A Nonlinear Scale-Space Filter by Physical Computation

Yiu-fai Wong*
Institute for Scientific Computing Research, L-426
Lawrence Livermore National Laboratory
Livermore, CA 94551
E-mail: wong@redhook.llnl.gov, Tel: (510) 422-3777

Abstract— Using maximum entropy principle and statistical mechanics, we derive and demonstrate a nonlinear scale-space filter. For each datum in a signal, a neighborhood of weighted data is used for scale-space clustering. The cluster center becomes the filter output. The filter is governed by a single scale parameter which dictates the spatial extent of nearby data used for clustering. This, together with the local characteristic of the signal, determine the scale parameter in the output space, which dictates the influences of these data on the output. This filter is thus completely unsupervised and data-driven. It provides a mechanism for a) removing noise; b) preserving edges and c) improved smoothing of nonimpulsive noise. This filter presents a new mechanism for detecting discontinuities differing from techniques based on local gradients and line processes. We demonstrate the filter using real images. This work shows that scale-space filtering, nonlinear filtering and scale-space clustering are closely related and provides a framework within which further image processing, image coding and computer vision problems can be investigated.

1 INTRODUCTION

One problem in image and speech processing is to filter out noise in the data while preserving the edges. This problem is also important in multiresolution systems because as resolution decreases, some information is necessarily classified as "noise" and discarded or relegated to the higher resolution description. If one desires the low-resolution description to resemble the original data, it is important to preserve as many edges as possible because edges carry

*Work performed while Wong was at Caltech and was supported in part by a grant from Pacific Bell to Caltech and the Caltech Communications Consortium.

so much information. From the coding point of view, do we need to represent edges by a set of linear components and then re-assemble them at the receiver end?

Without any *a priori* knowledge, however, it is not clear what constitutes noise or an edge. Intuitively, at a very fine scale, every pixel is a signal itself. At a very coarse scale, a nonlinear averaging effect comes into play. Thus, a sensible solution requires one to specify a level of description or scale. That is, one can only distinguish noise versus signal in the scale space.

This work presents a new approach to scale-space filtering. Preliminary results are reported in [1]. We derive a nonlinear clustering filter based on information theory and statistical mechanics. This filter is governed by a single scale parameter and has two additional features: 1) noise reduction and improved smoothing of nonimpulsive noise; and 2) edge preservation. These are achieved through saddle-node dynamics, which presents a new mechanism for edge preservation distinct from local gradients and line processes [2].

This paper is organized as follows: We first discuss related previous approaches, followed by a rigorous derivation of the clustering filter. We then discuss the scale parameters and fine details about the filter. Demonstrations using real data are then presented. This interdisciplinary approach provides a vastly different scale-space filtering which may find applications in vision problems such as edge detection.

2 RELATED PREVIOUS WORK

Gaussian scale-space filtering was introduced in [3] and has generated substantial interest in the computer vision community [4, 5]. But it suffers the major difficulty of blurring across the edges. This drawback is not surprising because the scale parameter is, in fact, a blurring parameter. Clearly, one can treat the Gaussian as the point spread function of the sensors used to detect the input signals. However, from the viewpoint of the human visual system, an image at a coarse scale, though blurred and missing the fine details, still have well-defined boundaries and edges.

To prevent blurring across the edges, anisotropic diffusion [6] and its variants were introduced such that diffusion is very slow across adjacent pixels with large gradients. There are at least three difficulties with these methods: 1) Impulsive noise will have large gradients in all directions and hence will not be removed; 2) the schemes tend to produce piece-wise constant surfaces and thus the images appear blocky; 3) It is not clear how scale is defined.

Median filtering was first proposed by Tukey [7] for smoothing data. It has been used for speech processing and image restoration. Its generalized versions, order-statistic (OS) filters [8] and variations have also been widely studied and applied. However, it is known that median or OS filters do not provide sufficient smoothing for nonimpulsive noise. Another recent approach was the line processes introduced in [2]. A line process is a binary variable which signals the presence of a discontinuity and inhibits the interaction between the neighboring pixels. This idea has been applied to both segmentation and data smoothing in the form of weak membrane models [9].

3 A CLUSTERING FILTER

Seeing things in scale space inevitably involves the removal of signal which could include noise. Once a scale is selected, the task then is to estimate what one sees, i.e. the output. Presumably, the output should be a good representative of the original data. This process can be cast as robust estimation. A cluster center is naturally a robust estimate of the data associated with it. To give a robust estimate, one can either exclude the outliers from the cluster membership or assigning them negligible weights in the final estimate. However, it is known that clustering depends on scale too. One can either treat every point as a cluster itself at a very fine scale or the whole data set as a single cluster at a very coarse scale. Hence, scale-space clustering and scale-space filtering are naturally related.

We now present the essential ideas of a nonlinear scale-space filter [1]. For the sake of exposition and clarity, we are only dealing with images in this paper although this filter is applicable to any kind of spatial or temporal data with continuous attributes.

Let x be the coordinate of a pixel in an image and y its real-valued attribute. Since the pixels are highly redundant due to spatial correlation, we can in principle estimate the true pixel value y at position x by its neighboring pixels. Common sense tells us that the data points near (x, y) should give more information while the data points far away should give less. To make this estimate robust, the information should be spread among the neighboring data. If we treat the contributions to the determination of y from the neighboring data as a probability distribution, this means that the entropy of this probability distribution should be maximized subject to linear cost constraints.

Say we are given a neighborhood of data points $S = \{(x_i, y_i) : i = 1, \ldots, N\}$. Let P_i denote the contribution of (x_i, y_i) to (x, y). To specify the cost function, we note that the input and output domains should be treated differently because the former is fixed and known but the latter is really environment-dependent. Thus, let the cost function have two components $e_x(x_i)$ and $e_y(y_i)$. A straightforward application of the maximum entropy principle and Lagrangian multiplier gives

$$P_i = e^{-\alpha e_x(x_i) - \beta e_y(y_i)}/Z \qquad (1)$$

where $Z = \sum_i e^{-\alpha e_x(x_i) - \beta e_y(y_i)}$. To make the connection with statistical mechanics, we call $\beta = 1/T$ the "inverse temperature" and Z the "partition function". Naturally, the "free energy" is defined as

$$F = -\frac{1}{\beta} \log Z. \tag{2}$$

Let us choose the cost functions $e_x(x_i) = (x - x_i)^2$ and $e_y(y_i) = (y - y_i)^2$. At equilibrium, a thermodynamical system always settles into a state which minimizes its free energy. That is, we want $\partial F / \partial y = 0$, or equivalently

$$y = \sum_i \frac{y_i w_i e^{-\beta(y_i - y)^2}}{\sum_j w_j e^{-\beta(y_j - y)^2}}, \tag{3}$$

the weighted mean of the data, where $w_i = e^{-\alpha(x_i - x)^2}$. One can easily extend the above analysis to data with multiple attributes.

We note that the general framework of using statistical mechanics for optimization has been investigated by various researchers [10, 2, 11].

3.1 TWO SCALES

From [12, 13], we can interpret y as a cluster center given the data y_is and w_i as either the *weight* given to y_i or the *relative multiplicity* of y_i. Note that the data used for clustering differ from pixel to pixel because the w_is differ.

Let us now explain what α means. When $\alpha = \infty$, $w_i = 1$ if $x_i = x$ and 0 otherwise. Every pixel is preserved perfectly. Conversely, a small α implies that more neighbors of x can contribute. Hence, α is a measure of scale in the input space.

Given a scale α in the input space, β is identical to the scale parameter in the scale-space clustering algorithm [12, 13]. When $\beta = \infty$, every datum y_i with nonzero w_i is a solution of (3). As β decreases, the number of solution decreases. Eventually when $\beta = 0$, there is only one solution. Hence, β is a measure of scale in the output space.

Given a fixed α and allow β to vary from a large to a small value, there are critical temperatures β_c's at which the fixed points of (3) become unstable and bifurcate. This results in the merging of the clusters. The merging of clusters as β varies can be rigorously analyzed using local bifurcation techniques [14]. It may thus be helpful to summarize the relevant results in [12] regarding the dynamics of (3) here.

1. Given any initial y, the iterative map always converges.
2. There are two types of local bifurcations for (3) as β varies: saddle-node and pitchfork bifurcations.
3. A saddle-node indicates an inhomogeneous spatial distribution in the data, while a pitchfork bifurcation indicates uniformly spaced or clustered data but with a high degree of symmetry with respect to scale.

The particular estimate y one obtains depends on the initial y and β. Intuitively, β should be related to the variations of the pixel intensities in the spatial window defined by the scale α in the input domain. In fact, the variation in the allowed output space (pixel intensity) should be inversely proportional to β. Therefore, a natural method to initialize y and choose β is as follows:

Let $\bar{y} = \sum_i y_i w_i / \sum_i w_i$ and $\sigma_y^2 = \sum_i (y_i - \bar{y})^2 w_i / \sum_i w_i$. Intuitively, $\bar{y}$ is the average of the pixel values in the neighborhood of x, while σ_y^2 is a measure of its local variance. A small σ_y^2 implies that the local region is quite smooth, while a large σ_y^2 implies that the local region may contain outliers or an edge. Hence for every pixel location, one can choose $\beta = (2\sigma_y^2)^{-1}$. The initial y can be set to $\bar{y}$. The converged value is the filter output.

Next, we demonstrate some properties of this filter on real data.

4 DEMONSTRATIONS

Example 1. The circles in Figure 1 show a "hat" of height 0.5, corrupted by a Gaussian noise of variance 0.001 and two very noisy pieces of data. The solid line shows the clustering filtered output with $\alpha = 1/32$. One can see that edges are well preserved and the non-edge portions are very smooth. The dashed line shows the output when a median filter with window of size 9 is used. The dot-dashed line is the Gaussian filtered output. The superiority of our method over both median filtering and Gaussian filtering is clearly demonstrated.

Figure 2a shows a series of filtered outputs for $\alpha = 1/2, 1/8, 1/32$, and $1/128$ respectively. The data is shown on the lowest plot. Figure 2b shows the outputs when Gaussians with identical α's are used. We note two differences: 1) Gaussian filtering simply smoothes across the data while the clustering filter can preserve the "bumps" much better; 2) Gaussian filtering also reduces the contrast much faster than the cluster filtering.

Real Data: Figure 3a shows the famous Lena image. Figure 3b shows the image obtained by the clustering filter with $\alpha = 0.125$ (standard deviation equals 2 pixels). Details of smaller scale are averaged, but the edges of larger scale are preserved very well. Figure 3c shows what we see at a scale of $\alpha = (2 \times 4^2)^{-1}$. One see smooth blobs but the edges are sharp. Our results differ vastly from those produced by diffusion schemes, which produce blocky and piece-wise constant images.

The above examples demonstrates that the clustering filter not only preserves edges, it also removes impulsive noise and provides improved smoothing of nonimpulsive noise.

It is important to note that although our formulation starts with two parameters, the clustering filter has, in effect, only one scale parameter. The other scale parameter is determined by the local variations in the data. This data-driven selection of β gives a family of adaptive filters. We remark that

β can be further modified to suit one's needs.

Figure 4a shows a picture of Einstein. Figure 4b shows a difference image by subtracting the filtered image from the original image and then taking the absolute values which are then stretched over 8 bit resolution. The peak signal-to-noise ratio (PSNR) is 0.235 dB, and yet one can recognize Einstein instantly. Thus, SNR does not say much about perception, especially face perception.

Intuitively, a difference image roughly contains the removed fine details in the images. If the clustering filter is indeed a more appropriate scale-space filtering scheme, then the difference should contain the signals of finer scale which are removed/lost as one looks at the image at a scale α. That is, one has to pay great attention to retain these signals. Indeed, Figure 4b shows that such details consist of the boundaries and the small noisy signals. Hence, both the filtered and the difference images generated by the clustering filter contain tremendous amount of information about the original image. This decomposition is achieved through a nonlinear mechanism.

The crucial difference between Gaussian filtering and our scheme is the nonlinear feed-back term $e^{-\beta(y_i-y)^2}$ in (3). There are intuitive arguments why scale space filtering is neither associative nor commutative. For example, suppose we quantize an interval into 5 levels and then further quantize the 5 levels into 2 levels. A signal passing through this two-stage quantizer differs from its passing through a one-stage quantizer with 2 levels. In scale-space, a "hard" decision has to be made. Moreover, the anisotropy is handled by nonlinear dynamics though the cost function is isotropic. The above comparison shows that the clustering filter constitutes a new scale-space filter.

5 SUMMARY

We have derived and demonstrated a nonlinear clustering filter based on information theory and statistical mechanics. The filter has only one parameter, which governs the size of the spatial or temporal neighborhood of data that can influence the filter output. This, together with local characteristics of the signal, determine the scale parameter in the output space. The result is adaptive filtering of the data in a completely unsupervised manner. The relation of this filter to scale-space clustering has been emphasized. The filter is shown to be able to remove noise, preserve edges and smooth data. It also presents a new mechanism to preserve discontinuities in the data.

Comparisons with Gaussian filtering and median filtering demonstrated that this clustering filter gives better performance than the other techniques. This filter may be a more appropriate approach to scale-space filtering. This work shows that scale-space filtering, nonlinear filtering and scale-space clustering are closely related and provides a framework within which further image processing, image coding and computer vision problems can be investigated. This filter has been applied to noise filtering [15] and image compression [16].

6 Acknowledgement

I am indebted to Edward C. Posner for his encouragements for this research. I thank Carver Mead for teaching me the magic of exponentials.

References

[1] Yiu-fai Wong, "A Clustering Filter for Scale-space Filtering and Image Restoration", in *Proc. IEEE Conf. Comp. Vis. and Pattern Recog.*, 1993.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. and Mach. Intell.*, 6, 721–741, 1984.

[3] A. Witkin, "Scale-space filtering," in *Int. Joint Conf. Artificial Intelligence*, Karlsruhe, West Germany, 1983, 1019–1021.

[4] Jan J. Koenderink, "The Structure of Images," *Biol. Cybern.*, 50, 363-370, 1984.

[5] A. Yuille and T. Poggio, "Scaling theorems for zero crossings," *IEEE Trans. Patt. Anal. and Mach. Intell.*, 8, 15-25, 1986.

[6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Patt. Anal. and Mach. Intell.*, 12, 629-639, 1990.

[7] J.W. Tukey, *Exploratory Data Analysis*, Reading, Addison-Wesley, MA, 1977.

[8] A.C. Bovik, T.S. Huang and D.C. Munson, Jr., "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, 31, 1342-1350, 1983.

[9] A. Blake and A. Zisserman, *Visual Reconstruction.* MIT Press, Cambridge, MA, 1987.

[10] J.J. Hopfield and D.W. Tank "Neural Computation of Decisions in Optimization Problems," *Biol. Cybern.*, 52, 141-152, 1986.

[11] R. Durbin and D. Willshaw, "An analogue approach to the traveling salesman problem using an elastic net method," *Nature*, 326, 689-691, 1987.

[12] Yiu-fai Wong, *Towards a Simple and Fast Learning and Classification System*, Ph.D Thesis, Caltech, Electrical Engineering, January, 1992.

[13] Yiu-fai Wong, "Clustering Data by Melting," *Neural Computation*, vol. 5, 89-104, 1993.

[14] S. Wiggins, *Introductions to Applied Nonlinear Dynamical Systems and Chaos,* Springer-Verlag, New York, 1990.

[15] Yiu-fai Wong, "Facial Lift by Clustering Filter," submitted, 1993.

[16] Yiu-fai Wong and Ming-Chieh Lee, "Image Compression by A Nonlinear Multiresolution System," accepted for *IEEE Workshop on Visual Signal Processing and Communications*, 1993.

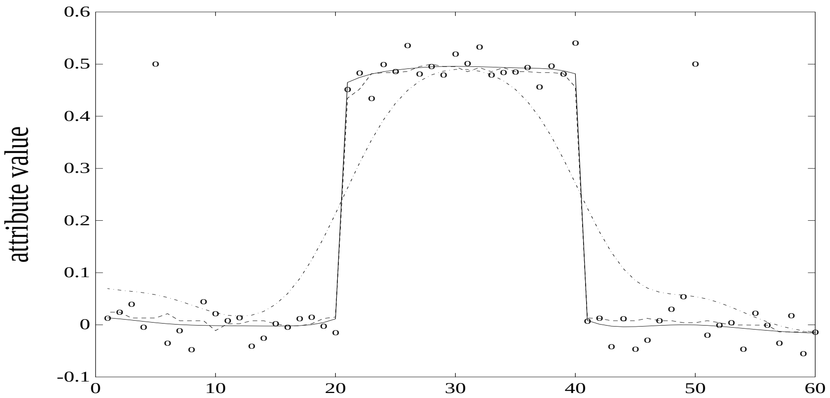


Figure 1. A corrupted hat and various filtered outputs.

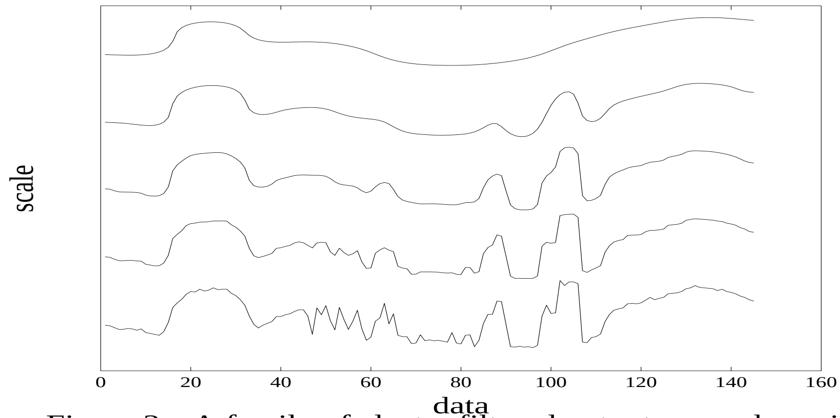


Figure 2a. A family of cluster filtered outputs as scale varies.

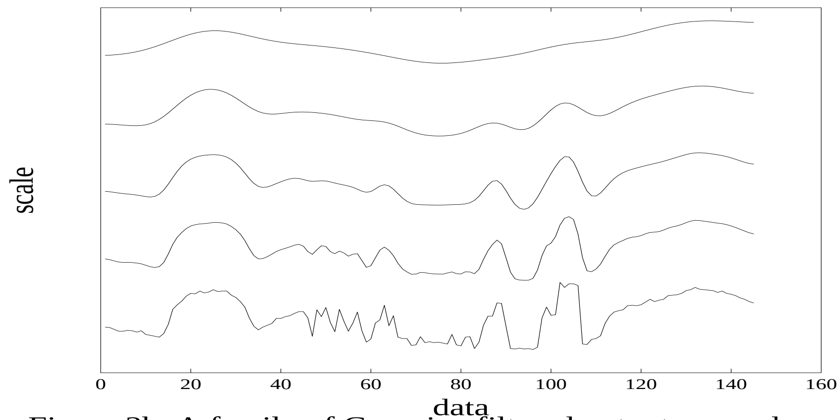


Figure 2b. A family of Gaussian filtered outputs as scale varies.

Figure 3a. Lena Image

Figure 3b. Cluster filtered image of Fig. 3a.

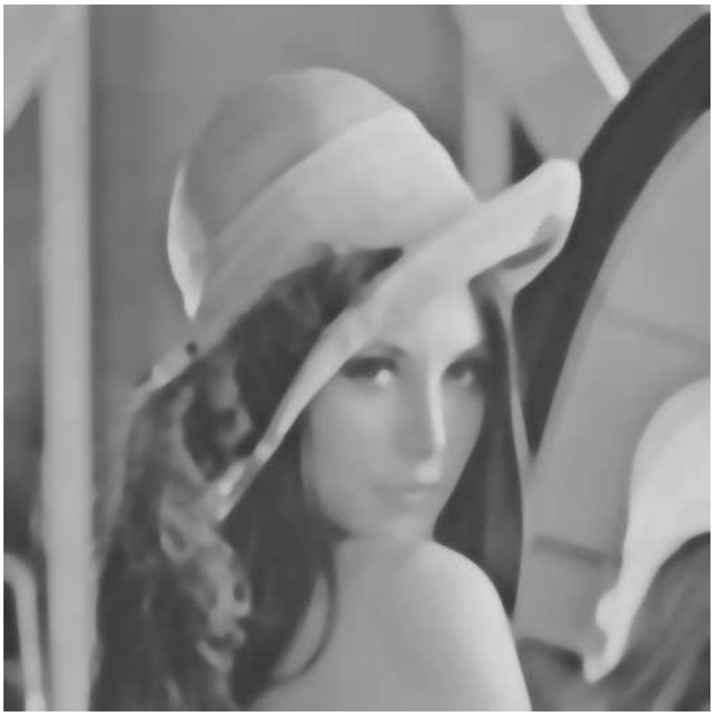
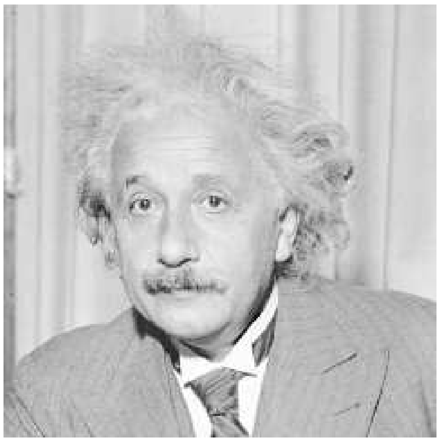


Figure 3c. Cluster filtered image at a coarser scale.



a b

Figure 4. (a) Original Einstein image.
(b) Difference image for cluster filtered image.