Adaptive anisotropic denoising:
a bootstrapped procedure

John Aldo Lee[1], Arnaud de Decker[2], and Michel Verleysen[2] *

1- Université catholique de Louvain (UCL), Machine Learning Group,
Place du Levant 3, B-1348 Louvain-la-Neuve, Belgium

2- UCL, Imagerie Moléculaire et Radiothérapie Expérimentale,
Avenue Hippocrate 54, B-1200 Bruxelles, Belgium

Abstract. Signal denoising proves to be important in many domains such as pattern recognition and image analysis. This paper investigates several refinements of adaptive local filters that rely on local mode finding. These spatial filters are anisotropic and offer the advantage of attenuating noise without smoothing salient signal features such as discontinuities or other sharp transitions. In particular, a bootstrapped procedure is developed and leads to an improvement of the denoising quality without increasing the computational complexity. Experiments with an artificial benchmark allow the quantification of the performance gain.

1   Introduction

Denoising of multidimensional signals such as images is a well known problem and can be achieved through different paradigms. A classical strategy consists in damping or trimming high-frequency components, by means of either a Fourier or wavelet transform. In the spatial domain, popular approaches rely on partial differential equations [1, 2], on regularisation techniques such as total variation [3], or on local regression [4]. Denoising can also be formulated as a mode finding problem, which can be solved by robust statistics [5] or probability density estimation [6]. Many surveys [7, 8, 9] connect these different approaches.

This paper focus on mode finding techniques and proposes a bootstrapped procedure. The objective is to enhance denoising by introducing some refinements such as a local and data-driven adjustment of the main meta-parameter.

The rest of this paper is organised as follows. Section 2 describes some usual denoising tools that rely on local mode finding. Next, Section 3 details the proposed bootstrapped scheme. Section 4 gives experimental results and comments them. Finally, Section 5 draws the conclusions and sketches some perspectives for future work.

2   Denoising as local mode finding problem

Let us denote collected data by $(\mathbf{X}, \mathbf{Y})$, where each pair $(\mathbf{x}_i, \mathbf{y}_i)_{1 \leq i \leq N}$ consists of a noisy measurement $\mathbf{y}_i$ that is performed at some given location $\mathbf{x}_i$. Under the assumption of i.i.d. Gaussian noise, each measurement can be written as

$\mathbf{y}_i \sim G(\boldsymbol{\mu}_i, \nu)$, where $\boldsymbol{\mu}_i$ is the unknown noisefree value and ν is the noise standard deviation. Any statistical estimation of $\boldsymbol{\mu}_i$ must overcome the fact that a single measurement is available at each location. For this purpose, if the signal is locally constant, then we can compute a local average that minimises

$$E(\boldsymbol{\Upsilon}, \mathbf{Y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \|\boldsymbol{v}_i - \mathbf{y}_j\|^2 \;\;,$$

where $K_{ij}^{\sigma}(\mathbf{X}) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/(2\sigma^2))$ is a spatial kernel. Equating the derivative w.r.t. $\boldsymbol{v}_i$ leads to the closed-form solution

$$\boldsymbol{v}_i = \frac{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \, \mathbf{y}_j}{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X})} \;\;.$$

This corresponds to classical Gaussian smoothing. Parameter σ must be tuned in order to attenuate noise without excessive smoothing of salient features such as discontinuities or high-gradient transitions.

If the signal is piecewise constant, then the local distribution is multimodal and a more appropriate objective function is

$$E(\boldsymbol{\Upsilon}, \mathbf{Y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp\left(-\frac{\|\boldsymbol{v}_i - \mathbf{y}_j\|^2}{2\rho^2}\right) \;\;.$$

In line with the idea behind robust statistics [5], the standard least squares are replaced with a radiometric kernel with width ρ. With an appropriate normalisation, the last objective function also corresponds to a Parzen window density estimator [6]. Finding any local maximum of $E(\boldsymbol{\Upsilon}, \mathbf{Y}, \mathbf{X})$ turns out to identify a mode of the local signal distribution. In practice, this is achieved by equating the derivative w.r.t. $\boldsymbol{v}_i$ with zero. Although a closed-form solution does not exist, it is easy to obtain a fixed-point update that is written as

$$\boldsymbol{v}_i^{k+1} = \frac{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \mathbf{y}_j\|^2/(2\rho^2)) \, \mathbf{y}_j}{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \mathbf{y}_j\|^2/(2\rho^2))}$$

and initialised with $\boldsymbol{v}_i^0 = \mathbf{y}_i$. This procedure is known as the local M-smoother [9] (LMS). In contrast to Gaussian smoothing, the LMS computes a local weighted average with potentially anisotropic contributions.

The image processing community also uses an intuitive update rule that is known as bilateral filtering [10] (BF) and written as

$$\boldsymbol{v}_i^{k+1} = \frac{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2/(2\rho^2)) \, \boldsymbol{v}_j^k}{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2/(2\rho^2))} \;\;.$$

As can be seen, the noisy data is used solely in the initialisation step; the following updates involve only previously filtered values. Dropping the spatial weights

$K_{ij}^{\sigma}(\mathbf{X})$ reduces BF to the popular mean-shift procedure [11] that is used in clustering and image segmentation. The update actually results from the fixed-point minimisation of

$$E(\mathbf{\Upsilon}, \mathbf{Y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp\left(-\frac{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|^2}{2\rho^2}\right) \ ,$$

which admits the trivial solution $\boldsymbol{v}_i = \boldsymbol{v}_j$ for $1 \leq i, j \leq N$. Early stopping with some adequate criterion prevents convergence on a flat signal.

3 Bootstrapped anisotropic local least squares

The good performance of bilateral filtering can be explained by looking at the argument of the radiometric kernel. It involves the comparison of the free variable with filtered values instead of noisy ones, the former ones being supposedly more reliable than the latter ones. On the other hand, convergence on a trivial solution results from averaging filtered values instead of the original ones. As a workaround, we propose to maximise

$$E(\mathbf{\Upsilon}^{k+1}, \mathbf{Y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|\boldsymbol{v}_i^{k+1} - \mathbf{y}_j\|^2 K_{ij}^{\sigma}(\mathbf{X}) \exp\left(-\frac{\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2}{2\rho^2}\right) \ .$$

This objective function is quadratic w.r.t. $\boldsymbol{v}_i^{k+1}$ and it entails some sort of bootstrapping: we assume that some previously filtered values $\boldsymbol{v}_i^k$ are available. Just as with the other filters, we start the process with $\boldsymbol{v}_i^0 = \mathbf{y}_i$. A closed-form solution is expressed by

$$\boldsymbol{v}_i^{k+1} = \frac{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2/(2\rho^2)) \, \mathbf{y}_j^k}{\sum_{j=1}^{N} K_{ij}^{\sigma}(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2/(2\rho^2))} \ .$$

Convergence of this update rule cannot be guaranteed. Nevertheless, it is easy to see that this filter is strictly local, as the weighted average always resorts to the initial noisy data. This eliminates any trivial solution such as a flat signal, in the same way as the fidelity term in total variation denoising [3].

The idea of comparing the free variable to filtered values leads to a further refinement if we observe that the variance of the filtered value $\boldsymbol{v}_i^k$ is smaller than that of the noisy signal $\mathbf{y}_i$, provided $k > 0$. Therefore the radiometric width should be adapted according to the variance of $\boldsymbol{v}_i^k$. More specifically, we propose to individualise ρ for each datum and to decrease it in order to hopefully better separate partially overlapping modes. If we consider that all previously defined filters can written as either $\mathbf{\Upsilon}^{k+1} = \mathbf{Y}\mathbf{W}^k$ or $\mathbf{\Upsilon}^{k+1} = \mathbf{\Upsilon}^k\mathbf{W}^k$, where $\mathbf{W}^k = [\mathbf{w}_i^k]_{1 \leq i \leq N}$ is a square matrix that contains all filter weights after the kth update. This reformulation allows us to write

$$E(\mathbf{W}^{k+1}, \mathbf{Y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|\mathbf{Y}\mathbf{w}_i^{k+1} - \mathbf{y}_j\|^2 K_{ij}^{\sigma}(\mathbf{X}) \exp\left(-\frac{\|\mathbf{Y}(\mathbf{w}_i^k - \mathbf{w}_j^k)\|^2}{2\rho^2\|\mathbf{w}_j^k\|^2}\right) \ .$$

The denominator in the radiometric kernel relies on the assumption that if two independent random variables u and v have their variance equal to ν_u^2 and ν_v^2 respectively, then the variance of $\alpha u + \beta v$ is equal to $\alpha^2 \nu_u^2 + \beta^2 \nu_v^2$. Maximising the last objective function leads to

$$w_{ij}^{k+1} = \frac{K_{ij}^\sigma(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2/(2\rho^2\|\mathbf{w}_j^k\|^2))}{\sum_{j=1}^{N} K_{ij}^\sigma(\mathbf{X}) \exp(-\|\boldsymbol{v}_i^k - \boldsymbol{v}_j^k\|^2/(2\rho^2\|\mathbf{w}_j^k\|^2))} \ ,$$

where $\mathbf{W}^0$ is initialised to the identity matrix. In practice, matrix $\mathbf{W}^k$ does not need to be stored in memory, as the update involves only the products $\mathbf{Y}\mathbf{w}_i^k = \boldsymbol{v}_i^k$ and the norms $\|\mathbf{w}_i^k\|$. A last refinement consists in controlling the spatial kernel width at each update. Plugging $\sigma^k = \sigma\sqrt{k+1}$ into $K_{ij}^\sigma(\mathbf{X})$ allows us to simulate the diffusion process of bilateral filtering without involving the filtered values.

4 Experiments

In order to evaluate the denoising quality, we use an artificial benchmark image [12] comprised of 64^2 pixels and illustrated in Fig. 1. Transitions between the various areas are slightly blurred, exactly as in a real image. Noisy images are obtained by adding random i.i.d. Gaussian perturbations with unit variance (the pixel intensity range is $[1, 10]$). An example is shown in Fig. 1 as well. Quantitative assessment is provided by comparing the denoised image with the noisefree reference, by means of the root mean square error (RMSE). RMSE results are averaged over 50 images with different noise realisations (the average is performed before applying the square root). As the proposed filter is intended to be applied in biomedical applications, signal recovery is of prime importance and perceptual criteria are thus not considered.
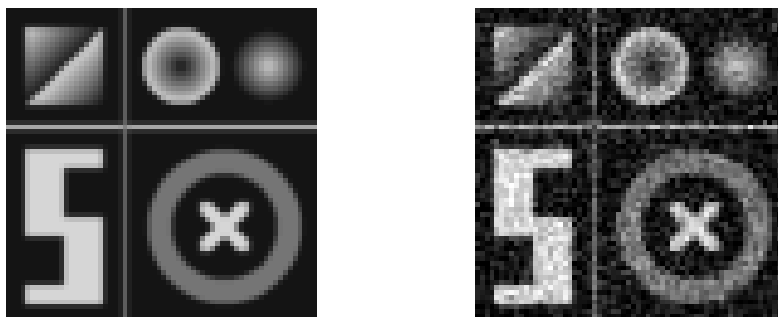

Fig. 1: Benchmark image without and with unit variance i.i.d. Gaussian noise.

Figure 2 reports a first experiment. RMSE curves are drawn w.r.t. σ for Gaussian smoothing, and ρ for filters with a radiometric kernel. For these filters, σ is fixed and four updates are performed. As expected, the outcome of the first update is the same for all of them. Afterwards, it is noteworthy that the optimal value of ρ is shifted to the left for filters with a non-adaptive radiometric width,

and to the right for the others. Subsequent updates increase the RMSE, which can be explained by the slight blur in the image. Divergence proves to be stronger for BF. The RMSE minima in Fig. 2 are 0.673, 0.545, 0.554, 0.536, 0.500, and 0.493 respectively.
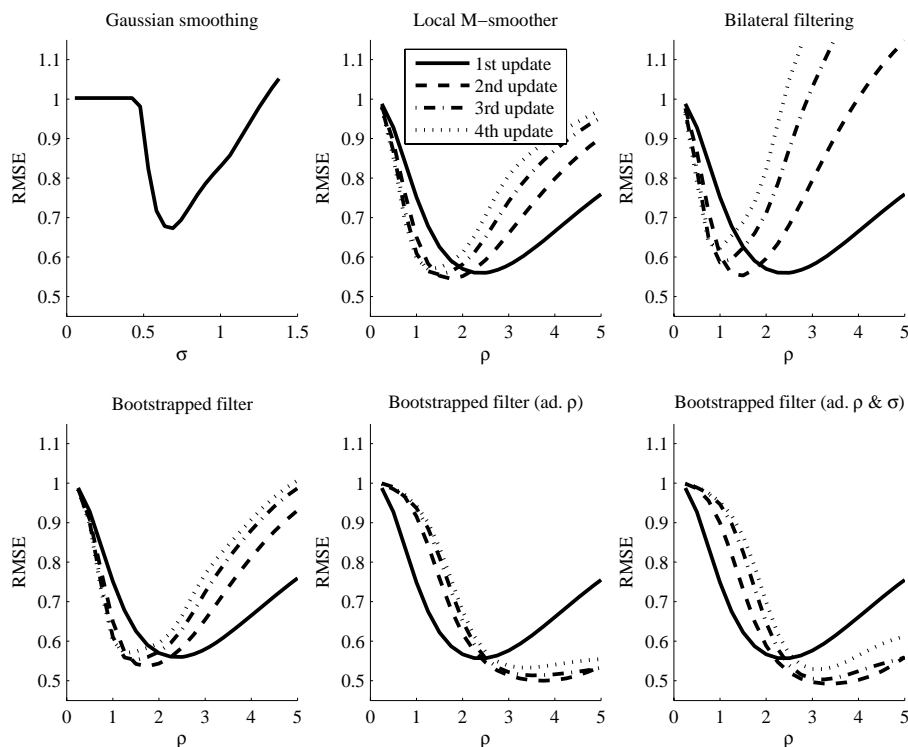

Fig. 2: RMSE curves w.r.t. σ and ρ for the benchmark image.

Table 1 reports a second experiments, in which we determine the values of σ and ρ that minimise the RMSE after two updates. This confirms that quality is improved by (i) comparing filtered values rather than noisy ones in the radiometric kernel, (ii) averaging noisy values rather than filtered ones, (iii) individualising and adapting the radiometric width, and (iv) simulating a diffusion process by increasing the spatial width. The most noticeable performance enhancement results from the combination of (i), (ii), and (iii).

5   Conclusions

Local mode finding proves to be a powerful tool for signal denoising. However, classical implementations such as the local M-smoother and bilateral filtering are optimal only if the mode overlap is minimal, which rarely happens in practice. We have addressed the issue of mode overlap by developing a bootstrapped

	σ	ρ	RMSE
Local M-smoother	1.0948	1.9482	0.5381
Bilateral filtering	0.8680	1.8227	0.5352
Bootstrapped filter	1.1031	1.9113	0.5280
Bootstrapped filter (adaptive ρ)	1.3395	3.8995	0.4972
Bootstrapped filter (adaptive ρ and σ)	1.1477	3.5924	0.4905

Table 1: Values of σ and ρ that minimise the RMSE for the benchmark image.

filter and by introducing a locally adaptive radiometric width. Higher denoising quality has been experimentally confirmed. Future work will investigate more complex ways to adjust the radiometric width by means of local statistics.

References

[1] P. Perona and J. Malik. Scale-space and edge-detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[2] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, 1998.

[3] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[4] I. Gijbels, A. Lambert, and P. Qiu. Edge-preserving image denoising and estimation of discontinuous surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1075–1087, 2006.

[5] P. J. Huber. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1981.

[6] Parzen E. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33:1065–1076, 1962.

[7] D. Barash. A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):844–847, 2002.

[8] G. Winkler, V. Aurich, K. Hahn, A. Martin, and K. Rodenacker. Noise reduction in images: Some recent edge-preserving methods. *Pattern Recognition and Image Analysis*, 9(4):749–766, 1999.

[9] P. Mrazek, J. Weickert, and A. Bruhn. On robust estimation and smoothing with spatial and tonal kernels. In R. Klette, R. Kozera, L. Noakes, and J. Weickert, editors, *Geometric Properties for Incomplete data*, volume 31 of *Computational Imaging and Vision*, pages 335–352. 2006.

[10] M. Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing*, 11(10):1141–1151, 2002.

[11] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[12] J.A. Lee, X. Geets, V. Grégoire, and A. Bol. Edge-preserving filtering of images with low photon counts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:1014–1027, 2008.