Robust anisotropic diffusion: Connections between robust statistics, line processing, and anisotropic diffusion

M. J. Black,[1] G. Sapiro,[2] D. Marimont,[1] and D. Heeger[3]

[1] Xerox PARC, Palo Alto, CA 94034
[2] HP Labs, Palo Alto, CA 94304
[3] Stanford University, Stanford, CA 94305

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. Finally, connections between robust estimation and line processing provide a framework to introduce spatial coherence in anisotropic diffusion flows.

1 Introduction

Since the elegant formulation of anisotropic diffusion introduced by Perona and Malik [5], 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, 6] 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 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, can be found in [2]. The connections to line processing present a framework for introducing spatial coherence in anisotropic diffusion [2]. Extensions of this theory to vector-valued images (e.g., color) and to robust sharpening are discussed in detail in [3].

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 $I(x, y, 0) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ 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. Perona and Malik [5] replaced the classical isotropic diffusion equation with $\frac{\partial I(x,y,t)}{\partial t} = \mathrm{div}(g(\| \nabla I \|)\nabla I)$, 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 $I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(\nabla I_{s,p}) \nabla I_{s,p}$, 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. They linearly approximated the image gradient (magnitude) in a particular direction as $\nabla I_{s,p} = I_p - I_s^t, \;\; p \in \eta_s$.

3 Robust estimation

We assume that an image is a piecewise constant function that has been corrupted by zero-mean Gaussian noise with small variance. The differences between pairs of corrupted pixels in the same (originally constant) region will be small and normally distributed. This is not true for pixels across edges. In estimating the brightness of the image on one side of a brightness discontinuity, measurements from the other side should be "rejected" as "outliers" that violate the statistical assumptions. The field of robust statistics [4] is concerned with estimation problems such as this in which the data contains gross errors, or outliers.

Motivated then by robust statistics, we wish to find an image I that satisfies the optimization criterion $\min_I \sum_{s \in I} \sum_{p \in \eta_s} \rho(I_p - I_s, \sigma)$, where $\rho(\cdot)$ is a robust error norm and σ is a "scale" parameter related to the rejection of outliers, and is discussed in [2]. To minimize this the intensity at each pixel must be "close" to those of its neighbors. This equation can be solved by gradient descent: $I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} \psi(I_p - I_s^t, \sigma)$, 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 is critical. To analyze the behavior of a given ρ-function, we consider its derivative ψ, which is proportional to the *influence function* [4]. This function characterizes the bias that a particular measurement has on the solution. To increase robustness and *reject* outliers, the ρ-function must be more forgiving about outliers than the classical quadratic norm; that is, it should increase less rapidly than x^2.

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. The continuous form of the robust estimation problem can be posed as $\min_I \int_\Omega \rho(\| \nabla I \|) d\Omega$, where Ω is the domain of the image and where we have omitted σ for notational convenience. One way to minimize this energy is via gradient descent: $\frac{\partial I(x,y,t)}{\partial t} = \operatorname{div}\left(\rho'(\| \nabla I \|)\frac{\nabla I}{\|\nabla I\|}\right)$. By defining $g(x) \doteq \frac{\rho'(x)}{x}$, we obtain the straightforward relation between image reconstruction via robust estimation and image reconstruction via anisotropic diffusion.

Perona and Malik suggested two different edge stopping functions ($g(\cdot)$) which can be viewed in the robust statistical framework by converting them into their related ρ-functions. It is easy to show for example that the function $g(x) = 1/(1 + x^2/K^2)$ proposed by Perona and Malik corresponds to the Lorentzian norm of robust statistics [2].

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? 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, such as *Tukey's biweight*

$$\rho(x,\sigma) = \begin{cases} \frac{x^2}{\sigma^2} - \frac{x^4}{\sigma^4} + \frac{x^6}{3\sigma^6} & |x| \le \sigma \\ \frac{1}{3} & \text{otherwise} \end{cases} \qquad \psi(x,\sigma) = \begin{cases} x(1-(x/\sigma)^2)^2 & |x| \le \sigma, \\ 0 & \text{otherwise.} \end{cases}$$

A detailed comparison of these and other norms, showing the preference for Tukey's biweight, can be found in [2]. Figure 1 compares these functions and the results of diffusing with each of them. The value of σ and λ are estimated automatically from the image gradients [2]. By examining the shape of the ψ-function, edges in the image can be interprated as occuring at locations where the gradient is treated as an outlier.

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.* [7] proposed a formulation that is equivalent to using the L_1 norm. You *et al.* [8] 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.

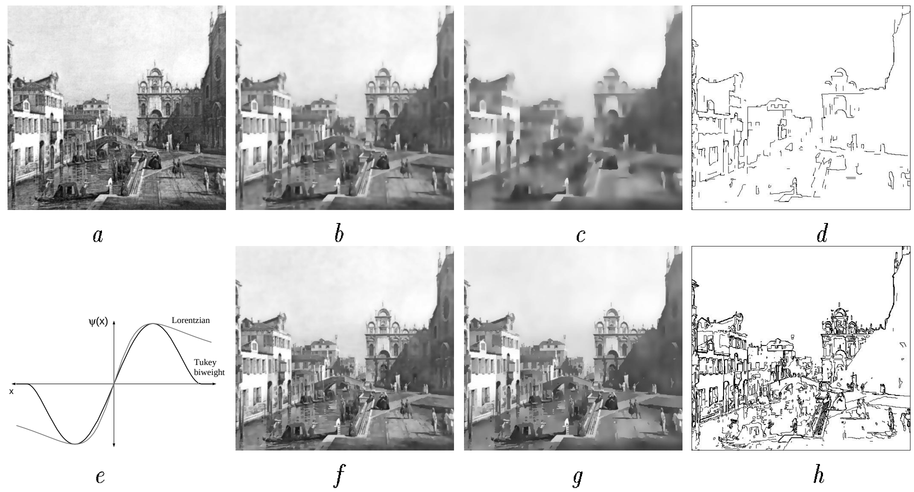

Fig. 1. *Comparison of the Perona-Malik (Lorentzian) function and the Tukey function.* (a) *original image;* (e) *comparison of Lorentzian and Tukey ψ-functions;* (b,f) *after 100 iterations of the Lorentzian and Tukey formulations respectively;* (c,g) *after 500 iterations;* (d,h) *edges obtained from the outliers after 500 iterations of Lorentzian and Tukey respectively.*

The connection between anisotropic diffusion and robust statistics is further exploited in [2] to take a diffusion equation with an edge stopping function $g(x)$ and convert it into an equivalent diffusion problem with an *explicit* “line process”. Making the line process explicit allows constraints on the spatial coherence of edges to be introduced into the diffusion equation.

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 Trans. IP*, to appear.
3. M. Black, G. Sapiro, D. Marimont, D. Heeger, “Robust anisotropic diffusion and sharpening of scalar and vector images,” January 1997, submitted.
4. 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.
5. P. Perona and J. Malik. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, pp. 629-639, 1990.
6. B. Romeny (Ed.). *Geometry Driven diffusion in Computer Vision*, Kluwer, 1994.
7. L. I. Rudin, S. Osher, and E. Fatemi. *Physica D* **60**, pp. 259-268, 1992.
8. Y. L. You, W. Xu, A. Tannenbaum, and M. Kaveh. *IEEE Trans. IP* **5**, pp. 1539-1553, 1996.

This article was processed using the LaTeX macro package with LLNCS style