Edges as Outliers: Anisotropic Smoothing using Local Image Statistics

Michael J. Black[1] and Guillermo Sapiro[2]

[1] Xerox Palo Alto Research Center,
3333 Coyote Hill Road, Palo Alto, CA 94304 USA.
black@parc.xerox.com
[2] Dept. of Electrical and Computer Eng., University of Minnesota
200 Union Street SE, Minneapolis, MN 55455.
guille@ece.umn.edu

Abstract. Edges are viewed as statistical outliers with respect to local image gradient magnitudes. Within local image regions we compute a robust statistical measure of the gradient variation and use this in an anisotropic diffusion framework to determine a spatially varying "edge-stopping" parameter σ. We show how to determine this parameter for two edge-stopping functions described in the literature (Perona-Malik and Tukey). Smoothing of the image is related the local texture and in regions of low texture, small gradient values may be treated as edges whereas in regions of high texture, large gradient magnitudes are necessary before an edge is preserved. Intuitively these results have similarities with human perceptual phenomena such as masking and "popout". Results are shown on a variety of standard images.

1 Introduction

Anisotropic diffusion has been widely used for "edge-preserving" smoothing of images. Little attention, however, has been paid to defining exactly what is meant by an "edge." In the traditional formulation of Perona and Malik [8], edges are related to pixels with large gradient magnitudes and an anisotropic smoothing function is one that inhibits smoothing across such boundaries. The effect of this smoothing is determined by some parameter, σ, which implicitly defines what is meant by an edge. This paper address how this σ parameter can be determined automatically from the image data in such a way that edges correspond to statistical outliers with respect to local image gradients. With this method, σ varies across the image and hence, what is considered to be an edge is dependent on local statistical properties of the image.

Consider for example the image in Figure 1. Regions **A** and **B** illustrate areas where there is little gradient variation and the fairly small gradient magnitudes of the features are locally significant. Intuitively, we would say that the eyebrow and the shoulder crease are significant image structures. In contrast, region **C** is

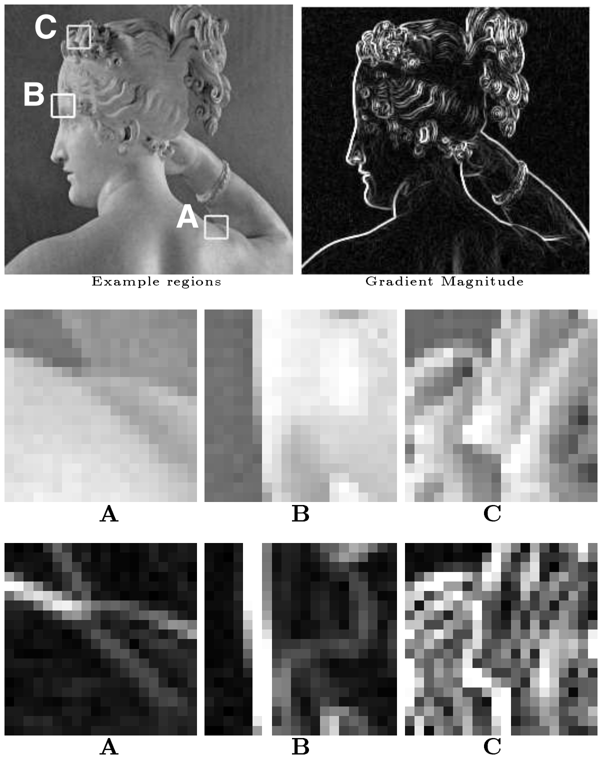


Fig. 1. Consider the image regions (**A, B, C**) in the upper left image. The middle row shows each image region in detail while the bottom row shows the gradient magnitude for each region. The faint image structures in regions **A** and **B** are statistically significant with respect the variation of intensity within the regions. The same variation in the highly textured region **C** would not be statistically significant due to the much larger image variation.

highly textured and there is a great deal of variation in the gradient magnitudes. Intuitively, in this region, the gradient magnitudes of features like those in regions **A** and **B** might be considered insignificant. To be considered an edge in region **C** we would like the gradient magnitude to be much larger.

Here we adopt the robust statistical interpretation of anisotropic diffusion elaborated in [1]. Anisotropic diffusion is viewed as a robust statistical procedure that estimates a piecewise smooth image from noisy input data. This work formalized the relationship between the "edge-stopping" function in the anisotropic diffusion equation and the error norm and influence function in a robust estimation framework. This robust statistical interpretation provides a principled means for defining and detecting the boundaries (edges) between the piecewise smooth regions in an image that has been smoothed with anisotropic diffusion. Edges are considered statistical outliers in this framework.

The robust statistical approach also provides a framework to locally define edges and stopping functions, as demonstrated in this paper (see [11] for a different approach to spatially adaptive anisotropic diffusion). In particular, the σ parameter in the edge stopping function has also a statistical interpretation. This statistical interpretation gives, among other properties, a completely automatic diffusion algorithm, since all the parameters are computed from the image. Our approach is to compute a statistically robust *local* measure of the brightness variation within image regions. From this we obtain a local definition of edges and a space-variant edge stopping function.

2 Review

We briefly review the traditional anisotropic diffusion formulation as presented by Perona and Malik [8].

2.1 Anisotropic diffusion: Perona-Malik formulation

Diffusion algorithms smooth images via a partial differential equation (PDE). For example, consider applying the isotropic diffusion equation (the heat equation) given by $\frac{\partial I(x,y,t)}{\partial t} = \mathrm{div}(\nabla I)$, using the original (degraded/noisy) image $I(x,y,0)$ as the initial condition, where $I(x,y,0) : \mathbb{R}^2 \to \mathbb{R}^+$ is an image in the continuous domain, (x,y) specifies spatial position, t is an artificial time parameter, and where ∇I is the image gradient. Modifying the image according to this isotropic diffusion equation is equivalent to filtering the image with a Gaussian filter.

Perona and Malik [8] replace the classical isotropic diffusion equation with

$$\frac{\partial I(x,y,t)}{\partial t} = \mathrm{div}(g(\| \nabla I \|, \sigma)\nabla I), \tag{1}$$

where $\| \nabla I \|$ is the gradient magnitude, and $g(\| \nabla I \|)$ is an "edge-stopping" function and σ is a scale parameter. This function is chosen to satisfy $g(x,\sigma) \to 0$ when $x \to \infty$ so that the diffusion is "stopped" across edges.

2.2 Perona-Malik discrete formulation

Perona and Malik discretized their anisotropic diffusion equation as follows:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(\nabla I_{s,p}, \sigma) \nabla I_{s,p}, \tag{2}$$

where I_s^t is a discretely-sampled image, s denotes the pixel position in a discrete, two-dimensional grid, and t now denotes discrete time steps (iterations). The constant $\lambda \in \mathbb{R}^+$ is a scalar that determines the rate of diffusion, η_s represents the spatial neighborhood of pixel s, and $|\eta_s|$ is the number of neighbors (usually

4, except at the image boundaries). Perona and Malik linearly approximated the image gradient (magnitude) in a particular direction as

$$\nabla I_{s,p} = I_p - I_s^t, \quad p \in \eta_s \tag{3}$$

Qualitatively, the effect of anisotropic diffusion is to smooth the original image while preserving brightness discontinuities. The choice of $g(x, \sigma)$ and the value of σ can greatly affect the extent to which discontinuities are preserved.

2.3 Related Work

In related work, a number of authors have explored the estimation of the scale at which to estimate edges in images [2, 5]. These methods find the optimal local scale for detecting edges with Gaussian filters; they do not explicitly use local image statistics. The approach described here might be augmented using these ideas to determine the size of the local area within which to compute image statistics.

Marimont and Rubner [6] computed local statistics of zero-crossings and used these to define the probability of a pixel belonging to an edge. Liang and Wang [4] also used the statistics of zero-crossings to set a local noise measure in an anisotropic diffusion formulation.

In contrast, the work here provides a robust statistical view which allows a principled choice of both the g-function and the scale parameter. Related to this is work on human perception that models feature saliency using a statistical test for outliers [9].

3 Robust Statistical View

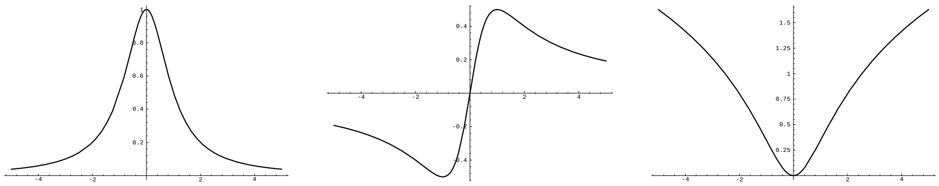
For the majority of pixels in Figure 1 **A**, the image gradient values can be approximately modeled as being constant (zero) with random Gaussian noise. The large gradient values due to the image feature however are statistical "outliers" [3] with respect to the Gaussian distribution; the distribution of these outliers is unknown. We seek a function $g(x, \sigma)$ and a scale parameter σ that will appropriately smooth the image when the variation in the gradient is roughly Gaussian and will inhibit smoothing when the gradient can be viewed as an outlier.

First we need to relate the form of the g-functions used for anisotropic diffusion to the tools used in robust statistics (see [1] for details). From a robust statistical perspective the goal of anisotropic smoothing is to iteratively find an image I that satisfies the following optimization criterion:

$$\min_I \sum_{s \in I} \sum_{p \in \eta_s} \rho(I_p - I_s, \sigma) \tag{4}$$

where $\rho(\cdot)$ is a robust error function and σ is a "scale" parameter.

In this formulation large image differences $|I_p - I_s|$ are assumed to be outliers which should not have a large effect on the solution. To analyze the behavior of



$g(x, \sigma)$ $\qquad$ $g(x, \sigma)x = \psi(x, \sigma)$ $\qquad$ $\rho(x, \sigma)$

Fig. 2. Lorentzian error norm and the Perona-Malik g stopping function.

a given ρ-function with respect to outliers, we consider its derivative (denoted ψ), which is proportional to its *influence function* [3]. This function characterizes the bias that a particular measurement has on the solution and by analyzing the shape of this function we can infer the behavior of a particular robust ρ-function with respect to outliers.

In [1] (see also [12] for a related approach) it was shown that

$$g(x, \sigma)x = \psi(x, \sigma) = \rho'(x, \sigma). \tag{5}$$

This relationship means that we can analyze the behavior of a particular anisotropic edge-stopping function g in terms of its outlier rejection properties by examining the influence function ψ.

For example, consider the edge-stopping function proposed by Perona and Malik [8]

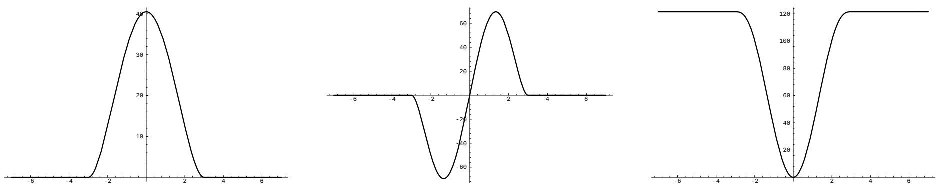
$$g(x, \sigma)x = \frac{2x}{2 + \frac{x^2}{\sigma^2}} = \psi(x, \sigma), \tag{6}$$

where $\psi(x, \sigma) = \rho'(x, \sigma)$. We can compute ρ by integrating $g(x, \sigma)x$ with respect to x to derive

$$\int g(x, \sigma)x \; dx = \sigma^2 \log\left(1 + \frac{1}{2}\left(\frac{x^2}{\sigma^2}\right)\right) = \rho(x, \sigma). \tag{7}$$

This function $\rho(x, \sigma)$ is proportional to the Lorentzian error norm use in robust statistics and $g(x)x = \rho'(x) = \psi(x)$ is proportional to the influence function (Figure 2).

The function $g(x, \sigma)$ acts as a "weight" and from the plot in Figure 2 we can see that small values of x (i.e. small gradient magnitudes) will receive high weight. As we move out to the tails of this function it flattens out and the weight assigned to some large x will be roughly the same as the weight assigned to some nearby $x + \epsilon$. This behavior is visible in the shape of the ψ-function which reaches a peak and then begins to descend. Outlying values of x beyond a point receive roughly equivalent weights and hence there is little preference for one outlying value over another. In this sense outliers have little "influence" on the solution. In the anisotropic diffusion context, $g(x, \sigma)x$ will be relatively small for outliers and, hence, each iteration in (2) will produce only a small change in the image.



$g(x,\sigma)$ $\qquad$ $g(x,\sigma)x = \psi(x,\sigma)$ $\qquad$ $\rho(x,\sigma)$

Fig. 3. Tukey's biweight.

In [1] a more robust edge stopping function was derived from Tukey's biweight ρ-error:

$$\rho(x,\sigma) = \begin{cases} \frac{x^2}{\sigma^2} - \frac{x^4}{\sigma^4} + \frac{x^6}{3\sigma^6} & |x| \leq \sigma \\ \frac{1}{3} & \text{otherwise} \end{cases} \tag{8}$$

$$\psi(x,\sigma) = \begin{cases} x(1-(x/\sigma)^2)^2 & |x| \leq \sigma, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$g(x,\sigma) = \begin{cases} \frac{1}{2}(1-(x/\sigma)^2)^2 & |x| \leq \sigma, \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The functions $g(x,\sigma)$, $\psi(x,\sigma)$ and $\rho(x,\sigma)$ are plotted in Figure 3: The influence of outliers drops off more rapidly than with the Lorentzian function and the influence goes to zero after a fixed value (a hard redescending function). These properties result in sharper boundaries than obtained with the Perona-Malik/Lorentzian function [1].

4 Local Measure of Edges

Both functions defined in the previous section reduce the influence of large gradient magnitudes on the smoothed image. The point at which gradient values begin to be treated as outliers is dependent on the parameter σ. In this section we consider how to globally and *locally* compute an estimate of σ directly from the image gradients. The main idea is that σ should characterize the variance of the majority of the data within a region. So, for example in Figure 1 **A**, σ should characterize the amount of variation in the gradients at all locations except where the feature is located. Outliers will then be determined relative to this background variation.

In deriving σ we appeal to tools from robust statistics to automatically estimate the "robust scale," σ_e, of the image as [10]

$$\begin{aligned} \sigma_e &= 1.4826 \,\mathrm{MAD}(\nabla I) \\ &= 1.4826 \,\mathrm{median}_I(\| \nabla I - \mathrm{median}_I(\| \nabla I \|) \|) \end{aligned} \tag{11}$$

where "MAD" denotes the median absolute deviation and the constant is derived from the fact that the MAD of a zero-mean normal distribution with unit

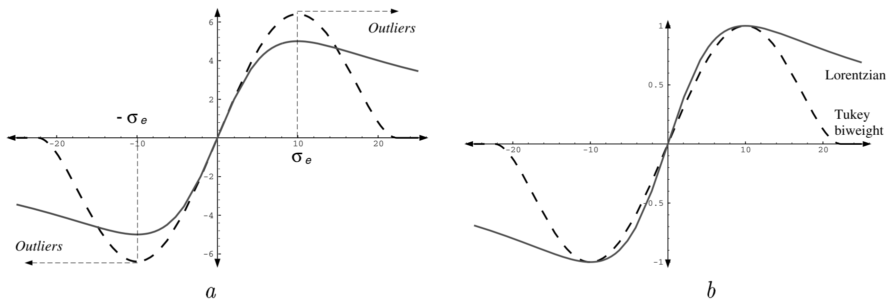


Fig. 4. Lorentzian and Tukey ψ-functions. *(a)* values of σ chosen as a function of σ_e so that outlier "rejection" begins at the same value for each function; *(b)* the functions aligned and scaled.

variance is $0.6745 = 1/1.4826$. We consider σ_e to be the gradient magnitude at which outliers begin to be downweighted.

We choose values for the scale parameters σ to dilate each of the influence functions so that they begin rejecting outliers at the same value: σ_e. The point where the influence of outliers first begins to decrease occurs when the derivative of the ψ-function is zero. For the Lorentzian ρ-function this occurs at $\sigma_e = \sqrt{2}\,\sigma$ and for the Tukey function it occurs at $\sigma_e = \sigma/\sqrt{5}$. Defining σ with respect to σ_e in this way we plot the influence functions for a range of values of x in Figure 4*a*. Note how each function begins reducing the influence of measurements at the same point.

We also scale the influence functions so that they return values in the same range. To do this we take λ in (2) to be one over the value of $\psi(\sigma_e, \sigma)$. The scaled ψ-functions are plotted in Figure 4*b*.

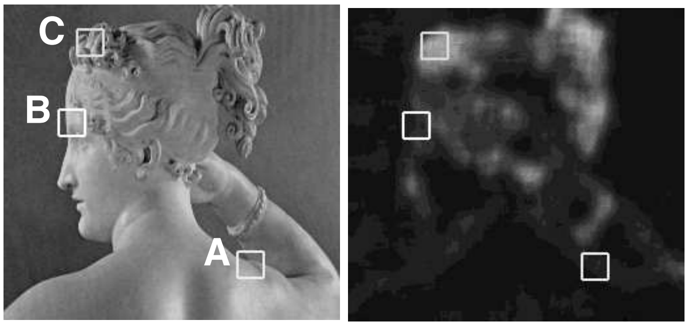
Now we can directly compare the results of anisotropic smoothing with the different edge-stopping functions. The Tukey function gives *zero* weight to outliers whose magnitude is above a certain value while the Lorentzian (or Perona-Malik) downweights outliers but still gives them some weight.

4.1 Spatially Varying σ

In previous work we took the region for computing σ_e to be the entire image. This approach works well when edges are distributed homogeneously across the image but this is rarely the case. Here we explore the computation of this measure in image patches. In particular we consider computing a local scale $\sigma_l(x, y)$, which is a function of spatial position, in $n \times n$ pixel patches at every location in the image. We take this value to be the larger of the σ_e estimated for the entire image and the value in the local patch. Then $\sigma_l(x, y)$ is defined as

$$\sigma_l(x, y) = \max(\sigma_e, 1.4826 \ \mathrm{MAD}_{-\frac{n}{2} \le i,j \le \frac{n}{2}}(\nabla I_{x+i,y+j})). \tag{12}$$

In practice σ_e provides a reasonable lower bound on the overall spatial image variation and the setting of σ_l to be the maximum of the global and local variation prevents the amplification of noise in relatively homogeneous image regions.



a *b*

Fig. 5. Local estimate of scale, $\sigma_l(x, y)$. Bright areas in *(b)* correspond to larger values of σ_l.

Figure 5 shows the results of estimating σ_l in 15×15 pixel patches. Bright areas have higher values of σ_l and correspond to more textured image regions.

To see the effects of the spatially varying σ_l consider the results in Figure 6. The images show the results of applying diffusion using the $g(x, \sigma)$ corresponding to the Tukey biweight function. The top row uses a fixed value of σ_e estimated over the entire image while the bottom row shows the results with a spatially varying σ_l. We can detect edges in the smoothed images very simply by detecting those points that are treated as outliers by the given ρ-function. Figure 6 shows the outliers (edge points) in each of the images, where outliers are given by those points having $|\nabla I_{x,y}| > \sigma_e(x, y)$ (global, first row) or $|\nabla I_{x,y}| > \sigma_l(x, y)$ (local, second row).

In areas containing little texture the results are identical since in these areas the sigma estimated locally is likely to be less than σ_e and hence, σ_l is set to σ_e. The differences become apparent in the textured regions of the image. A detail is shown for a region of hair. With a fixed global σ_e, discontinuities are detected densely in the hair region as the large gradients are considered outliers with respect to the rest of the image which has relatively few large gradients. With the spatially varying σ_l, these regions are smoothed more heavily and only the statistically significant discontinuities remain.

5 Experimental Results

In this section we test the spatially varying smoothing method with both the Lorentzian and Tukey g-functions. Figure 7 compares the results for the Tukey function at 500 iterations and the Lorentzian at 50 iterations. The Lorentzian

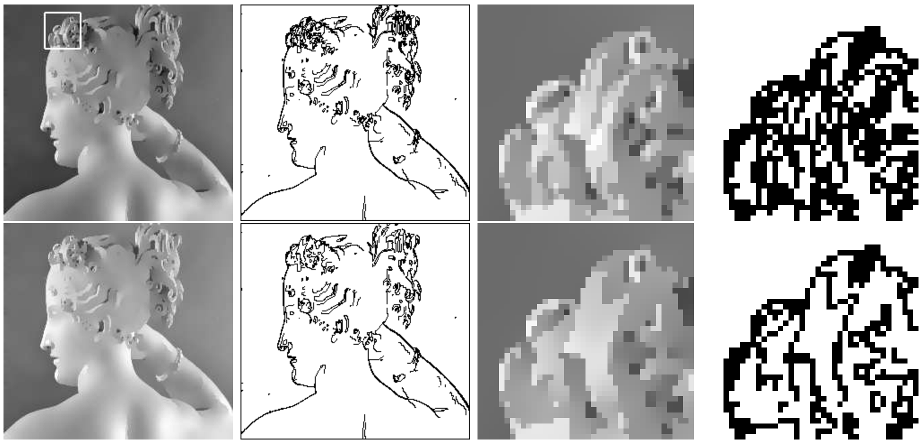


Fig. 6. Anisotropic smoothing with the Tukey function (500 iterations). Top row shows smoothing with a fixed value of σ_e. Bottom row shows a spatially varying σ_l.

must be stopped sooner as, unlike the Tukey function, outliers have a finite influence and hence the image will eventually become oversmoothed. In both cases note that the edges detected in the highly textured regions have a spatial density similar to that of other regions of image structure.

Figure 8 shows a more textured image. Note that the highest scale values correspond to the steps in the lower middle portion of the image. The discontinuities here are smoothed while the boundaries of the people against a relatively uniform background are preserved. One can also see in this image the difference between the Lorentzian and Tukey functions in that the Tukey g-function results in sharper brightness discontinuities.

The Magnetic Resonance image in Figure 9 is challenging because there are areas of high contrast as well as detailed brain structures of very low contrast. No single scale term will suffice for an image such as this. The results with the Tukey function preserve much of the fine detail and the detected edges reveal structure in both the high and low contrast regions.

6 Conclusions

One of the crucial steps in anisotropic diffusion is to define an edge, and from this definition, an edge stopping function. Several attempts have been reported in the literature, mainly dealing with global definitions. In this paper we have addressed the search for a local definition of edges. We have described a simple method for determining a spatially varying scale function based on robust statistical techniques. From this, we have provided a local definition of edges and a space-varying edge stopping function.

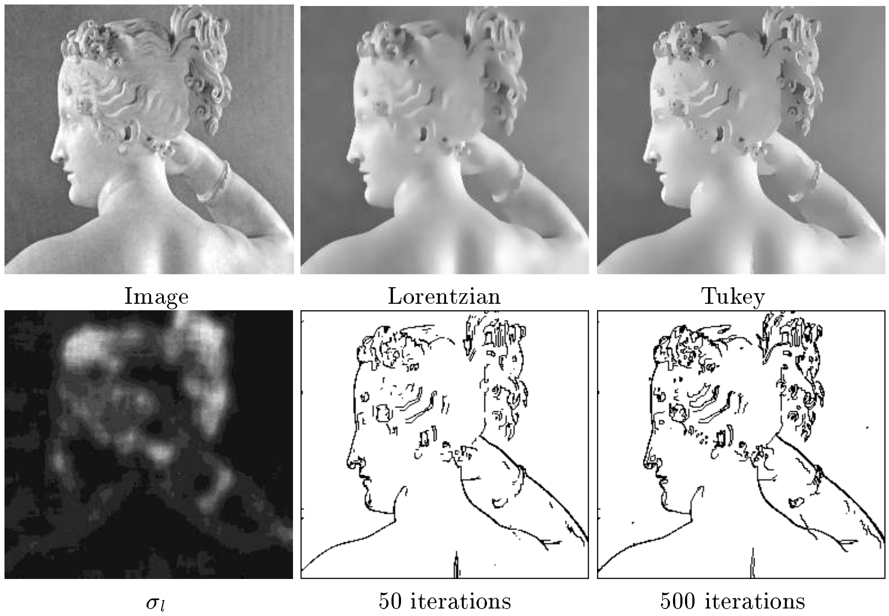


Fig. 7. Results for both the Perona-Malik (Lorentzian) function and the Tukey function.

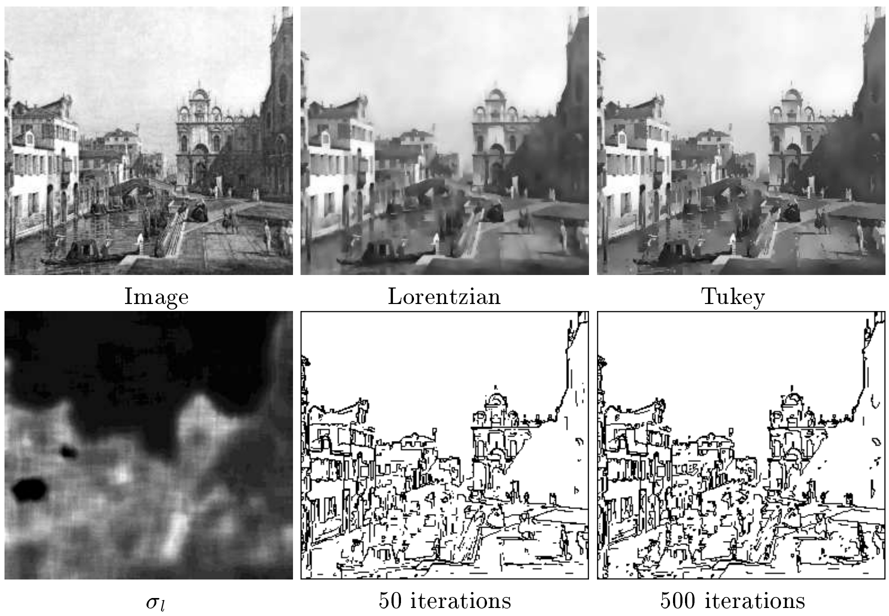


Fig. 8. Results for both the Perona-Malik (Lorentzian) function and the Tukey function.

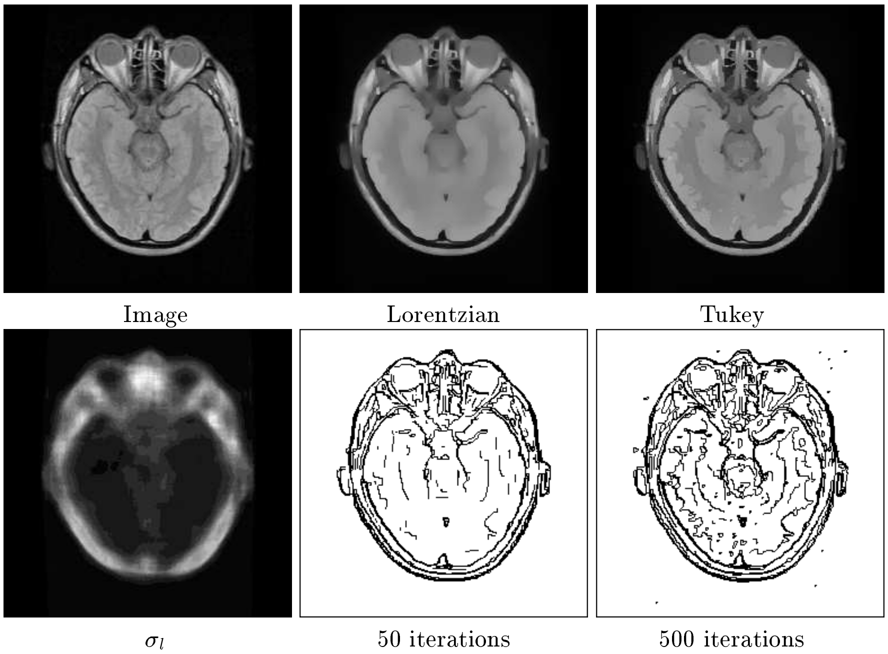


Fig. 9. Magnetic Resonance Image. Results for both the Perona-Malik (Lorentzian) function and the Tukey function.

A number of topics remain open. First, the only parameter left in the proposed anisotropic diffusion algorithm is the size of the window within which σ_l is computed. This also should be space-variant, and needs to be automatically determined from the image itself.

We are interested in comparing the output of our simple local edge detector with others as for example those proposed by Perona [7] or Elder and Zucker [2]. They use much more sophisticated techniques that might not be computationally efficient if the goal is to compute stopping functions for anisotropic diffusion. On the other hand, a more accurate computation of edges might be crucial for anisotropic diffusion applications like enhancing medical images.

Finally, it would be interesting to explore the relationship to human perception of image features which can be effected by the local image statistics [9].

Acknowledgements. GS is partially supported by a grant from the Office of Naval Research ONR-N00014-97-1-0509, the Office of Naval Research Young Investigator Award, the Presidential Early Career Awards for Scientists and Engineers (PECASE), and by the National Science Foundation Learning and Intelligent Systems Program (LIS).

References

1. M. J. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. on Image Processing*, 7(3):421–432, 1998.
2. J. H. Elder and S. W. Zucker. Scale space localization, blur, and contour-based image coding. In *Proc. Computer Vision and Pattern Recognition, CVPR-96*, pages 27–34, San Francisco, June 1996.
3. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York, NY, 1986.
4. P. Liang and Y. F. Wang. Local scale controlled ansiotropic diffusion with local noise estimate for image smoothing and edge detection. In *Proceedings of the International Conference on Computer Vision*, pages 193–200, Mumbai, India, January 1998.
5. T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–154, 1998.
6. D. H. Marimont and Y. Rubner. A probabilistic framework for edge detection and scale selection. In *Proceedings of the International Conference on Computer Vision*, pages 207–214, Mumbai, India, January 1998.
7. P. Perona. Steerable-scalable kernals for edge detection and junction analysis. In G. Sandini, editor, *Proc. of Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS-Series*, pages 3–23. Springer-Verlag, May 1992.
8. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.
9. R. Rosenholtz. General-purpose localization of textured image regions. In *Advances in Neural Information Processing Systems, 11*, 1999.
10. P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection.* John Wiley & Sons, New York, 1987.
11. D. Strong, P. Blomgren, and T. F. Chan. Spatially adaptive local feature-driven total variation minimizing image restoration. Technical Report 97-32, UCLA-CAM Report, July 1997.
12. Y. L. You, W. Xu, A. Tannenbaum, and M. Kaveh. Behavioral analysis of anisotropic diffusion in image processing. *IEEE Trans. Image Processing*, 5:1539–1553, 1996.