ROBUST ANISOTROPIC DIFFUSION AND SHARPENING OF SCALAR AND VECTOR IMAGES

*Michael Black**, *Guillermo Sapiro***, *David Marimont**, *and David Heeger****

*Xerox PARC, **U. of Minnesota, ***Stanford U.
{black,marimont}@xerox.com, guille@ee.umn.edu, heeger@white.stanford.edu

ABSTRACT

Relations between anisotropic diffusion and robust statistics are described in this paper. We show that anisotropic diffusion can be seen as a robust estimation procedure that estimates a piecewise smooth image from a noisy input image. The "edge-stopping" function in the anisotropic diffusion equation is closely related to the error norm and influence function in the robust estimation framework. This connection leads to a new "edge-stopping" function based on *Tukey's biweight* robust estimator, that preserves sharper boundaries than previous formulations and improves the automatic stopping of the diffusion. The robust statistical interpretation also provides a means for detecting the boundaries (edges) between the piecewise smooth regions in the image. We extend the framework to vector-valued images and show applications to robust image sharpening.

Key words: Anisotropic diffusion, robust statistics, Riemannian geometry, color and texture images.

1. INTRODUCTION

Since the elegant formulation of anisotropic diffusion introduced by Perona and Malik [9], a considerable amount of research has been devoted to the theoretical and practical understanding of this and related methods for image enhancement. See [1, 2, 10] and references therein. In this paper we develop a statistical interpretation of anisotropic diffusion, specifically, from the point of view of robust statistics. We show that the Perona-Malik [9] diffusion equation is equivalent to a robust procedure that estimates a piecewise constant image from a noisy input image. The "edge-stopping" function in the anisotropic diffusion equation is closely related to the error norm and influence function in the robust estimation framework. We exploit this robust statistical interpretation of anisotropic diffusion to choose alternative robust error norms, and hence, alternative "edge-stopping" functions. In particular, we propose a new "edge-stopping" function based on *Tukey's biweight* robust error norm, which preserves sharper boundaries than previous formulations and improves the automatic stopping of the diffusion. The robust statistical interpretation also provides a means for detecting the boundaries (edges) between the piecewise constant regions. These boundaries are considered to be "outliers" in the robust estimation framework. Edges in a smoothed image are, therefore, very simply detected as those points that are treated as outliers. Details, examples, and extensions, including connections to line processing and techniques as those described in [4, 5], can be found in [2]. We also show, following [13], extensions of this method to color images and vector data in general, and describe the use of the theory here presented for image sharpening.

GS is partially supported by an ONR grant. Part of this work was performed while GS was at HP Labs, California.

2. ANISOTROPIC DIFFUSION

Diffusion algorithms remove noise from an image by modifying the image 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 $I(x,y,0) : I\!R^2 \rightarrow I\!R^+$ 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.

Perona and Malik [9] replaced the classical isotropic diffusion equation with

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

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

Perona and Malik discretized their anisotropic diffusion equation as follows:

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

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 $\lambda \in I\!R^+$ 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. They linearly approximated the image gradient (magnitude) in a particular direction as $\nabla I_{s,p} = I_p - I_s^t, \;\; p \in \eta_s$

Qualitatively, the effect of anisotropic diffusion is to smooth the original image while preserving brightness discontinuities. As we will see, the choice of $g(\cdot)$ can greatly affect the extent to which discontinuities are preserved.

0-8186-8183-7/97 $10.00

3. ROBUST ESTIMATION

Our goal is to develop a statistical interpretation of the Perona-Malik anisotropic diffusion equation. Toward that end, we adopt an oversimplified statistical model of an image. In particular, we assume that a given input image is a piecewise constant function that has been corrupted by zero-mean Gaussian noise with small variance.

Consider the image intensity differences, $I_p - I_s$, between pixel s and its neighboring pixels p. Within one of the piecewise constant image regions, these neighbor differences will be small, zero-mean, and normally distributed. Hence, an optimal estimator for the "true" value of the image intensity I_s at pixel s minimizes the square of the neighbor differences. This is equivalent to choosing I_s to be the mean of the neighboring intensity values.

The neighbor differences will not be normally distributed, however, for an image region that includes a boundary (intensity discontinuity). The intensity values of the neighbors of an edge pixel s are drawn from two different populations, and in estimating the "true" intensity value at s we want to include only those neighbors that belong to the same population. With respect to our assumption of Gaussian noise within each constant region, if p and s belong to two different sides of the edge, the neighbor difference $I_p - I_s$ can be viewed as an *outlier* because it does not conform to the statistical assumptions.

The field of robust statistics [6, 7] is concerned precisely with estimation problems in which the data contains gross errors, or outliers. See for example [2] for references to the applications of robust statistics to image processing and computer vision.

Motivated then by robust statistics, we wish to find an image I that satisfies the following optimization criterion:

$$\min_I \sum_{s \in I} \sum_{p \in \eta_s} \rho(I_p - I_s, \sigma), \tag{3}$$

where $\rho(\cdot)$ is a robust error norm and σ is a "scale" parameter that will be discussed further below. To minimize (3), the intensity at each pixel must be "close" to those of its neighbors. As we shall see, an appropriate choice of the ρ-function allows us to minimize the effect of the outliers at the boundaries between piecewise constant image regions.

Equation (3) can be solved by gradient descent:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} \psi(I_p - I_s^t, \sigma), \tag{4}$$

where $\psi(\cdot) = \rho'(\cdot)$, and t again denotes the iteration. The update is carried out simultaneously at every pixel s.

The specific choice of the robust error norm or ρ-function in (3) is critical. To analyze the behavior of a given ρ-function, we consider its derivative ψ, which is proportional to the *influence function* [6]. This function characterizes the bias that a particular measurement has on the solution. For example, the quadratic ρ-function has a linear ψ-function.

If the distribution of values $(I_p - I_s^t)$ in every neighborhood is a zero-mean Gaussian, then $\rho(x, \sigma) = x^2/\sigma^2$ provides an optimal local estimate of I_s^t. This *least-squares* estimate of I_s^t is, however, very sensitive to outliers because the influence function increases linearly and without bound. For a quadratic ρ, I_s^{t+1} is assigned to be the mean of the neighboring intensity values I_p. When these values come from different populations (across a boundary) the mean is not representative of either population, and the image is blurred too much. Hence, the quadratic gives outliers (large values of $|\nabla I_{s,p}|$) too much *influence*.

To increase robustness and *reject* outliers, the ρ-function must be more forgiving about outliers; that is, it should increase less rapidly than x^2. For example, consider the *Lorentzian* error norm (see Figure 1):

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2}. \tag{5}$$

Examination of the ψ-function reveals that, when the absolute value of the gradient magnitude increases beyond a fixed point determined by the scale parameter σ, its influence is reduced. We refer to this as a *redescending* influence function [6].

4. ROBUST STATISTICS AND ANISOTROPIC DIFFUSION

We now explore the relationship between robust statistics and anisotropic diffusion by showing how to convert back and forth between the formulations. Recall the continuous anisotropic diffusion equation (1). The continuous form of the robust estimation problem in (3) can be posed as:

$$\min_I \int_\Omega \rho(\| \nabla I \|) d\Omega, \tag{6}$$

where Ω is the domain of the image and where we have omitted σ for notational convenience. One way to minimize (6) is via gradient descent using the calculus of variations:

$$\frac{\partial I(x, y, t)}{\partial t} = \operatorname{div}\left(\rho'(\| \nabla I \|) \frac{\nabla I}{\| \nabla I \|}\right). \tag{7}$$

By defining $g(x) \doteq \frac{\rho'(x)}{x}$, we obtain the straightforward relation between image reconstruction via robust estimation (6) and image reconstruction via anisotropic diffusion (1). (See for example [9, 14] for previous uses of this relation.)

The same relationship holds for the discrete formulation; compare (2) and (4) with $\psi(x) = \rho'(x) = g(x)\,x$. Note that additional terms will appear in the gradient descent equation if the magnitude of the image gradient is discretized in a nonlinear fashion. In the remainder of this paper we proceed with the discrete formulation as given in previous section. The basic results we present hold for the continuous domain as well.

Perona and Malik suggested two different edge stopping functions ($g(\cdot)$) in their anisotropic diffusion equation. Each of these can be viewed in the robust statistical framework by converting the $g(\cdot)$ functions into the related ρ-functions. They first suggested $g(x) = \frac{1}{1 + \frac{x^2}{K^2}}$, for a positive constant K. It is easy to see, [2, 8, 14], that this edge stopping function corresponds to the Lorentzian norm of robust statistics. The other g-function proposed by Perona and Malik is related to the robust error norm proposed by Leclerc.

5. EXPLOITING THE RELATIONSHIP

The above derivations demonstrate that anisotropic diffusion is the gradient descent of an estimation problem with a familiar robust error norm. What's the advantage of knowing this connection? We answer this question now.

While the Lorentzian is more robust than the quadratic norm, its influence does not descend all the way to zero. We can choose a more "robust" norm from the robust statistics literature which does descend to zero, as the *Tukey's biweight*, whose influence function is plotted in Figure 1:

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

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

Another error norm from the robust statistics literature, is Huber's *minimax* norm [7] (see also [12, 14]), whose influence function is also plotted in Figure 1:

$$\rho(x,\sigma) = \begin{cases} x^2/2\sigma + \sigma/2 & |x| \le \sigma, \\ |x| & |x| > \sigma, \end{cases} \quad (10)$$

$$\psi(x,\sigma) = \begin{cases} x/\sigma, & |x| \le \sigma, \\ \text{sign}(x) & |x| > \sigma. \end{cases} \quad (11)$$

We would like to compare the influence (ψ-function) of these three norms, but a direct comparison requires that we dilate and scale the functions to make them as similar as possible.

First, we need to determine how large the image gradient can be before we consider it to be an outlier. We appeal to tools from robust statistics to automatically estimate the "robust variance," σ_e, of the image as the MAD [11]:

$$\sigma_e = 1.4826 \text{ median}_I(\| \nabla I - \text{median}_I(\| \nabla I \|) \|) \quad (12)$$

where the constant is derived from the fact that the MAD of a zero-mean normal distribution with unit variance is $0.6745 = 1/1.4826$. For a discrete image, the robust variance, σ_e, is computed using the gradient magnitude approximation introduced before.

Second, we choose values for the scale parameters σ to dilate each of the three 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 modified L_1 or Huber's norm this occurs at $\sigma_e = \sigma$. For the Lorentzian norm it occurs at $\sigma_e = \sqrt{2}\,\sigma$ and for the Tukey norm 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 1*a*. Note how each function now begins reducing the influence of measurements at the same point.

Third, we scale the three 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 1*b*.

Now we can compare the three error norms directly. The modified L_1 norm gives all outliers a constant weight of one while the Tukey norm gives *zero* weight to outliers whose magnitude is above a certain value. The Lorentzian (or Perona-Malik) norm is in between the other two. Based on the shape of $\psi(\cdot)$ we would predict that diffusing with the Tukey norm produces sharper boundaries than diffusing with the Lorentzian (standard Perona-Malik) norm, and that both produce sharper boundaries than the modified L_1 norm. We can also see how the choice of function affects the "stopping" behavior of the diffusion; given a piecewise constant image where all discontinuities are above a threshold, the Tukey function will leave the image unchanged whereas the other two functions will not.

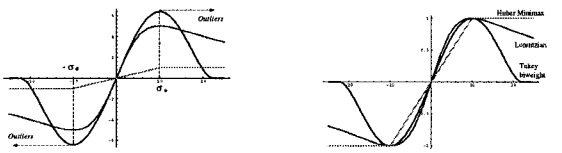

Figure 1: *Lorentzian, Tukey, and Huber ψ-functions. Left: Values of σ chosen as a function of σ_e so that outlier "rejection" begins at the same value for each function. Right: The functions aligned and scaled. (This is a color figure.)*

These predictions are born out experimentally, as can be seen in Figure 2. The figure compares the results of diffusing with the Lorentzian and the Tukey functions. The value of $\sigma_e = 10.278$ was estimated automatically using (12) and the values of σ and λ for each function were defined with respect to σ_e as described above. The figure shows the diffused image after 500 iterations of each method. Observe how the Tukey function results in sharper discontinuities. Note that 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 2 shows the outliers (edge points) in each of the images, that is, pixels where $|\nabla I_{s,p}| > \sigma_e$.

It is interesting to note that common robust error norms have frequently been proposed in the literature without mentioning the motivation from robust statistics. For example, Rudin *et al.* [12] proposed a formulation that is equivalent to using the L_1 norm. You *et al.* [14] explored a variety of anisotropic diffusion equations and reported better results for some than for others. In addition to their own explanation for this, their results are predicted, following the development presented here, by the robustness of the various error norms they use.

6. ROBUST ANISOTROPIC SHARPENING

The basic idea behind image sharpening is to add high frequencies to an image. That is, the sharpened image $\hat{I}$ is obtained from the blurred image I via $\hat{I} = I + H(I)$, where $H(\cdot)$ represents some kind of high-pass filter operation, e.g., the Laplacian. The problem with this approach, denoted as *unsharp masking*, is that it also enhances noise, and not only edges. To solve this problem, in [3] the author proposes to mask $H(I)$ with an edge detector. Since the framework here described is natural to detect edges, it is natural as well to accomplish this task. The basic idea is then to perform anisotropic diffusion on the Image I, robustly detect edges

based on outliers, and then mask $H(I)$ using these edges. An example is given in Figure 3.

7. VECTOR-VALUED IMAGES

The extension of the results presented above for vector-valued images follows the framework introduced in [13]. The basic idea is that the gradient direction $\frac{\nabla I}{\|\nabla I\|}$ and the gradient magnitude $\| \nabla I \|$ are replaced by concepts derived from the first fundamental form of the vector image. The direction of maximal change θ_+ ("the gradient direction") of the vector data is given by the eigenvector of this fundamental form corresponding to the maximal eigenvalue λ_+, and the value of the maximal change ("the gradient magnitude") is given by a function of both eigenvalues, that is, $f(\lambda_+, \lambda_-)$. Note that θ_+, λ_+, and λ_- depend on all the components of the vector-valued image.

To extend the robust anisotropic diffusion approach to vector data, we have basically two possibilities. One, [13], if to formulate the problem as the minimization of

$$\int \rho(\lambda_+, \lambda_-) d\Omega,$$

selecting ρ to be the Tukey's robust function. The gradient descent of this variational problem will give a system of coupled anisotropic diffusion equations. The second option if to derive directly the anisotropic equation from (7), and evolve each one of the image components I_i according to

$$\frac{\partial I_i}{\partial t} = \text{div}\left(\psi(\lambda_+, \lambda_-)\begin{pmatrix}\cos\theta_+ \\ \sin\theta_+\end{pmatrix}\right),$$

where ψ is the Tukey's influence function.

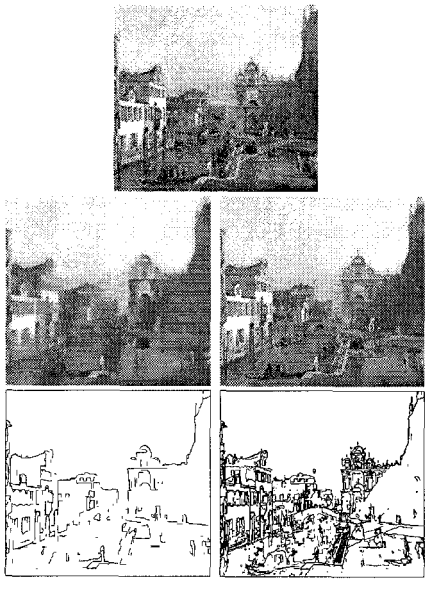

Figure 2: *Comparison of the Perona-Malik (Lorentzian) function (left) and the Tukey function (right) after 500 iterations. The first row shows the original image. The last row shows the edges obtained from the outliers.*

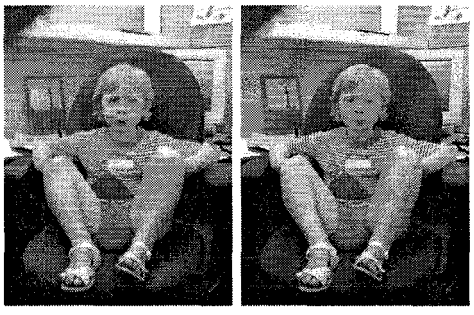

Figure 3: *Original image (left) and result of sharpening with robust masking. (This is a color figure).*

8. REFERENCES

[1] L. Alvarez, P. L. Lions, and J. M. Morel. *SIAM-JNA* **29**, pp. 845-866, 1992.

[2] M. Black, G. Sapiro, D. Marimont, D. Heeger, "Robust anisotropic diffusion," *IEEE-IP*, to appear.

[3] K. Castleman, *Digital Image Processing*, Prentice-Hall, 1996.

[4] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE-IP*, to appear.

[5] D. Geman and C. Yang, *IEEE-IP* 4:7, pp. 932-946, 1995.

[6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York, 1986.

[7] P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.

[8] N. Nordström, *Image and Vision Computing* **8**, pp. 318-327, 1990.

[9] P. Perona and J. Malik. *IEEE-PAMI* **12**, pp. 629-639, 1990.

[10] B. Romeny (Ed.). *Geometry Driven diffusion in Computer Vision*, Kluwer, 1994.

[11] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, 1987.

[12] L. I. Rudin, S. Osher, and E. Fatemi. *Physica D* **60**, pp. 259-268, 1992.

[13] G. Sapiro and D. Ringach, *IEEE-IP* **5**, pp. 1582-1586, 1996.

[14] Y. L. You, W. Xu, A. Tannenbaum, and M. Kaveh. *IEEE-IP* **5**, pp. 1539-1553, 1996.