IEEE Trans. Image Processing, to appear

Robust Anisotropic Diffusion

Michael J. Black*, **Guillermo Sapiro**†, **David Marimont***, **and David Heeger**‡

**Xerox Palo Alto Research Center,* Palo Alto, CA 94304
†*Hewlett-Packard Labs,* Palo Alto, CA 94304
‡*Dept. of Psychology,* Stanford Univ., Stanford, CA 94305

Abstract

Relations between anisotropic diffusion and robust statistics are described in this paper. Specifically, 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 an image that has been smoothed with anisotropic diffusion. Additionally, we derive a relationship between anisotropic diffusion and regularization with line processes. Adding constraints on the spatial organization of the line processes allows us to develop new anisotropic diffusion equations that result in a qualitative improvement in the continuity of edges.*

Key words: Anisotropic diffusion, robust statistics, line processes.

1 Introduction

Since the elegant formulation of anisotropic diffusion introduced by Perona and Malik [38] (see [15] for very early work in this topic), a considerable amount of research has been devoted to the theoretical and practical understanding of this and related methods for image enhancement. Research in this area has been oriented toward understanding the mathematical properties of anisotropic diffusion and related variational formulations [4, 11, 25, 38, 50], developing related well-posed and stable equations [2, 3, 11, 21, 35, 40, 50], extending and modifying anisotropic diffusion for fast and accurate implementations, modifying the diffusion equations for specific applications [20], and studying the relations between anisotropic diffusion and other image processing operations [41, 45].

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 [38] 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 in an image that has been smoothed with anisotropic diffusion. The boundaries between the piecewise constant regions 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.

We also show (following [6]) that, for a particular class of robust error norms, anisotropic diffusion is equivalent to regularization with an explicit line process. The advantage of the line-process formulation is that we can add constraints on the spatial organization of the edges. We demonstrate that adding such constraints to the Perona-Malik diffusion equation results in a qualitative improvement in the continuity of edges.

2 Background

2.1 Anisotropic diffusion: Perona-Malik formulation

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 $\frac{\partial I(x,y,t)}{\partial t} = \text{div}(\nabla I)$, using the original (degraded/noisy) image $I(x, y, 0)$ as the initial condition, where $I(x, y, 0) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ 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 [38] replaced the classical isotropic diffusion equation with

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

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.

As mentioned in the Introduction, (1) motivated a large number of researchers to study the mathematical properties of this type of equation, as well as its numerical implementation and adaptation to specific applications. The stability of the equation was the particular concern of extensive research, e.g., [3, 11, 25, 38, 50]. In this paper, we present equations that are modifications of (1); we do not discuss the stability of these modified equations because the stability results can be obtained from the mentioned references. Briefly, however, we should point out that stability problems will typically be solved, or at least moderated, by the spatial regularization and temporal delays introduced by the numerical methods for computing the gradient in $g(\| \nabla I \|)$ [11, 25, 37].

2.2 Perona-Malik discrete formulation

Perona and Malik discretized their anisotropic diffusion equation as follows:

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

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

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

Figure 8 shows examples of applying this equation to an image, using two different choices for the edge-stopping function, $g(\cdot)$. 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. Understanding this is one of the main goals of this paper.

2.3 A statistical view

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. In [50], the authors presented interesting theoretical (and practical) analysis of the behavior of anisotropic diffusion for piecewise constant images. We will return later to comment on their results.

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

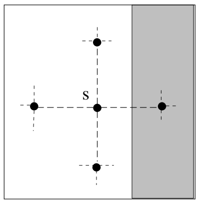


Figure 1: Local neighborhood of pixels at a boundary (intensity discontinuity).

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.

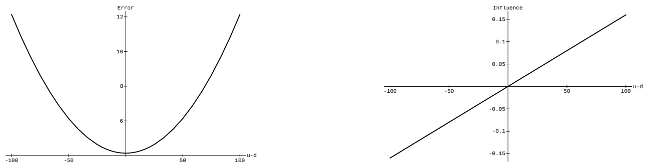
The neighbor differences will not be normally distributed, however, for an image region that includes a boundary (intensity discontinuity). Consider, for example, the image region illustrated in Figure 1. The intensity values of the neighbors of pixel s are drawn from two different populations, and in estimating the "true" intensity value at s we want to include only those neighbors that belong to the same population. In particular, the pixel labelled p is on the wrong side of the boundary so I_p will skew the estimate of I_s significantly. With respect to our assumption of Gaussian noise within each constant region, the neighbor difference $I_p - I_s$ can be viewed as an *outlier* because it does not conform to the statistical assumptions.

2.4 Robust estimation

The field of robust statistics [22, 24] is concerned with estimation problems in which the data contains gross errors, or outliers.

Many robust statistical techniques have been applied to standard problems in computer vision [1, 32, 44]. There are robust approaches for performing local image smoothing [5], image reconstruction [19, 17], blur classification [13], surface reconstruction [47], segmentation [31], pose estimation [26], edge detection [28], structure from motion or stereo [48, 49], optical flow estimation [7, 8, 43], and regularization with line processes [6]. For further details see [22] or, for a review of the applications of robust statistics in computer vision, see [32].

The problem of estimating a piecewise constant (or smooth) image from noisy data can also be



$\rho(x, \sigma)$ $\psi(x, \sigma)$

Figure 2: Least-squares (quadratic) error norm.

posed using the tools of robust statistics. We wish to find an image I that satisfies the following optimization criterion:

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

where $\rho(\cdot)$ is a robust error norm and σ is a "scale" parameter that will be discussed further below. To minimize (4), 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, $(I_p - I_s)$, at the boundaries between piecewise constant image regions.

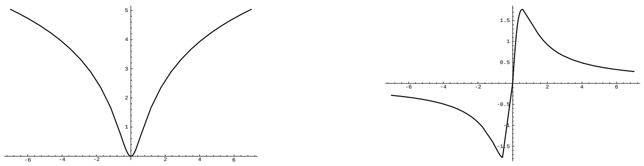
Equation (4) can be solved by gradient descent

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} \psi(I_p - I_s^t, \sigma) \tag{5}$$

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 (4) is critical. To analyze the behavior of a given ρ-function, we consider its derivative (denoted ψ), which is proportional to the *influence function* [22]. 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 (see Figure 2). For a quadratic ρ, I_s^{t+1} is assigned to be the mean of the neighboring intensity



$\rho(x, \sigma)$ $\psi(x, \sigma)$

Figure 3: Lorentzian error norm.

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 following *Lorentzian* error norm plotted in Figure 3:

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2}. \tag{6}$$

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 [22].[1] If a particular local difference, $\nabla I_{s,p} = I_p - I_s^t$, has a large magnitude then the value of $\psi(\nabla I_{s,p})$ will be small and therefore that measurement will have little effect on the update of I_s^{t+1} in (5).

3 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:

$$\frac{\partial I(x, y, t)}{\partial t} = \operatorname{div}(g(\| \nabla I \|)\, \nabla I). \tag{7}$$

[1]Some authors reserve the term redescending to describe functions for which $\psi(x) = 0$ for $|x| > r$ for some finite constant r but we will use the term more generally.

The continuous form of the robust estimation problem in (4) can be posed as:

$$\min_I \int_\Omega \rho(\| \nabla I \|) d\Omega, \tag{8}$$

where Ω is the domain of the image and where we have omitted σ for notational convenience. One way to minimize (8) is via gradient descent using the calculus of variations (see for example [21, 36, 38, 50]):

$$\frac{\partial I(x,y,t)}{\partial t} = \text{div}\left(\rho'(\| \nabla I \|) \frac{\nabla I}{\| \nabla I \|}\right). \tag{9}$$

By defining

$$g(x) \doteq \frac{\rho'(x)}{x}, \tag{10}$$

we obtain the straightforward relation between image reconstruction via robust estimation (8) and image reconstruction via anisotropic diffusion (7). You *et al.* [50] show and make extensive use of this important relation in their analysis (we will comment on their results later in this paper).

The same relationship holds for the discrete formulation; compare (2) and (5) 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.

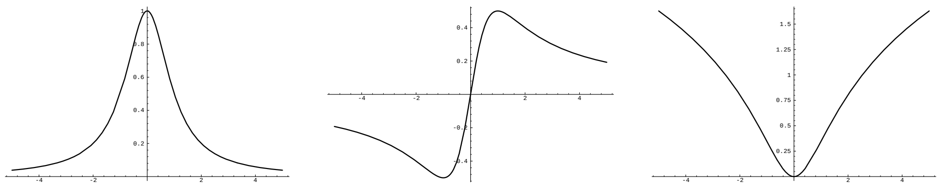
Perona and Malik suggested two different edge stopping $g(\cdot)$ functions in their anisotropic diffusion equation. Each of these can be viewed in the robust statistical framework by converting the $g(\cdot)$ functions into the related ρ-functions.

Perona and Malik first suggested

$$g(x) = \frac{1}{1 + \frac{x^2}{K^2}}, \tag{11}$$

for a positive constant K. We want to find a ρ-function such that the iterative solution of the diffusion equation and the robust statistical equation are equivalent. Letting $K^2 = 2\sigma^2$, we have

$$g(x)x = \frac{2x}{2 + \frac{x^2}{\sigma^2}} = \psi(x, \sigma), \tag{12}$$



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

Figure 4: Lorentzian error norm and the Perona-Malik g stopping function.

where $\psi(x, \sigma) = \rho'(x, \sigma)$. Integrating $g(x)x$ with respect to x gives:

$$\int g(x)x \ dx = \sigma^2 \log \left(1 + \frac{1}{2}\left(\frac{x^2}{\sigma^2}\right)\right) = \rho(x). \tag{13}$$

This function $\rho(x)$ is proportional to the Lorentzian error norm introduced in the previous section, and $g(x)x = \rho'(x) = \psi(x)$ is proportional to the influence function of the error norm; see Figure 4. Iteratively solving (5) with a Lorentzian for ρ is therefore equivalent to the discrete Perona-Malik formulation of anisotropic diffusion. This relation was previously pointed out in [50].

The same treatment can be used to recover a ρ-function for the other g-function proposed by Perona and Malik

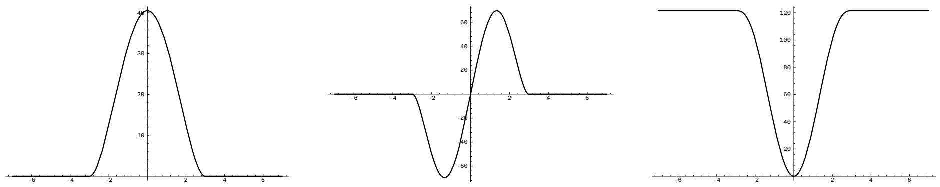
$$g(x) = e^{-\frac{x^2}{k^2}}. \tag{14}$$

The resulting ρ-function is related to the robust error norm proposed by Leclerc [27]. The derivation is straightforward and is omitted here.

4 Exploiting the relationship

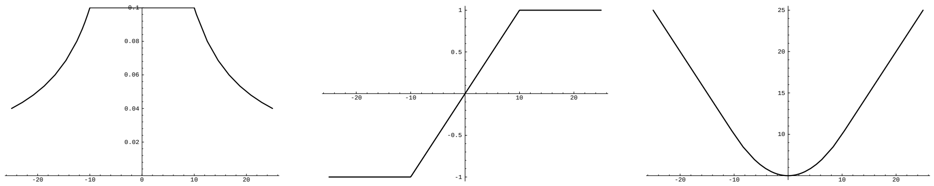
The above derivations demonstrate that anisotropic diffusion is the gradient descent of an estimation problem with a familiar robust error norm. What's the advantage of knowing this connection? We argue that the robust statistical interpretation gives us a broader context within which to evaluate, compare, and choose between alternative diffusion equations. It also provides tools for automatically determining what should be considered an outlier (an "edge"). In this section we illustrate these connections with an example.

While the Lorentzian is more robust than the L_2 (quadratic) norm, its influence does not descend



$g(x)$ $\quad$ $g(x)x = \psi(x)$ $\quad$ $\rho(x)$

Figure 5: Tukey's biweight.



$g(x)$ $\quad$ $g(x)x = \psi(x)$ $\quad$ $\rho(x)$

Figure 6: Huber's minmax estimator (modification of the L_1 norm).

all the way to zero. We can choose a more "robust" norm from the robust statistics literature which does descend to zero. The *Tukey's biweight*, for example, is plotted along with its influence function in Figure 5:

$$\rho(x,\sigma) = \begin{cases} \frac{x^2}{\sigma^2} - \frac{x^4}{\sigma^4} + \frac{x^6}{3\sigma^6} & |x| \le \sigma \\ \frac{1}{3} & \text{otherwise} \end{cases} \tag{15}$$

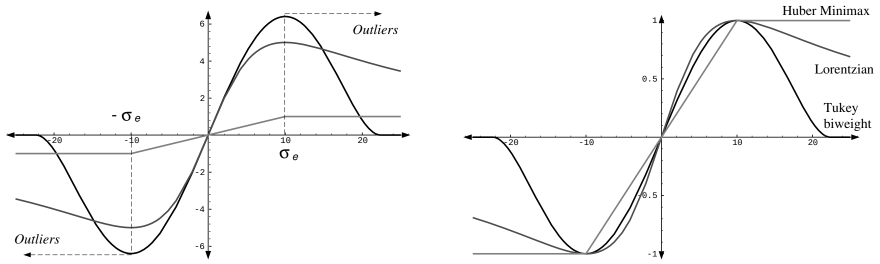
$$\psi(x,\sigma) = \begin{cases} x(1-(x/\sigma)^2)^2 & |x| \le \sigma, \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

$$g(x,\sigma) = \begin{cases} \frac{1}{2}(1-(x/\sigma)^2)^2 & |x| \le \sigma, \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Another error norm from the robust statistics literature, Huber's *minimax* norm [24] (see also [40, 50]), is plotted along with its influence function in Figure 6. Huber's minmax norm is equivalent to the L_1 norm for large values. But, for normally distributed data, the L_1 norm produces estimates with higher variance than the optimal L_2 (quadratic) norm, so Huber's minmax norm is designed to be quadratic for small values:

$$\rho(x,\sigma) = \begin{cases} x^2/2\sigma + \sigma/2 & |x| \le \sigma, \\ |x| & |x| > \sigma, \end{cases} \tag{18}$$

$$\psi(x,\sigma) = \begin{cases} x/\sigma, & |x| \le \sigma, \\ \text{sign}(x) & |x| > \sigma. \end{cases} \tag{19}$$



a *b*

Figure 7: Lorentzian, Tukey, and Huber 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.

$$g(x, \sigma) = \begin{cases} 1/\sigma, & |x| \leq \sigma, \\ \text{sign}(x)/x & |x| > \sigma. \end{cases} \tag{20}$$

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 scale," σ_e, of the image as [39]

$$\begin{aligned} \sigma_e &= 1.4826 \ \text{MAD}(\nabla I) \\ &= 1.4826 \ \text{median}_I(\| \nabla I - \text{median}_I(\| \nabla I \|) \|) \end{aligned} \tag{21}$$

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 variance is $0.6745 = 1/1.4826$. For a discrete image, the robust scale, 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 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 7a. 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 7b.

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 correctly 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.

These predictions are born out experimentally, as can be seen in Figure 8. The figure compares the results of diffusing with the Lorentzian $g(\cdot)$ function and the Tukey g function. The value of $\sigma_e = 10.278$ was estimated automatically using (21) 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 100 iterations of each method. Observe how the Tukey function results in sharper discontinuities.

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 9 shows the outliers (edge points) in each of the images, where $|\nabla I_{s,p}| > \sigma_e$.

Finally, Figure 10 illustrates the behavior of the two functions in the limit (shown for 500 iterations). The Perona-Malik formulation continues to smooth the image while the Tukey version has effectively "stopped."

These examples illustrate how ideas from robust statistics can be used to evaluate and compare

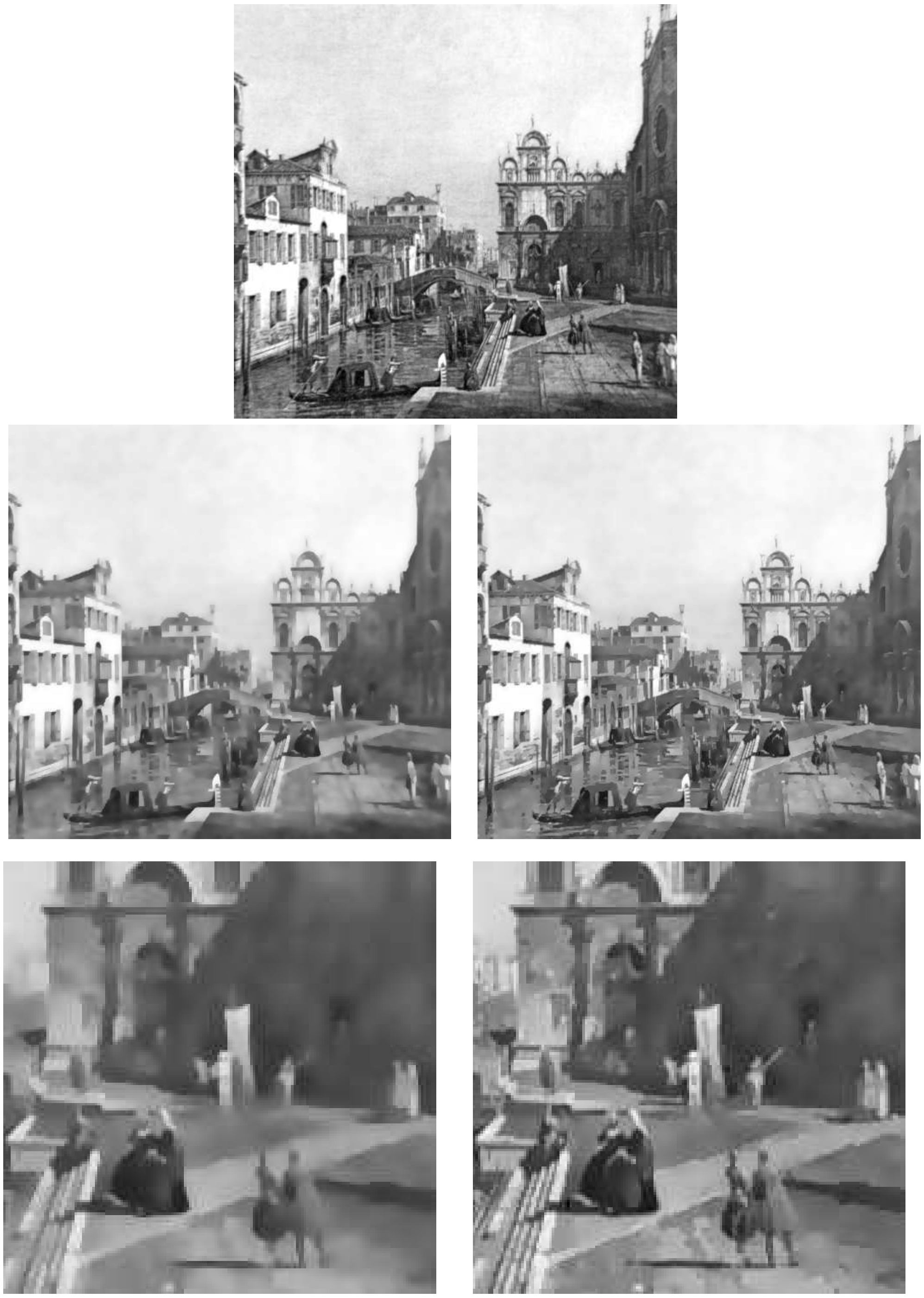


Figure 8: Comparison of the Perona-Malik (Lorentzian) function (left) and the Tukey function (right) after 100 iterations. Top: original image. Middle: diffused images. Bottom: magnified regions of diffused images.

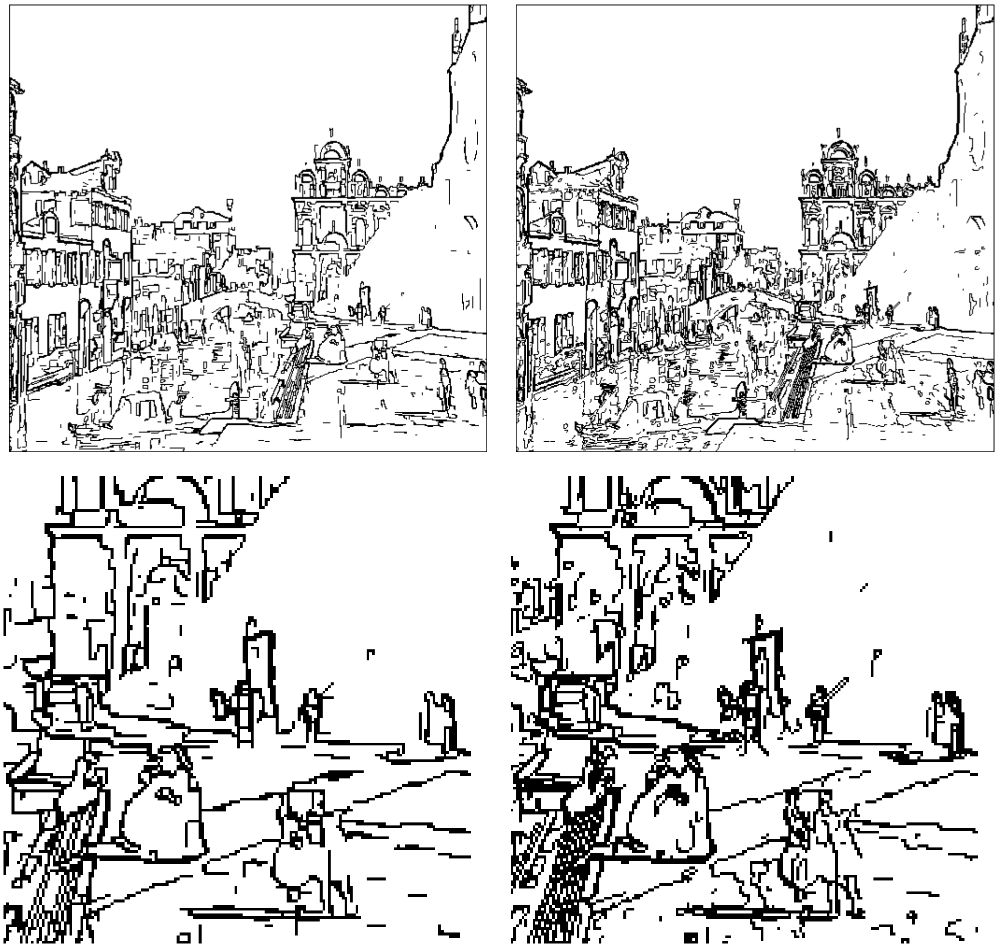


Figure 9: Comparison of edges (outliers) for the Perona-Malik (Lorentzian) function (left) and the Tukey function (right) after 100 iterations. Bottom row shows a magnified region.

different g-functions and how new functions can be chosen in a principled way. See [6] for other robust ρ-functions which could be used for anisotropic diffusion. See also [16] for related work connecting anisotropic diffusion, the mean-field ρ-function, and binary line processes.

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.* [40] proposed a formulation that is equivalent to using the L_1 norm. You *et al.* [50] 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. Moreover, some of their theoretical

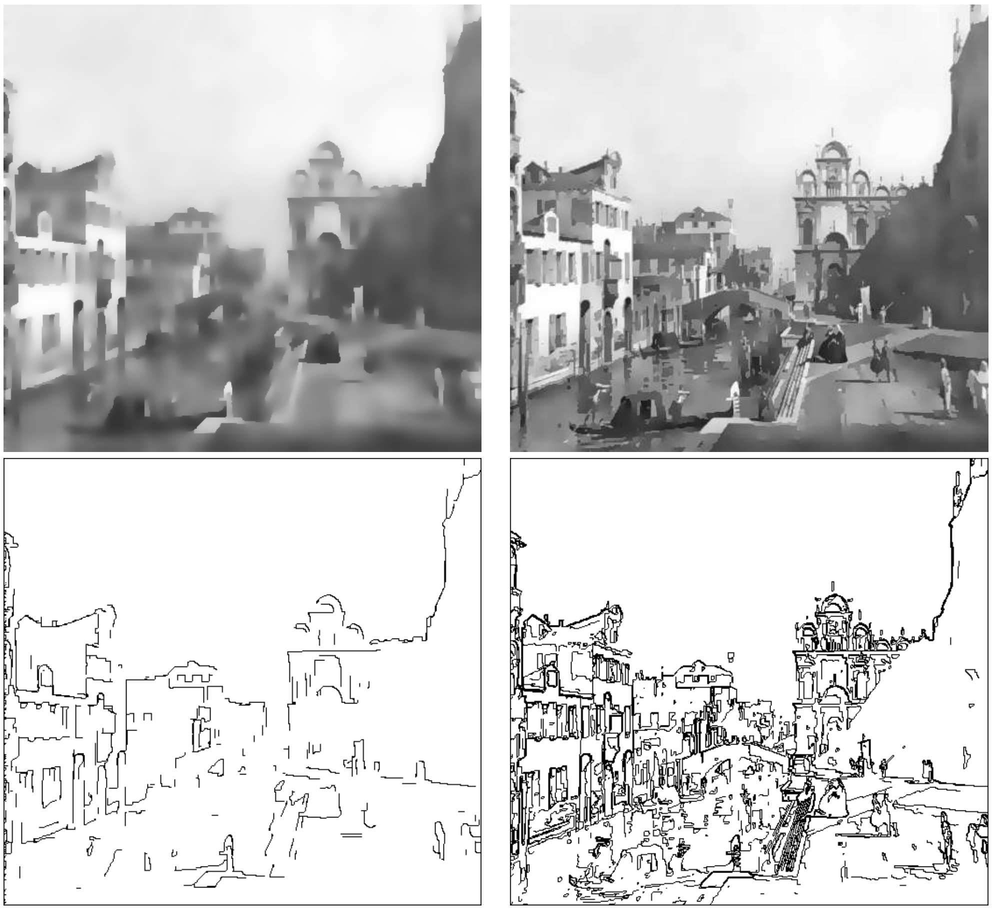


Figure 10: Comparison of the Perona-Malik (Lorentzian) function (left) and the Tukey function (right) after 500 iterations.

results, e.g., Theorem 1 and Theorem 3, are easily interpreted based on the concept of influence functions. Finally, Mead and colleagues [23, 30] have used analog VLSI (aVLSI) technology to build hardware devices that perform regularization. The aVLSI circuits behave much like a resistive grid, except that the resistors are replaced with "robust" resistors made up of several transistors. Each such resistive grid circuit is equivalent to using a different robust error norm.

5 Robust estimation and line processes

This section derives the relationship between anisotropic diffusion and regularization with line processes. The connection between robust statistics and line processes has been explored elsewhere;

see [6] for details and examples as well as [9, 12, 16, 19, 18] for recent related results. While we work with the discrete formulation here, it is easy to verify that the connections hold for the continuous formulation as well.

Recall that the robust formulation of the smoothing problem was posed as the minimization of

$$E(I) = \sum_s E(I_s), \tag{22}$$

where

$$E(I_s) = \sum_{p \in \eta_s} \rho(I_p - I_s, \sigma). \tag{23}$$

There is an alternative, equivalent, formulation of this problem that makes use of an explicit *line process* in the minimization:

$$E(I, \mathbf{l}) = \sum_s E(I_s, \mathbf{l}), \tag{24}$$

where

$$E(I_s, \mathbf{l}) = \sum_{p \in \eta_s} \left[\frac{1}{2\sigma^2}(I_p - I_s)^2 l_{s,p} + P(l_{s,p}) \right] \tag{25}$$

and where $l_{s,p} \in \mathbf{l}$ are analog line processes ($0 \leq l_{s,p} \leq 1$) [16, 17]. The line process indicates the presence (l close to 0) or absence (l close to 1) of discontinuities or *outliers*. The last term, $P(l_{s,p})$, *penalizes* the introduction of line processes between pixels s and p. This penalty term goes to zero when $l_{s,p} \to 1$ and is large (usually approaching 1) when $l_{s,p} \to 0$.

One benefit of the line-process approach is that the "outliers" are made explicit and therefore can be manipulated. For example, as we will see in Section 5.1 we can add constraints on these variables which encourage specific types of spatial organizations of the line processes.

Numerous authors have shown how to convert a line-process formulation into the robust formulation with a ρ-function by minimizing over the line variables [9, 16, 18]. That is

$$\rho(x) = \min_{0 \leq l \leq 1} E(x, l),$$

where

$$E(x, l) = (x^2 l + P(l)).$$

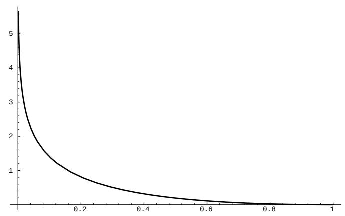


Figure 11: Lorentzian (Perona-Malik) penalty function, $P(l)$, $0 \le l \le 1$.

For our purposes here it is more interesting to consider the other direction: can we convert a robust estimation problem into an equivalent line-process problem? We have already shown how to convert a diffusion problem with a $g(\cdot)$ function into a robust estimation problem. If we can make the connection between robust ρ-functions and line processes then we will be able to take a diffusion formulation like the Perona-Malik equation and construct an equivalent line process formulation.

Then, our goal is to take a function $\rho(x)$ and construct a new function, $E(x, l) = (x^2 l + P(l))$, such that the solution at the minimum is unchanged. Clearly the penalty term $P(\cdot)$ will have to depend in some way on $\rho(\cdot)$. By taking derivatives with respect to x and l, it can be shown that the condition on $P(l)$ for the two minimization problems to be equivalent is given by:

$$-x^2 = P'\left(\frac{\psi(x)}{2x}\right).$$

By integrating this equation we obtain the desired line process penalty function $P(l)$. See [6] for details on the explicit computation of this integral. There are a number of conditions on the form of ρ that must be satisfied in order to recover the line process, but as described in [6], these conditions do in fact hold for many of the redescending ρ-functions of interest.

In the case of the Lorentzian norm, it can be shown that $P(l) = l - 1 - \log l$; see Figure 11. Hence, the equivalent line-process formulation of the Perona-Malik equation is:

$$E(I_s, \mathbf{l}) = \sum_{p \in \eta_s} \left[\frac{1}{2\sigma^2}(I_p - I_s)^2 l_{s,p} + l_{s,p} - 1 - \log l_{s,p}\right]. \tag{26}$$

Differentiating with respect to I_s and l gives the following iterative equations for minimizing

$$\begin{array}{ccc} s\circ & | & p\circ \\ u\circ & | & v\circ \end{array} \qquad\qquad s\circ \;|\; p\circ \;|\; u\circ$$

$$C_{hyst} \qquad\qquad\qquad C_{supp}$$

Figure 12: Cliques for spatial interaction constraints (up to rotation) at a site s. The circles indicate pixel locations and the bars indicate discontinuities between pixels. The C_{hyst} cliques are used for hysteresis and the C_{supp} cliques are used for non-maxima suppression.

$E(I_s, \mathbf{l})$:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} l_{s,p} \nabla I_{s,p} \tag{27}$$

$$l_{s,p} = \frac{2\sigma^2}{2\sigma^2 + \nabla I_{s,p}^2}. \tag{28}$$

Note that these equations are equivalent to the discrete Perona-Malik diffusion equations. In particular, $l_{s,p}$ is precisely equal to $g(\|\nabla I_{s,p}\|)$.

5.1 Spatial organization of outliers

One advantage of the connection between anisotropic diffusion and line processes, obtained through the connection of both techniques to robust statistics, is the possibility of improving anisotropic flows by the explicit design of line processes with spatial coherence. In the classical Perona-Malik flow, which relies on the Lorentzian error norm, there is no spatial coherence imposed on the detected outliers; see Figure 9. Since the outlier process is explicit in the formulation of the line processing (Equation (25)), we can add additional constraints on its spatial organization. While numerous authors have proposed spatial coherence constraints for discrete line processes [14, 17, 18, 33], we need to generalize these results to the case of analog line processes [6].

We consider two kinds of interaction terms, *hysteresis* [10] and *non-maximum suppression* [34]. Other common types of interactions (for example corners) can be modeled in a similar way. The hysteresis term assists in the formation of unbroken contours while the non-maximum suppression term inhibits multiple responses to a single edge present in the data. Hysteresis lowers the penalty for creating extended edges and non-maximum suppression increases the penalty for creating edges that are more that one pixel wide.

We consider a very simple neighborhood system as illustrated in Figure 12. We define a new term that penalizes the configurations on the right of the figure and rewards those on the left. This term, E_I, encodes our prior assumptions about the organization of spatial discontinuities

$$E_I(l_{s,p}) = \alpha \left[-\epsilon_1 \sum_{C_{hyst}} (1 - l_{s,p})(1 - l_{u,v}) \; + \; \epsilon_2 \sum_{C_{supp}} (1 - l_{s,p})(1 - l_{p,u}) \right] \tag{29}$$

where the parameters ϵ_1 and ϵ_2 assume values in the interval $[0, 1]$ and α controls the importance of the spatial interaction term. These parameters were chosen by hand to compute the results reported in this paper.

Starting with the Lorentzian norm, the new error term with constraints on the line processes becomes

$$E(I, \mathbf{l}) = \sum_s E(\nabla I_{s,p}, l_{s,p}), \tag{30}$$

where

$$E(\nabla I_{s,p}, l_{s,p}) = \frac{1}{2\sigma^2} \nabla I_{s,p}^2 l_{s,p} + l_{s,p} - 1 - \log l_{s,p} + E_I(l_{s,p}).$$

Differentiating this equation with respect to I and l gives the following update equations:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} l_{s,p} \nabla I_{s,p}, \tag{31}$$

$$l_{s,p}^{t+1} = \frac{2\sigma^2}{2\sigma^2 (1 + \epsilon_1 \sum_{C_{hyst}} (1 - l_{u,v}^t) - \epsilon_2 \sum_{C_{supp}} (1 - l_{p,u}^t)) + \nabla I_{s,p}^2}. \tag{32}$$

Without the additional spatial constraints, the line process formulation was identical to the original Perona-Malik formulation. In contrast, note here that the value of the line process is dependent on neighboring values.

To see the effect of spatial constraints on the interpretation of discontinuities consider the simple example in Figure 13. The original image is shown on the left. The next column shows values of the line process for the standard Perona-Malik diffusion equation, i.e., without the additional spatial

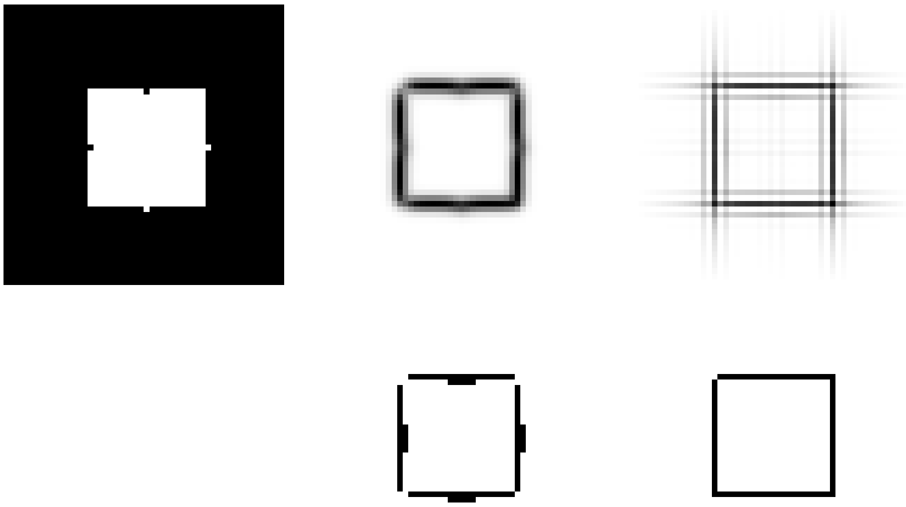


Figure 13: Anisotropic diffusion with spatial organization of outliers. Left: input image. Middle: line process for Perona-Malik (bottom - thresholded). Right: Perona-Malik line process with spatial coherence (bottom - thresholded).

coherence constraint. The value of the line process at each point is taken to be the product, $l_{s,h} l_{s,v}$, of the horizontal and vertical line processes. Dark values correspond to likely discontinuities. The bottom image in the column shows a thresholded version of the top image. Note how the line process is "diffuse" and how the anomalous pixels on the edges of the square produce distortions in the line process.

The column on the right of Figure 13 shows the results when spatial coherence constraints are added. Note how the line process is no longer affected by the anomalous pixels; these are ignored in favor of a straight edge. Note also the hashed, or grid-like, pattern present in the image at the top right. This pattern reflects the simple notion of spatial coherence embodied in E_I that encourages horizontal and vertical edges and discourages edges that are more than one pixel wide.

Figure 14 shows the result of applying the spatial coherence constraints on a real image (see top of Figure 8 for the input). The image on the left of Figure 14 is the result of the diffusion while the image on the right shows the line process values (note the grid-like structure appears here as well).

Figure 14: Anisotropic diffusion with spatially coherent outliers. Left: smoothed image. Right: the value of the line process at each point taken to be the product, $l_{s,h} l_{s,v}$, of the horizontal and vertical line process at s.

Figure 15 compares edges obtained from the standard Perona-Malik diffusion equation with those obtained by adding the spatial coherence constraints. Recall that we interpret edges to be gradient outliers where $|\nabla I_{s,p}| > \sigma_e$. This is equivalent to defining outliers as locations, s, at which the line process, $l(\nabla I_{s,p}, \sigma)$, is less than the value of the line process when $\nabla I_{s,p} = \sigma_e$. Notice that we obtain more coherent edges by adding the spatial coherence constraints.

The line process formulation is more general than the standard anisotropic diffusion equation. Here we have shown only some simple examples of how spatial coherence can be added to diffusion. More sophisticated spatial coherence constraints could be designed to act over larger neighborhoods and/or to encourage edges at a greater variety of orientations

6 Vector-valued images

The extension of the results presented above to vector-valued images is straightforward following the framework introduced in [42]. 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

Figure 15: Edges obtained with Perona-Malik (left); and Perona-Malik with additional spatial coherence in the line processes (right). Lower images show details on the gondola.

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 a number of possibilities. The first possibility, [42], is to formulate the problem as the minimization of

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

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. A second option if to derive directly the anisotropic equation 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. In addition, we can introduce spatial organization of outliers, with information from all the channels. Examples for color and texture data will be reported elsewhere.

7 Concluding remarks

In this paper, we have shown connections between three popular techniques for image reconstruction: anisotropic diffusion, robust statistics, and regularization with line processes. The relations were obtained via simple algebraic operations. These connections make it possible to analyze, design, implement, and interpret anisotropic diffusion using the tools of robust statistics. These connections also make it possible to add spatial coherence to anisotropic diffusion using line processes.

We have demonstrated the practical benefits of this connection by showing how the theory of influence functions can be used to chose "edge-stopping" functions that are more "robust". In particular, we proposed 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 constant image regions, and for selecting the "scale" parameter of the "edge-stopping" functions automatically. Finally, we demonstrated that adding simple spatial coherence constraints edges can improve the continuity of the recovered edges.

A number of related issues remain to be investigated: (1) extending our techniques to vector-valued images; (2) investigating further spatial coherence constraints and their relationship to the robust statistics framework, both from a theoretical and practical viewpoint; (3) connecting formally the results on stability of anisotropic diffusion equations to the theory of influence functions; and (4) extending the rationale for applying robust statistics to anisotropic diffusion from the case of piecewise-constant images to more general ones.

Acknowledgments

We are grateful to Prof. David Fleet for his careful reading of the manuscript and his many helpful suggestions.

References

[1] Proc. Int. Workshop on Robust Computer Vision, Seattle, WA, October 1990.

[2] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mechanics* **123**, 1993.

[3] L. Alvarez, P. L. Lions, and J. M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM-JNA* **29**, pp. 845-866, 1992.

[4] G. Aubert and L. Vese. A variational method for image recovery. *SIAM J. of App. Math.* to appear.

[5] P. J. Besl, J. B. Birch, and L. T. Watson. Robust window operators. In *Proc. Int. Conf. on Comp. Vision, ICCV-88*, pages 591–600, 1988.

[6] M. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision* **19**, pp. 57-92, July 1996.

[7] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. Computer Vision and Pattern Recognition, CVPR-91*, pages 296–302, Maui, Hawaii, June 1991.

[8] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. Int. Conf. on Computer Vision, ICCV-93*, pages 231–236, Berlin, Germany, May 1993.

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

[10] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[11] F. Catte, P-L. Lions, J-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM-JNA* **29**, pp.182-193, 1992.

[12] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans.Image Processing*, to appear.

[13] D. S. Chen and B. G. Schunck. Robust statistical methods for building classification procedures. In *Proc. Int. Workshop on Robust Computer Vision*, pages 72–85, Seattle, WA, October 1990.

[14] P. B. Chou and C. M. Brown. The theory and practice of Bayesian image labeling. *Int. Journal of Computer Vision*, 4(3):185–210, 1990.

[15] D. Gabor. Information theory in electron microscopy. *Laboratory Investigation* **14**, pp. 801-807, 1965.

[16] D. Geiger and A. Yuille, A common framework for image segmentation. *Int. J. of Comp. Vision* **6**, pp. 227-243, 1991.

[17] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[18] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):376–383, March 1992.

[19] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing* 4:7, pp. 932-946, 1995.

[20] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Trans. on Medical Imaging* **11**, pp. 221-232, 1992.

[21] F. Guichard. Multiscale analysis of movies: Theory and algorithms. *PhD Dissertation*, CEREMADE-Paris, 1993.

[22] 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.

[23] J. G. Harris, C. Koch, E. Staats, and J. Luo, Analog hardware for detecting discontinuities in early vision. *Int. J. Comp. Vision* **4**, pp. 211-223, 1990.

[24] P. J. Huber. *Robust Statistics*. John Wiley and Sons, New York, NY, 1981.

[25] S. Kichenassamy. Edge localization via backward parabolic and hyperbolic PDE. University of Minnesota, pre-print, 1996.

[26] R. Kumar and A. R. Hanson. Analysis of different robust methods for pose refinement. In *Proc. Int. Workshop on Robust Computer Vision*, pages 167–182, Seattle, WA, October 1990.

[27] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3(1):73–102, 1989.

[28] L. Lui, B. G. Schunck, and C. C. Meyer. On robust edge detection. In *Proc. Int. Workshop on Robust Computer Vision*, pages 261–286, Seattle, WA, October 1990.

[29] R. Malladi and J. A. Sethian. Image processing: Flows under min/max curvature and mean curvature. *Graphical Models and Image Processing*, 58(2):127–141, March 1996.

[30] C. Mead. *Analog VLSI and Neural Systems*, Addison-Wesley, New York, 1989.

[31] P. Meer, D. Mintz, and A. Rosenfeld. Robust recovery of piecewise polynomial image structure. In *Proc. Int. Workshop on Robust Computer Vision*, pages 109–126, Seattle, WA, October 1990.

[32] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim. Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[33] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(2):220–228, March 1987.

[34] R. Nevatia and K. R. Babu. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13:257–269, July 1980.

[35] M. Nitzberg and T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE-PAMI* **14**, pp. 826-833, 1992.

[36] N. Nordström, "Biased anisotropic diffusion: A unified regularization and diffusion approach to edge detection," *Image and Vision Computing* **8**, pp. 318-327, 1990.

[37] S. Osher and L. I. Rudin, Feature-oriented image enhancement using shock filters. *SIAM-JNA* **27**, pp. 919-940, 1990.

[38] 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.

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

[40] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D* **60**, pp. 259-268, 1992.

[41] G. Sapiro, From active contours to anisotropic diffusion: Relations between basic PDE's in image processing. *Proc. ICIP*, Lausanne, September 1996.

[42] G. Sapiro and D. Ringach, Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. on Image Processing* **5**, pp. 1582-1586, 1996.

[43] B. G. Schunck. Image flow segmentation and estimation by constraint line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(10):1010–1027, October 1989.

[44] B. G. Schunck. Robust computational vision. In *Proc. Int. Workshop on Robust Computer Vision*, pages 1–18, Seattle, WA, October 1990.

[45] J. Shah. A common framework for curve evolution, segmentation, and anisotropic diffusion. *Proc. CVPR*, San Francisco, June 1996.

[46] D. Shulman and J. Hervé. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–85, Irvine, CA, March 1989. IEEE Computer Society Press.

[47] S. S. Sinha and B. G. Schunck. A two-stage algorithm for discontinuity-preserving surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1):36–55, January 1992.

[48] A. P. Tirumalai, B. G. Schunck, and R. C. Jain. Robust dynamic stereo for incremental disparity map refinement. In *Proc. Int. Workshop on Robust Computer Vision*, pages 412–434, Seattle, WA, October 1990.

[49] J. Weng and P. Cohen. Robust motion and structure estimation using stereo vision. In *Proc. Int. Workshop on Robust Computer Vision*, pages 367–388, Seattle, WA, October 1990.

[50] Y. L. You, W. Xu, A. Tannenbaum, and M. Kaveh. Bahavioral analysis of anisotropic diffusion in image processing. *IEEE Trans. IP* **5**, pp. 1539-1553, 1996.

[51] J. Weickert. Non-linear diffusion scale-spaces: From the continuous to the discrete setting. *Proc. ICAOS '96*, pp. 111-118, Paris, June 1996.