A novel anisotropic local polynomial estimator based on directional multiscale optimizations

Alessandro Foi[◇*], Vladimir Katkovnik[*], Karen Egiazarian[*] and Jaakko Astola[*]
[*]**Institute of Signal Processing, Tampere University of Technology, Finland**
[◇]**Dipartimento di Matematica, Politecnico di Milano, Italy**

Abstract A novel anisotropic estimator for image restoration is presented. The proposed approach originates from the geometric idea of a starshaped estimation neighborhood topology. In this perspective, an optimal adaptation is achieved by selecting in a pointwise fashion the ideal starshaped neighborhood for the estimation point. In practice, this neighborhood is approximated by a sectorial structure composed by conical sectors of adaptive size. Special varying-scale kernels, supported on these sectors, are exploited in order to bring the original geometrical problem to a practical multiscale optimization.

It is proposed to use this adaptive estimator iteratively. This recursion results in the anisotropic enlargement of the estimation neighborhood, an effect that can be interpreted as a special diffusion process.

The resulting estimators are truly anisotropic, providing clean and accurate edge adaptation and excellent restoration performance. Their implementation is fast as it is based on simple convolutions and scalar optimizations. Although we focus on image processing, the approach is general and can be extended to higher-dimensional data.

1. Motivation and idea

We consider the denoising problem of restoration of the image intensity y from the noisy observations $z(x) = y(x) + \sigma\,\eta(x)$, $\eta \sim \mathcal{N}(0,1)$. Our main intention is to develop algorithms efficient for highly anisotropic images.

When estimating y, a trade-off between noise suppression (variance) and smoothing (bias) has to be considered. Usual images are nonstationary, often characterized by localized features. Therefore, images should be treated adaptively: for example, one would achieve a higher noise suppression where the original image is smooth than in the vicinity of sharp transitions such as edges, where oversmoothing should be avoided. So, the desired balance between variance and bias depends on the image's local features. How to control this balance is a key problem in adaptive signal processing. A novel strategy to achieve such adaptation is presented in this paper.

1.1 Estimates with support optimization

Let $X \subset \mathbb{R}^2$ (or $\subset \mathbb{Z}^2$) be the image domain. Consider a conventional kernel estimator (filter) in the form

$$\hat{y}(x) = \int 1_{U_x}(x-\upsilon)z(\upsilon)d\upsilon = \int 1_{\tilde{U}_x}(\upsilon)z(\upsilon)d\upsilon = \int_{\tilde{U}_x} z(\upsilon)d\upsilon\,/\,\mu\,(U_x)\,, \tag{1.1}$$

where U_x is a neighborhood of the origin, and the uniform smoothing kernel 1_{U_x} has support U_x and constant value $1/\mu(U_x)$ on U_x ($\mu(U_x)$ stands for the Lebesgue measure of U_x). We use the decoration $\sim$ to denote the translated and mirrored neighborhood about the reference point x, $\tilde{U}_x(\cdot) = U_x(x-\cdot)$, distinguishing it from U_x that is *always* about the origin. The term *neighborhood* (of a point) is used in a generic sense, meaning a simply connected set (containing the point). Relations between sets are always considered up to a null-set.

Bias and variance of the estimate (1.1) are, respectively, $m_{\hat{y}}(x) = y(x) - \int 1_{\tilde{U}_x}(\upsilon)y(\upsilon)d\upsilon$ and $\sigma^2_{\hat{y}}(x) = \sigma^2/\mu\,(U_x)$. The *ideal* support U^*_x, yielding the best mean squared error, can be found by minimization of the quadratic risk $l_{\hat{y}}(x)$:

$$U^*_x = \arg\min_{U_x} l_{\hat{y}}(x), \qquad l_{\hat{y}}(x) = m^2_{\hat{y}}(x) + \sigma^2_{\hat{y}}(x). \tag{1.2}$$

Thus $\hat{y}(x) = \int 1_{U^*_x}(x-\upsilon)z(\upsilon)d\upsilon$ is the best local mean estimate of $y(x)$. The optimization (1.2) can be quite difficult to achieve. In order to make it practical further specifications of the problem are required.

1.1.1 Starshaped unbiased estimates and the $\mathfrak{U}_x$ topology

We discuss here a simplified model, which will serve as a ground for the development of a more general approach. Let y be a binary black-and-white image, i.e. $y(x) \in \{0,1\}\ \forall x$, and let us restrict our attention to *starshaped unbiased* estimates. It means that we consider only sets U_x which are starshaped with respect to the origin and such that $m_{\hat{y}}(x) = 0$.

The best estimate is obtained by minimization of the variance only or, equivalently, by maximization (with respect to the set inclusion $\subset$) of the set U_x. Unbiasedness holds if and only if $y(\upsilon) = y(x)$ for almost every $\upsilon \in \tilde{U}_x$. Under mild regularity assumptions on y (e.g. piecewise regular boundary of level sets), such equality has to hold for every $\upsilon \in \tilde{U}_x$. Thus, the best unbiased estimate corresponds to the largest starshaped U_x such that $y(\tilde{U}_x(\upsilon)) = y(x)\ \forall\upsilon$. This procedure can be formalized nicely in a topological manner. Let $\mathfrak{U}_x$ be the topology constituted by all sets U_x such that: (i) $U_x \setminus \{0\}$ is an open set in the Euclidean topology, (ii) U_x is starshaped with respect to 0 and (iii) $y(x-\upsilon) = y(x)\ \ \forall\upsilon \in U_x$. The maximum (w.r.t. $\subset$) element in $\mathfrak{U}_x$ corresponds to the ideal starshaped unbiased estimate of $y(x)$, $U^*_x = \max\mathfrak{U}_x$. This suggests a risk minimization strategy based on a progressive set enlargement within this topology. It may be achieved also by "decomposing" $\mathfrak{U}_x$ as follows. Let $\{S_i\}_{i=1}^K$ be a collection of K starshaped neighborhoods of the origin such that $\cup_{i=1}^K S_i = \mathbb{R}^2$ (e.g. a collection of conical sectors). Then, $\mathfrak{U}^{S_i}_x = \{U^{S_i}_x = U_x \cap S_i : U_x \in \mathfrak{U}_x\}$ are also topologies, and $U^*_x = \max\mathfrak{U}_x = \cup_{i=1}^K \max\mathfrak{U}^{S_i}_x$. It means that the optimization can be performed independently on each "subcomponent" $\mathfrak{U}^{S_i}_x$. Examples of the ideal $\tilde{U}^*_x$ are given in Figure 1 for two images: the characteristic function of an open disc and the "Cheese" image. Although different points x', x'' may

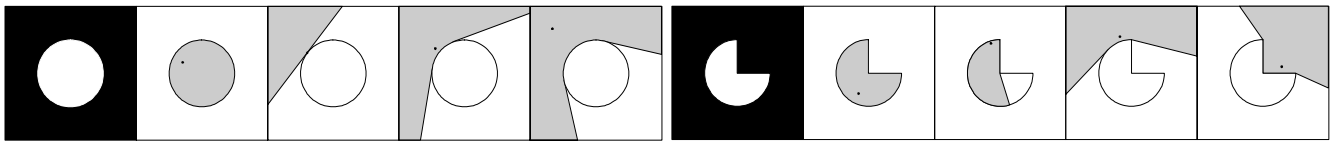

Figure 1: Examples of the ideal starshaped neighborhoods $\tilde{U}_x^*$ resulting from $U_x^* = \max \mathfrak{U}_x$.

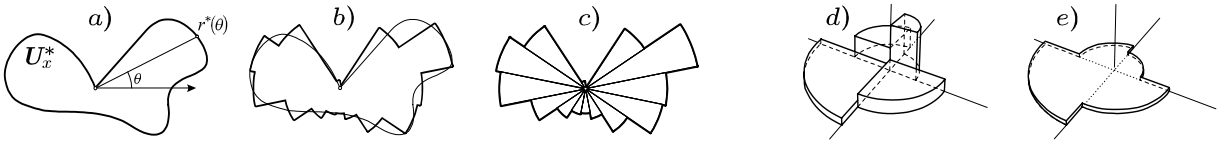

Figure 2: Piecewise constant approximation of $r^*(\theta)$ and its representation by varying size sectors (*a-b-c*); adaptive fusing of uniform sectorial kernels $g_{h^*(x,\theta_i),\theta_i}$ produces the equivalent uniform anisotropic kernel g_x^* (*d-e*).

have $\tilde{U}_{x'}^* = \tilde{U}_{x''}^*$, the corresponding $U_{x'}^*$ and $U_{x''}^*$ are not equal, and in both examples each point x has its own different ideal neighborhood U_x^*. Adapting perfectly to the edges, they are typically non-convex and their shape can be rather complex.

Despite the apparent simplicity of these speculations, the practical realization of this approach can be still hard to achieve, since the function y is usually unknown and only its noisy observation z is available. In particular, unless y is known to belong to some very specific class, ensuring unbiasedness is not possible, and biased estimates have to be considered.

1.2 Estimates with kernel scale optimization

Another way to adapt to the signal's varying local features, following the majority of multiscale techniques, is to use kernels equipped with a scale parameter h (e.g. $g_h(\cdot) = g(\cdot/h)/h^2$). This estimate can be presented in the form $\hat{y}(x) = \int g_{h(x)}(x - v)z(v)dv$. The scale optimization can be formulated, similarly to (1.2), as $h^*(x) = \arg\min_h l_{\hat{y}}$. The bias and variance are, respectively, $m_{\hat{y}}(x) = y(x) - \int g_{h(x)}(x-v)y(v)dv$, and $\sigma_{\hat{y}}^2(x) = \sigma^2 \int g_{h(x)}^2(v)dv$. This kind of optimization is known to be practical and can give good results through algorithms of reasonable complexity (e.g. Mallat, 1999).

When the support of the kernel g_h is bounded, the scale parameter $h(x)$ controls the size of the neighborhood for estimation at the point x. The support of the optimal scale kernel $g_{h^*(x)}$ can be thought as an approximation of the optimal U_x^* considered in (1.2). However, traditional kernels have supports of simple convex geometry (square, rectangle, circle, oval, etc.) whereas the optimal neighborhoods can be quite complex, especially near edges or corners. Thus, this approximation of U_x^* can be quite poor.

2. Anisotropic estimator based on adaptive directional scale

A reasonable compromise between the geometrical approach discussed in section 1.1 and the above kernel-based method, is obtained with a directional adaptive scale estimator (Katkovnik et al., 2004). The considerations from section 1.1.1 shed some insight on how this sort of compromise is produced and clarify the geometrical properties of the estimator.

The starshapedness of U_x^* allows to describe this set using polar coordinates: there exists a function $r^*(\theta)$, $\theta \in [0, 2\pi)$ (see Figure 2a), such that $U_x^* = \{v \in X, v = (v_1, v_2) = (r_v \cos\theta_v, r_v \sin\theta_v) : r_v < r^*(\theta_v)\}$. Instinctively, one may assume some sort of continuity of $r^*(\theta)$ with respect to its argument. This regularity, however, fails in the vicinity of edges where, as in the examples shown in Figure 1, $r^*(\theta)$ presents sharp transitions. This irregular behaviour is a direct manifestation of the *anisotropy* of y or, roughly speaking, that the function's properties are different in different directions. The most natural model, allowing good approximation of such rapid transitions and also discontinuities is to assume $r^*(\theta)$ as a piecewise constant function of its angular argument, i.e. assuming that the optimal neighborhood U_x^* has a sectorial structure, as shown in Figure 2(*b-c*).

In our approach we exploit this sectorial decomposition. A collection of directional *LPA* (local polynomial approximation, see Fan & Gijbels, 1996) kernels $\{g_{h,\theta_i}\}_{h\in H, i=1,\dots,K}$ supported on such sectors is designed. Each kernel is characterized by a direction θ_k and a scale parameter h. The corresponding estimate is the convolution $\hat{y}_{h,\theta_i}(x) = (g_{h,\theta_i} \circledast z)(x)$. The statistical *ICI* rule (Goldenshluger & Nemirovski, 1997; Katkovnik, 1999) is used to select a pointwise optimal scale $h^*(x,\theta_i) \approx r^*(\theta_i)$ for each direction. Let $\hat{y}_{h^*(x,\theta_i),\theta_i}(x)$ be the directional optimal scale estimate and $\sigma_i^2(x)$ its variance. All these estimates can be fused in the final one as follows:

$$\hat{y}(x) = \sum\nolimits_i \lambda(x,\theta_i)\hat{y}_{h^*(x,\theta_i),\theta_i}(x), \qquad \lambda(x,\theta_i) = \sigma_i^{-2}(x) / \sum\nolimits_j \sigma_j^{-2}(x). \tag{2.1}$$

The weights $\lambda(x,\theta_i)$ in the above convex combination are data-driven adaptive, as $\sigma_i^{-2}(x)$ depend on the adaptive $h^*(x,\theta_i)$. The estimate (2.1) is equivalent to the adaptive anisotropic kernel estimate $\hat{y}(x) = \int g_x^*(x-v)z(v)dv$, where $g_x^* = \sum_i \lambda(x,\theta_i) g_{h^*(x,\theta_i),\theta_i}$. When uniform kernels g_{h,θ_k} are used, the adaptive weights $\lambda(x,\theta_i)$ make so that also the anisotropic kernel g_x^* is uniform on its support, i.e. $g_x^* = 1_{\cup_i \operatorname{supp} g_{h^*(x,\theta_i),\theta_i}}$, as shown in Figure 2(*d-e*).

Figure 3 shows the estimation neighborhoods resulting from the proposed anisotropic *LPA-ICI* approach for noisy images ($\sigma = 0.1$). A comparison with Figure 1 shows the similarity between the previous ideal example and this concrete case.

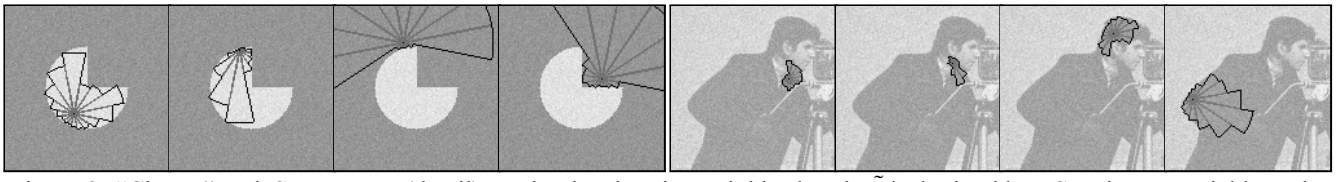

Figure 3: "Cheese" and *Cameraman* (detail): optimal estimation neighborhoods $\tilde{U}_x^*$ obtained by *ICI* using sectorial kernels.

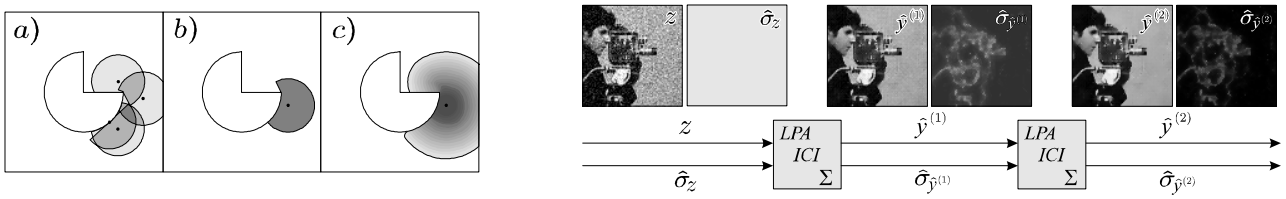

Figure 4: Recursive *LPA-ICI*: estimation neighborhood's fattening (left), and layout of algorithm's implementation (right).

2.1 Recursive *LPA-ICI* filtering

The idea behind this procedure is to apply recursively the anisotropic *LPA-ICI* algorithm, filtering the final output $\hat{y}$ (2.1) once or many times over again. Denoting by $\mathcal{LI}$ the overall anisotropic *LPA-ICI* filter, this recursion is expressed as follows:

$$z^{(1)} = z, \qquad \hat{y}^{(l)} = \mathcal{LI}(z^{(l)}), \qquad z^{(l+1)} = \hat{y}^{(l)}, \qquad l = 1, 2, \dots \quad . \tag{2.2}$$

Expanding (2.2), in order to explicitly write $\hat{y}^{(l)}$ with respect to the initial observations z, we obtain

$$\hat{y}^{(l)}(x) = \int g_x^{*(l)}(x-v)\hat{y}^{(l-1)}(v)dv = \int\left(\int\cdots\int\left(\tilde{g}_x^{*(l)}(v^{(1)})\tilde{g}_{v^{(1)}}^{*(l-1)}(v^{(2)})\cdots\tilde{g}_{v^{(l-1)}}^{*(1)}(v^{(l)})\right)dv^{(1)}\ldots dv^{(l-1)}\right)z(v^{(l)})dv^{(l)}, \tag{2.3}$$

where $g_x^{*(l)}$ is the anisotropic kernel at the l-th iteration, $\tilde{g}_x^{*(l)}(\cdot) = g_x^{*(l)}(x-\cdot)$, and $v^{(i)}$ are auxiliary variables.

2.1.1 Estimation neighborhood's enlargement

Under the simple settings discussed in section 1.1.1, the ideal U_x^* does not depend on the observed signal z, but rather only on the (unknown) signal y. When a second iteration is performed in (2.2), the ideal neighborhood for estimating $y(x)$ from $z^{(2)} = \hat{y}^{(1)}$ is again the same U_x^* as in the first iteration. Since this applies to all iterations, the whole process is described by replacing all kernels $\tilde{g}_t^{*(\cdot)}$ with $1_{\tilde{U}_t^*}$ in (2.3). Despite the ideal neighborhood U_x^* is always the same for all l, the support of the resulting kernel that is used for integration against $z(v^{(l)})$ in the right hand side of (2.3) may grow at every iteration. For example, at the second iteration the estimation support with respect to the initial observations z is $\mathrm{supp}\int 1_{\tilde{U}_x^*}(v)1_{\tilde{U}_v^*}(\cdot)dv = \cup_{v\in\tilde{U}_x^*}\tilde{U}_v^*$. This is illustrated in Figure 4(left), with (a) some ideal starshaped neighborhoods $\tilde{U}_v^*$ corresponding to points v belonging to, (b) the ideal neighborhood $\tilde{U}_x^*$ of the estimation point x, and (c) the resulting *enlarged* neighborhood of x, $\cup_{v\in\tilde{U}_x^*}\tilde{U}_v^*$, obtained by the second iteration of the adaptive algorithm. Such sets are not necessarily starshaped w.r.t. x.

If the ideal neighborhoods were translation-invariant, $U_x^* = U^* \ \forall x$, then (2.3) would take the simple convolutional form $\hat{y}^{(l)}(x) = (1_{U^*} \circledast \cdots \circledast 1_{U^*} \circledast z)(x)$, where convolution between kernels is repeated l-1 times. This resembles other iterative constructions, such as the Gaussian/Laplacian pyramids or wavelet-type projections (e.g. Mallat, 1999), where multiscale filtering is obtained by recursively convolving the observations against the same filter.

In general, however, formula (2.3) cannot be written in a simple convolutional form, because the adaptive kernels are not translation-invariant. Nevertheless, the considerations previously given about the enlargement of ideal neighborhoods hold similarly for the supports of the fused kernels. This anisotropic propagation of estimation neighborhoods realizes a diffusion flow similar to the non-linear anisotropic diffusion (Perona & Malik, 1990), but intrinsically robust to noise because of the *ICI*-based adaptive scale. Regardless of their linear appearance, (2.3), as well as (2.1), are also non-linear estimators. The non-linearity is introduced by the adaptive selection of the directional scale $h^*(x,\theta_i)$.

2.1.2 Variance of l-th iteration's estimates

Let $G_{x,h,\theta_i}^{(l)}(\cdot) = \int\cdots\int(g_{h,\theta_i}(x-v^{(1)})\tilde{g}_{v^{(1)}}^{*(l-1)}(v^{(2)})\cdots\tilde{g}_{v^{(l-1)}}^{*(1)}(\cdot))dv^{(1)}\ldots dv^{(l-1)}$, then, the standard deviation of the estimate $\hat{y}_{h,\theta_i}^{(l)}(x)$, needed in order to use the *ICI* rule to select the optimal scale $h^*(x,\theta_i)$ at the l-th iteration, is $\sigma\|G_{x,h,\theta_i}^{(l)}\|_2$. However, its calculation is computationally quite complex, and it requires also a good deal of computer memory. These technical reasons limit the direct and accurate implementation of the recursive system (2.2). It would be appealing to use a simpler construction, where each step is performed without keeping track of the previous iterations, i.e. using the $\mathcal{LI}$ operator as a "black box", with a pair of inputs (observations and their standard deviations) and a pair of outputs (estimates and their standard deviations), as shown in Figure 4(right).

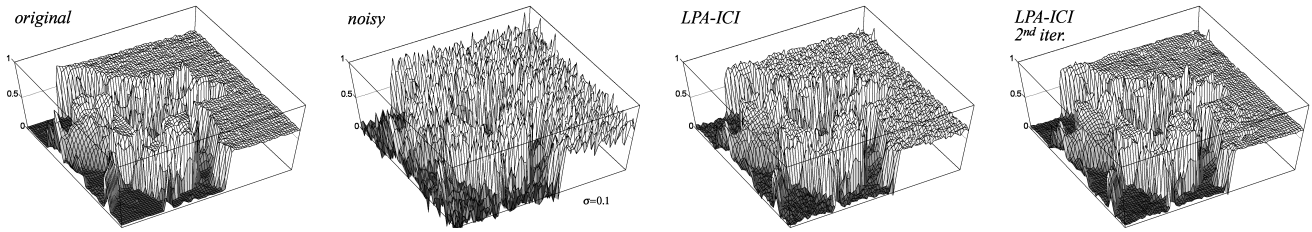

Figure 5: Fragment of the *Cameraman* image: from left to right, original, noisy image, *LPA-ICI* estimate (first iteration), recursive *LPA-ICI* estimate (second iteration). Further iterations of the recursive procedure yield visually identical estimates.

2.1.3 Implementation

The residual noise in the estimate $\hat{y}$ from (2.1) is no longer uncorrelated (estimation neighborhoods may overlap with each other) nor its standard deviation is a constant (estimation neighborhoods are adaptive), as shown in Figure 4(right). The expression for its variance is $\hat{\sigma}^2_{\hat{y}}(x) = 1/\sum_i \sigma_i^{-2}(x)$. If we assume that this residual noise is uncorrelated, the standard deviation of the directional estimates $\hat{y}^{(2)}_{h,\theta_i}(x)$ for the second stage of the recursive algorithm would be simply calculated as the convolution $(g^2_{h,\theta_i} \circledast \hat{\sigma}^2_{\hat{y}^{(1)}})^{1/2}$, avoiding the use of the complicated kernel $G^{(2)}_{x,h,\theta_i}$. This reasoning may be extended to further iterations, assuming that the noise in $\hat{y}^{(l)}$ is always uncorrelated. However, as this assumption does not hold, the quality of estimation deteriorates, and typically results in oversmoothing of details in the image. It turns out, for low-order kernels, that a simple compensating factor for the standard deviation can effectively reduce this degeneration. This modification of the calculation of the variance may be interpreted as an attempt to filter out only the white component of the residual noise.

After setting the initial conditions $y^{(0)} = z$ and $\hat{\sigma}^{(0)}_y \equiv \sigma$, the l-th recursive step of the modified recursive algorithm is

$$\hat{y}^{(l)} = \mathcal{LI}(\hat{y}^{(l-1)}), \qquad \hat{\sigma}_{\hat{y}^{(l)}} = \left(\sum\nolimits_i \left(\hat{\sigma}^{(l)}_i\right)^{-2}\right)^{-1/2}, \qquad l = 1, 2, \dots \quad ,$$

where $\hat{\sigma}^{(l)}_i = \hat{\sigma}_{\hat{y}^{(l)}_{h^*(x,\theta_i),\theta_i}}$, $\hat{\sigma}_{\hat{y}^{(l)}_{h,\theta_i}} = \alpha(g^2_{h,\theta_i} \circledast \hat{\sigma}^2_{\hat{y}^{(l-1)}})^{1/2}$, and $\alpha < 1$ being the fixed correcting factor.

In spite of the striking simplicity of the modification, simulation results show that it enables *ICI* to properly select the adaptive scale. Moreover, convergence of the above recursive system is easily guaranteed, since $\hat{\sigma}^{(l)}_{\hat{y}} = \mathcal{O}(\alpha^l) \to 0$. More precisely, since $\|g^*_x\|_2 \le 1$, there exist a constant c such that $|\hat{y}^{(l)}(x) - \hat{y}^{(l+1)}(x)| < c\hat{\sigma}^{(l)}_{\hat{y}}(x) \le c\alpha^l\sigma$. This implies that $\hat{y}^{(l)}(x)$ is a Cauchy sequence. Qualitatively, the actual convergence rate of the algorithm depends on $\mu\left(U^*_x\right) \approx \|g^*_x\|_2^{-1}$, and usually the algorithm reaches a numerical steady-state already after three iterations. The proposed recursive method can be used for accurate detail-preserving image denoising, segmentation and edge detection applications.

Table 1 shows the *ISNR* and *MAE* (ℓ^1-distance) results for the restoration of the *Cameraman* image, corrupted by additive Gaussian white noise, $\sigma = 0.1$. Zero-order uniform kernels for a total of eight directions and four scales, $h \in \{1, 2, 3, 5\}$, were used with fixed $\alpha = 2/3$. These results are illustrated (for a fragment of the image) in Figure 5. The table shows a fast convergence of the iterations and criteria values attesting the high quality of the filtering.

iteration #	noisy	1	2	3	4	5	6
ISNR (dB)	0	7.361	8.098	8.119	8.120	8.120	8.120
255∗*MAE*	20.38	7.894	6.597	6.538	6.535	6.535	6.535

Table 1: *ISNR* and *MAE* results for the *Cameraman* image denoising experiment (σ=0.1, *SNR*=14.39dB).

The use of higher order kernel mixtures, together with a more refined update of the standard deviations and a larger set of scales allows to achieve, for the same experiment, an *ISNR* of 7.50, 8.23 and 8.47dB at the first, second and third iteration, respectively. A similar performance cannot be achieved by the non-recursive algorithm.

References

FAN, J. & GIJBELS, I. 1996. Local polynomial modelling and its application, Chapman and Hall, London.

GOLDENSHLUGER, A. & NEMIROVSKI, A. 1997. On spatial adaptive estimation of nonparametric regression, *Math. Meth. Statistics*, **6**, 135-170.

KATKOVNIK, V. 1999. A new method for varying adaptive bandwidth selection, *IEEE Trans. on Signal Proc.*, **47**(9), 2567-2571.

KATKOVNIK, V., EGIAZARIAN K., & ASTOLA, J. 2002. Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule, *J. of Mathematical Imaging and Vision*, **16**(3), 223-235.

KATKOVNIK, V., FOI, A., EGIAZARIAN, K., & ASTOLA, J. 2004. Directional varying scale approximations for anisotropic signal processing, *Proceedings of XII European Signal Processing Conference*, EUSIPCO 2004, Vienna, Austria, 6-10 Sept. 2004, 101-104.

MALLAT, S., 1999. A wavelet tour of signal processing, second edition, Academic Press, New York.

PERONA, P., & MALIK, J. 1990. Scale-space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7), 629-639.