

# An Adaptive Window Approach for Image Smoothing and Structures Preserving

Charles Kervrann

IRISA - INRIA Rennes / INRA - Mathématiques et Informatique Appliquées  
Campus de Beaulieu, 35042 Rennes Cedex, France  
`ckervran@irisa.fr`

**Abstract.** A novel adaptive smoothing approach is proposed for image smoothing and discontinuities preservation. The method is based on a locally piecewise constant modeling of the image with an adaptive choice of a window around each pixel. The adaptive smoothing technique associates with each pixel the weighted sum of data points within the window. We describe a statistical method for choosing the optimal window size, in a manner that varies at each pixel, with an adaptive choice of weights for every pair of pixels in the window. We further investigate how the I-divergence could be used to stop the algorithm. It is worth noting the proposed technique is data-driven and fully adaptive. Simulation results show that our algorithm yields promising smoothing results on a variety of real images.

## 1 Introduction

The problem of image recovering is to reduce undesirable distortions and noise while preserving important features such as homogeneous regions, discontinuities, edges and textures. Popular image smoothing algorithms are therefore nonlinear to reduce the amount of smoothing near abrupt changes, i.e. edges. Most of them are based on discrete [3] or continuous [18, 21] energy functionals minimization since they are designed to explicitly account for the image geometry. In recent years, a large number of partial differential equations (PDE) and variational methods, including anisotropic diffusion [20, 4] and total variation (TV) minimization [21], have shown impressive results to tackle the problem of edge-preserving smoothing [5, 4, 6, 17] and to separate images into noise, texture and piecewise smooth components [16, 19].

In this paper, we also address the adaptive image smoothing problem and present a nonparametric estimation method that smooth homogeneous regions and inhibits smoothing in the neighborhood of discontinuities. The proposed *adaptive window approach* differs from previous energy minimization-based methods [3, 18, 21]. It is conceptually very simple being based on the key idea of estimating a regression function with an adaptive choice of the window size (neighborhood) in which the unknown function is well approximated by a constant. At each pixel, we estimate the regression function by iteratively growing a window and adaptively weighting input data to achieve an optimal compromise between

the bias and variance [14, 15, 13]. The motivation behind this nonparametric estimation approach is to use a well-established theory in statistics [10, 14] for adaptive smoothing, yielding to non-iterative algorithms for 2D-3D imaging. The proposed algorithm complexity is actually controlled by simply restricting the size of the larger window and setting the window growing factor. In contrast to most digital diffusion-based filtering processes for which the input noisy image is “abandoned” after the first iteration [20, 4], the adaptive window approach recycles at each step the original data. Other related works to our approach are nonlinear Gaussian filters (iterative or non-iterative bilateral filtering [12, 22, 7, 1, 23]) that essentially average values within a local window but changes the weights according to local differences in the intensity [12, 22, 7, 1, 23]. However, these weighted schemes use a static window size which can be arbitrarily large, in the both spatial and range domains. Our structure-adaptive smoothing also works in the joint spatial-range domain but has a more powerful adaptation to the local structure of the data since the size of the window and internal parameters are computed from local image statistics.

## 2 A statistical nonparametric approach

We observe the function  $u$  with some additive errors  $\xi_i$ :

$$Y_i = u(x_i) + \xi_i, \quad i = 1, \dots, n, \quad (1)$$

where  $x_i \in \mathbb{R}^2$  represents the spatial coordinates of the discrete image domain  $\mathcal{S}$  of  $n$  pixels and  $Y_i \in \mathbb{R}$  is the observed intensity at location  $x_i$ . We suppose the errors  $\xi_i$  to be independent and distributed zero-mean random variables with unknown variances, i.e.  $\text{var}(\xi_i) = \sigma_i^2$ .

### 2.1 Image model and basic idea

A classical nonparametric estimation approach is based on the structural assumption that regression function  $u(x)$  is constant in a local neighborhood in the vicinity of a point  $x$ . An important question under such an approach is first how to determine for each pixel the size and shape of the neighborhood under concern from image data. The regression function  $u(x)$  can be then estimated from the observations lying in the estimated neighborhood of  $x$  by a local maximum likelihood (ML) method.

Our procedure is iterative and uses this idea. First, suppose we are given a local window  $\mathcal{W}_i^{(0)}$  containing the point of estimation  $x_i$ . By  $\hat{u}_i^{(0)}$  we denote an approximation of  $\hat{u}^{(0)}(x_i)$ . We can calculate an initial ML estimate  $\hat{u}_i^{(0)}$  at point  $x_i$  (and its variance  $\hat{\vartheta}_i^{(0)}$ ) by averaging observations over a small neighborhood  $\mathcal{W}_i^{(0)}$  of  $x_i$  as

$$\hat{u}_i^{(0)} = \frac{1}{|\mathcal{W}_i^{(0)}|} \sum_{x_j \in \mathcal{W}_i^{(0)}} Y_j \quad \text{and} \quad \hat{\vartheta}_i^{(0)} = \frac{\hat{\sigma}_i^2}{|\mathcal{W}_i^{(0)}|} \quad (2)$$

where  $\hat{\sigma}_i^2$  is a pilot estimator which can be plugged in place of the noise variance  $\sigma_i^2$  and  $|\mathcal{W}_i^{(0)}|$  denotes the number of points  $x_j \in \mathcal{W}_i^{(0)}$ . At the next iteration, a larger neighborhood  $\mathcal{W}_i^{(1)}$  with  $\mathcal{W}_i^{(0)} \subset \mathcal{W}_i^{(1)}$  centered at  $x_i$  is considered, and every point  $x_j$  from  $\mathcal{W}_i^{(1)}$  gets a weight  $w_{ij}^{(1)}$  which is defined by comparing the estimates  $\hat{u}_i^{(0)}$  and  $\hat{u}_j^{(0)}$  obtained at the first iteration. Then we recalculate the estimate  $\hat{u}_i^{(1)}$  as the weighted average of data points lying in the neighborhood  $\mathcal{W}_i^{(1)}$ . We continue this way, growing with  $k$  the considered neighborhood  $\mathcal{W}_i^{(k)}$ ; for each  $k \geq 1$ , the ML estimate  $\hat{u}_i^{(k)}$  and its variance are given by:

$$\hat{u}_i^{(k)} = \sum_{x_j \in \mathcal{W}_i^{(k)}} w_{ij}^{(k)} Y_j \quad \text{and} \quad \hat{\vartheta}_i^{(k)} = \hat{\sigma}_i^2 \sum_{x_j \in \mathcal{W}_i^{(k)}} \left[ w_{ij}^{(k)} \right]^2 \quad (3)$$

where weights  $w_{ij}^{(k)}$  are continuous variables ( $0 \leq w_{ij}^{(k)} \leq 1$ ), computed by comparison of the preceding estimates  $\hat{u}_i^{(k-1)}$  and  $\hat{u}_j^{(k-1)}$ . Note we can also write  $\hat{u}_i^{(k)} = \sum_{j=1}^n 1_{\{x_j \in \mathcal{W}_i^{(k)}\}} w_{ij}^{(k)} Y_j$  where  $1_{\{x_j \in \mathcal{W}_i^{(k)}\}}$  is the spatial rectangular kernel. In the following, we use a spatial rectangular kernel (square windows) for mathematical convenience, but the method can be easily extended to the case of a more usual spatial Gaussian kernel [12, 22, 7, 1, 23]. Moreover, we choose an optimal window for each pixel  $x_i$  by comparing the new estimate  $\hat{u}_i^{(k)}$  with the estimate  $\hat{u}_i^{(k-1)}$  obtained at the preceding iteration. Finally, a global convergence criterion is introduced to stop the estimation procedure.

At this level, an important connection between our method and local robust M-estimation [7] should be mentioned. In Equation (3), the weight function  $w_{ij}^{(k)}$  does not depend on input data but are only calculated from neighboring local estimates, which contributes to the regularization performance. In addition, the method is able to recover a piecewise smooth image even the underlying image model is locally constant as it is confirmed in our experiments (see Fig. 2). The approach is similar also (at least in spirit) to twofold weighting schemes employed in the bilateral filtering [22, 1, 23] and *mean shift*-based algorithms [8].

## 2.2 Adaptive weights

In our approach, we may decide on the basis of the estimates  $\hat{u}_i^{(k-1)}$  and  $\hat{u}_j^{(k-1)}$ , whether a point  $x_i$  and  $x_j \in \mathcal{W}_i^{(k)}$  are in the same region and then prevent significant discontinuities oversmoothing. In the local Gaussian case, significance is measured using a contrast  $|\hat{u}_i^{(k-1)} - \hat{u}_j^{(k-1)}|$ . If this contrast is high compared to  $\sqrt{\hat{\vartheta}_i^{(k-1)}}$ , then  $x_j$  is an outlier and should not participate to the estimation of  $\hat{u}_i^{(k)}$  and  $w_{ij}^{(k)} \rightarrow 0$ .

Hence, motivated by the robustness and smoothing properties of the Huber M-estimator in the probabilistic approach of image denoising [2], we introduce

the following related weight function (but other weight functions are possible [4, 23])

$$w_{ij}^{(k)} = \frac{g_{ij}^{(k)}}{\sum_{x_j \in \mathcal{W}_i^{(k)}} g_{ij}^{(k)}}, \quad g_{ij}^{(k)} = \begin{cases} 1 & \text{if } \left| \widehat{u}_i^{(k-1)} - \widehat{u}_j^{(k-1)} \right| \leq \lambda \sqrt{\widehat{\vartheta}_i^{(k-1)}} \\ \frac{\lambda \sqrt{\widehat{\vartheta}_i^{(k-1)}}}{\left| \widehat{u}_i^{(k-1)} - \widehat{u}_j^{(k-1)} \right|} & \text{otherwise.} \end{cases} \quad (4)$$

Here  $\lambda \sqrt{\widehat{\vartheta}_i^{(k-1)}}$  is related to the spatially varying fraction of contamination of the Gaussian distribution: for the majority of points  $x_j \in \mathcal{W}_i$ , the values  $\widehat{u}_i^{(k-1)} - \widehat{u}_j^{(k-1)}$  can be approximatively modeled as being constant (zero) with random Gaussian noise. Hence  $\lambda$  is an appropriate quantile of the standard normal distribution and depends on the level of noise in images. In our experiments, we arbitrarily set  $\lambda = 3$  according to the well known “rule of 3 sigma”.

### 2.3 Optimal window selection

Statistical inference under such a structural assumption focuses on searching for every point  $x_i$  the largest neighborhood (window)  $\mathcal{W}_i$  where the hypothesis of structural homogeneity is not rejected. In other words, we seek to estimate a regression function  $u$  from the data, while having deal with a so-called nuisance parameter, that is the neighborhood. The classical measure of the closeness of the estimator  $\widehat{u}$  obtained in the window  $\mathcal{W}_i$  to its target value  $u$  is the mean squared error (MSE) which is decomposed into the sum of the squared bias  $[\text{Bias}(\widehat{u}_i)]^2$  and variance  $\widehat{\vartheta}_i$ .

As explained before, we should choose a window that achieves an optimal compromise between the squared bias and variance for each pixel. Accordingly, we make the reasonable assumption that the squared bias is an increasing function of the neighborhood size and the variance is a decreasing function of the neighborhood size. Then, in order to minimize the MSE we search for the window where the squared bias and the variance of the estimate are equal. The corresponding critical MSE is ( $E[\cdot]$  denotes the mathematical expectation):

$$\text{MSE} = E \left[ \widehat{u}_i^{(k)} - u(x_i) \right]^2 = \left[ \text{Bias} \left( \widehat{u}_i^{(k)} \right) \right]^2 + \widehat{\vartheta}_i^{(k)} = 2\widehat{\vartheta}_i^{(k)}. \quad (5)$$

Now, let us introduce a finite set of windows  $\{\mathcal{W}_i^{(0)}, \dots, \mathcal{W}_i^{(k_M)}\}$  centered at  $x_i \in \mathcal{S}$ , with  $\mathcal{W}_i^{(k)} \subset \mathcal{W}_i^{(k+1)}$ , starting with a small  $\mathcal{W}_i^{(0)}$  and the corresponding estimates  $\widehat{u}_i^{(0)}$  of the true image  $u(x_i)$ . Denote by  $\widehat{k}_i$  the ideal window size corresponding to the minimum value of the pointwise MSE at location  $x_i$ . Accordingly,  $\mathcal{W}_i^{(\widehat{k}_i)}$  can be obtained according to the following statistical pointwise rule [14, 15, 13]:

$$\widehat{k}_i = \max \left\{ k : \forall k' < k : \left| \widehat{u}_i^{(k)} - \widehat{u}_i^{(k')} \right|^2 \leq 8\widehat{\vartheta}_i^{(k')} \right\}. \quad (6)$$

In other words, as long as successive estimates  $\hat{u}_i^{(k)}$  stay close to each other, we decide that the bias is small and the size of the estimation window can be increased to improve the estimation of the constant model (and to decrease the variance of the estimate  $\hat{u}_i^{(k)}$ ). If an estimated point  $\hat{u}_i^{(k')}$  appears far from the previous ones, we interpret this as the dominance of the bias over the variance term. For each pixel, the detection of this transition enables to determine the critical size that balances bias and variance. Note the choice of the detection threshold in (6) between  $2^2$  and  $4^2$  does not change the result of the algorithm significantly.

## 2.4 Global stopping rule

A stopping rule can be used to save computing time if two successive solutions are very close and prevent an useless setting of the larger window size. Here, we adopt the so-called Csiszár's I-divergence [9] to detect global convergence:

$$I(\hat{u}^{(k)}, \hat{u}^{(k+1)}) = \sum_{i=1}^n \left[ \hat{u}_i^{(k)} \log \frac{\hat{u}_i^{(k)}}{\hat{u}_i^{(k+1)}} - \hat{u}_i^{(k)} + \hat{u}_i^{(k+1)} \right]. \quad (7)$$

We choose this criterion to obtain the distance between succeeding iterations since the decorrelation criterion proposed in [17] tends to underestimate the number of necessary iterations in our framework. In addition, the I-divergence criterion can be used for a variety of algorithms, as it does not directly depend on the restoration method. In practice, the I-divergence is normalized with its maximal occurring value at iteration  $k = 0$ . When  $I(\hat{u}^{(k)}, \hat{u}^{(k+1)})$  sinks under a threshold (of the order  $10^{-3}$  for typical images) that sufficiently represents convergence, the algorithm is stopped at the final iteration  $k_c = k$ , with  $k_c \leq k_M$ . Finally, the window size increases at each iteration  $k$  if the global convergence criterion is not met (or  $k \leq k_M$ ) without changing the estimate  $\hat{u}_i^{(k)}$  if the point-wise rule (6) has already been violated at  $x_i$ , i.e.  $\hat{u}_i^{(k)} = \hat{u}_i^{(\hat{k}_i)}$  if  $\hat{k}_i < k$ . If the rule (6) has not been violated at  $x_i$ , we have  $k_i = k$  where  $k$  is the current iteration of the algorithm.

## 3 Implementation

The key ingredient of the procedure is an increasing sequence of neighborhoods  $\mathcal{W}_i^{(k)}$ ,  $k = 0, 1, \dots, k_M$  with  $\mathcal{W}_i^{(k)} \subset \mathcal{W}_i^{(k+1)}$  centered at each image pixel  $x_i$ . In what follows,  $|\mathcal{W}_i^{(k)}|$  denotes the number of points  $x_j$  in  $\mathcal{W}_i^{(k)}$ , i.e.  $|\mathcal{W}_i^{(k)}| = \#\{x_j \in \mathcal{W}_i^{(k)}\}$ . In our experiments, we arbitrarily use neighborhoods  $\mathcal{W}^{(k)}$  corresponding to successive square windows of size  $|\mathcal{W}^{(k)}| = (2k + 1) \times (2k + 1)$  pixels with  $k = 0, 1, 2, \dots, k_M$ .

The estimation procedure described in Section 2.1 relies also on the preliminary local estimation of the noise variance. This estimation is an (off-line) pre-processing step to initialize the adaptive smoothing procedure. In most applications, the noise variance  $\sigma_i^2$  at each image point  $x_i$  is unknown and an

estimate  $\hat{\sigma}_i^2$  can be obtained from data as

$$\hat{\sigma}_i^2 = \frac{1}{|\mathcal{B}_i|} \sum_{x_j \in \mathcal{B}_i} \varepsilon_j^2 \quad (8)$$

where  $\mathcal{B}_i$  denotes a square block of pixels centered at  $x_i$  and local residuals  $\varepsilon_j$  are defined as (we note  $Y_{j_1, j_2}$  the observation  $Y_j$  at site  $j = (j_1, j_2)$ ) [10]:

$$\varepsilon_j = \frac{4Y_{j_1, j_2} - (Y_{j_1+1, j_2} + Y_{j_1-1, j_2} + Y_{j_1, j_2+1} + Y_{j_1, j_2-1})}{\sqrt{20}}. \quad (9)$$

In presence of discontinuities in the block  $\mathcal{B}_i$ , a local estimate of the noise variance based on robust statistics is preferable. In this framework, high discontinuities within the region  $\mathcal{B}_i$  correspond to statistical outliers with respect to local image contrasts. As in [4], we suggest to define  $\hat{\sigma}_i^2 = \max(\hat{\sigma}^2, \hat{\sigma}_{\mathcal{B}_i}^2)$  where  $\hat{\sigma}^2$  and  $\hat{\sigma}_{\mathcal{B}_i}^2$  are respectively the robust estimates of the noise variance computed for the entire image and within a local block centered at point  $x_i$ . We propose the following robust estimates for  $\hat{\sigma}$  and  $\hat{\sigma}_{\mathcal{B}_i}$ :

$$\hat{\sigma} = 1.4826 \text{ median} (| |\underline{\varepsilon}_{\mathcal{S}}| - \text{median}|\underline{\varepsilon}_{\mathcal{S}}| |) \quad (10)$$

$$\hat{\sigma}_{\mathcal{B}_i} = 1.4826 \text{ median} (| |\underline{\varepsilon}_{\mathcal{B}_i}| - \text{median}|\underline{\varepsilon}_{\mathcal{B}_i}| |) \quad (11)$$

where  $\underline{\varepsilon}_{\mathcal{S}} = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$  is the set of  $n$  local residuals of the entire image,  $\underline{\varepsilon}_{\mathcal{B}_i}$  is the set of  $|\mathcal{B}_i|$  local residuals within the local block  $\mathcal{B}_i$  and the constant is derived from the fact that the median absolute deviation of a zero-mean normal distribution with unit variance is  $0.6745 = 1./1.4826$ . The local estimation  $\hat{\sigma}_{\mathcal{B}_i}^2$  of noise variance is useful for filtering off spatially varying textures. The global estimate of the noise variance provides also a reasonable lower bound and prevents the amplification of noise in relatively homogeneous areas. From a practical point of view, we have explored the computation of the local estimate of the noise variance within blocks  $\mathcal{B}_i$  of size  $(2k_M + 1) \times (2k_M + 1)$ , i.e.  $\mathcal{B}_i = \mathcal{W}_i^{(k_M)}$  at every location  $x_i$  in the image.

Below, we give the proposed algorithm:

- **Initialization:** For each point  $x_i$ , we calculate initial estimates  $\hat{u}_i^{(0)}$  and  $\hat{v}_i^{(0)}$  using Equation (2) and set  $k = 1$ . We naturally choose  $|\mathcal{W}^{(0)}| = 1$ , i.e. the initial local neighborhood  $\mathcal{W}^{(0)}$  contains only  $x_i$ . Here  $\hat{\sigma}_i^2$  is the noise variance robustly estimated at each point  $x_i$  from data as it has been explained before.
- **Estimation:** For all  $x_j$  in  $\mathcal{W}_i^{(k)}$ , we compute weights  $w_{ij}^{(k)}$  using Equation (4) and new estimates  $\hat{u}_i^{(k)}$  and  $\hat{v}_i^{(k)}$  using Equation (3).
- **Pointwise control:** After the estimate  $\hat{u}_i^{(k)}$  has been computed, we compare it to the previous estimates  $\hat{u}_i^{(k')}$  at the same point  $x_i$  for all  $k' < k$ . If the pointwise rule (6) is violated at iteration  $k$ , we do not accept  $\hat{u}_i^{(k)}$  and keep the estimates  $\hat{u}_i^{(k-1)}$  from the preceding iteration as the final estimate at  $x_i$  (i.e.  $\hat{k}_i = k - 1$  at point  $x_i$ ). The estimate at  $x_i$  is unchanged if  $k > \hat{k}_i$ .

- **Convergence:** Stop the procedure if  $k = k_M$  or if  $I(\hat{u}^{(k)}, \hat{u}^{(k+1)}) < 10^{-3}$ , otherwise increase  $k$  by 1 and continue with the estimation step. We use the parameter  $k_M$  to bound the numerical complexity. As expected, increasing  $k_M$  allows for additional variance reduction in homogeneous regions but usually does not change the estimates in the neighborhood of discontinuities. In our experiments,  $k_M = 15$  satisfies a good compromise and over-estimates the number of necessary iterations.

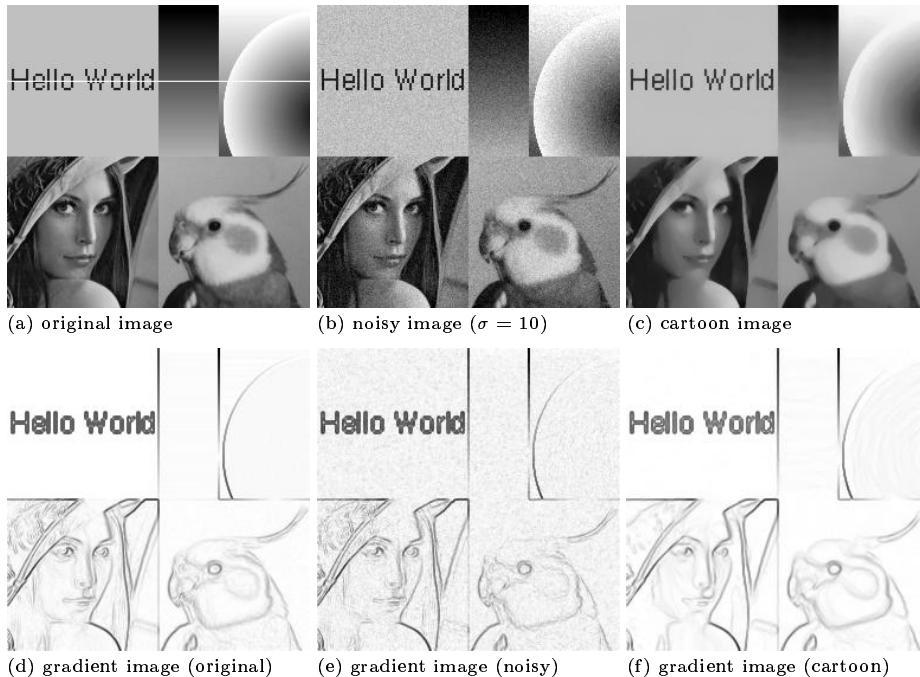
Using this algorithm, it can be shown that the average gray level of the smoothed image is not affected by the adaptive window procedure, i.e.  $\frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \hat{u}_i^{(\hat{k}_i)}$  and the *extremum principle* is guaranteed:  $\min_j Y_j \leq \hat{u}^{(k)} \leq \max_j Y_j$ ,  $\forall x_i \in \mathcal{S}$ ,  $\forall k \leq k_M$ .

## 4 Image decomposition

The proposed adaptive window approach is able to extract a piecewise smooth image, but significant textures and oscillating patterns are actually excluded. The purpose of this section is to briefly show that a relatively simple modification of the above global procedure yields an algorithm able to separate images into noise, texture and piecewise smooth (or *cartoon*) parts. In general the image-signal is assumed to be corrupted by an additive zero-mean white Gaussian noise with a constant variance. Therefore, the key idea is first to remove noise from the input image by setting  $\hat{\sigma}_i = \hat{\sigma}$  in the estimation procedure described in Section 3.1. If the original image consists of three additive components (noise, texture, *cartoon*) [16, 19, 11], the texture component is simply obtained by computing the difference between the noise-free image and the piecewise smooth image. The piecewise smooth image is estimated as described earlier in the paper, i.e by considering local estimations of the noise-texture variances  $\hat{\sigma}_i^2$  in the procedure. In areas containing little texture this difference is close to zero since  $\hat{\sigma}_{\mathcal{B}_i}^2$  is likely to be less than  $\hat{\sigma}^2$  in these areas. While the experimental results are promising using this simple mechanism, we do not pretend that it is able to decompose any images into the three main components under concern ([16, 19, 11]).

## 5 Experiments

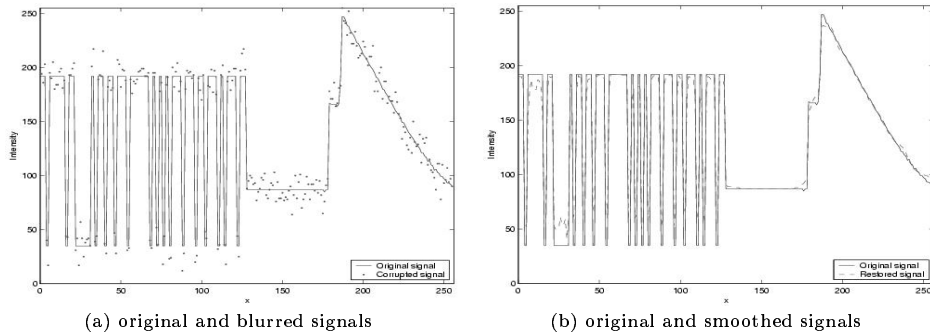
The potential of the adaptive window method is first shown on a synthetic-natural image (Fig. 1a) corrupted by an additive white-Gaussian noise (Fig. 1b, PSNR = 28.3 db,  $\sigma = 10$ ). In Fig. 1c, the noise and small textures are reduced in a natural manner and significant geometric features such as corners and boundaries, and original contrasts are visually well preserved (PSNR = 31 db). In this experiment, the final iteration  $k_c = 11$  was determined autonomously by the I-divergence criterion ( $k_M = 15$ ). To enhance the visibility of the results, we have computed the gradient norm on each of the three images; in Figs. 1d-f, high gradient values are coded with black and zero gradient with white. To better appreciate the smoothing process, a horizontal cross-section marked in white



**Fig. 1.** An example of adaptive smoothing results.

in Fig. 1a is graphically depicted in Fig. 2. The abrupt changes are correctly located and satisfying smooth variations of the signal are recovered (Fig. 2b). The processing of a  $256 \times 256$  image required typically 15 seconds on a PC (2.6 Ghz, Pentium IV) using a standard C++ implementation of the algorithm.

In a second example, the effects of the adaptive window approach are illustrated using the well-known  $512 \times 512$  *Barbara* image (Fig. 3). It turns out that the results of the adaptive window approach and TV minimization are visually similar when they are applied to the original image corrupted by an additive white-Gaussian noise (Fig. 3a,  $\sigma = 10$ ). In this experiment, the noise-free image shown in Fig. 3b, has been obtained by setting  $\hat{\sigma}_i = \hat{\sigma}$  in the adaptive window procedure. Additionally, our method is also able to reliably estimate the piecewise smooth component as shown in Fig. 3c and Fig. 3e. Small textures on clothes in the original image are nearly eliminated after  $k_c = 11$  iterations (automatically detected). In Fig. 3f, the TV minimizing process [21] does not completely eliminate small textures without blurring edges even if the balance between the fidelity and regularizing terms are modified. Finally, the estimated texture component shown in Fig. 3d correspond to the difference between the piecewise smooth image (Fig. 3b) and the noise-free image (Fig. 3a). In Fig. 4, the horizontal cross-section marked in white in Fig. 3a is depicted to better appreciate the image decomposition. The method provides also some additional structural information about the image. Figure 5a shows the results of local estimations of the noise-texture variance  $\hat{\sigma}_i^2$  within local windows ( $k_M = 15$ ). Dark



**Fig. 2.** A horizontal cross-section corresponding to line 128 drawn in white in Fig. 1a.

areas have higher values of  $\hat{\sigma}_i^2$  and correspond to more textured image regions; bright areas correspond to uniform regions, i.e.  $\hat{\sigma}_i^2 = \hat{\sigma}^2$ . Figure 5b shows the locations and sizes of optimal estimation windows; we have coded small windows with black and large windows with white. As expected, small windows are in the neighborhood of image gradients shown in Fig. 5e. The histogram of the windows sizes is also shown in Fig. 5f. Finally, Figures 5c and 5d show respectively the values  $\{|\mathcal{W}_i^{(\hat{k}_i)}|^{-1} \sum_{x_j \in \mathcal{W}_i^{(\hat{k}_i)}} g_{ij}^{(\hat{k}_i)}\}$  and  $\{\hat{\vartheta}_i^{(\hat{k}_i)}\}$  where dark values correspond to image discontinuities (Fig. 5c) and to low-confidence estimates (Fig. 5d).

We have also tested the algorithm (also implemented for processing 3D data) on a 3D confocal fluorescence microscopy image that contain complex structures. A typical 2D image taken from the 3D stack of 80 images depicts glomeruli in the antennal lobes of the moth olfactory system (Fig. 6a). The smoothed 3D image have larger homogeneous areas than the original 3D image (Fig. 6b). The corresponding noise-free texture image is shown in Fig. 6c. Here, this image decomposition is useful to extract regions/volumes of interest.

## 6 Summary and conclusions

We have described a novel feature-preserving adaptive smoothing algorithm where local statistics and variable window sizes are jointly used. Since  $|\mathcal{W}^{(k)}|$  grows exponentially in our set-up, the whole complexity of the proposed algorithm is of order  $O(n|\mathcal{W}^{(k_c)}|)$  if an image contains  $n$  pixels and  $k_c \leq k_M$ . In addition, the proposed smoothing scheme provides an alternative method to the anisotropic diffusion and bilateral filtering or energy minimization methods. An advantage of the method is that internal parameters can be easily calibrated using statistical arguments. Experimental results show demonstrate its potential for image decomposition into noise, texture and piecewise smooth components.

## References

1. D. Barash. A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 24(6): 844-847, 2002.

2. A. Ben Hamza, H. Krim. A variational approach to maximum a posteriori estimation for image denoising. In *Proc. EMMCVPR'01*, LNCS 2134, pp. 19-34, Sophia-Antipolis, France, 2001.
3. A. Blake, A. Zisserman. *Visual Reconstruction*, MIT Press, 1987.
4. M.J. Black, G. Sapiro. Edges as outliers: Anisotropic smoothing using local image statistics. In *Proc. Scale-Space'99*, LNCS 1682, pp. 259-270, Kerkyra, Greece, 1999.
5. F. Catte, P.-L. Lions, J.-M. Morel, T. Coll. Image selective smoothing and edge-detection by nonlinear diffusion. *SIAM J. Numerical Analysis*, 29(1): 182-193, 1992.
6. T. Chan, S. Osher, J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2): 231-241, 2001.
7. C.K. Chu, K. Glad, F. Godtlielsen, J.S. Marron. Edge-preserving smoothers for image processing. *J. Am. Stat. Ass.*, 93(442): 526-555., 1998.
8. D. Comaniciu, P. Meer. Mean-shift: a robust approach toward feature space analysis. *IEEE Trans. Patt. Anal. Mach. Intell.*, 24(5): 603-619, 2002.
9. I. Csizsár. Why least squares and maximum entropy ? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.*, 19: 2032-2066, 1991.
10. T. Gasser, L. Sroka, C. Jennen Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73: 625-633, 1986.
11. G. Gilboa, N. Sochen, Y.Y. Zeevi. Texture preserving variational denoising using an adaptive fidelity term. In *Proc. VLSM'03*, Nice, France, 2003.
12. F. Godtlielsen, E. Spjotvoll, J.S. Marron. A nonlinear Gaussian filter applied to images with discontinuities, *J. Nonparametric Statistics*, 8: 21-43, 1997.
13. A. Juditsky. Wavelet estimators: adapting to unknown smoothness. *Math. Methods of Statistics*, 1:1-20, 1997.
14. O. Lepski. Asymptotically minimax adaptive estimation 1: uppers bounds. *SIAM J. Theory of Prob. and Appl.*, 36(4): 654-659, 1991.
15. M. Maurizot, P. Bouthemy, B. Delyon, A. Juditski, J.-M. Odobez. Determination of singular points in 2D deformable flow fields. In *IEEE Int. Conf. Image Processing*, Washington DC, 1995.
16. Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations, *University Lecture Series*, 22, AMS 2002.
17. P. Mrazek. Selection of optimal stopping time for nonlinear diffusion filtering. *Int. J. Comp. Vis.*, 52(2/3): 189-203, 2003.
18. D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and variational problems. *Comm. Pure and Appl. Math.*, 42(5): 577-685, 1989.
19. S. Osher, A. Solé, L. Vese. Image decomposition and restoration using total variation minimization and the  $H^{-1}$  norm. *Multiscale Model. Simul.*, 1(3): 349-370, 2003.
20. P. Perona, J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12(7): 629-239, 1990.
21. L. Rudin, S. Osher, E. Fatemi. Nonlinear Total Variation based noise removal algorithms. *Physica D*, 60: 259-268, 1992.
22. C. Tomasi, R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Int Conf. Comp. Vis. (ICCV'98)*, pp. 839-846, Bombay, India, 1998.
23. J. van de Weijer, R. van den Boomgaard. Local mode filtering. In *Proc. Comp. Vis. Patt. Recogn. (CVPR'01)*, vol. II, pp. 428-433, Kauai, Hawaii, USA, 2001.

## Acknowledgments

We thank Sileye Ba for his contribution to this work.



(a) noisy image ( $\sigma = 10$ )



(b) noise removal using our method



(c) cartoon-like image using our method



(d) noise-free texture component

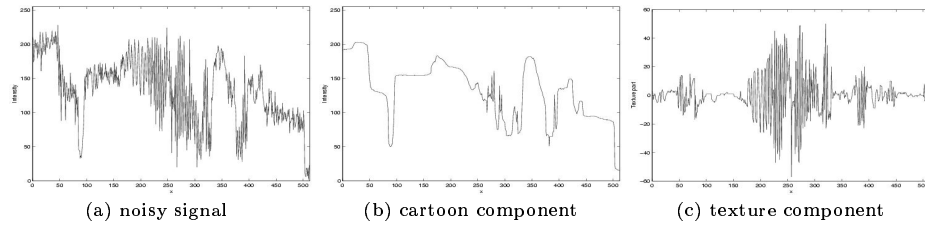


(e) part of the cartoon-like image using our method

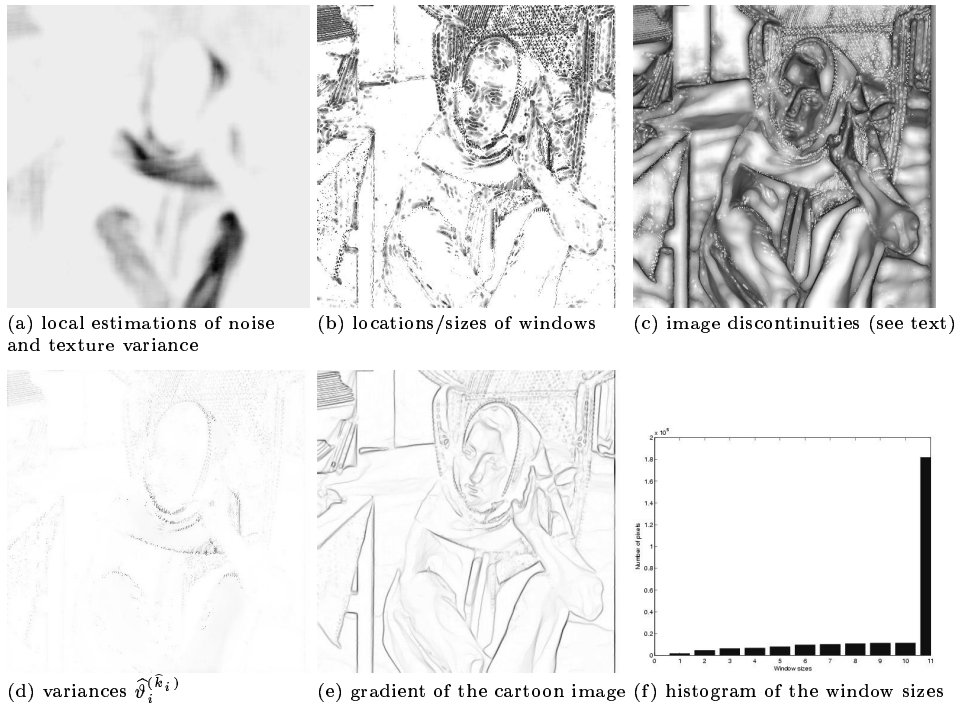


(f) part of the cartoon-like image by TV minimization [21]

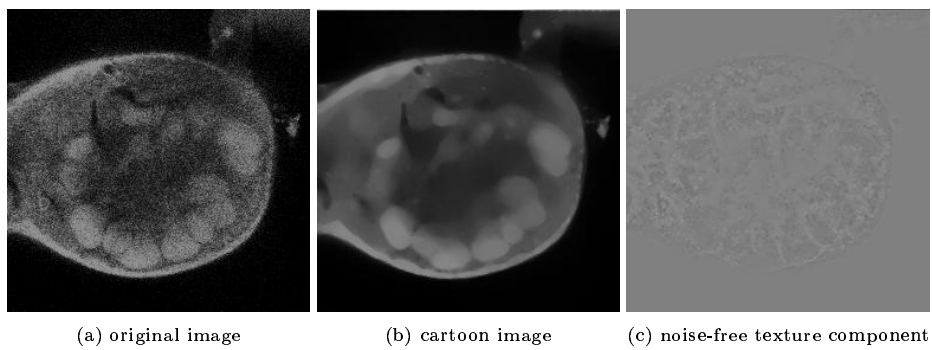
**Fig. 3.** Processing of the noisy  $512 \times 512$  *Barbara* image.



**Fig. 4.** A horizontal cross-section corresponding to line 256 drawn in Fig. 3a.



**Fig. 5.** Results of the cartoon-like  $512 \times 512$  *Barbara* image.



**Fig. 6.** Decomposition of a 3D confocal microscopy image.