Image Statistics and Anisotropic Diffusion

Hanno Scharr [*‡]   **Michael J. Black** [†]   **Horst W. Haussecker** [*]

[*]Intel Research, 2200 Mission College Blvd, Santa Clara, CA 95054, USA
[†]Department of Computer Science, Brown University, Box 1910, Providence, RI 02912, USA
[‡]ICG III, Research Center Jülich, 52425 Jülich, Germany

Abstract

Many sensing techniques and image processing applications are characterized by noisy, or corrupted, image data. Anisotropic diffusion is a popular, and theoretically well understood, technique for denoising such images. Diffusion approaches however require the selection of an "edge stopping" function, the definition of which is typically ad hoc. We exploit and extend recent work on the statistics of natural images to define principled edge stopping functions for different types of imagery. We consider a variety of anisotropic diffusion schemes and note that they compute spatial derivatives at fixed scales from which we estimate the appropriate algorithm-specific image statistics. Going beyond traditional work on image statistics, we also model the statistics of the eigenvalues of the local structure tensor. Novel edge-stopping functions are derived from these image statistics giving a principled way of formulating anisotropic diffusion problems in which all edge-stopping parameters are learned from training data.

1 Introduction

We consider the problem of reconstructing a high-quality image from a single noisy example. Optimal solutions to this problem require knowledge of the noise statistics and the spatial statistics of the "true" image. We focus on situations in which we can *learn* these image statistics from examples and then show how they can be used for denoising. In doing so we make connections between Bayesian inference, image statistics, and anisotropic diffusion. These insights lead to new edge-stopping functions for diffusion algorithms which are derived from a sound statistical basis. In particular, the parameters of these algorithms are automatically learned from training data.

Anisotropic diffusion has become a widely used technique with a well understood computational theory (see e.g. [27]). It has been used for various applications (e.g. [3, 11, 25]) and there are many different implementation approaches. Besides the simple, conventional, two-level explicit finite-difference schemes [10, 23, 27, 29] as used here, there are, for instance, semi-implicit approaches [9], multi-grid methods [1] or adaptive finite element techniques [4].

Central to the technique is a "diffusivity" or "edge stopping" function that controls the degree of image smoothing. Despite the mathematical rigor of current anisotropic diffusion techniques, the choice of this function is typically ad hoc. Additionally, all of these functions have parameters (e.g. scale parameters) which influence the enhancement result and must be tuned in some way (typically by hand). Black and Sapiro [7] used local image statistics to automatically set some of the relevant parameters. We go a step further and use learned image statistics to design both the edge-stopping function and its parameters in a principled, automatic, way. While we focus on images of man-made scenes here, the method is general and can be applied to other classes of imagery such as natural scenes where the spatial statistics are easily learned from training data.

The statistics of natural images have recently received a great deal of attention [17, 18, 19, 21, 22] and a number of authors have shown how they can be exploited for denoising [8, 30], edge detection [13, 17], and tracking [24]. Most work has looked at the spatial statistics of images in terms of empirical probability distributions (normalized histograms) of first derivative filters. Many authors have observed the non-Gaussian nature and scale invariance of these statistics. We observe that, if these empirical statistics are to be exploited in an algorithm, then they must correspond to the computational implementation of the filters in that algorithm. In particular, anisotropic diffusion is often formulated as in terms of a continuous partial differential equation (PDE) which is then discretized. Different discretizations result in algorithms with different filter kernels.

In a probabilistic model, these spatial statistics represent the prior probability of image filter responses. In addition to this prior term we can model a likelihood, or data term that encodes the conditional probability of observing a noisy image given the true image. Using training data with corresponding smooth and noisy image pairs, we learn a statistical model of the deviations between the pairs; i.e. the (non-Gaussian) observation noise.

The prior and the likelihood could then be exploited in a Bayesian formulation of the image reconstruction problem. Optimization of such a formulation would give a max-
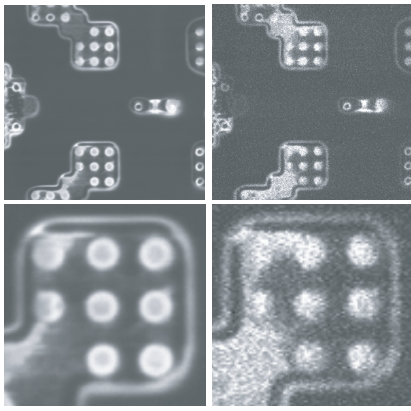
1

Figure 1: Example images (see text). Top left: high quality image. Top right: lower quality image. Below: sub-region of images above.

Figure 2: Empirical image noise statistics for silicon chip images. **(a)** distribution of image noise ($f_i - g_i$). **(b)** log of image noise distribution. **(c, d)** log probability of horizontal and vertical image derivatives.

imum a posteriori estimate of the true image but would require expensive stochastic optimization methods such as simulated annealing. Instead we formulate a non-optimal, but deterministic, problem and minimize the negative logarithm of the Bayesian formulation. We note that this corresponds to a standard objective function formulation used for anisotropic diffusion. Abandoning a Bayesian interpretation, we are free to formulate this objective function either *with or without* a data term. While the data term helps prevent the smoothed image from deviating inappropriately from the original, diffusion schemes commonly omit it.

Our automatic approach derives a statistically motivated edge-stopping function from training data and achieves competitive results compared with traditional, hand-tuned, methods. We start with a particular discretization of a continuous PDE and then compute the correct *algorithm specific* statistics from a set of training images. For images of "man-made" scenes we find similar statistics to those previously reported [14, 18] and show that a probabilistic mixture of a t-distribution and a normal distribution models this data well. Furthermore, we extend results in image statistics to capture locally oriented image structure by modeling the eigenvalues of the structure tensor. We then develop a novel mechanism for converting the spatial statistics into the appropriate edge-stopping function. We illustrate the application of these methods with various diffusion formulations and apply them to various images of man-made scenes.

2 Image Statistics

Our approach exploits learned models of image statistics. While there has been a great deal of recent work on modeling the statistics of natural scenes, here we first focus on a specific class of images for which we can carefully model the appropriate statistics and measure the results. The re-
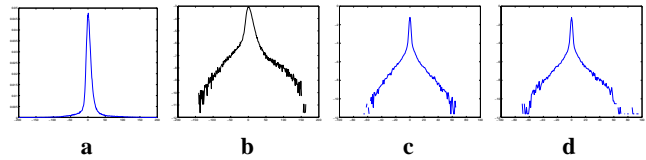
sults here however are applicable to general imagery and we draw connections to natural image statistics.

Consider the images of a silicon structure in Fig. 1 which are produced by a direct read-write imaging system. The image on the left is the result of a high quality "scan" and will form what we consider the "ground truth". The image acquisition process is destructive and the higher quality the scan the more material is removed. Consequently it may be desirable to work with lower-quality, less destructive, images such as the one on the right[1]. From the low-quality image, the task is to reconstruct a high quality one such that it approximates the ground truth[2]. This particular example has the nice property that the noise comes from a real, but unknown, process and we know approximately what the recovered image should look like.

We begin by measuring the deviation between the high quality and noisy images. Let **g** be the "true" image and let **f** be the measured image. We can define the generative model of the image values f_i at pixel i as $f_i = g_i + \eta$ where η represents the image noise which is distributed according to some, as yet, unknown distribution that we will measure from our training images. Fig. 2**a** shows the empirical distribution (normalized histogram) of pixel intensity differences, $f_i - g_i$, between the images. This distribution has a characteristic shape which is made more readily apparent by looking at the empirical log probability in Fig. 2**b**.

We also model the spatial image statistics of **g** which, for this initial example, we compute using simple neighborhood differences. This corresponds to a simple derivative filter and, as we will see later, more appropriate filter kernels can be exploited for anisotropic filtering. The log marginal probability of horizontal and vertical derivative filter responses are shown in Fig. 2**c** and Fig. 2**d** respectively.

One might suspect that the spatial statistics observed here are somehow special to this type of image data. It is interesting to observe, however, that these statistics are common to wide class of images of man-made objects. Consider the images of buildings in Fig. 3**a**, and **b** [3]. Lee and Mum-

---

[1]Similar situations arise with other imaging technology such as low-dosage X-ray or functional MRI.

[2]The high-quality scan is, itself, only an approximation to the "true" image which is not observable.

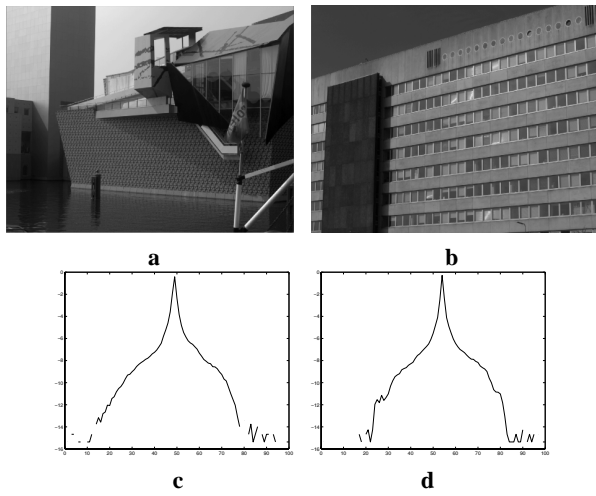[3]Images are from the natural image collection [26] available from:

Figure 3: Images of "man-made" scenes (**a**, **b**). Log probability of horizontal and vertical image derivatives (**c**, **d**).
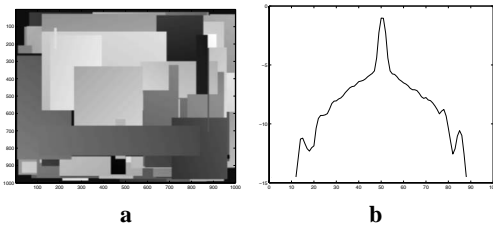

Figure 4: (**a**) Synthetic image with statistics similar to man-made scenes (similar to the "dead-leave" model [18]). (**b**) Empirical log probability of the horizontal and vertical derivatives.

ford [18] note that images such as these have characteristic spatial statistics and that these are distinct from "natural" images of, for example, vegetation, water, and sky. The log probability of horizontal and vertical derivatives in these images are shown in Fig. 3**c** and **d**. Note the similarity with the spatial statistics in Fig. 2 **c** and **d**.

Lee an Mumford explore a "dead leaves" model of image formation that can account for these and other spatial statistics. We take a similar approach. Consider the synthetic image in Fig. 4**a**. This was generated by choosing a randomly sized rectangle, in a random location, with a linear brightness function the parameters of which were chosen uniformly over reasonable ranges. The result is a patchwork of overlapping rectangles with smooth brightness gradients. The image was then smoothed to simulate the optical blurring of the image discontinuities and then Gaussian noise was added. The resulting spatial statistics have qualitatively the same form as those observed above in images of man-made objects (Fig. 4**b**). In particular, they have a peak near zero and then broad "shoulders."

http://hlab.phys.rug.nl/archive.html

2.1 A Mixture Model of Man-Made Image Statistics

In this synthetic case above, the filter outputs come from one of two populations. If there was no additive Gaussian noise, the smoothed piecewise constant image would give rise to image statistics that can be well approximated by a t-distribution [14]. Huang and Mumford fit this t-distribution to natural images and find that t should be approximately 2.6. Here the image derivatives are perturbed by the addition of Gaussian noise within the uniform regions. Consequently we propose to model the empirical distribution as a probabilistic mixture of a t-distribution and a Gaussian.

Let X be a random variable representing the filter response for an image (data difference, horizontal derivative, or vertical derivative). Formally, the probability of a filter response, $p(X = x)$, is given by the mixture model

$$p(X = x) = \frac{w}{Z}(1 + \frac{x^2}{\sigma_1^2})^{-t} + (1 - w)\frac{1}{\sqrt{2\pi}\sigma_2}\exp(-\frac{x^2}{2\sigma_2^2}) \quad (1)$$

where $0 \leq w \leq 1$ is the mixture proportion of the two terms and Z is a normalization factor insuring that the first term integrates to 1.

The first term in the mixture is the t-distribution which captures the edge statistics and the second term captures the Gaussian noise. We note that the data noise statistics for the image in Figure 1 are also well modeled by this mixture and we exploit this below. This mixture model has four free parameters which we fit to the marginal distributions: the mixture proportion w, the degree t, and the two scale terms σ_1 and σ_2.

We now formulate the problem of recovering the image $\mathbf{g}$ from $\mathbf{f}$ as the maximization of

$$p(\mathbf{g}|\mathbf{f}) \propto \prod_i \left(p(f_i|g_i) \prod_{j=1}^{J} p(\mathbf{n}_j \nabla g_i) \right) \quad (2)$$

where $p(\mathbf{g}|\mathbf{f})$ approximates the posterior probability of the image $\mathbf{g}$ conditioned on the observed, noisy, image $\mathbf{f}$. The likelihood term, $p(f_i|g_i)$, at every pixel, i, is defined by our measured image statistics (e.g. Fig. 2**a**). The spatial prior term exploits a Markov Random Field assumption [12] which defines the prior in terms of local neighborhood properties. Here it is defined in terms of the spatial derivatives, ∇g_i, at a pixel i, in J different directions $\mathbf{n}_j$, and uses the learned image statistics to assess the prior probability.

2.2 Formulating an Objective Function

Maximizing the posterior probability in Eq. 2 is challenging. Rather than seek an exact, global, solution, we re-cast the problem as one of minimizing the negative logarithm of Eq. 2. Our goal here is to construct a simple objective func-
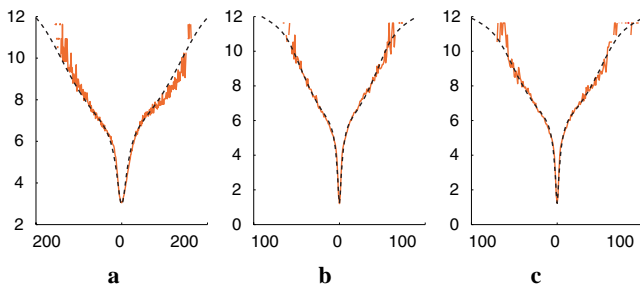
3

Figure 5: Mixture model fit to image statistics. Smooth (dashed) plots show the fit of the mixture of a t-distribution and a Gaussian to the image statistics from Fig. 2. The negative logarithm is shown here to illustrate the shape of the ρ-function. (**a**) data noise statistics. (**b**, **c**) horizontal and vertical derivative statistics (computed via neighbor differences). In fitting the mixture model we find the degree of the t-distribution to be: $t_{data} = 1.43$, $t_{dx} = 1.27$, $t_{dy} = 1.22$.

tion that can be minimized using gradient descent. That is, we formulate an objective function

$$E(\mathbf{g}) = -\sum_i \left(\log(p(f_i|g_i)) + \lambda \sum_{j=1}^{J} \log(p(\mathbf{n}_j \nabla g_i)) \right) \quad (3)$$

where we have introduced a weight term λ which accounts for the confidence one has in the different model terms; this is common in regularization approaches. In a pure diffusion scheme, the data term and λ are dropped but one must set a parameter to stop the algorithm. Either scheme results in one free parameter that must be determined manually.

Recall that the likelihood and the prior terms are defined as mixture models. Consequently, the log probability here does not have a simple form. What we seek is a function, $\rho(x) \approx -\log(p(x))$, that is computationally tractable. A variety of ρ-functions have been proposed for similar image smoothing problems (see [5] for a review and [6] for the application of such ρ-functions to diffusion processes).

In contrast to previous work where ad hoc ρ-functions are used, here we construct ρ-functions that are specific to the measured image statistics. Toward that end, we first fit the mixture model to the empirical distributions (using Matlab's lsqnonlin function). Fig. 5 shows the fitted model overlaid on the negative logarithm of these distributions.

To construct a computationally tractable approximation to the logarithm of this mixture model we model it using a third order B-spline. The spline can be evaluated and differentiated easily at low numerical cost. Using these novel ρ-functions, our objective function finally becomes

$$E(\mathbf{g}) = \sum_i \left(\rho_0(g_i - f_i) + \lambda \sum_{j=1}^{J} \rho_j(|\mathbf{n}_j \nabla g_i|) \right). \quad (4)$$

Note, we observe the image statistics to be symmetric and

thus can use either $\mathbf{n}_j \nabla_\sigma g_i$ or $|\mathbf{n}_j \nabla_\sigma g_i|$ in the spatial term.

Local optimization of Eq. 4 is performed using gradient descent. While simple and effective, this does not guarantee convergence to a global minimum. An analysis of the resulting computational scheme reveals that it is a discrete implementation of a diffusion process. We study the numerical properties of this scheme in the following section. This provides a connection to the diffusion literature where numerically stable diffusion processes have been studied in detail. The question is: What is a good scheme and how is it related to the objective function? Or: How can we obtain the benefits of a numerically stable diffusion scheme and still exploit the local image statistics? The key idea will be to pose a computational scheme, determine the corresponding filter kernels it implies, and then compute the image statistics *appropriate to that kernel*. The result, as described in the following section, is a statistically and numerically sound scheme in which the edge stopping function and its parameters are learned from image statistics.

3 Diffusion

In this section we will draw the connection between image statistics and anisotropic diffusion.

From image statistics to anisotropic diffusion.

We derive a partial differential equation (PDE) from the objective function in Eq. 4. Toward this end, we convert the discrete problem into the continuous formulation

$$E(g, \nabla g) = \int_\Omega \rho_0(g(\mathbf{x}) - f(\mathbf{x})) + \lambda \sum_{j=1}^{J} \rho_j(|\mathbf{n}_j \nabla_\sigma g(\mathbf{x})|) \mathrm{d}\mathbf{x}$$

where $\mathbf{x}$ represents an (x, y) image position over the image domain Ω. Rather than the discrete images $\mathbf{g}$ and $\mathbf{f}$ we now work with continuous functions g and f. As before, $\mathbf{n}_j$ is a unit vector indicating the neighbor direction and $\nabla_\sigma g = \int_\Omega \omega(\mathbf{x} - \mathbf{x}') \nabla g(\mathbf{x}') \mathrm{d}\mathbf{x}'$ is a smoothed version of (the ideal) ∇g where $\omega(\cdot)$ is the smoothing kernel. Note that the smoothing denoted by σ is *implicit* in the choice of filter kernel used to compute $\nabla_\sigma g$.

We minimize E via the calculus of variations, which leads us to the following optimality criterion

$$0 = \lambda \nabla_\sigma^T \sum_{j=1}^{J} \psi_j(|\mathbf{n}_j^T \nabla_\sigma g(\mathbf{x})|) \mathbf{n}_j \mathbf{n}_j^T \nabla_\sigma g(\mathbf{x}) - \rho_0'(g(\mathbf{x}) - f(\mathbf{x})) \quad (5)$$

where ρ' denotes the derivative of ρ and we define

$$\psi_j(x) = \rho_j'(x)/x. \quad (6)$$

Note that the ρ_j (or ψ_j) functions may differ in each direction, j, as they represent the appropriate image statistics. $\psi(x)$ is the classic edge-stopping function [6, 20].

4

Now, what does Eq. 5 tell us? Let us explore two examples using different neighborhoods. In a 4-neighborhood the two neighbor directions are $\mathbf{n}_1 = (1,0)^T$ and $\mathbf{n}_2 = (0,1)^T$. The first term in Eq. 5 can be rewritten as an anisotropic kernel oriented along the coordinate axes

$$\nabla_\sigma^T \begin{pmatrix} \psi_1(|\partial_{x,\sigma}g|) & 0 \\ 0 & \psi_2(|\partial_{y,\sigma}g|) \end{pmatrix} \nabla_\sigma g.$$

Alternatively we might use an 8-neighborhood where we add the diagonal neighbor directions $\mathbf{n}_3 = (\sqrt{2}/2, \sqrt{2}/2)^T$ and $\mathbf{n}_4 = (\sqrt{2}/2, -\sqrt{2}/2)^T$. This gives fully anisotropic diffusion as the first term in Eq. 5 becomes

$$\nabla_\sigma^T \begin{pmatrix} \psi_1 + \frac{1}{2}(\psi_3 + \psi_4) & \frac{1}{2}(\psi_3 - \psi_4) \\ \frac{1}{2}(\psi_3 - \psi_4) & \psi_2 + \frac{1}{2}(\psi_3 + \psi_4) \end{pmatrix} \nabla_\sigma g.$$

with an abbreviated notation $\psi_i \hat{=} \psi_i(|\mathbf{n}_i^T \nabla_\sigma g|)$. In general the smoothness term is of the form

$$\nabla_\sigma^T \mathbf{D} \nabla_\sigma g \tag{7}$$

where $\mathbf{D}$ is a symmetric 2×2 *diffusion tensor*.

In the 4-neighborhood case the diagonal diffusion tensor results in filter kernels oriented along the coordinate axes. As a result it is not rotationally invariant and may introduce artifacts. Using the 8-neighborhood we can construct any symmetric 2×2 diffusion tensor and thus are able to rotate the kernel along any direction.

Please note that the numerical properties of the derivative operator introduced above are carried over to ∇_σ in this term. Thus there is a one to one correspondence between the operator used for neighborhood statistics and the numerical scheme used to implement the diffusion.

The optimality criterion derived in Eq. 5 can be considered to be the steady state solution of

$$\lambda \nabla_\sigma^T \mathbf{D} \nabla_\sigma g - \rho_0'(g - f) = \partial_t g \tag{8}$$

using the notation from Eq. 7. This is an anisotropic diffusion equation with a source term. It means that we evolve an initial image, $g^{(0)}$, under this diffusion. For implementation convenience we choose simple explicit discretization of the temporal derivative. The spatial term $\nabla_\sigma^T \mathbf{D} \nabla_\sigma$ is discretized using convolution filters for the derivatives. The update scheme now reads

$$g^{(t+\tau)} = (1 + \tau \lambda \nabla_\sigma^T \mathbf{D} \nabla_\sigma) g^{(t)} - \tau \rho_0'(g^{(t)} - f) \tag{9}$$

with the time step size τ. This parameter τ has little effect on the solution, as long as it is small enough.

Above, we assumed simple $[1, -1]$ derivative filters. More generally, we adopt the filters

$$\partial_x = \frac{1}{32}\begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix}, \quad \partial_y = \frac{1}{32}\begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix} \tag{10}$$

which are recommended in [23, 29]. Thus we get two update schemes using a 4 neighborhood. We denote the one using $[1, -1]$ by **2tap** and the one using the 3×3-filters above by **3×3**. We refer to [23, 29] for related numerical issues, including the choice of τ, stability and performance.

Image statistics and the structure tensor.

Typically, the diffusion tensor for rotationally invariant anisotropic diffusion is not constructed via an 8-neighborhood, but rather via the structure tensor

$$\mathbf{J} = G_\phi * \nabla_\sigma g \nabla_\sigma g^T \tag{11}$$

$$= \sum_j \mu_j \mathbf{e}_j \mathbf{e}_j^T \tag{12}$$

where G_ϕ is a Gaussian kernel with variance ϕ, μ_j are the eigenvalues, and $\mathbf{e}_j$ are the normalized eigenvectors of $\mathbf{J}$ (c.f. [23, 27, 29]). The eigenvalues are sorted in descending order by size and can be expressed by

$$\mu_k = G_\phi * (\mathbf{e}_j^T \nabla_\sigma g)^2. \tag{13}$$

The diffusion tensor $\mathbf{D}$ is now given by

$$\mathbf{D} = \sum_j d(\mu_j) \mathbf{e}_j \mathbf{e}_j^T \tag{14}$$

where the function d is often chosen in an ad hoc way.

Adopting our approach, we derive d in the same way we derived ψ in Eq. 6 above. We start by building a normalized histogram for each eigenvalue and thus estimate the empirical probabilities $p_j(\mu_j)$. As before, the negative logarithm of this gives $\rho_j(\mu_j)$ and the objective function we derive is

$$E(g, \nabla g) = \int_\Omega \rho_0(g(\mathbf{x}) - f(\mathbf{x})) + \lambda \sum_j \rho_j(\mu_j) \mathrm{d}\mathbf{x}.$$

Via calculus of variations we derive the optimality criterion

$$0 = \lambda \nabla_\sigma^T \sum_j \rho_j'(\mu_j) \mathbf{e}_j \mathbf{e}_j^T 2 G_\phi * \nabla_\sigma g(\mathbf{x}) - \rho_0'(g(\mathbf{x}) - f(\mathbf{x}))$$

$$= \lambda \nabla_\sigma^T \mathbf{D} \nabla_\sigma g_\phi(\mathbf{x}) - \rho_0'(g(\mathbf{x}) - f(\mathbf{x})) \tag{15}$$

with the notation $g_\phi := G_\phi * g$ and the diffusion tensor

$$\mathbf{D} = 2 \sum_j \rho_j'(\mu_j) \mathbf{e}_j \mathbf{e}_j^T.$$

This criterion is very similar to Eq. 5, but features the following differences: 1. Smoothing is performed along local eigenvectors $\mathbf{e}$ of the structure tensor $\mathbf{J}$ rather than fixed directions $\mathbf{n}$. 2. As μ depends on squared derivatives, the diffusivity terms in $\mathbf{D}$ are built using ρ' instead of ψ (see Eq. 6). 3. The smoothing kernel G_ϕ is carried over to the gradient of g.

Interestingly, the statistics of the eigenvalues μ_j, can also be modeled by the mixture model from Sec. 2.1. We note,
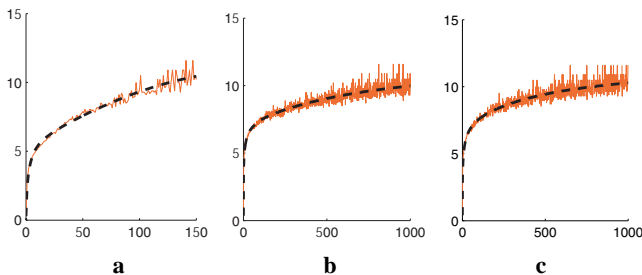
IEEE
COMPUTER
SOCIETY

Figure 6: Image statistics (negative logarithm) using eigenvalues μ computed with the derivative filters given in Eq. 10. Smooth (dashed) plots show the fit of the mixture model. **a**: μ_1. **b**: μ_2 using a 3×3 binomial G_ϕ. **c**: μ_1 for $\phi = 0$ (μ_2 is not used in this case).

however, that $\mu_1 = (\nabla_\sigma g)^2$ when $\phi = 0$ and consequently we replace x^2 in Eq. 1 by μ (note that μ is always positive or zero)

$$p(\mu) = \frac{w}{Z}(1 + \frac{\mu}{\sigma_1^2})^{-t} + (1-w)\frac{1}{\sqrt{2\pi}\sigma_2}\exp(-\frac{\mu}{2\sigma_2^2}). \quad (16)$$

The negative logarithm of the histograms of the eigenvalues and their fits by this mixture model are plotted in Fig. 6. This provides a novel view of oriented image statistics in terms of the structure tensor.

This optimality criterion is similar to that of standard usual anisotropic diffusion with the only difference being that the diffusion update $\nabla^T \mathbf{D} \nabla g$ is derived from data smoothed by G_ϕ at each time step. Note that, when we choose $\phi = 0$, the criterion leads to isotropic nonlinear diffusion rather than anisotropic diffusion. This is due to the fact, that, in this case, $\nabla_\sigma g$ is the first eigenvector of $\mathbf{J}$ and thus of $\mathbf{D}$. Consequently $\mathbf{D} \nabla g = \rho_1'(\mu_1) \nabla g$ and the diffusion tensor collapses to a scalar diffusivity.

For the implementation of this diffusion, we use the scheme proposed in [23, 29] with the 3×3-derivative filters in Eq. 10 and introduce G_ϕ as needed. For G_ϕ we use either 1 or a 3×3-binomial filter and refer to the resulting schemes as **ST1** and **ST3**. The update scheme then reads as

$$g^{(t+\tau)} = (g^{(t)} + \tau\lambda\nabla_\sigma^T \mathbf{D}\nabla_\sigma G_\phi * g^{(t)}) - \tau\rho_0'(g^{(t)} - f). \quad (17)$$

Summary of the method.

To summarize the approach, one first chooses a diffusion scheme. The discretization of the continuous problem then defines the appropriate smoothing and derivative kernels. These are then used in conjunction with training data to build the empirical prior distributions, $p(\mathbf{n}_j \nabla_\sigma g_i)$ or $p(\mu_j)$. In the case of "man-made" images, we fit the proposed mixture model to the empirical distribution (other image data may require different models). We then approximate the negative log of this model with a B-spline to derive practical ρ, ρ' and ψ functions for all terms. Finally, Eq. 9 or

Eq. 17 give the deterministic update scheme.

4 Results

We illustrate the ideas here with denoising results for a variety of "man-made" images and compare different update schemes described in Sec. 3. Evaluating the performance of anisotropic diffusion schemes is difficult, consequently we focus here on demonstrating how various schemes behave with the automatically constructed diffusivities. We do not claim "better" reconstructions than previous methods. Rather our focus is on providing previous diffusion schemes with statistically grounded, and automatically generated, diffusivities. The results are qualitatively what one would expect but have the advantage of not requiring the hand tuning of parameters.

Fig. 1 presents noisy and "clean" micro-machining images. Our goal is recover the clean image by smoothing the noisy image in Fig. 1 in a way that exploits our knowledge of the image noise and underlying spatial statistics. We compute the data noise statistics and, for the **2tap**, **3×3**, **ST1** and **ST3** schemes, the appropriate spatial statistics. From these we derive the 4 optimization schemes in which the edge-stopping-functions are automatically determined.

We also compare these methods with a usual scheme [23, 29] in which G_ϕ does not appear outside the tensor. Here we use the edge-stopping functions derived from the structure tensor (**ST3**) and refer to this scheme as **AD3**.

Fig. 8 shows a variety of reconstruction results with, and without, the data term. With the data term, we must set the weight parameter λ (see Eq. 3) and we illustrate results for various settings. In the pure diffusion case we must set a stopping time T (here set to be $T = 1000$). For all schemes that use the derivative kernels from Eq. 10, we observe checkerboard-like artifacts as known from [29]. They can be suppressed using a fourth order term (compare [23] and Fig. 7). This term can also be set automatically via a statistical prior term with $\triangle g$ in the same way as the diffusion itself, but this is beyond the scope of this paper.

For the results without a data term (left column) we observe a smoother behavior than for the other columns. As expected, the smaller we choose λ, the lower the influence of the smoothing term and the closer the solution stays to the input data.

We have argued that images such as those in Fig. 1 and Fig. 3 have similar spatial statistics. While we do not know the data noise for the images in Fig. 3 we can still exploit the learned image statistics for anisotropic smoothing that is appropriate to this class of images. Here we run the diffusion without a source term and show the results for stopping time $T = 1000$. Details of results for two sub-regions of the image in Fig. 3**a** are shown in Fig. 9. We used the spatial statistics derived from the noiseless image in Fig. 1.
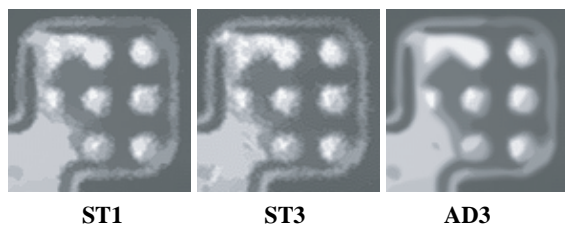
6

ST1   ST3   AD3

Figure 7: Denoising results using the estimated ρ-functions, $\lambda = 0.5$, and additional 5% of a fourth order correction term [23]: (compare Fig. 8 for results without this term).

The learned edge-stopping functions (ρ' or ψ depending on the algorithm) do a good job of preserving gray-level discontinuities. As expected we observe staircase artifacts where non-horizontal, non-vertical edges occur (see Fig. 9, rows 2 and 3) for the non-rotation-invariant schemes **2tap** and **3×3**.

5   Summary, Conclusion and Outlook

In this paper we develop and exploit the connection between image statistics and image reconstruction via anisotropic diffusion. This connection yields a principled way to automatically construct an edge-stopping function using the appropriate image statistics. In particular, we have considered a class of man-made images for which we show the image statistics to be well modeled by a mixture of a t-distribution and a normal distribution. Moreover, we modeled the statistics of both standard derivative kernels and the eigenvalues of the structure tensor. In recent work we have observed that the mixture model proposed here can also be applied to images of natural scenes.

From the statistical models we derive novel ρ-functions which can be used in a gradient descent optimization scheme. While many such ρ-functions have been proposed, to our knowledge, this is the first time one has been motivated by image statistics in this way. While more complex than previous such functions we derive a computationally tractable form by modeling the ρ-function (and its related diffusivity $\psi(x)$) using B-splines. It is worth noting that these ρ-functions are specialized to particular image classes and even to particular image dimensions as appropriate.

The coupling between the image statistics and a particular diffusion method is quite deep. In particular, we showed that the properties of derivative operators used to build the spatial prior (or smoothness) term are carried over from the diffusion scheme; i.e. the derivatives used to compute the image statistics and the derivatives used in the diffusion have to act on the same scale. This leads to the notion of algorithm-specific image statistics as introduced here.

We are currently exploring the use of image statistics to distinguish between different classes of images (e.g. vege-
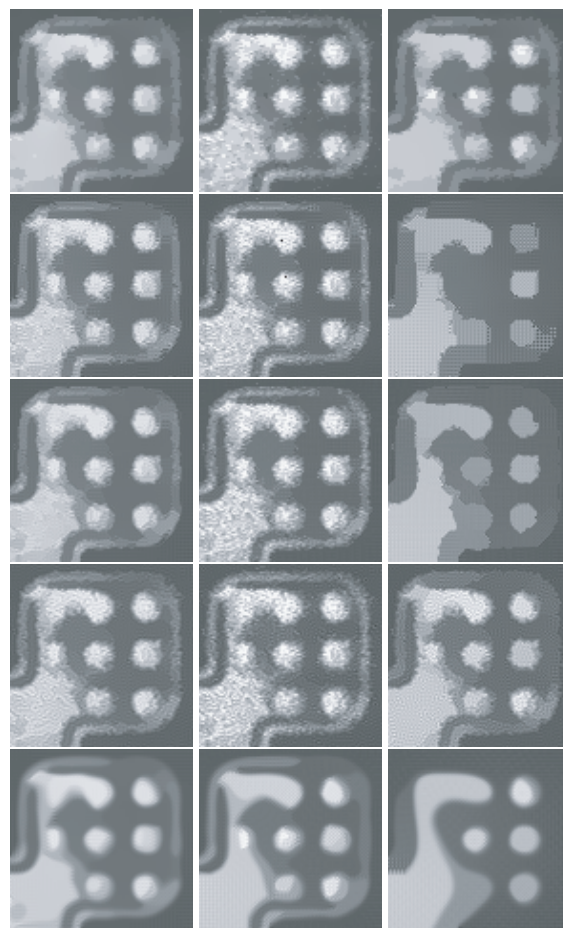

Figure 8: Denoising results using ρ-functions estimated from the images shown in Fig. 1. Top to bottom: results using **2tap**, **3×3**, **ST1**, **ST3**, and **AD3**. Left column: results without a data term, $T = 1000$. Middle and right columns: results using $\lambda = 0.5$ and $\lambda = 1$.

tation versus man-made) with the goal of adapting the diffusion method to the local statistics (c.f. [2, 7]). In this way we hope to automatically add higher-level contextual information into the diffusion process.

References

[1] S. Acton, *Multigrid anisotropic diffusion*, IEEE Trans. Image Proc., Vol. 7, 280–291, 1998.

[2] S. Arridge and A. Simmons, *Multi-spectral probabilistic diffusion using Bayesian classification*, Scale-Space'97, LNCS 1252, pp, 224–235, 1997.

[3] I. Bajla, I. Holländer, *Nonlinear filtering of magnetic resonance tomograms by geometry-driven diffusion*, Machine Vision and Applications, Vol. 10, 243–255, 1998.
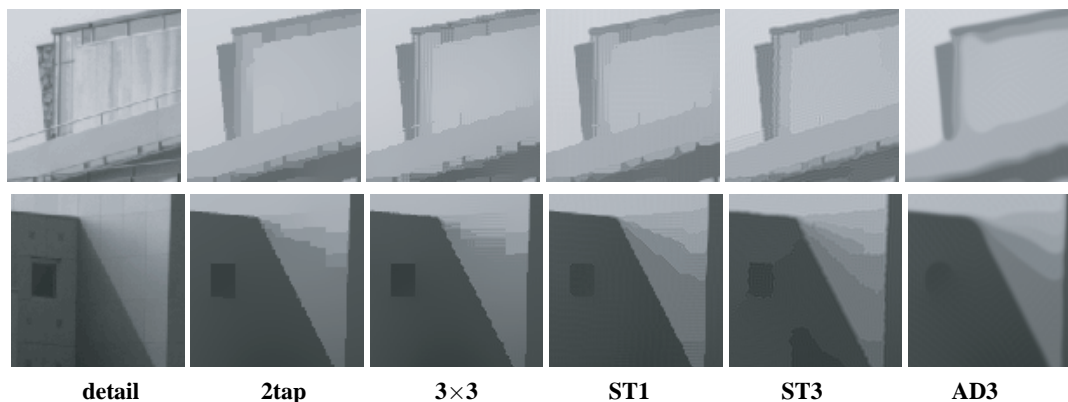
7

| detail | 2tap | 3×3 | ST1 | ST3 | AD3 |

Figure 9: Results of smoothing Fig. 3**a** using diffusion without a source term.

[4] E. Bänsch, K. Mikula, *A coarsening finite element strategy in image selective smoothing*, Computation and Visualization in Science, Vol. 1, 53–61, 1997.

[5] M. Black, A. Rangarajan, *On the unification of line processes, outlier rejection, and robust statistics with applications in early vision*, IJCV, 19(1):57–92, Jul. 1996

[6] M. Black, G. Sapiro, D. Marimont, D. Heeger, *Robust anisotropic diffusion*, IEEE Trans. Image Proc., 7(3):421–432, Mar. 1998.

[7] M. Black, G. Sapiro, *Edges as outliers: Anisotropic smoothing using local image statistics*, Scale-Space'99, LNCS 1682, Springer, Sept. 1999, pp. 259–270.

[8] R. Buccigrossi, E. Simoncelli, *Image compression via joint statistical characterization in the wavelet domain*, IEEE Trans. Image Proc., 8(12):1688–1701, 1999.

[9] F. Catté, P.-L. Lions, J.-M. Morel, T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., Vol. 29, 182–193, 1992.

[10] G.-H. Cottet, L. Germain, *Image processing through reaction combined with nonlinear diffusion*, Math. Comp., Vol. 61, 659–673, 1993.

[11] A. Frangakis, R. Hegerl, *Nonlinear anisotropic diffusion in three-dimensional electron microscopy*, Scale-Space'99, LNCS 1682, Springer, 386–397, 1999.

[12] S. Geman, D. Geman. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. PAMI*, 6(6):721–741, Nov. 1984.

[13] D. Geman, B. Jedynak, *An active testing model for tracking roads in satellite images* PAMI, 18(1):1–14, 1996

[14] J. Huang, D. Mumford, *Statistics of natural images and models*, CVPR, pp. 541–547, 1999.

[15] B. Jähne,, *Digital Image Processing*. Springer, 1997.

[16] B. Jähne, H. Scharr, S. Körkel, *Principles of filter design*, Handbook on Comp. Vision and Apps., Vol. 2, Academic Press, San Diego, 125–152, 1999.

[17] S. Konishi, A. Yuille, J. Coughlan, S. Zhu, *Fundamental bounds on edge detection: An information theoretic evaluation of different edge cues*, CVPR, pp. 573–579, 1999.

[18] A. Lee, D. Mumford, J. Huang, *Occlusion models for natural images: A statistical study of a scale-invariant dead-leaves model*, IJCV, 41(1/2):35–59, 2001.

[19] B. Olshausen, D. Field, *Natural image statistics and efficient coding*, Comp. Neural Sys., 7(2):333–339, 1996.

[20] P. Perona, J. Malik, *Scale space and edge detection using anisotropic diffusion*, PAMI, Vol. 12, 629–639, 1990.

[21] J. Portilla, E. Simoncelli, *A parametric texture model based on joint statistics of complex wavelet coefficients*, IJCV, 40(1):49–71, Dec 2000.

[22] D. Ruderman, *The statistics of natural images*, Network, 5(4):517–548, 1994.

[23] H. Scharr, H. Spies, D. Uttenweiler, B. Jähne, *Noise suppression for image sequence analysis by spatio-temporal anisotropic diffusion*, Image Sequence Analysis to Investigate Dynamic Processes, LNCS, Springer, 2003, in press.

[24] H. Sidenbladh, M. Black, *Learning image statistics for Bayesian tracking*, ICCV, Vol. II, pp. 709–716, 2001.

[25] A. Solé, A. López, C. Cañero, P. Radeva, J. Saludes, *Crease enhancement diffusion*, M.I. Torres, A. Sanfeliu (Eds.), Pattern Recognition and Image Analysis (VIII NSPRIA, Bilbao, 1999), Vol. 1, pp. 279–286, 1997.

[26] J. van Hateren, A. van der Schaaf, *Independent component filters of natural images compared with simple cells in primary visual cortex*, Proc. R. Soc. Lond. B 265:359–366, 1998.

[27] J. Weickert, *Anisotropic diffusion in image processing*, Teubner, Stuttgart, 1998.

[28] J. Weickert, *Nonlinear diffusion filtering*, Handbook on Comp. Vis. and Apps., Vol. 2, Acad. Press, 423–450, 1999.

[29] J. Weickert, H. Scharr, *A Scheme for Coherence-Enhancing Diffusion Filtering with Optimized Rotation Invariance*, J. Visual Communication and Image Representation, Special Issue On PDEs In Image Proc., Comp. Vis., And Computer Graphics, Academic Press 13(1/2):103–118, 2002.

[30] S. Zhu, D. Mumford, *Prior learning and Gibbs reaction-diffusion*, PAMI, 19(11):1236–1250, 1997.

8