Unsupervised Patch-Based Image Regularization and Representation

Charles Kervrann[1,2] and Jérôme Boulanger[1,2]

[1] IRISA/INRIA Rennes, Projet Vista, Campus Universitaire de Beaulieu,
35 042 Rennes Cedex, France
[2] INRA - MIA, Domaine de Vilvert, 78 352 Jouy-en-Josas, France

Abstract. A novel adaptive and patch-based approach is proposed for image regularization and representation. The method is unsupervised and based on a pointwise selection of small image patches of fixed size in the variable *neighborhood* of each pixel. The main idea is to associate with each pixel the weighted sum of data points within an adaptive neighborhood and to use image patches to take into account complex spatial interactions in images. In this paper, we consider the problem of the adaptive neighborhood selection in a manner that it balances the accuracy of the estimator and the stochastic error, at each spatial position. Moreover, we propose a practical algorithm with no hidden parameter for image regularization that uses no library of image patches and no training algorithm. The method is applied to both artificially corrupted and real images and the performance is very close, and in some cases even surpasses, to that of the best published denoising methods.

1 Introduction

Most of the more efficient regularization methods are based on energy functionals minimization since they are designed to explicitly account for the image geometry, involving the adjustment of global weights that balance the contribution of prior smoothness terms and a fidelity term [23, 28]. Thus, related partial differential equations (PDE) and variational methods have shown impressive results to tackle the problem of edge-preserving smoothing [24, 28, 32] and more recently the problem of image decomposition [1]. Moreover, other smoothing algorithms aggregate information over a neighborhood of fixed size, based on two basic criteria: a spatial criterion to select points in the vicinity of the current point and a brightness criterion in order to choose only points which are similar in some sense. In view of this generic approach, a typical filter is the sigma filter [19] and a continuous version of this filter gives the well-known nonlinear Gaussian filter [14]. Finally, if we substitute a Gaussian window to the hard disk-shaped window around the current position, we get variants of the *bilateral filtering* [31], controlled by setting the standard deviations in both spatial and brightness domains. Nevertheless, as effective as bilateral filtering and variants, they lacked a theoretical basis and some connections to better understood methods have been

investigated. In particular, the relationships between bilateral filtering and iterative *mean-shift* algorithm, local mode filtering, clustering, local M-estimators, non-linear diffusion, regularization approaches combining nonlocal data and nonlocal smoothness terms, and Beltrami flow, can be found in [33, 11, 3, 22, 29].

Nevertheless, we note that all cited methods have a relatively small number of regularity parameters that control the global amount of smoothing being performed. They are usually chosen to give a good and global visual impression and are sometimes heuristically chosen [31]. Furthermore, when local characteristics of the data differ significantly across the image domain, selecting local smoothing parameters seems more satisfying and, for instance, has been addressed in [4, 13, 5, 8]. But, what makes image regularization very hard, is that natural images often contain many irrelevant objects. To develop better image enhancement algorithms that can deal with such a structured noise, we need explicit models for the many regularities and geometries seen in local patterns. This corresponds to another line of work which consists in modeling non-local pairwise interactions from training data [35] or a library of natural image patches [12, 27]. The idea is to improve the traditional Markov random field (MRF) models by learning potential functions from examples and extended neighborhoods for computer vision applications [35, 12, 27]. In our framework, we will also assume that small image patches in the semi-local neighborhood of a point contains the essential process required for local smoothing. Thus, the proposed patch-based regularization approach is conceptually very simple being based on the key idea of iteratively increasing a window at each pixel and adaptively weighting input data. The data points with a similar patch to the central patch will have larger weights in the average. We use 7×7 or 9×9 image patches to compute these weights since they are able to capture most of local geometric patterns and texels seen in images. Note also that, it has been experimentally confirmed that intuitive exemplar-based approaches are fearsome for 2D texture synthesis [10] and image inpainting [34, 9]. Nevertheless, we propose here a theoretical framework for choosing a semi-local neighborhood adapted to each pixel. This neighborhood which could be large, is chosen to balance the accuracy of the pointwise estimator and the stochastic error, at each spatial position [20]. This adaptation method is a kind of change-point detection procedure, initiated by Lepskii [20]. By introducing spatial adaptivity, we extend the work earlier described in [7] which can be considered as an extension of bilateral filtering [31] to image patches. The related works to our approach are the unsupervised recent *non-local means* algorithm [7], nonlinear Gaussian filters [31, 33, 22] and statistical smoothing schemes [25, 16, 17], but are enhanced via incorporating either a variable window scheme or patch-based weights. Finally, to our knowledge, the more related competitive methods for image denoising, are recent wavelet-based methods [30, 26]. In our experiments, we have then reported the results when these methods are applied to a commonly-used image dataset [26]. We show that the performance of our method surpasses most of the already published and very competitive denoising methods [30, 26, 27, 7].

2   Patch-Based Approach

Consider the following basic image model: $Y_i = u(\mathbf{x}_i) + \varepsilon_i, i = 1, \ldots, |G|$ where $\mathbf{x}_i \in G \subset \mathbb{R}^d$, $d \geq 2$, represents the spatial coordinates of the discrete image domain of $|G|$ pixels, and $Y_i \in \mathbb{R}_+$ is the observed intensity at location $\mathbf{x}_i$. We suppose the errors ε_i to be independent, distributed Gaussian zero-mean random variables with unknown variances σ^2. In order to recover $u : \mathbb{R}^d \to \mathbb{R}$ from noisy observations, we suppose there exists repetitive patterns in the semi-local neighborhood of a point $\mathbf{x}_i$. In particular, we assume that the unknown image $u(\mathbf{x}_i)$ can be calculated as the weighted average of input data over a variable neighborhood Δ_i around that pixel $\mathbf{x}_i$. Henceforth, the points $\mathbf{x}_j \in \Delta_i$ with a similar regularized patch $\mathbf{u}_j$ to the reference regularized image patch $\mathbf{u}_i$ will have larger weights in the average. Now, we just point out that our ambition is not to learn generic image priors from a database of image patches as already described in [35, 12, 27], but we just consider image patches as non-local image features, and adapt kernel regression techniques for image regularization.

For simplicity, an image patch $\mathbf{u}_i$ is modeled as a fixed size square window of $p \times p$ pixels centered at $\mathbf{x}_i$. In what follows, $\mathbf{u}_i$ will denote indifferently a patch or a vector of p^2 elements where the pixels are concatenated along a fixed lexicographic ordering. As with all exemplar-based techniques, the size of image patches must be specified in advance [10, 34, 9, 7]. We shall see that a 7×7 or 9×9 patch is able to take care of the local geometries and texture in the image while removing undesirable distortions. In addition, the proposed approach requires no training step and may be then considered as *unsupervised*. This makes the method somewhat more attractive for many applications.

Another important question under such an estimation approach is how to determine the size and shape of the variable neighborhood (or window) Δ_i at each pixel, from image data. The selected window must be different at each pixel to take into account the inhomogeneous smoothness of the image. For the sake of parsimony, the set $\mathcal{N}_\Delta$ of admissible neighborhoods will be arbitrarily chosen as a geometric grid of nested square windows $\mathcal{N}_\Delta = \{\Delta_{i,n} : |\Delta_{i,n}| = (2^n+1) \times (2^n+1), n = 1, \ldots, N_\Delta\}$, where $|\Delta_{i,n}| = \#\{\mathbf{x}_j \in \Delta_{i,n}\}$ is the cardinality of $\Delta_{i,n}$ and N_Δ is the number of elements of $\mathcal{N}_\Delta$. For technical reasons, we will require the following conditions: $\Delta_{i,n}$ is centered at $\mathbf{x}_i$ and $\Delta_{i,n} \subset \Delta_{i,n+1}$. Finally, we focus on the *local L_2 risk* as an objective criterion to guide the optimal selection of the smoothing window for constructing the "best" possible estimator. This optimization will be mainly accomplished by starting, at each pixel, with a small window $\Delta_{i,0}$ as a *pilot* estimate, and increasing $\Delta_{i,n}$ with n. The use of variable and overlapping windows combined with adaptive weights contributes to the regularization performance with no block effect.

Adaptive estimation procedure. The proposed procedure is iterative [25, 17] and works as follows. At the initialization, we choose a local window $\Delta_{i,0}$ containing only the point of estimation $\mathbf{x}_i$ ($|\Delta_{i,0}| = 1$). A first estimate $\widehat{u}_{i,0}$ (and its variance $\widehat{v}_{i,0}^2 = \mathrm{Var}(\widehat{u}_{i,0})$) is then given by: $\widehat{u}_{i,0} = Y_i$ and $\widehat{v}_{i,0}^2 = \widehat{\sigma}^2$ where an estimated variance $\widehat{\sigma}^2$ has been plugged in place of σ^2 since the variance

of errors are supposed to be unknown. At the next iteration, a larger window $\Delta_{i,1}$ with $\Delta_{i,0} \subset \Delta_{i,1}$ centered at $\mathbf{x}_i$ is considered. Every point $\mathbf{x}_j$ from $\Delta_{i,1}$ gets a weight $\pi_{i\sim j,1}$ defined by comparing pairs of $p \times p$ regularized patches $\widehat{\mathbf{u}}_{i,0} = \left(\widehat{u}_{i,0}^{(1)}, \cdots, \widehat{u}_{i,0}^{(p^2)}\right)^T$ and $\widehat{\mathbf{u}}_{j,0} = \left(\widehat{u}_{j,0}^{(1)}, \cdots, \widehat{u}_{j,0}^{(p^2)}\right)^T$ obtained at the first iteration. Note that p is fixed for all the pixels in the image. As usual, the points $\mathbf{x}_j$ with a similar patch to $\widehat{\mathbf{u}}_{i,0}$ will have weights close to 1 and 0 otherwise. Then we recalculate an new estimate $\widehat{u}_{i,1}$ defined as the *weighted average* of data points lying in the neighborhood $\Delta_{i,1}$. We continue this way, increasing with n the considered window $\Delta_{i,n}$ while $n \le N_\Delta$ where N_Δ denotes the maximal number of iterations of the algorithm. For each $n \ge 1$, the studied maximum likelihood (ML) estimator $\widehat{u}_{i,n}$ and its variance $\widehat{v}_{i,n}^2$ can be then represented as

$$\widehat{u}_{i,n} = \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\, Y_j, \quad \widehat{v}_{i,n}^2 = \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \left[\pi_{i\sim j,n}\right]^2 \tag{1}$$

where the weights $\pi_{i\sim j,n}$ are continuous variables and satisfy the usual conditions $0 \le \pi_{i\sim j,n} \le 1$ and $\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} = 1$. In our modeling, these weights are computed from pairs of regularized $p \times p$ patches $\widehat{\mathbf{u}}_{i,n-1}$ and $\widehat{\mathbf{u}}_{j,n-1}$ obtained at iteration $n-1$ and p is fixed for all the pixels in the image. In what follows, n will coincide with the iteration and we will use $\widehat{n}(\mathbf{x}_i)$ to designate the index of the "best" window $\widehat{\Delta}(\mathbf{x}_i) \overset{def}{=} \widehat{\Delta}_{i,\widehat{n}(\mathbf{x}_i)}$ and the "best" estimate $\widehat{u}(\mathbf{x}_i) \overset{def}{=} \widehat{u}_{i,\widehat{n}(\mathbf{x}_i)}$. Among all non-rejected window $\Delta_{i,n}$ from $\mathcal{N}_\Delta$, the optimal window is chosen as

$$\widehat{\Delta}(\mathbf{x}_i) = \arg \max_{\Delta_{i,n} \in \mathcal{N}_\Delta} \left\{ |\Delta_{i,n}| : |\widehat{u}_{i,n} - \widehat{u}_{i,n'}| \le \varrho\, \widehat{v}_{i,n'}, \text{ for all } 1 \le n' < n \right\}$$

where ϱ is a positive constant. Throughout this paper, we shall see the rational behind this pointwise statistical rule and the proposed strategy that updates the pointwise estimator when the neighborhood increases at each iteration [25].

Adaptive weights. In order to compute the similarity of between patches $\widehat{\mathbf{u}}_{i,n}$ and $\widehat{\mathbf{u}}_{j,n}$, a distance must be first considered. In [10, 34, 9, 7], several authors showed that the L_2 distance $\|\widehat{\mathbf{u}}_{i,n} - \widehat{\mathbf{u}}_{j,n}\|^2$ is a reliable measure to compare image patches. To make a decision, we have rather used a normalized distance

$$\begin{aligned} \text{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1}) \;=\; & \frac{1}{2}\left[(\widehat{\mathbf{u}}_{i,n-1} - \widehat{\mathbf{u}}_{j,n-1})^T \widehat{\mathbf{V}}_{i,n-1}^{-1} (\widehat{\mathbf{u}}_{i,n-1} - \widehat{\mathbf{u}}_{j,n-1}) \right. \\ & \left. + \; (\widehat{\mathbf{u}}_{j,n-1} - \widehat{\mathbf{u}}_{i,n-1})^T \widehat{\mathbf{V}}_{j,n-1}^{-1} (\widehat{\mathbf{u}}_{j,n-1} - \widehat{\mathbf{u}}_{i,n-1}) \right] \end{aligned} \tag{2}$$

where $\widehat{\mathbf{V}}_{\cdot,n-1}$ is $p^2 \times p^2$ diagonal matrix of the form (the symbol "$\cdot$" is used to denote a spatial position)

$$\widehat{\mathbf{V}}_{\cdot,n-1} = \begin{pmatrix} \left(\widehat{v}_{\cdot,n-1}^{(1)}\right)^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \left(\widehat{v}_{\cdot,n-1}^{(p^2)}\right)^2 \end{pmatrix}$$

and $\widehat{v}^{(\ell)}_{\cdot,n-1}, \ell = 1, \cdots, p^2$, is the local standard deviation of the estimator $\widehat{u}^{(\ell)}_{\cdot,n-1}$, and the index ℓ is used to denote a spatial position in an image patch $\widehat{\mathbf{u}}_{\cdot,n-1} = \left(\widehat{u}^{(1)}_{\cdot,n-1}, \cdots, \widehat{u}^{(\ell)}_{\cdot,n-1}, \cdots, \widehat{u}^{(p^2)}_{\cdot,n-1}\right)^T$. Moreover, we used a symmetrized distance to test both the hypotheses that $\mathbf{x}_j$ belongs to the region $\Delta_{i,n}$ and $\mathbf{x}_i$ belongs to the region $\Delta_{j,n}$, at the same time. Accordingly, the hypothesis $\widehat{\mathbf{u}}_{i,n-1}$ and $\widehat{\mathbf{u}}_{j,n-1}$ are similar, is accepted if the distance is small, i.e. $\mathrm{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1}) \leq \lambda_\alpha$. In our modeling, the parameter $\lambda_\alpha \in \mathbb{R}_+$ is chosen as a quantile of a $\chi^2_{p^2,1-\alpha}$ distribution with p^2 degrees of freedom, and controls the probability of type I error for the hypothesis of two points to belong to the same region: $\mathbb{P}\{\mathrm{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1}) \leq \lambda_\alpha\} = 1 - \alpha$. All these tests ($|\Delta_{i,n}|$ tests) have to be performed at a very high significance level, our experience suggesting to use a $1 - \alpha = 0.99$-quantile. Henceforth, we introduce the following commonly-used weight function

$$\pi_{i\sim j,n} = \frac{K\left(\lambda_\alpha^{-1}\,\mathrm{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)}{\displaystyle\sum_{\mathbf{x}_j \in \Delta_{i,n}} K\left(\lambda_\alpha^{-1}\,\mathrm{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)} \tag{3}$$

with $K(\cdot)$ denoting a monotone decreasing function, e.g. a kernel $K(x) = \exp(-x/2)$. Due to the fast decay of the exponential kernel, large distances between estimated patches lead to nearly zero weights. Note that the use of weights enables to relax the structural assumption the neighborhood is a variable square window.

An "ideal" smoothing window. In this section, we address the problem of the automatic selection of the window $\Delta_{i,\cdot}$ adapted for each pixel $\mathbf{x}_i$. It is well understood that the local smoothness varies significantly for point to point in the image and *global risk* measures cannot wholly reflect the performance of estimators at a point. Then, a classical way to measure the performance of the estimator $\widehat{u}_{i,n}$ to its target value $u(\mathbf{x}_i)$ is to choose the *local L_2 risk*, which is explicitly decomposed into the sum of the squared bias $\widehat{b}^2_{i,n}$ and variance $\widehat{v}^2_{i,n}$:

$$\left[\mathbb{E}|\widehat{u}_{i,n} - u(\mathbf{x}_i)|^2\right]^{1/2} = |\widehat{b}^2_{i,n} + \widehat{v}^2_{i,n}|^{1/2}. \tag{4}$$

Our goal is then to minimize this *local L_2 risk* with respect to the size of the window $\Delta_{i,n}$, at each pixel in the image. Actually, the optimal solution explicitly depends on the smoothness of the "true" function $u(\mathbf{x}_i)$ which is unknown, and so, of less practical interest [16]. A natural way to bring some further understanding of the situation is then to individually analyze the behavior of the bias and variance terms when $\Delta_{i,n}$ increases or decreases with n as follows:

- The bias term $\widehat{b}_{i,n} = \mathbb{E}\left[\widehat{u}_{i,n} - u(\mathbf{x}_i)\right]$ is nonrandom and characterizes the accuracy of approximation of the function u at the point $\mathbf{x}_i$. As it explicitly depends on the unknown function $u(\mathbf{x}_i)$, its behavior is doubtful. Nevertheless, if we use the geometric inequality $|\mathbf{x}_j - \mathbf{x}_i| \leq \frac{\sqrt{2}}{2}|\Delta_{i,n}|^{1/2}$ and assume that there exists a real constant $0 < C_1 < \infty$ such that

$|u(\mathbf{x}_j) - u(\mathbf{x}_i)| \leq C_1|\mathbf{x}_j - \mathbf{x}_i|$, then $|\widehat{b}_{i,n}| \leq \frac{C_1}{\sqrt{2}}|\Delta_{i,n}|^{1/2}$. Accordingly, $|\widehat{b}_{i,n}|^2$ is of the order $O(|\Delta_{i,n}|)$ and typically increases when $\Delta_{i,n}$ increases (see also [16]).

– The behavior of the variance term is just opposite. The errors are independent and the stochastic term $\widehat{v}_{i,n}^2$ can be exactly computed on the basis of observations. Since $0 \leq \pi_{i\sim j,n} \leq 1$ and $\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} = 1$, it follows that $\widehat{\sigma}^2|\Delta_{i,n}|^{-1} \leq \widehat{v}_{i,n}^2 \leq \widehat{\sigma}^2$. In addition, we can reasonably assume that there exits a constant $0 \leq \gamma^2 \leq 1$ such that $\widehat{v}_{i,n}^2 \approx \widehat{\sigma}^2|\Delta_{i,n}|^{-\gamma^2}$. Accordingly, as $\Delta_{i,n}$ increases, more data is used to construct the estimate $\widehat{u}_{i,n}$, and so $\widehat{v}_{i,n}^2$ decreases.

Therefore, the bias and standard deviation terms are monotonous functions with opposite behaviors. In order to approximately minimize the local L_2 risk with respect to $|\Delta_{i,n}|$, a natural idea would be to minimize an upper bound of the form

$$\mathbb{E}|\widehat{u}_{i,n} - u(\mathbf{x}_i)|^2 \leq \frac{C_1^2}{2}|\Delta_{i,n}| + \widehat{\sigma}^2|\Delta_{i,n}|^{-\gamma^2}.$$

Unfortunately, the size of the optimal window defined as $|\Delta^\star(\mathbf{x}_i)| = \left[\frac{2\gamma^2\widehat{\sigma}^2}{C_1^2}\right]^{\frac{1}{\gamma^2+1}}$ cannot be used in practice since C_1 and γ are unknown. However, for this optimal solution $|\Delta^\star(\mathbf{x}_i)|$, it can be easily shown that the ratio between the optimal bias $b^\star(\mathbf{x}_i)$ and the optimal standard deviation $v^\star(\mathbf{x}_i)$ is not image dependent, i.e. $|b^\star(\mathbf{x}_i)| \leq \gamma v^\star(\mathbf{x}_i)$. Accordingly, the ideal window will be chosen as the largest window $\Delta_{i,n}$ such that $\widehat{b}_{i,n}$ is still not larger than $\gamma\widehat{v}_{i,n}$, for some real value $0 \leq \gamma^2 \leq 1$: $\Delta^\star(\mathbf{x}_i) = \sup_{\Delta_{i,n}\in\mathcal{N}_\Delta}\{|\Delta_{i,n}| : \widehat{b}_{i,n} \leq \gamma\widehat{v}_{i,n}\}$.

Now, we just point out that the estimator $\widehat{u}_{i,n}$ is usually decomposed as $\widehat{u}_{i,n} = u(\mathbf{x}_i) + \widehat{b}_{i,n} + \nu_i$ where $\nu_i \sim \mathcal{N}(0, \mathbb{E}[\nu_i^2])$. Hence, $\mathbb{E}[\widehat{u}_{i,n}] = u(\mathbf{x}_i) + \widehat{b}_{i,n}$, $\mathbb{E}[\nu_i^2] = \mathbb{E}[|\widehat{u}_{i,n} - u(\mathbf{x}_i) - \widehat{b}_{i,n}|^2] \overset{def}{=} \widehat{v}_{i,n}^2$ and the following inequality $|\widehat{u}_{i,n} - u(\mathbf{x}_i)| \leq \widehat{b}_{i,n} + \varkappa\widehat{v}_{i,n}$ holds with a high probability for $0 < \varkappa < \infty$. Accordingly, we can modify the previous definition of the *ideal* window as follows

$$\Delta^\star(\mathbf{x}_i) = \sup_{\Delta_{i,n}\in\mathcal{N}_\Delta}\{|\Delta_{i,n}| : |\widehat{u}_{i,n} - u(\mathbf{x}_i)| \leq (\gamma + \varkappa)\widehat{v}_{i,n}\}, \tag{5}$$

which depends no longer on $\widehat{b}_{i,n}$. In the next section, we shall see that a practical *data-driven* window selector based on this definition of $\Delta^\star(\mathbf{x}_i)$ which is yet related to the ideal and unobserved function $u(\mathbf{x}_i)$, can actually be derived.

A data-driven local window selector. In our approach, the collection of estimators $\{\widehat{u}_{i,1}, \ldots, \widehat{u}(\mathbf{x}_i)\}$ is naturally ordered in the direction of increasing $|\Delta_{i,n}|$ where $\widehat{u}(\mathbf{x}_i)$ can be thought as the best possible estimator with the smallest variance. A selection procedure can be then described based on pairwise comparisons of an essentially one-dimensional family of competing estimators $\widehat{u}_{i,n}$. In this modeling, the differences $\widehat{u}_{i,n} - \widehat{u}_{i,n'}$ are Gaussian random variables with known variances $\text{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'}) \leq \widehat{v}_{i,n'}^2$ with $1 \leq n' < n$, and expectations equal to the bias differences $\widehat{b}_{i,n} - \widehat{b}_{i,n'}$. From the definition of $\Delta^\star(\mathbf{x}_i)$ (see (5)),

we derive $|\widehat{u}_{i,n'} - \widehat{u}_{i,n}| \leq (2\gamma + \varkappa)\widehat{v}_{i,n'}$, $1 \leq n' < n$, and, among all good candidates $\{\widehat{u}_{i,n}\}$ satisfying this inequality, one choose the one with the smallest variance $\widehat{v}_{i,n}^2$. Following the above discussion, a window selector will be then based on the following pointwise rule [20, 21]:

$$\widehat{\Delta}(\mathbf{x}_i) = \arg \max_{\Delta_{i,n} \in \mathcal{N}_\Delta} \left\{ |\Delta_{i,n}| : |\widehat{u}_{i,n} - \widehat{u}_{i,n'}| \leq \varrho\,\widehat{v}_{i,n'}, \text{ for all } 1 \leq n' < n \right\} \quad (6)$$

where $\varrho = (2\gamma + \varkappa)$. This rule actually ensures the balance between the stochastic term and the bias term, and means that we take the largest window such that the estimators $\widehat{u}_{i,n}$ and $\widehat{u}_{i,n'}$ are not too different, in some sense, for all $1 \leq n' < n$ (see [18]). Hence, if an estimated point $\widehat{u}_{i,n'}$ appears far from the previous ones, this means that the bias is already too large and the window $\Delta_{i,n}$ is not a good one. This idea underlying our construction definitely belongs to Lepskii [20, 21].

Implementation. At the initialization, we naturally choose $|\Delta_{i,0}| = 1$, the fixed size of $p \times p$ patches and choose the number of iterations N_Δ. In addition, the noise variance $\widehat{\sigma}^2$ is robustly estimated from input data (see [17]). To complete

---

Algorithm.   Let $\{p, \alpha, \varrho, N_\Delta\}$ be the parameters

---

Initialization: compute $\widehat{\sigma}^2$ and set $\widehat{u}_{i,0} = Y_i$ and $\widehat{v}_{i,0}^2 = \widehat{\sigma}^2$ for each $\mathbf{x}_i \in G$.

Repeat

- for each $\mathbf{x}_i \in G$
  - compute

$$\pi_{i\sim j,n} = \frac{K\left(\lambda_\alpha^{-1} \operatorname{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)}{\displaystyle\sum_{\mathbf{x}_j \in \Delta_{i,n}} K\left(\lambda_\alpha^{-1} \operatorname{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)}$$

$$\widehat{u}_{i,n} = \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\, Y_j, \quad \widehat{v}_{i,n}^2 = \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} [\pi_{i\sim j,n}]^2$$

  - choose the window as

$$\widehat{\Delta}(\mathbf{x}_i) = \arg \max_{\Delta_{i,n} \in \mathcal{N}_\Delta} \left\{ |\Delta_{i,n}| : |\widehat{u}_{i,n} - \widehat{u}_{i,n'}| \leq \varrho\,\widehat{v}_{i,n'}, \text{ for all } 1 \leq n' < n \right\}.$$

    If this rule is violated at iteration n, we do not accept $\widehat{u}_{i,n}$ and keep the estimate $\widehat{u}_{i,n-1}$ as the final estimate at $\mathbf{x}_i$, i.e. $\widehat{u}(\mathbf{x}_i) = \widehat{u}_{i,n-1}$ and $\widehat{n}(\mathbf{x}_i) = n - 1$. This estimate is unchanged at the next iterations and $\mathbf{x}_i$ is "frozen".
- increment n

while $n \leq N_\Delta$.

Fig. 1. Patch-based image regularization algorithm

the procedure, we choose $\varrho \in [2, 4]$ in order to get a good accuracy for the pointwise estimator (see [18] for the proof) and λ_α as a $1 - \alpha = 0.99$-quantile of a $\chi^2_{p^2, 1-\alpha}$ distribution. Finally, the complexity of the algorithm given in Fig. 1, is of the order $p \times p \times |G| \times (|\Delta_{\cdot,1}| + \ldots + |\Delta_{\cdot,N_\Delta}|)$.

3 Experimental Results

Our results were measured by the peak signal-to-noise ratio (PSNR) in decibels (db) as $\text{PSNR} = 10 \log_{10}(255^2/\text{MSE})$ with $\text{MSE} = |G|^{-1}\sum_{\mathbf{x}_i \in G}(u_o(\mathbf{x}_i) - \widehat{u}(\mathbf{x}_i))^2$ where u_0 is the noise-free original image. We have done simulations on a commonly-used set of images available at http://decsai.ugr.es/~javier/denoise/test_images/ and described in [26]. In all our experiments, we have chosen image patches of 9×9 pixels and set the algorithm parameters as follows: $\lambda_{0.01} = \chi^2_{81,0.01} = 113.5$, $\varrho = 3$ and $N_\Delta = 4$ (see [18]). The processing of a 256×256 image required typically about 1 minute ($p = 9$) on a PC (2.0 Ghz, Pentium IV) using a C++ implementation. The potential of the estimation method is mainly illustrated with the 512×512 *lena* image (Fig. 2a) corrupted by an additive white-Gaussian noise (WGN) (Fig. 2b, PSNR = 22.13 db, $\sigma = 20$). In Fig. 2c, the noise is reduced in a natural manner and significant geometric features, fine textures, and original contrasts are visually well recovered with no undesirable
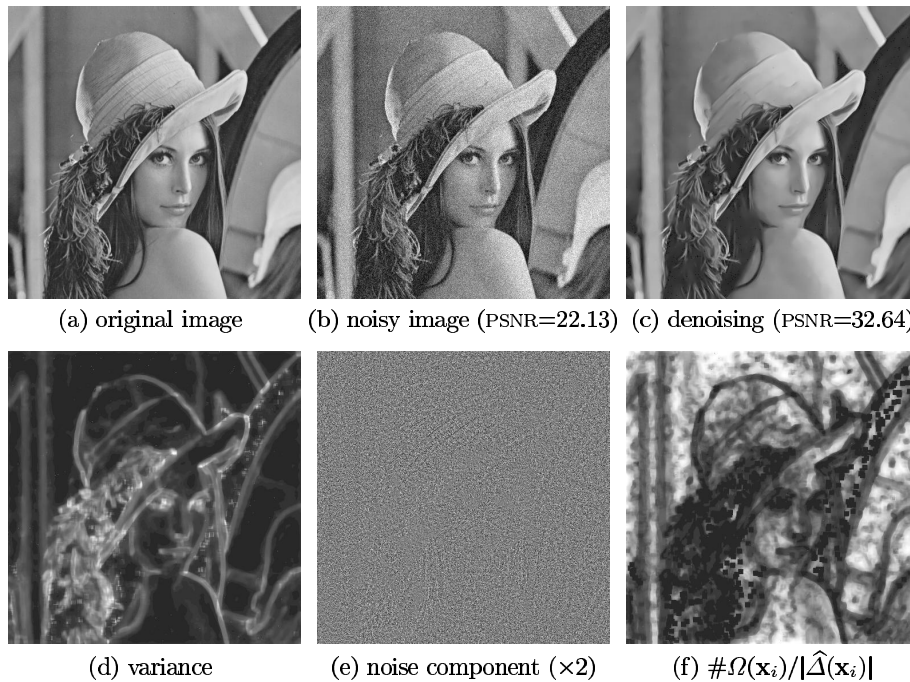

(a) original image   (b) noisy image (PSNR=22.13)   (c) denoising (PSNR=32.64)

(d) variance   (e) noise component (×2)   (f) $\#\Omega(\mathbf{x}_i)/|\widehat{\Delta}(\mathbf{x}_i)|$

Fig. 2. Denoising of the artificially corrupted *lena* image (WGN, $\sigma = 20$)

(a) our method   (b) NL-means filter   (c) Fields-of-Experts   (d) BLS-GSM

Fig. 3. Denoising results on the noisy *lena* image (WGN, $\sigma = 20$). a) Our method (PSNR = 32.64), b) NL-means filter [7] (PSNR=31.09), c) Fields-of-Experts [27] (PSNR=31.92), d) wavelet-based denoising method (BLS-GSM) [26] (PSNR=32.66).
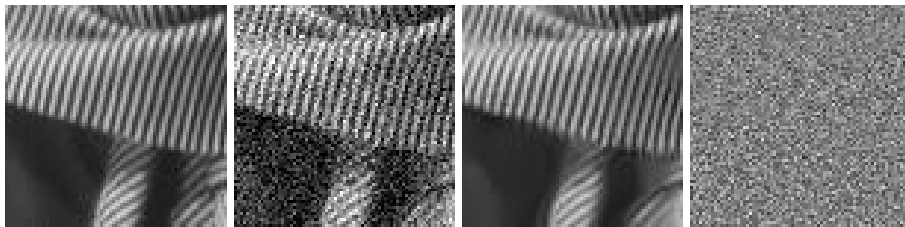

Fig. 4. Detail of the *barbara* image. From left to right: original image, artificially corrupted image (WGN, $\sigma = 30$), result of our patch-based method, difference between the noisy image and the regularized image (noise component).

artifacts (PSNR = 32.64 db). The noise component is shown in Fig. 2e (magnification factor of 2) and has been estimated by calculating the difference between the noisy image (Fig. 2b) and the recovered image (Fig. 2c). The estimated noise component contains few geometric structures and is similar to a simulated white Gaussian noise. To better appreciate the accuracy of the denoising process, the variance $\widehat{v}^2(\mathbf{x}_i)$ of the pointwise estimator $\widehat{u}(\mathbf{x}_i)$ is shown in Fig. 2d where dark values correspond to high-confidence estimates. As expected, pixels with a low level of confidence are located in the neighborhood of image discontinuities. Figure 2f shows the probability of a patch $\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)$:
$$\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i) \text{ occurring in } \widehat{\Delta}(\mathbf{x}_i)\} \stackrel{def}{=} \#\Omega(\mathbf{x}_i)/|\widehat{\Delta}(\mathbf{x}_i)|$$ where $\Omega(\mathbf{x}_i)$ is used to denote the set $\{\mathbf{x}_j \in \widehat{\Delta}(\mathbf{x}_i) : \text{dist}(\widehat{\mathbf{u}}(\mathbf{x}_i), \widehat{\mathbf{u}}(\mathbf{x}_j)) \leq \lambda_\alpha\}$. In Fig. 2f, dark values correspond low probabilities of occurrence and, it is confirmed that repetitive patterns in the neighborhood of image discontinuities are mainly located along image edges. Our approach is also compared to the *non-local means* algorithm [7] using 7×7 image patches and a fixed search window of 21×21 pixels as recommended by the authors: the visual impression and the numerical results are improved using our algorithm (see Fig. 3b). Finally, we have compared the performance of our method to the Wiener filtering (WF) (Matlab function `wiener2`) and other competitive methods [28, 31, 24], including recent patch-based approaches [7, 27] and pointwise adaptive estimation approaches [25, 17]. We point out that, visually
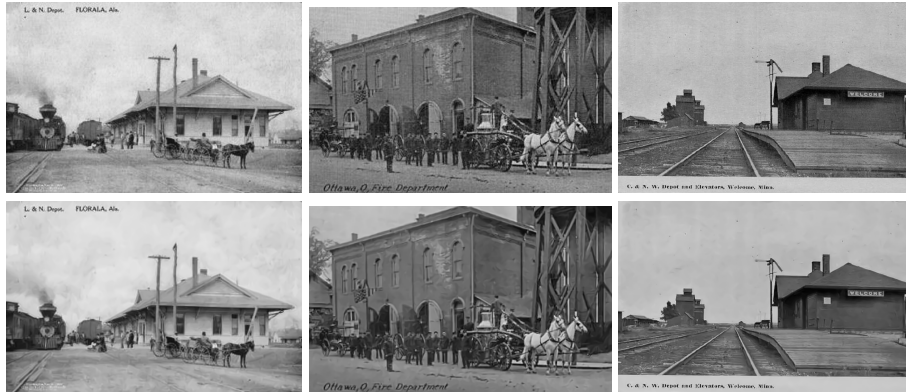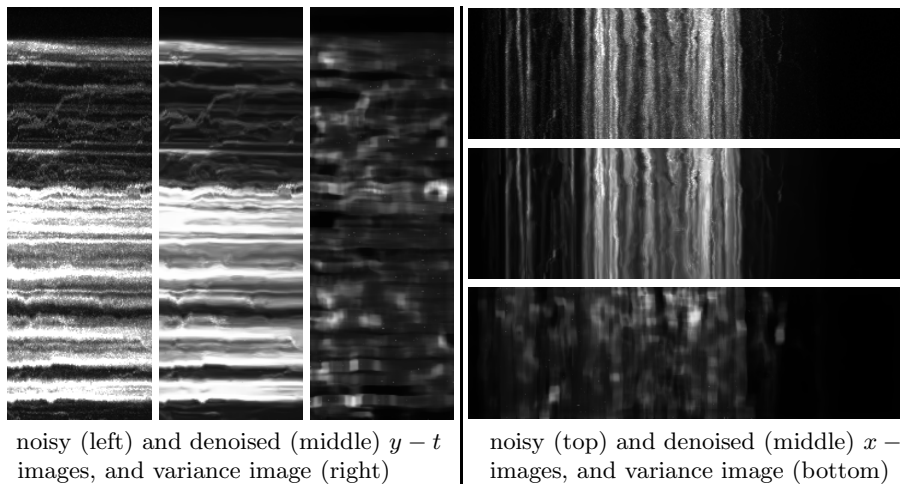
Fig. 5. Real noisy photographs (top) and restoration results (bottom)

noisy (left) and denoised (middle) $y - t$ images, and variance image (right)

noisy (top) and denoised (middle) $x - t$ images, and variance image (bottom)

Fig. 6. Results on a 2D image depicting trajectories of vesicles of transport in the spatio-temporal planes $y - t$ (left) and $x - t$ (right) (analysis of Rab proteins involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum).

and quantitatively, our very simple and unsupervised algorithm favorably compares to any of these denoising algorithms, including the more sophisticated and best known wavelet-based denoising methods [30, 26] (see Fig. 3d) and learned filters-based denoising methods [27] (see Fig. 3c). In Table 1 (a), we reported the best available published PSNR results for the same image dataset [26] ; we note that our method nearly outperforms any of the tested methods in terms of PSNR. Also, if the PSNR gains are marginal for some images, the visual difference can be significant as shown in Fig. 3 where less artifacts are visible using our method (see also Fig. 4). Nevertheless, other competitive unsupervised patch-based methods exist (e.g. see [15, 2]), but we did not report the results on this

Table 1. Performances of our method when applied to test noisy (WGN) images

Image σ/PSNR	Lena 20/22.13	Barbara 20/22.18	Boats 20/22.17	House 20/22.11	Peppers 20/22.19
Our method	**32.64**	**30.37**	**30.12**	**32.90**	**30.59**
Buades *et al.* [7]	31.09	29.38	28.60	31.54	29.05
Kervrann [17]	30.54	26.50	28.01	30.70	28.23
Polzehl *et al.* [25]	29.74	26.05	27.74	30.31	28.40
Portilla *et al.* [26]	32.66	30.32	30.38	32.39	30.31
Roth *et al.* [27]	31.92	28.32	29.85	32.17	30.58
Rudin *et al.* [28]	30.48	27.07	29.02	31.03	28.51
Starck *et al.* [30]	31.95	-	-	-	-
Tomasi *et al.* [31]	30.26	27.02	28.41	30.01	28.88
Wiener filering	28.51	26.99	27.97	28.74	28.10

(a) Performances of denoising algorithms when applied to test noisy (WGN) images.

σ / PSNR	Lena 512^2	Barbara 512^2	Boats 512^2	House 256^2	Peppers 256^2
5 / 34.15	37.91	37.12	36.14	37.62	37.34
10 / 28.13	35.18	33.79	33.09	35.26	34.07
15 / 24.61	33.70	31.80	31.44	34.08	32.13
20 / 22.11	32.64	30.37	30.12	32.90	30.59
25 / 2017	31.73	29.24	29.20	32.22	29.73
50 / 14.15	28.38	24.09	25.93	28.67	25.29
75 / 10.63	25.51	22.10	23.69	25.49	22.31
100 / 8.13	23.32	20.64	21.78	23.08	20.51

patch size	Lena 512^2	Barbara 512^2	Boats 512^2	House 256^2	Peppers 256^2
3×3	32.13	28.97	29.86	32.69	30.86
5×5	32.52	29.97	30.15	33.05	30.98
7×7	32.63	30.27	30.17	33.03	30.80
9×9	32.64	30.37	30.12	32.90	30.59

(b) Performances of our method ($p = 9, N_\Delta = 4, \alpha = 0.01$) for different signal-to-noise ratios (WGN).

(c) Performances of our method ($N_\Delta = 4, \alpha = 0.01$) for different patch sizes (WGN, $\sigma = 20$)

image dataset since they are not available. These methods must be considered for future comparisons. To complete the experiments, Table 1 (b) shows the PSNR values using our patch-based regularization method when applied to this set of test images for a wide range of noise variance. Moreover, we have also examined some complementary aspects of our approach. Table 1 (c) shows the PSNR values obtained by varying the patch size. Note the PSNR values are close for every patch size and the optimal patch size depends on the image contents. Nevertheless, a patch 9×9 seems appropriate in most cases and a smaller patch can be considered for processing piecewise smooth images.

In the second part of experiments, the effects of the patch-based regularization is approach are illustrated on real old photographs. The set of parameters is unchanged for processing all these test images: $p = 9, N_\Delta = 4, \alpha = 0.01$. In most cases, a good compromise between the amount of smoothing and preservation of edges and textures is automatically reached. In that case, the noise variance $\widehat{\sigma}^2$ is automatically estimated from image data. The reconstruction of images is respectively shown in Fig. 5. Note that geometric structures are well preserved and the noise component corresponding to homogeneous artifacts is removed.

Finally, we have applied the patch-based restoration method to noisy 2D images extracted from a temporal 2D+time ($xy - t$) sequence of 120 microscopy images in intracellular biology, showing a large number of small fluorescently labeled moving vesicles in regions close to the Golgi apparatus (courtesy of Curie Institute). The reading of observed trajectories is easier if the patch-based

estimation method is applied to both noisy $x-t$ or $y-t$ projection images shown in Fig. 6 (see also [6]).

4   Conclusion

We have described a novel adaptive regularization algorithm where patch-based weights and variable window sizes are jointly used. The use of variable and overlapping windows contributes to the regularization performance with no block effect, enhances the flexibility of the resulting local regularizers and make them possible to cope well with spatial inhomogeneities in natural images. An advantage of the method is that internal parameters can be easily chosen and are relatively stable. The algorithm is able to regularize both piecewise-smooth and textured natural images since they contain enough redundancy. Actually, the performance of our simple algorithm is very close to that of the best already published denoising methods. In the future, we plan to study the automatic patch size selection to better adapt to image contents.

References

1. Aujol, J.F., Aubert, G., Blanc-Fraud, L., Chambolle, A.: Image decomposition into a bounded variation component and an oscillating component. J. Math. Imag. Vis. **22** (2005) 71-88
2. Awate, S.P., Whitaker, R.T.: Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In: Proc. CVPR'05, San Diego (2005)
3. Barash, D., Comaniciu, D.: A Common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. Image Vis. Comp. **22** (2004) 73-81
4. Black, M.J., Sapiro, G.: Edges as outliers: anisotropic smoothing using local image statistics. In: Proc. Scale-Space'99, Kerkyra (1999)
5. Brox, T., Weickert, J.: A TV flow based local scale measure for texture discrimination. In: Proc. ECCV'04, Prague (2004)
6. Boulanger, J., Kervrann, C., Bouthemy, P.: Adaptive spatio-temporal restoration for 4D fluoresence microscopic imaging. In: Proc. MICCAI'05, Palm Springs (2005)
7. Buades, A., Coll, B., Morel, J.M.: Image denoising by non-local averaging. In: Proc. CVPR'05, San Diego (2005)
8. Comaniciu, D., Ramesh, V., Meer, P.: The variable bandwidth mean-shift and data-driven scale selection. In: Proc. ICCV'01, Vancouver (2001)
9. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based inpainting. IEEE T. Image Process. **13** (2004) 1200-1212
10. Efros, A., Leung, T.: Texture synthesis by non-parametric sampling. In: Proc. ICCV'99, Kerkyra (1999)
11. Elad, M.: On the bilateral filter and ways to improve it. IEEE T. Image Process. **11** (2002) 1141-1151
12. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. Int. J. Comp. Vis. **40** (2000) 25-47
13. Gilboa, G., Sochen, N., Zeevi, Y.Y.: Texture preserving variational denoising using an adaptive fidelity term. In: Proc. VLSM'03, Nice (2003)

14. Godtliebsen, F., Spjotvoll, E., Marron, J.S.: A nonlinear Gaussian filter applied to images with discontinuities. J. Nonparametric Stat. **8** (1997) 21-43
15. Jojic, N., Frey, B., Kannan, A.: Epitomic analysis of appearance and shape. In: Proc. ICCV'03, Nice (2003)
16. Katkovnik, V., Egiazarian, K., Astola, J.: Adaptive window size image denoising based on intersection of confidence intervals (ICI) rule. J. Math. Imag. Vis. **16** (2002) 223-235
17. Kervrann, C.: An adaptive window approach for image smoothing and structures preserving. In: Proc. ECCV'04, Prague (2004)
18. Kervrann, C., Boulanger, J.: Local adaptivity to variable smoothness for exemplar-based image denoising and representation. INRIA RR-5624 (2005)
19. Lee, J.S.: Digital image smoothing and the sigma filter. Comp. Vis. Graph. Image Process. **24** (1983) 255-269
20. Lepskii, O.: On a problem of adaptive estimation on white Gaussian noise. Th. Prob. Appl. **35** (1980) 454-466
21. Lepskii, O.V., Mammen, E., Spokoiny, V.G.: Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. Ann. Stat. **25** (1997) 929-947
22. Mrazek, P., Weickert, J., Bruhn, A.: On robust estimation and smoothing with spatial and tonal kernels. Preprint 51, U. Bremen (2004)
23. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and variational problems. Comm. Pure and Appl. Math. **42** (1989) 577-685
24. Perona. P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE T. Patt. Anal. Mach. Intell. **12** (1990) 629-239
25. Polzehl, J., Spokoiny, V.: Adaptive weights smoothing with application to image restoration. J. Roy. Stat. Soc. B **62** (2000) 335-354
26. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE T. Image Process. **12** (2003) 1338-1351
27. Roth, S., Black, M.J.: Fields of experts: a framework for learning image priors with applications. In: Proc. CVPR'05, San Diego (2005)
28. Rudin, L., Osher, S., Fatemi, E.: Nonlinear Total Variation based noise removal algorithms. Physica D (2992) **60** (1992) 259-268
29. Sochen, N., Kimmel, R., Bruckstein, A.M.: Diffusions and confusions in signal and image processing. J. Math. Imag. Vis. **14** (2001) 237-244
30. Starck, J.L., Candes, E., Donoho, D.L.: The Curvelet transform for image denoising. IEEE T. Image Process. **11** (2002) 670-684
31. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. ICCV'98, Bombay (1998)
32. Tschumperlé, D.: Curvature-preserving regularization of multi-valued images using PDE's. In: Proc. ECCV'06, Graz (2006)
33. van de Weijer, J., van den Boomgaard, R.: Local mode filtering. In: Proc. CVPR'01, Kauai (2001)
34. Zhang, D., Wang, Z.: Image information restoration based on long-range correlation. IEEE T. Circ. Syst. Video Technol. **12** (2002) 331-341
35. Zhu, S.C., Wu, Y., Mumford, D.: Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. Int. J. Comp. Vis. **27** (1998) 107-126