorithms for well-known artistic styles and media. At the same time, there has been substantial interest in pure abstraction techniques, i.e., methods for creating a version of an input scene with markedly less detail than a conventionally rendered image or a photograph. While early abstraction processes such as that of Haeberli [1] sought a specifically painterly look with explicit strokes, the more cartoon-like abstraction presented by DeCarlo and Santella [2] has become the norm for NPR abstraction. Image abstraction mechanisms based both on segmentation and on direct filtering sought to remove detail, yielding images characterized by large uniform regions and smooth gradients, free from texture.

Yet while it is often taken for granted that texture should be removed while undertaking image abstraction, it is far from clear that painted or otherwise artistically created images lack texture. Artists often seek to convey the material properties of surfaces, such as roughness, and do so by careful use of strokes or other drawing primitives. Some NPR algorithms seek to introduce texture into images where it was formerly absent: for example, the watercolorization methods of Bousseau et al. [3, 4] introduce high-frequency texture to represent pigment granulation. Various example-based techniques (e.g., Zhao and Zhu's "Sisley" [5]) introduce extra texture also, seeking to recreate large-scale image structures by placing large numbers of textured strokes. In both of these examples, texture is introduced without reference to the textural properties of the original image; instead, the texture is introduced to mimic the appearance of the medium rather than portraying texture details from the scene. In this work, we attempt to abstract input images while retaining small-scale and medium-scale details; we use a conservative smoothing process that preferentially avoids smoothing across image edges, even very weak edges. Our mechanism preserves the *locally strongest* details, eschewing the flattening effect common in image-based abstraction techniques. The abstracted images are textured if and where the original image was textured. Very small features, whether texture elements or image details, fade into their surroundings without vanishing; sometimes this gives the impression of a 'glazing' effect resembling watercolor. More importantly, the image retains a sense of the surface structure visible in the original image; preservation of irregular silhouettes, internal edges, color variation, and intermediate-sized details all contribute to this effect. Our results demonstrate detail preservation using natural textures including examples of bark, hair, fur, and rock surfaces, among others.

Numerous dedicated algorithms for image abstraction have been proposed over the years, some of which we discuss in more detail below. This paper presents a variation on the geodesic filter that has been specifically designed for abstraction of textured images, where the textured regions are abstracted but retain their irregular shapes and ragged edges. The algorithm builds a dedicated mask for each image pixel, taking the nearest n pixels according to a new "cumulative range" variation of geodesic distance. The problem of texture indication has been of longstanding interest to the NPR community, and our mask customization process offers some insight into how texture indication can be achieved. Results of our abstraction process are shown in Figure 1.

The main contribution of this work is the *cumulative range geodesic filter* (CRGF) and a discussion of some of its properties. The bulk of the paper is devoted to an exposition of the definition and characteristics of the filter and a comparison of its output to that of other generic abstraction methods. This portion of the work was previously presented at NPAR [6]. The current extended version of the paper includes explorations of spatial variations of the level of abstraction using the proposed filter. We discuss mask sizes based on simple spatial calculations, semantically meaningful user-defined regions, and some techniques for automated detail variation to produce additional visual effects, such as apparent contrast heightening. In addition, we present the results of applying the CRGF to video; the robustness of the method allows effective creation of abstracted video simply by applying the method to individual video frames. Before we discuss the CRGF in detail, however, we first review some of the existing methods for image abstraction.

2. Background

Some work in image abstraction depends on an initial segmentation of the image; for example, DeCarlo and Santella use an automatic hierarchical segmentation [2]. However, since our approach is filter-based, we will emphasize filter-based approaches in our survey.

The two concepts most relevant to this paper are the bilateral filter [7] and geodesic filtering [8]. The bilateral filter involves computing a custom arrangement of weights for each pixel, where the distance of the central pixel to each other pixel is a combination of spatial distance and colorspace distance. Geodesic filtering involves treating a 2D image as a 3D surface and computing distances from a pixel or group of pixels over the

Figure 1: Some large-scale results. Clockwise from upper left: stranger; cliff; forked tree; ranch; autumn.

manifold [9]. Our proposed filter uses a geodesic distance where the incremental horizontal distance from a starting pixel is proportional to the colorspace distance of the current pixel to the original pixel. Results of applying the filter are shown in Figure 1.

The bilateral filter has been used explicitly for abstraction purposes, notably by Winnemöller et al. [10] whose iterated approximation allowed real-time video processing; Winnemöller et al. also noted texture indication as a goal, though this effort was only partly successful. The domain transform presented by Gastal and Oliviera [11] allows real-time edge-aware operations; they presented it as an acceleration of bilateral filtering and demonstrated a wide range of effects. Real-time video processing was also an advantage of the geodesic-based formulation presented by Criminisi et al. [8], one aspect of which was image abstraction. Our naive implementation is far from real time, but we believe that the visual effect we present cannot readily be obtained by other known techniques.

Orzan et al. undertook edge-preserving filtering [12] based on the Poisson equation, with the explicit intention of removing small-scale, weak edges. We share their goal of faithful edge preservation; like ours, their output images contain gradients, which are suppressed in approaches favoring flattening. However, weak edges make frequent appearances in image texture, and elimination of weak edges is a point of departure between our goals and theirs. Strong edges are generally preserved in the flow-based process of Kang et al. [13], in which custom smoothing kernels align to local edge tangents, but weak edges are eliminated and region boundaries are simplified. Kass and Solomon [14] used local histograms for abstraction effects, effectively preserving strong edges and medium-scale details. GradientShop [15] is a versatile system operating in the gradient domain; when employed for abstraction purposes, weak edges and textures (e.g., hair) are abstracted away, replaced by longer-range gradients. A related endeavour was reported by Olsen and Gooch [16], who used a sequence of linear filters to prepare for the creation of an edge image for subsequent stylization and compression; again, weak edges are eliminated and eventually replaced by gradients in this approach.

Much work in edge-preserving abstraction has been undertaken based on the Kuwahara filter [17, 18, 19, 20], a nonlinear edge-preserving filter. Such work has not sought to preserve texture; in fact, Papari et al. [17]

have the stated goal of eliminating texture. As we argued above, texture removal should not be held out as the sine qua non of abstraction: only certain styles of imagery are texture-free. In the remainder of this paper, we demonstrate a style of abstraction in which the textures are muted and abstracted to a degree but, by design, preserved sufficiently to be communicated to the viewer.

3. Algorithm

Bilateral filtering uses a fixed mask shape with custom weights for each pixel within the mask, depending on the mask's location. In contrast, we propose a filtering process which creates a customized mask shape for each individual pixel but with uniform weights within the mask. The mask consists of the n pixels nearest to the centre, where "nearest" is with respect to a particular distance formulation, explained in detail below, that incorporates aspects of both bilateral and geodesic filters. Edge preservation, even of weak edges, is a natural consequence.

Let I be an image, whose pixel values are greyscale or color; we refer to "intensity" below without loss of generality, but in practice compute color distances in RGB space. Distances between arbitrary pixels are computed as the infimum of path costs among possible paths connecting them. More formally, using notation adapted from Criminisi et al. [8], for pixels a and b, the distance $d(a,b)$ is defined as follows:

$$d(a,b) = \inf_{\Gamma \in P(a,b)} \int_0^{l(\Gamma)} C(\Gamma, s)\, \mathrm{d}s \qquad (1)$$

where the integration occurs over the arc length s of the path, from 0 to the total path length $l(\Gamma)$. The function $C(\Gamma, s)$ is the infinitesimal cost of proceeding along path Γ at s and is given by

$$C(\Gamma, s) = |I(a) - I(\Gamma(s))| + \gamma |\nabla I(\Gamma(s)) \cdot \Gamma'(s)|. \qquad (2)$$

In equation 2 we wrote $I(a)$ to indicate the intensity at the beginning of the path, but could equivalently have written $I(\Gamma(0))$.

In equation 1, $P(a,b)$ refers to the ensemble of paths linking a and b; Γ is one such path, parameterized by arc length s. Note that $\Gamma(s) \in \Re^2$ is a location in the image plane. The image intensity at a location x is given by $I(x)$. The formulation $\nabla I(\Gamma(s)) \cdot \Gamma'(s)$ is the component of the image intensity gradient parallel to the path direction. The parameter γ weights the relative importance of local edge-crossing versus deviation from the original pixel color.

In practice, we do not use equations 1 and 2 directly. Rather, we use a front propagation method to add pixels to the mask one by one. The incremental distance when proceeding from pixel g to pixel h is

$$|I(h) - I_0| + \gamma |I(h) - I(g)| \qquad (3)$$

where $I_0 = I(a)$ is the intensity at the first pixel of the path, or equivalently, the intensity at the centre of the mask. The two terms of equation 3 map to the two terms of the right hand side of equation 2: the first term is the intensity difference between the original pixel and the incremental pixel, and the second term is the discrete version of the gradient along the direction of the path, i.e., simply the difference in intensities along the path increment. The formulation of equation 3 makes obvious the role of the original pixel intensity in controlling the overall region shape. Typically we use mask size (pixel count) to control the termination of mask expansion; in section 6 we also discuss the possibility of using cumulative distance to decide on mask size. Once the mask has been computed, we use a box filter over the defined domain to obtain the output color value.

Better intuition for the mask shape customization can be gleaned from Figure 2, which shows some example masks (n=180,γ=1). The original pixel is shown with a black circle and the region boundary with a heavy red line. Notice how the mask avoids crossing strong edges (left hand example) and how the mask can become very irregular in order to conform to highly irregular structures in the image (right hand example). Ultimately, the mask is a contiguous region, possibly with holes, consisting of those pixels whose cumulative distance to the starting pixel is smaller than that of other nearby pixels. In practice, this typically means pixels of similar colors, although note that a large mask size may mean that very dissimilar pixels can be included if insufficient numbers of similar pixels are reachable.

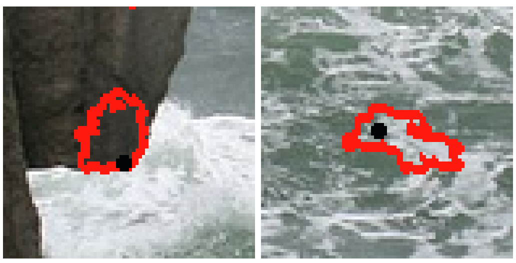

Figure 2: Irregular masks.

Figure 3 shows a visual comparison to geodesic flattening. (The original image is included as part of Figure 5.) Our result better preserves texture, as intended;

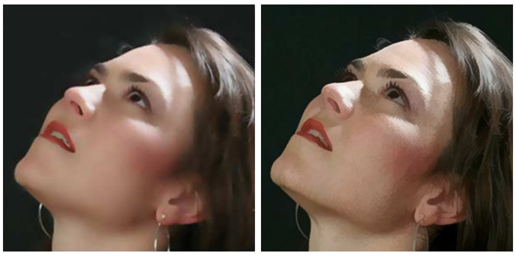

Figure 3: Comparison with geodesic flattening. Left: geodesic output; right: our result. Former image provided by Criminisi et al. [8].

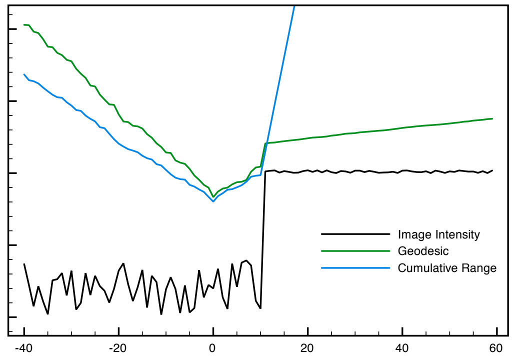

Figure 4: Comparison of geodesic vs. cumulative range distance.

pay particular attention to the hair above and behind the ear, which has been significantly blurred by geodesic flattening but is more gracefully abstracted with our technique. While the lack of smoothing of texture over the face may not be desirable in this instance, the texture is indeed present in the original, illustrating that the abstraction style chosen must be consonant with the user's aims.

The difference seen between the proposed method and geodesic smoothing is not an accident. Geodesic approaches (based on range distance plus integral of gradient magnitude) always penalize edges, but do not directly consider intensity difference from the source; our method does not always penalize gradients, but flat regions of intensity different from the source pixel are heavily penalized. Consider the illustration in Figure 4. The original intensity profile, beneath, contains an area of high variance adjacent to an area of low variance with a different average intensity. Distances from the pixel at x=0 are illustrated above. The geodesic distance (green) jumps suddenly when crossing the edge, but rises slowly thereafter, as local gradients are small. The cumulative range distance, however, rises quickly and steadily after crossing the edge. Conversely, the geodesic distance rises relatively quickly in the textured area, as numerous small gradients must be crossed; the cumulative distance rises much less quickly, since despite local variations the intensity does not stray too far from that of the starting pixel. Whether used to create a mask, as in the present proposal, or used to compute weights for a local averaging process, the cumulative range and geodesic distances treat texture in quite different ways that cannot be reconciled simply by selecting parameters or making other superficial changes.

In the following, we demonstrate the practical effect of the filter by applying it to several test images and showing comparisons to existing abstraction methods.

4. Results from Fixed Parameters

4.1. Images

Several examples were shown in Figure 1. The original images for our examples and comparisons can be seen in Figure 5. Since appreciating the results depends on careful attention to small-scale details, we recommend viewing these images on screen and at a high zoom level.

We chose sample images that contain large textured regions. The different types of texture are generally recognizable in the abstracted images. In the cliff image, perserving the texture is especially important since it offers a natural way to depict the motion of the water, apparent even in this still image. Areas which lack texture, such as the house in the ranch image, are more conventionally flattened. The remainder of this subsection discusses specific details of some of these images.

Figure 6 shows the basic texture abstraction and edge preservation properties of the algorithm. In the top image pair, notice how the texture is muted without being obliterated: the dynamic range has been locally compressed, but the details are still subtly present. In particular, irregular structures remain visible: e.g., the foam boundaries on the upper right. In the middle pair, the corrugated bark texture is still apparent, and the silhouette of the tree has been preserved. The bottom image has been selected to demonstrate strong edge preservation: the complex silhouette of the fabric is unchanged while the higher-frequency details within the body of the cap are abstracted.

Thin but extended linear features can be difficult to preserve using other approaches; segmentation-based approaches find them particularly challenging. However, our method quite naturally maintains linear features: as long as enough similar-colored pixels lie in a

Figure 5: Original images for the abstractions used throughout this paper.

contiguous region, regardless of shape, the masks can stretch to collect them. Figure 7 illustrates preservation of linear features. While the masses of leaves and the main part of the beard are abstracted, individual tree branches and hairs remain visible, hinting at the overall structure without portraying it explicitly. For a non-texture example of linear feature preservation, look at the whiskers of the cat in Figure 12. The yellow leaves in Figure 7, combined with the slight paling of the smaller branches, provides a striking example of the watercolor 'glazing' effect alluded to earlier. The leaves of Figure 8 provide another example. Edge-preserving fading is discussed in more detail in the folowing subsection.

Figure 6: Texture abstraction and edge preservation. Top: wave detail from cliff image; middle: bark detail from forked tree image; bottom: detail of cap from stranger image. Left: originals; right: filtered.

Figure 7: Linear feature preservation. Above: detail from autumn image; below: detail of stranger's beard. Left: originals; right: filtered.

Figure 8: Subtle texture edges and irregular regions.

While Figure 6 has examples of preserving strong edges, we are also interested in *weak* edge preservation. Because of the custom mask shape, even weak edges can be preserved if they are locally the strongest. This is sometimes a subtle effect, but it is apparent in Figure 8, especially the spray near the cliff in the middle image. The water contains many low-intensity edges that are nonetheless maintained in the abstracted output. A larger-scale example of this is the shadow of the cliff, seen in full in Figure 1. Although the intensity of the shadow edge varies considerably from place to place, it always represents a sufficient change in contrast that the masks rarely straddle it. Figure 8 also contains further examples of irregular texture regions in the water and foliage. Particularly note the muted irregular structures in the abstracted mass of leaves, indicating the details but leaving it to the mind of the viewer to fill them in. The ragged silhouette helps considerably in furthering the illusion of detail.

4.2. Comparisons

In the following, we provide visual comparisons to some recent and relevant abstraction techniques: structure-preserving photo manipulation [12]; closest-mode histogram filtering [14]; the multiscale Kuwahara filter [20]; and bilateral filtering [7]. We previously gave a comparison to geodesic flattening [8]. Original images for the results in this section are part of Figure 5.

Figure 9: Comparison with structure-preserving photo manipulation. Left: result from Orzan et al. [12]; right: our result.

Figure 10: Comparison with closest-mode histogram filtering. Top: result from Kass and Solomon [14]; bottom: our result with $n = 300$.

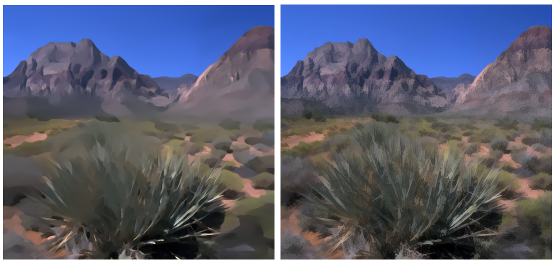

Figure 11: Comparison with multiscale Kuwahara filter. Left: Kuwahara filtering; right: our result. Former result provided by Kyprianidis [20].

Figure 9 compares our filter with structure-preserving photo manipulation [12]. Like us, Orzan et al. intend to faithfully preserve strong edges; the two methods are approximately equally successful at this. However, Orzan et al. deliberately seek to eliminate weaker edges, with the consequence that the detail on the petals, the texture on the fruit, and even an entire faint leaf (above the upper white flower) are largely removed. These textures are not particularly well defined in the original image, yet our method somewhat suggests them. A subtle point is the texture on the object behind the fruit: entirely smoothed by photo abstraction, nonetheless our result conveys a delicate sense of half-glimpsed surface detail.

Figure 10 gives a comparison between our method and closest-mode histogram filtering. Local histogram filters were proposed by Kass and Solomon [14]; the closest-mode variant has similar edge-preserving properties to ours. However, weaker edges are not preserved, such as the front wheel of the tractor, or the shading on the rim of the larger wheel. Conversely, small distinct objects are well preserved, such as the white marker posts in the field, on the middle left of the image; to our eye, these are somewhat distracting details ill-suited for preservation. In our result, the tractor wheels remain defined, while the markers have faded.

Figure 11 shows our attempt to deal with an especially difficult image, presented by Kyprianidis [20] as a failure case for multiscale Kuwahara filtering. While we agree that this is a challenging image and our result is not completely clear, we are able to maintain some separation between the bush and the middle ground and avoid the overblurring visible in the output of the Kuwahara filter. The textures on the distant mountains have been abstracted nicely.

Figure 12 compares a successful outcome from Kuwahara filtering to our approach. We have also included the bilateral filtering result in this figure. Compared to both single-scale and, especially, multiscale

Figure 12: Comparison with multiscale Kuwahara filter. Left to right: bilateral filter; anisotropic Kuwahara filtering; multiscale anisotropic Kuwahara filtering; proposed filter. The first three images are provided by Kyprianidis [20].

anisotropic Kuwahara filtering, our result better suggests the underlying texture. Our filter flattens the image but the resulting mostly-uniform regions have irregular boundaries and the silhouettes are ragged, better conveying a sense of the fur. The whiskers are also better preserved, particularly in the image's upper left where the contrast with the cat's fur is relatively low. More subtle surface details are preserved as well: the rough surface of the cat's tongue is suggested by the irregular internal edges, eliminated entirely by Kuwahara filtering.

With respect to the bilateral filter, our result appears more muted, having removed small high-contrast elements such as the black spots around the cat's nose and mouth. Similarly, contrast has been reduced in the fur across the cat's forehead. At the same time, our texture detail is in many places superior to the result from the bilateral filter: for example, subtle distinctions have been maintained on the sides of the cat's face. This attests to the difficulty of using a single parameter setting for the bilateral filter to preserve edges of different scales.

This difficulty is further illustrated in Figure 13, which compares the proposed filter and the bilateral filter. The original synthetic image is a series of ideal vertical step edges corrupted with uniform noise; the noise has higher magnitude at higher intensities. We show a horizontal cross-section, with the original data in light grey. With the settings used, the bilateral filter recovers intermediate intensity discontinuities well, but smaller edges are smeared out. At the same time, the

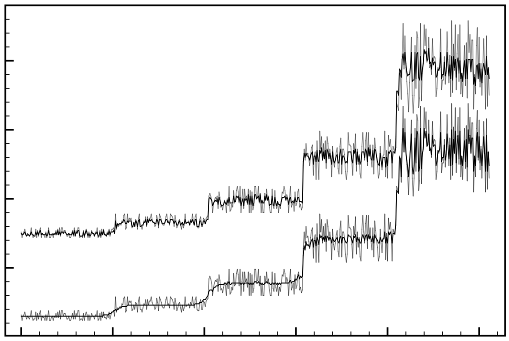

Figure 13: Response to synthetic data. Above: proposed filter; below: bilateral filter. Original noisy data shown in light grey.

larger quantities of noise are little smoothed. In general, the bilateral filter offers a tradeoff between preserving edges and preserving noise or fine detail. For applications to detailed images with complex and heterogeneous edge magnitudes, it is not possible to find a single setting for the bilateral filter to treat the entire image properly: in the example, the setting that works well for intermediate noise values oversmooths small edges and undersmooths large noise values. Conversely, the proposed filter automatically adapts to the noise levels. It has better edge localization for the smaller edges; although to some extent it treats the noise as edges to be preserved, the noise pixels are incoherent and its ability to preserve them is limited by the unavailability of sufficient numbers of pixels of similar color: i.e., incoherent noise is always attenuated. Larger noise values leave a larger residual noise after smoothing, but are still smoothed more than by bilateral filtering. Note that while the noise removal applications of the proposed filter may be limited, we are not proposing the method for noise removal proper, but rather abstraction; the small-scale intensity changes that are best preserved are those which are coherent over a region roughly the size of the mask or larger. Coherent intensity changes are exactly those which we do want to preserve.

4.3. Video

We applied the method on a frame-by-frame basis to video. Here, we report the outcome of this exercise. We chose a video sequence using two criteria: first, it should have a variety of scenes and activities; second, it should contain considerable texture and rich visual detail. We selected a short film about forests (created for 2011's "International Year of the Forest"[21]) which

Figure 14: Video frames

comprises a series of short clips, each typically only 5-15 seconds in length; the subject matter (forests of the world) naturally provides ubiquitous texture detail to demonstrate the effectiveness of our method. Individual frames (approximately 0.5 seconds apart) are shown in Figure 14.

On watching the video sequence in motion, one observes high coherence. Objects seem to move as units and there is minimal flicker. The film was uniformly abstracted with a mask size of 240, resulting in a medium to high level of abstraction, as can be seen in the sample frames in the figure. Smoke and dust clouds are neatly preserved and the motion of detailed amorphous phenomena such as spray in waterfalls is straightforwardly conveyed by the abstracted video.

Watching closely, one can observe a few spurious behaviors. Small objects such as leaves can take on the color of the background, with curious results when the background changes. Fine details can become blurred, akin to the treatment of the leaves at the top of the forked tree in Figures 1 and 17, and in motion this can give the impression of a solid surface rather than a collection of leaves and branches. However, these are minor limitations and overall we consider the application of the technique to video to be quite successful.

5. Parameter Variations

Here we discuss the effect of changing parameter settings, n and γ. The filter size n has the stronger effect on the results, and we devote most of our efforts to investigating it. We first show the results of changing it uniformly over the image, then describe some schemes for varying n on a pixel-by-pixel basis over the image, using spatial, semantic, or intensity information.

Figure 15 shows the effect of changing the filter size n. Unsurprisingly, larger masks produce more abstracted images: features of size larger than the mask can straightforwardly be preserved, while smaller features begin to disappear. However, the small features do not become blurred, but instead gradually fade. This effect is most apparent in the dark rock in the upper right of Figure 15.

The fading phenomenon is due to the asymmetry inherent in our formulation. If pixel b is part of the mask for pixel a, there is no guarantee that pixel a will belong to b's mask. In general $d(a, b) \neq d(b, a)$. In the context of the black rock, the behavior manifests as follows: when possible, the black pixels form masks which occupy only black pixels, but when the mask size is larger than the rock, part of the mask necessarily extends into the water, lightening the box filter output. Conversely, the lighter pixels near the rock have a huge area of light-colored pixels in which to form their masks, and they never need to extend into the black rock. Thus, though the rock fades, it has little influence on the color of the water pixels, and the shape of its outline is preserved.

Figure 16 shows the effect of changing γ. With low γ, the deviation from the origin dominates, and we have high detail preservation. With high γ, the mask tends to extend into nearby smooth regions regardless of whether they are similar to the original pixel's color, so detail tends to be lost. In general we prefer the look of low γ, but for specific images or more aggressive detail removal, higher γ may be preferred. Unless otherwise stated, all images in this paper were created with γ=1.

5.1. Spatial Variation of Mask Size

Thus far we have considered only constant mask size, relying on adaptive mask shape for detail preservation. However, mask size can be set independently at each pixel, leading to different effects such as improved contrast or emphasis. Here and in the following few subsections, we discuss some possibilities for adjusting mask size locally.

The first and simplest way to adjust mask size it to set mask size according to location in an image. Figure 17 shows two examples: in the left image, the abstraction level increases with distance from the bottom, and in the right image, the abstraction level increases with distance from the centre.

In the case of the tree image, the effect is to make the treetop seem to recede into the distance. The trunk itself appears solid, but the upper leaves are more ephemeral. In the case of the ruins, the effect is reminiscent of a depth-of-field blur effect, where attention is concentrated on the sharper foreground objects, but because of the inherent detail-preservation properties of the filter, we are able to achieve the effect in postprocessing without explicit segmentation. More complex scenes can benefit more strongly from segmentation, as we see in the next subsection.

5.2. Semantic Variation of Mask Size

We manually segmented images into regions of low salience and regions of high salience, then applied the abstraction process with two different mask sizes: $n = 40$ for high salience and $n = 400$ for low salience. Figure 18 shows the outcome of this process as well as comparisons with uniform abstraction. In the uniform low-abstraction case, the background is distracting, drawing attention away from the subjects. In the

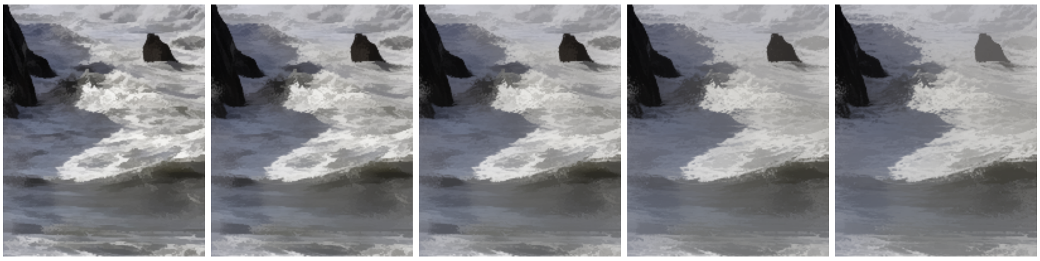

Figure 15: Effect of changing size of region. Left to right: region sizes 40, 80, 160, 480, and 800 pixels.

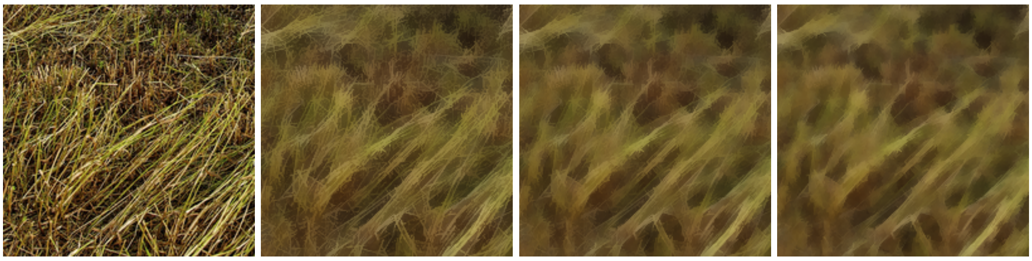

Figure 16: Effect of changing γ. Left to right: original, γ=1, γ=8, γ=64.

Figure 17: Spatial variation of mask size. Forked tree: n rises vertically from 20 to 400. Ruins: n rises radially from 25 to 400.

uniform high-abstraction case, the filtering is too aggressive: the subjects are flattened and the result unappealing. The masked result draws attention to the salient areas, while the background appears to recede, yet the context of the scene is still apparent. Still more aggressive filtering of the background, as with existing abstraction processes, would not be appropriate in these cases, since the backgrounds are highly detailed. The autumn forest behind the tourists, and the faces of the crowd and the textured sweater of the protestor, are still apparent after application of our filter; however, the contrast is muted and these areas no longer call the viewer's attention as strongly. Further, because some abstraction occurs everywhere, the boundaries of the semantic regions are not too jarring. In fact, the regions have been quite crudely drawn, not necessarily conforming exactly to the subjects, but this has not had a detrimental effect on the abstraction effect or the salience emphasis. This suggests that lighter-weight but inexact techniques for estimating salience, such as eye tracking, would be effective for this application.

5.3. *Intensity-based Variation of Mask Size*

Semantic variation is most worthwhile when the image contains distinct objects which can be segmented from the background, but is less feasible for textures or for tiny details. Here, we vary mask size based on image intensity in order to increase contrast or to better preserve specific details which can be distinguished by their intensity levels.

In our implementation of intensity-based mask size, we compute mask size n by linear interpolation between n_{min} and n_{max} according to a pixel's intensity distance from the least-abstract intensity level, say I_0. That is, $n(x,y) = n_{\text{min}} + N|I(x,y) - I_0| \times (n_{\text{max}} - n_{\text{min}})$, where N is a normalizing constant. For $I_0 = 0$ and I in the range $0 - 255$, we would have $N = 1/255$. In the examples, we use $n_{\text{max}} = 240$ and $n_{\text{min}} = 20$.

An example result is shown in Figure 19. Much of the structure of the rock wall is conveyed by the self-shadowing, but small dark shadow areas fade when uniform abstraction is applied, making it harder to appreciate the rock surface. Using intensity-based abstraction ($I_0 = 0$) better preserves these specific details, while still applying a strong abstraction effect to the image as a whole.

Another pair of examples appears in Figure 20. On the left, we set $I_0 = 255$ with the intent of better preserving the small light features, especially the kitten's whiskers. On the right we have $I_0 = 128$ to preserve intermediate-lightness detail in the waves. In both cases we retain a strong abstraction effect while preserving details of lightness close to the selected value.

Figure 19: Abstraction of "Sedona" image. Above: uniform abstraction. Below: less abstraction of dark pixels.

Figure 21 compares intensity-based abstraction to uniform abstraction in selected regions of the images. In the Sedona image, we can see the surface of the rock wall much more clearly, owing to the preservation of small dark areas. In the kitten image, the whiskers stand out more clearly from the surrounding fur. In the cliff image, small details of the wave structure has been preserved, leading to an effect of texture indication.

As described, the effectiveness of simple intensity-based mask resizing depends crucially on finding a single intensity value that captures the details of interest. In many images, no such value exists. A more elaborate scheme computes mask size based on a pixel's deviation from the local average, thus preserving all details: $n = n_{\text{max}} - k \times |I(x,y) - \bar{I}|$, where the amount of detail is parameterized by k. The value of n is clamped below by a value n_{min}. Applying this technique gives an abstraction where small details and sharp edges are preserved, giving the appearance of contrast enhancement. An example is shown in Figure 22.

Figure 18: Abstraction extent governed by semantic regions. Topmost: manually created regions; next: low abstraction everywhere; next: low abstraction within regions, high abstraction elsewhere; bottom: high abstraction everywhere.

Figure 20: “Kitten” and “cliff” with abstraction based on intensity.

Figure 21: Comparison of uniform abstraction (right) and intensity-based (left). Dark self-shadowing on the rock wall, the kitten’s white whiskers, and isolated grey shapes on the waves are all more prominent in the intensity-based result.

Figure 22: Comparison of uniform abstraction (above) to delta-based abstraction (below).

6. Controlling Mask Size with Cumulative Distance

In the previous sections, we exclusively relied on mask size (measured by counting pixels) to control the extent of abstraction. We varied mask size spatially or by intensity. However, subtly different abstraction effects emerge when we adjust mask size according to cumulative distance or, more meaningfully, the *rate of increase* of cumulative distance.

Replacing mask size with cumulative distance directly has the effect of limiting abstraction while not limiting runtime. Areas of high uniformity, least in need of smoothing, will see enormous masks; masks centred on outliers will be small. The former effect can be ameliorated by enforcing a maximum region size, regardless of distance; the latter effect implies that outliers will be little smoothed and is an inevitable effect of terminating mask expansion based on cumulative distance travelled. Even so, direct use of cumulative distance does not produce attractive results: generally uniform areas are oversmoothed, and outliers are undersmoothed. Some example results are shown in Figure 23. For the "country road" example, the high-contrast leaves in the foreground have been jarringly undersmoothed, while the trees in the background and the road in the middle distance are oversmoothed. In the "winter" example, the dark spaces between the snowy branches are undersmoothed, while the snow itself and the tree trunk are oversmoothed, destroying the subtle details formerly visible.

Sample profiles of cumulative distance with respect to mask size increase are shown in figure 24. The profiles are typically simple in shape. In smooth regions or uniformly textured regions, the profiles are approximately linear. In cases where the initial pixel is substantially different from its surroundings, we observe a convex curve: an initial steep increase in cost, slowing as the distance grows and more pixels are reachable with little incremental distance. Rarely, curves can be concave, with an initially small slope followed by a sharp increase in incremental cost; this shape is characteristic of masks originating withing a small solid-colored region distinct from its surroundings.

Examining cumulative distance as the mask grows provides a better basis to set the target mask size. Our intended policy can be summarized as follows:

- If the cumulative distance increases slowly, there is little local variation beneath the mask: every nearby pixel is similar to the original pixel. To improve performance and to preserve what little detail is there, we ought to use a small mask.

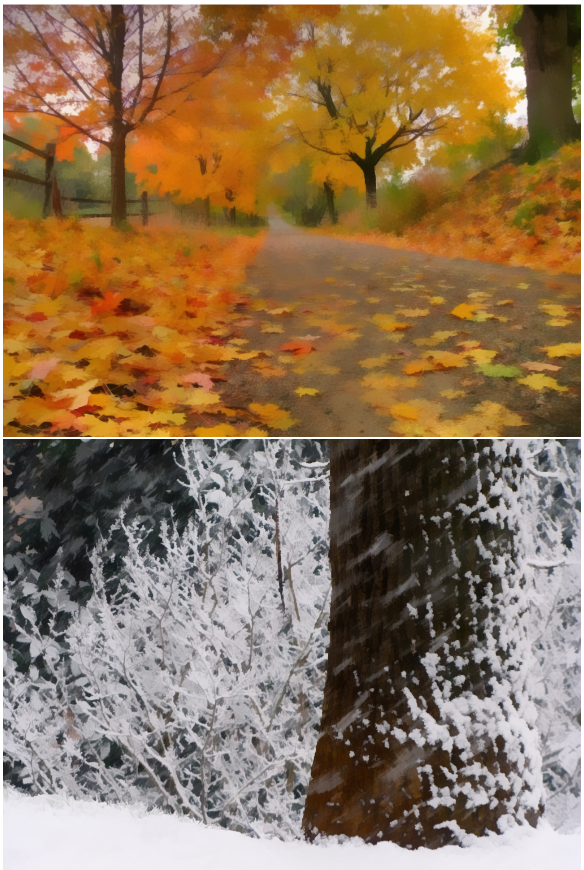

Figure 23: Filtering governed naively by distance. Above: country road; below: winter.

- If the cumulative distance increases quickly, the most similar nearby pixels are quite different from the original pixel. The original pixel was an outlier or part of a very small detail. We should use a very large mask size to eliminate the detail.

- If the cumulative distance increases at an intermediate rate, either the original pixel was slightly different from an otherwise-uniform surrounding region, or the surrounding region contains a heterogenous mixture of pixel colors. Mild abstraction with a mask of intermediate size is warranted.

We implemented this policy as follows. We follow the previously described process for every pixel using a small mask size, say $n = 40$. Then, we compute the average cumulative distance of the nth pixel; call this value D. The value D provides a simple benchmark to estimate whether distance is rising quickly or slowly relative to the rest of the image; we use $f \times D$ for a user-defined f, where a cumulative distance below fD is considered slow, and the ratio $r = d(n)/(fD)$ characterizes the rate of increase otherwise. The mask size is set to 1.5 times the original small mask size for $r < 1$, and $1 + r$ otherwise. A plausible range of f is approximately 0.2 to 0.8; $f = 0.5$ is usually effective. We cap the mask size at 500 pixels.

Results from this scheme are shown in Figure 25. These images show a good mix of abstraction and detail. Some small details, such as the foreground leaves from the country road result, or snow on the tree trunk in the winter image, are still highly visible; other details have been abstracted away. Subtle variations remain; the distant foliage in the country road image, and the bark in the winter image, are both suggested. Some specific details are shown in Figure 26. As can be seen in these examples, using a fixed distance value to terminate mask expansion sometimes oversmooths and sometimes undersmooths, while using the more elaborate policy we outline avoids both problems.

7. Discussion

The proposed filter is an edge-preserving abstraction method with the ability to preserve weak edges as well as strong ones, if the weak edges are locally the strongest. It preserves features of size approximately the size of the mask, where "size" measures the number of pixels but not necessarily the spatial extent. Preserved edges and features can be highly irregular in shape.

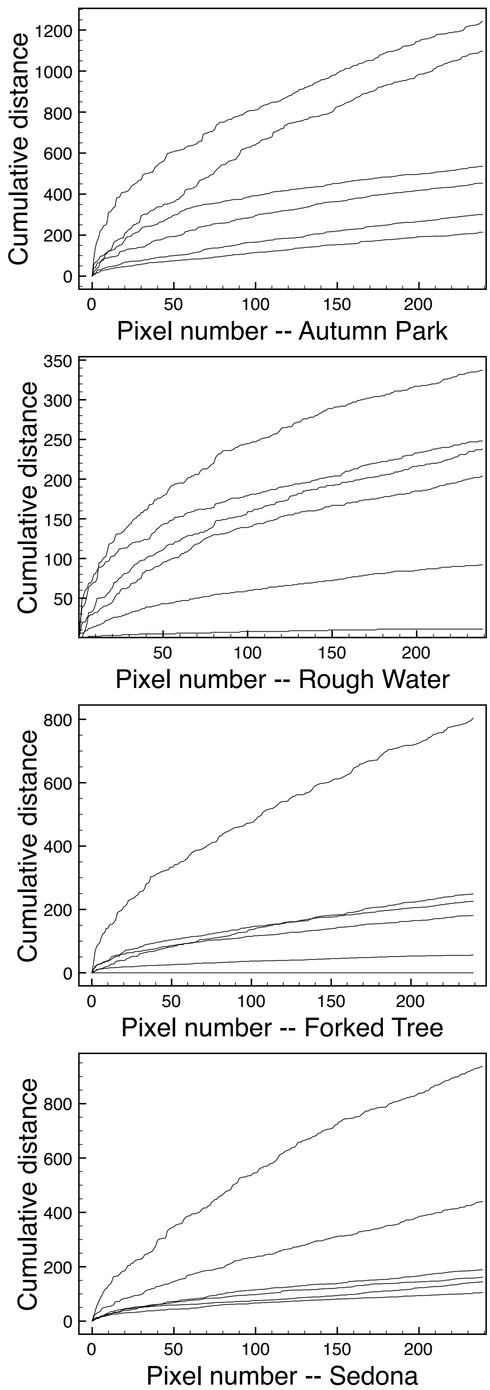

Figure 24: Sample plots of cumulative distance for different images. Clockwise from upper left: autumn park; rough water; Sedona; forked tree.

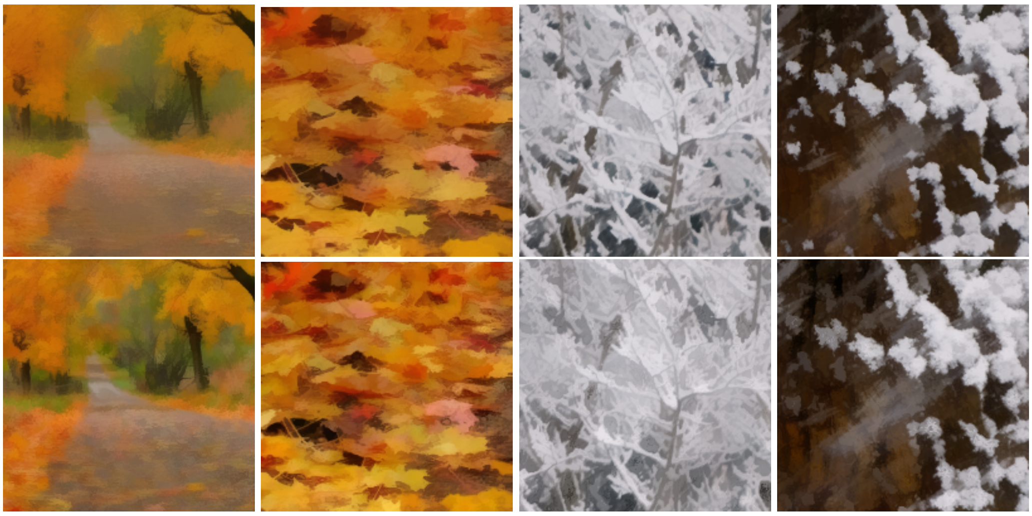

Figure 26: Details from "country road" and "winter" images. Above: filtering terminated by distance; below: mask size guided by distance-aware policy.

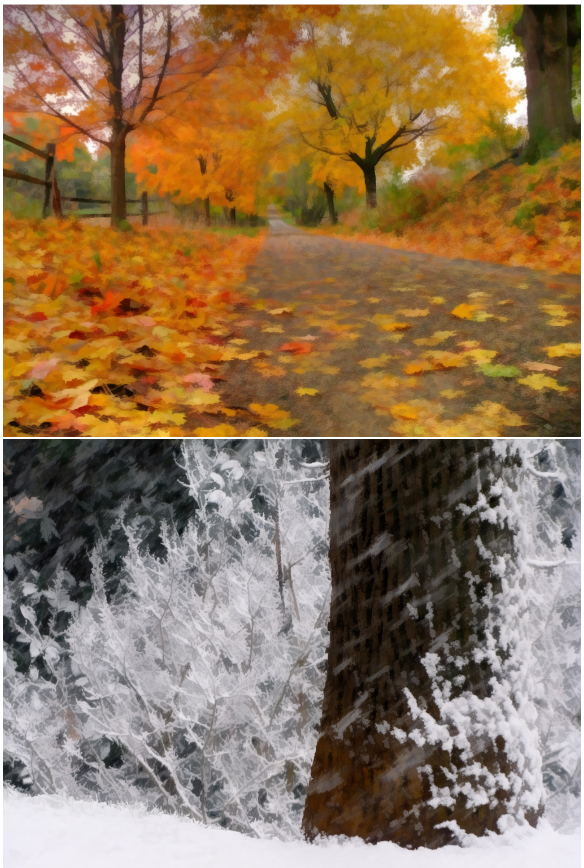

Figure 25: Filtering with mask size governed by estimated rate of change of distance. Above: country road ($f = 0.8$); below: winter ($f = 0.3$).

The main limitation of the proposed filter is its speed. Bilateral and geodesic filtering have benefitted from recent advances which make them extremely fast, but our unoptimized single-threaded prototype implementation is orders of magnitude slower: the speed is $O(kn)$ for a k-pixel image with mask size n. For typical images in this paper (roughly one megapixel), with $n = 160$, our processing time is 2-3 minutes, which includes time-consuming debugging and tracking processes.

The property of weak edge preservation is beneficial when weak edges represent real structure, but the algorithm also preserves noise to some extent. In practice this would be ameliorated by lightly preprocessing the input with another filter.

Finally, not all textures can be adequately preserved using the proposed approach. Very high frequency textures still tend to be suppressed. Figure 27 shows an example: the noise-like texture of the sand in the foreground is almost entirely removed, leaving only the shapes and shadows of the driftwood in the middle ground. This is quite like a traditional image abstraction result, but we consider it a failure case for our method given our objective of texture preservation. A user can potentially address this with automatic mask size variations, but tuning the parameters to achieve a particular effect may not be easy.

8. Conclusion

We have presented a novel variant of geodesic filtering, in which horizontal distance over the image manifold is locally stretched by the range distance to the

Figure 27: Left: original; right: filtered. The sand texture is not preserved.

origin. We made use of this distance to build custom masks; box filtering over such masks yields a texture-preserving abstraction effect. This is an effect rarely seen in past abstractions, which mainly concentrated on flattening the image and removing texture.

We showed images demonstrating the properties of the filter: its adherence to strong edges such as silhouettes; its ability to convey weak and irregular edges; its preservation of extended linear features; and its attenuation of isolated small features. A rarity in having been designed for texture abstraction, this method produces images visually distinct from those of other methods.

The main drawback to the proposed approach is its slow speed, so the obvious goal in future work is to address this. The existing approach can straightforwardly be parallelized to take advantage of multiple cores, and we would also like to investigate a GPU implementation and alternatives to naive front propagation. Improving the speed would make the application to video more feasible in practice; while already possible, it requires considerable patience at present.

Future work also includes exploring a multiscale version of the proposed filter and using the technique for texture and edge enhancement as well as abstraction. Finally, we would like to consider other distance functions so as to extend the range of stylization effects achievable within the geodesic framework.

Acknowledgements

Thanks to Hua Li and Herb Yang for suggestions as this work was being developed. Also thanks to the photographers who contributed images through Flickr: jocelyn (stranger), Ian Carroll (cliff), Paul Reynolds (kitten), Paulo Osorio (protest), Ben Sollis (autumn), Al_HikesAZ (ranch, Sedona), Autumn (country road), Nick_T (ruins), blmiers2 (autumn-orange). This work was supported by NSERC and by Carleton University.

References

[1] Haeberli, P.. Paint by numbers: abstract image representations. In: Proceedings of SIGGRAPH. SIGGRAPH '90; ACM; 1990, p. 207–214.

[2] DeCarlo, D., Santella, A.. Stylization and abstraction of photographs. ACM Trans Graph 2002;21:769–776.

[3] Bousseau, A., Kaplan, M., Thollot, J., Sillion, F.. Interactive watercolor rendering with temporal coherence and abstraction. In: International Symposium on Non-Photorealistic Animation and Rendering (NPAR). ACM; 2006, p. 141–149.

[4] Bousseau, A., Neyret, F., Thollot, J., Salesin, D.. Video watercolorization using bidirectional texture advection. ACM Transaction on Graphics (Proceedings of SIGGRAPH 2007) 2007;26(3).

[5] Zhao, M., Zhu, S.C.. Sisley the abstract painter. In: Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering. NPAR '10; ACM; 2010, p. 99–107.

[6] Mould, D.. Texture-preserving abstraction. In: International Symposium on Non-Photorealistic Animation and Rendering (NPAR). 2012, p. 75–82.

[7] Tomasi, C., Manduchi, R.. Bilateral filtering for gray and color images. In: Proceedings of the Sixth International Conference on Computer Vision. ICCV '98; Washington, DC, USA: IEEE Computer Society. ISBN 81-7319-221-9; 1998, p. 839–846. URL http://dl.acm.org/citation.cfm?id=938978.939190.

[8] Criminisi, A., Sharp, T., Rother, C., Pérez, P.. Geodesic image and video editing. ACM Trans Graph 2010;29(5):134:1–134:15. doi:10.1145/1857907.1857910. URL http://doi.acm.org/10.1145/1857907.1857910.

[9] Bai, X., Sapiro, G.. A geodesic framework for fast interactive image and video segmentation and matting. In: ICCV. 2007, p. 1–8.

[10] Winnemöller, H., Olsen, S.C., Gooch, B.. Real-time video abstraction. ACM Trans Graph 2006;25:1221–1226.

[11] Gastal, E.S.L., Oliveira, M.M.. Domain transform for edge-aware image and video processing. ACM Trans Graph 2011;30(4):69:1–69:12. doi:10.1145/2010324.1964964. URL http://doi.acm.org/10.1145/2010324.1964964.

[12] Orzan, A., Bousseau, A., Barla, P., Thollot, J.. Structure-preserving manipulation of photographs. In: Proceedings of the Symposium on Non-photorealistic Animation and Rendering. NPAR '07; ACM; 2007, p. 103–110.

[13] Kang, H., Lee, S., Chui, C.K.. Flow-based image abstraction. IEEE Trans Vis Comput Graph 2009;15(1):62–76.

[14] Kass, M., Solomon, J.. Smoothed local histogram filters. ACM Trans Graph 2010;29.

[15] Bhat, P., Zitnick, C.L., Cohen, M., Curless, B.. Gradientshop: A gradient-domain optimization framework for image and video filtering. ACM Trans Graph 2010;29(2):10:1–10:14. doi:10.1145/1731047.1731048. URL http://doi.acm.org/10.1145/1731047.1731048.

[16] Olsen, S., Gooch, B.. Image simplification and vectorization. In: Proceedings of the Symposium on Non-Photorealistic Animation and Rendering. NPAR '11; ACM. ISBN 978-1-4503-0907-3; 2011, p. 65–74. doi:10.1145/2024676.2024687. URL http://doi.acm.org/10.1145/2024676.2024687.

[17] Papari, G., Petkov, N., Campisi, P.. Artistic edge and corner enhancing smoothing. IEEE Transactions on Image Processing 2007;16(10):2449–2662.

[18] Kyprianidis, J.E., Kang, H., Döllner, J.. Image and video abstraction by anisotropic Kuwahara filtering. Computer Graphics Forum 2009;28(7):1955–1963. Special issue on Pacific Graphics 2009.

[19] Kyprianidis, J.E., Kang, H.. Image and video abstraction by coherence-enhancing filtering. Computer Graphics Forum 2011;30(2):593–602. Proceedings of Eurographics 2011.

[20] Kyprianidis, J.E.. Image and video abstraction by multi-scale anisotropic Kuwahara filtering. In: Proceedings of the Symposium on Non-Photorealistic Animation and Rendering. NPAR '11; ACM. ISBN 978-1-4503-0907-3; 2011, p. 55–64. doi:10.1145/2024676.2024686. URL http://doi.acm.org/10.1145/2024676.2024686.

[21] Arthus-Bertrand, Y.. Forests and people. 2011. URL http://offorestsandmen.org.