Edge-Preserving Image Denoising and Estimation of Discontinuous Surfaces

Irène Gijbels, Alexandre Lambert and Peihua Qiu

Abstract

In this paper, we are interested in the problem of estimating a discontinuous surface from noisy data. A novel procedure for this problem is proposed based on local linear kernel smoothing, in which local neighbourhoods are adapted to the local smoothness of the surface measured by the observed data. The procedure can therefore remove noise correctly in continuity regions of the surface, and preserve discontinuities at the same time. Since an image can be regarded as a surface of the image intensity function and such a surface has discontinuities at the outlines of objects, this procedure can be applied directly to image denoising. Numerical studies show that it works well in applications, compared to some existing procedures.

Index Terms

Corners, edges, jump-preserving estimation, local linear fit, noise, nonparametric regression, smoothing, surface fitting, weighted residual mean square.

I. Introduction

This work was supported in part by the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy), a grant from the National Science Foundation of Belgium, a grant from National Security Agency of USA, and a grant from National Science Foundation of USA. I. Gijbels is with Department of Mathematics, University Center of Statistics, University of Leuven; email: irene.gijbels@wis.kuleuven.be. A. Lambert is with Institut de Statistique, Université catholique de Louvain; e-mail: alambert@stat.ucl.ac.be. P. Qiu is with School of Statistics, University of Minnesota; e-mail: qiu@stat.umn.edu.

THIS paper presents a novel procedure for estimating discontinuous surfaces from noisy data. The procedure is constructed in the framework of statistical jump regression analysis (JRA), which is a research area handling regression models involving jumps and discontinuities [1]. Since an image can be regarded as a surface of the image intensity function, this work can be used directly for image restoration, especially for image denoising. Other possible applications include early vision [2] and computer aided design [3].

In the literature, there are many existing procedures for image restoration. Early work involves mainly Fourier and inverse Fourier transformations and other algebraic manipulations (see e.g., Chapter 5 of [4]). Image restoration can also be formulated as a Bayesian estimation problem with the Markov random field (MRF) modeling and the maximum *a posteriori* (MAP) estimation (e.g., [5], [6]). In these Bayesian image restoration procedures, the equivalence between Gibbs distributions and MRFs plays an important role [7], [8]. Geman and Geman [5] suggested using a stochastic relaxation algorithm and an annealing schedule for computing the MAP estimator of the true image. To simplify its computation, Besag [6] suggested the iterated conditional modes (ICM) algorithm. Numerous other generalizations and modifications exist in the literature, see [9]-[14].

Image denoising by robust estimation, adaptive smoothing, and bilateral filtering attracts much attention in the literature. Median filtering was first suggested in the statistical literature [15], and it has become a popular pre-smoothing tool in image processing, because it has a certain ability of preserving edges while removing noise [16], [17], [18], [19]-[21]. To avoid pixels located on two different sides of an edge segment from being averaged in local smoothing, Saint-Marc *et al.* [22] suggested an adaptive smoothing filter which can adapt to the edge structure of the image. This filter was further generalized to the bilateral filtering procedure by Tomasi and Manduchi [23]. Motivated by the relationship between a Gaussian operator and a linear diffusion equation [24], [25], Perona and Malik [26] suggested restoring images by nonlinear diffusion filtering. Barash [27] pointed out that both adaptive smoothing and bilateral filtering can be regarded as special cases of nonlinear diffusion filtering.

Image denoising based on wavelet transformation and thresholding is also an active research topic recently, after it is shown that the wavelet method is useful for recovering regression curves and surfaces from noisy data with jumps and some other local features preserved [28]. There are several different versions of wavelet transformation and thresholding schemes (e.g., [29]-[32]). Software packages are also available for general applications [33].

Another popular approach is to consider image restoration as an inverse problem solved by edge preserving regularization methods, i.e. by minimizing an objective function that enforces a roughness penalty, in addition to a term measuring fidelity of the estimator to the data (e.g. [34], [35], [36]).

Discontinuity-preserving surface estimation can be regarded as a more general problem than image denoising, in the sense that image pixels are usually regularly spaced in rows and columns but data points in the former problem could be irregularly spaced. A two-stage algorithm for discontinuity-preserving surface estimation was suggested by Sinha and Schunck [37]. Its first stage is to clean and grid the data, and its second stage is to estimate the surface with a discontinuity-preserving spline smoothing procedure. Discontinuity-preserving and viewpoint invariant surface estimation procedures have been proposed by several authors (e.g., [38], [39]).

Surface estimation from noisy data can be regarded as a special case of 2-D statistical regression problem (see Section II-A). There are several procedures in the statistical literature for estimating surfaces with possible jumps preserved. For example, Qiu [40] proposed a three-stage procedure for this purpose, in which jumps are preserved by fitting local principal component lines. Chu *et al.* [41] studied the sigma filter and M-smoother. Both of them are based on robust estimation and closely related to adaptive smoothing and bilateral filtering mentioned above. Polzehl and Spokoiny [42] proposed the adaptive weights smoothing algorithm, which is an iterative procedure in which the size of a neighbourhood is adaptive to the surface smoothness. Recently, Qiu [43] suggested estimating jump surfaces by local piecewise linear kernel smoothing.

The proposed discontinuity-preserving surface estimation procedure is based on local linear kernel smoothing. Its basic idea is that local neighbourhoods used in local smoothing should be adaptive to local features of the underlying surface, in the way that full neighbourhoods are used in continuity regions of the surface and half or even smaller neighbourhoods are used around edges. Of course, the local features of the underlying surface are unobservable. To overcome this difficulty, we propose a data-driven mechanism for making decisions whether a given point is close to edges. Comparing with the existing procedures mentioned above, the proposed procedure has the following major features. First, its surface estimator has an explicit formula, which has at least two benefits compared to some iterative procedures (e.g., [5], [6], [26], [41], [42]). One is that it is easy to compute, and the other is that it helps when studying its theoretical properties. Second, besides a bandwidth and a threshold parameter, this procedure has no other parameters to choose. Therefore, it is simple to use. Third, some existing procedures (e.g., [26], [23], [41]) have some ability for preserving edges, but the edges cannot be preserved completely due to the fact that pixels located on the other side, other than the side located by the given pixel, of the related edge segment still receive *some* weights in local smoothing, although such weights are usually small. This phenomenon is partially eliminated in the current procedure because half neighbourhoods are used around edges.

The rest of the article is organized as follows. Our surface estimation procedure is described in details in Section II. Some numerical comparisons with several existing procedures are presented in Section III. A related problem to preserve corners of edges in surface estimation is discussed in Section IV, and the surface estimation procedure is modified accordingly to handle this problem using more flexible neighbourhoods. In Section V, we illustrate the method using various datasets. The discontinuity-preserving surface estimation problem when the design points are random is discussed in Section VI. Finally, some concluding remarks are given in Section VII.

For convenience of the reader we posted the digital images as well as a more extended technical report version of the paper, at the website http://www.stat.ucl.ac.be/∼alambert.

II. Methodology

In this section, we first describe a statistical model for specifying the discontinuous surface estimation problem, and introduce a popular surface estimation procedure based on local linear kernel smoothing, which is appropriate for estimating continuous regression surfaces (Section II.A). Then our jump-preserving surface estimation procedure is introduced in some detail (Section II.B). In the proposed procedure, there are two parameters (i.e., a bandwidth and a threshold). A data-driven mechanism is discussed in the last part (Section II.C) for selecting their values properly in applications.

A. Surface estimation

Suppose that a 2-D regression model for discontinuous surface estimation is:

$$Z_i = m(X_i, Y_i) + \varepsilon_i,\ i = 1, \ldots, n. \tag{1}$$

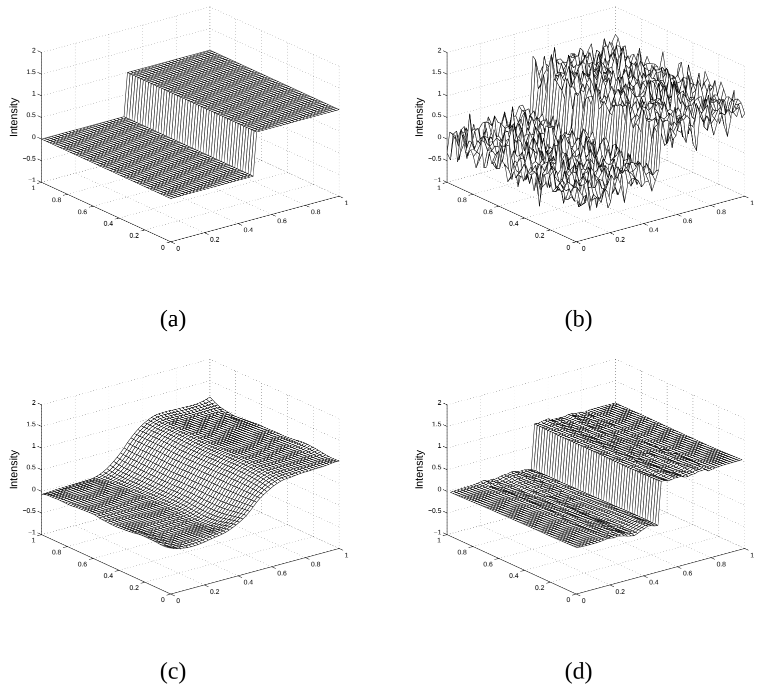
In (1), m is the true surface continuous in the design space $[0,1] \times [0,1]$ except along some discontinuity curves (i.e., edges); ε_i's are i.i.d. random errors with zero mean and finite variance σ^2; (X_i, Y_i)'s are design points, which are either i.i.d. r.v.'s from a bivariate distribution $f_{X,Y}(x, y)$ with support $[0,1] \times [0,1]$ (i.e., random design case), or, equispaced grid points on the same support (i.e., fixed design case). The 2-D model (1) with equispaced fixed design is appropriate for describing an image with $n = n_1 \times n_2$ pixels.

In the statistical literature, a popular surface estimation procedure is based on the following local linear kernel smoothing:

$$\begin{aligned}&(\widehat{a}_c(x,y), \widehat{a}_{c,x}(x,y), \widehat{a}_{c,y}(x,y))\\ =&\arg\min_{a,b,c} \sum_{i=1}^{n} (Z_i - a - b(X_i - x) - c(Y_i - y))^2 \cdot K_B((X_i - x), (Y_i - y)),\end{aligned} \tag{2}$$

where $K_B(x,y) = \frac{1}{|B|} K(B^{-1} \cdot (x,y)^t)$; B is a 2×2 global bandwidth matrix with determinant $|B|$, and $K(x,y)$ is a radially symmetric bivariate kernel function with compact support $\{(x,y) : x^2 + y^2 \leq 1\}$. In practice, we often use $K(x,y) = ((\exp(-(x^2 + y^2)/2) - \exp(-0.5))/(2\pi - 3\pi \exp(-0.5)))$ on $\{(x,y) : x^2 + y^2 \leq 1\}$ and 0 elsewhere, which is the truncated 2-D Gaussian density function. Using a 1^{st} order Taylor expansion of $m(X_i, Y_i)$ around the point (x,y): $m(X_i, Y_i) = m(x,y) + \frac{\partial m}{\partial x}(x,y)(X_i - x) + \frac{\partial m}{\partial y}(x,y)(Y_i - y) + \cdots$, one can show that the components of the triplet $(\widehat{a}_c(x,y), \widehat{a}_{c,x}(x,y), \widehat{a}_{c,y}(x,y))$ estimate, respectively, $m(x,y)$, $\frac{\partial m}{\partial x}(x,y)$, and $\frac{\partial m}{\partial y}(x,y)$. This means that we can obtain at the same time estimations of the surface $m(x,y)$ and its gradient $\nabla m(x,y) = (\frac{\partial m}{\partial x}, \frac{\partial m}{\partial y})(x,y)$. They are the so-called *local linear kernel estimators*. These estimators have explicit formulas. It has been demonstrated in the literature that they have some better properties in estimating continuous surfaces than some other local smoothing estimators, such as the local constant kernel estimator, especial near boundaries. See Fan and Gijbels [44] for more discussion.

If the surface m is discontinuous at a point (x,y), then the estimator $\widehat{a}_c(x,y)$ is not statistically consistent at this discontinuity point, since observations on both sides of the discontinuity are averaged in constructing $\widehat{a}_c(x,y)$. As a consequence, jumps around (x,y) are blurred. An illustration of this fact is depicted in Fig 1. Fig. 1(a) is the true *step* surface, 1(b) is a noisy version, and 1(c) is the local linear kernel estimator where the edge is clearly blurred. Therefore, the conventional local linear kernel smoothing procedure should be modified properly for estimating discontinuous surfaces with jumps preserved.



(a) (b) (c) (d)

Fig. 1. (a) True Step surface. (b) Noisy version of (a). (c) Local linear kernel surface estimator. (d) Proposed estimator

B. Edge-preserving surface estimation

An edge can be defined as a curve in the (X, Y) plane, along which the surface is discontinuous. Of course, the conventional estimator $\widehat{a}_c(x, y)$ is biased for estimating $m(x, y)$, if there is an edge in the neighbourhood of (x, y). Next, we present a solution to overcome this limitation.

By its definition, the gradient $(\frac{\partial m}{\partial x}, \frac{\partial m}{\partial y})$ indicates the direction of the maximal increase in m around (x, y). If the point (x, y) is on an edge segment, then the gradient direction would be asymptotically perpendicular to the tangent direction of the edge segment. The support of the kernel function K is then divided into two parts by a line passing the point (x, y) and perpendicular to the gradient direction $(\frac{\partial m}{\partial x}, \frac{\partial m}{\partial y})$. See Fig. 2 for a demonstration.

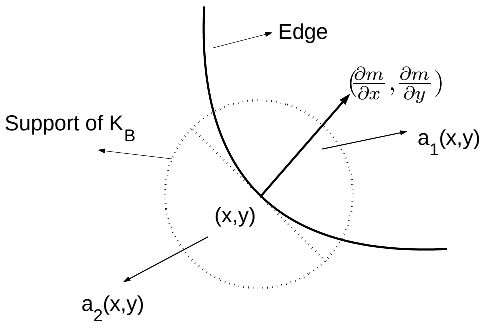

Edge
$(\frac{\partial m}{\partial x}, \frac{\partial m}{\partial y})$
Support of K$_B$
a$_1$(x,y)
(x,y)
a$_2$(x,y)

Fig. 2. Decomposition of the support of the kernel function along a direction perpendicular to the gradient direction of m at (x, y).

In the two parts, we define two *one-sided* local linear kernel estimators as follows:

$$\begin{aligned}&(\widehat{a}_j(x,y), \widehat{a}_{j,x}(x,y), \widehat{a}_{j,y}(x,y))\\ =&\arg\min_{a,b,c} \sum_{i=1}^{n} (Z_i - a - b(X_i - x) - c(Y_i - y))^2 \cdot K_B^{(j)}((X_i - x), (Y_i - y)),\end{aligned} \tag{3}$$

for $j = 1, 2$. In (3), $K_B^{(1)}$ and $K_B^{(2)}$ are the same as K_B in (2), except that their supports have been restricted to the two half-circles, as demonstrated by Fig. 2. Then, $\widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$ provide two one-sided estimators of $m(x, y)$.

By now, we have obtained three estimators for $m(x, y)$: the conventional estimator $\widehat{a}_c(x, y)$ and two one-sided estimators $\widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$. If there are no edge pixels in the neighbourhood of (x, y), then $\widehat{a}_c(x, y)$ should be selected for estimating $m(x, y)$, because it averages more observations around the point (x, y) and thus is more powerful in removing noise. If there is an edge segment around (x, y), then the conventional estimator $\widehat{a}_c(x, y)$ is not a good estimator of $m(x, y)$ any more, as explained at the end of Section II.A. In such a case, however, one of the two one-sided estimators $\widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$ should still estimate the surface well, because most observations used by this estimator are located on a single side of the edge segment, guaranteed by the statistical properties of the estimated gradient direction from (2).

In practice, the edge locations are usually unknown; so, we need to choose among the three estimators $\widehat{a}_c$, $\widehat{a}_1$ and $\widehat{a}_2$ in a data-driven way, which is discussed below.

The quality of the three estimators $\widehat{a}_c$, $\widehat{a}_1$ and $\widehat{a}_2$ can be measured by the Weighted Residual Mean Squares (WRMS) of the related fitted surfaces, defined by:

$$\begin{aligned}&\text{WRMS}_c(x,y)\\ =&\frac{1}{\sum_i K_B(i)}\sum_i \left[Z_i-\widehat{a}_c(x,y)-\widehat{a}_{c,x}(x,y)(X_i-x)-\widehat{a}_{c,y}(x,y)(Y_i-x)\right]^2 K_B(i),\end{aligned}$$

$$\begin{aligned}&\text{WRMS}_j(x,y)\\ =&\frac{1}{\sum_i K_B^{(j)}(i)}\sum_i \left[Z_i-\widehat{a}_j(x,y)-\widehat{a}_{j,x}(x,y)(X_i-x)-\widehat{a}_{j,y}(x,y)(Y_i-x)\right]^2 K_B^{(j)}(i),\end{aligned} \quad (4)$$

where $K_B(i)$ denotes $K_B((X_i-x),(Y_i-y))$, and $K_B^{(j)}(i)$ denotes $K_B^{(j)}((X_i-x),(Y_i-y))$, for $j=1,2$.

The behaviour of these quantities depends on whether there are edge pixels in the neighbourhood of the point (x,y). If there are no edge pixels in the neighbourhood, then all WRMS's are good estimators of the noise variance σ^2. Otherwise, those WRMS's who use data points on both sides of edge segments would be biased for estimating σ^2, and the bias would depend on the jump size and the Euclidean distance between the point (x,y) and the edge segments (see Qiu [43]).

Based on these results, our edge-preserving surface estimator is defined by:

$$\widehat{m}(x,y)=\begin{cases}\widehat{a}_c(x,y) & \text{if diff}(x,y)\le u\\ \widehat{a}_1(x,y) & \text{if diff}(x,y)>u \text{ and WRMS}_1(x,y)<\text{WRMS}_2(x,y)\\ \widehat{a}_2(x,y) & \text{if diff}(x,y)>u \text{ and WRMS}_1(x,y)>\text{WRMS}_2(x,y)\\ \frac{\widehat{a}_1(x,y)+\widehat{a}_2(x,y)}{2} & \text{if diff}(x,y)>u \text{ and WRMS}_1(x,y)=\text{WRMS}_2(x,y),\end{cases} \quad (5)$$

where u is a threshold value and

$$\text{diff}(x,y)=\max\{\text{WRMS}_c(x,y)-\text{WRMS}_1(x,y),\text{WRMS}_c(x,y)-\text{WRMS}_2(x,y)\}. \quad (6)$$

So our surface estimator $\widehat{m}(x, y)$ is defined by one of the three estimators: $\widehat{a}_c(x, y), \widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$, depending on whether there are edge pixels around (x, y), judged by the WRMS values. If we are in a continuity region of the surface, then all three WRMS's are close to σ^2, so that diff(x, y) is close to zero. On the other hand, if we are close to an edge segment, then one of the two one-sided WRMS's would be smaller than WRMS$_c$, and thus diff(x, y) would be relatively large. Therefore, diff(x, y) can be used to judge whether there are edge pixels around (x, y). See Gijbels et al. [45] for related discussion. In (5), the case WRMS$_1(x, y) =$ WRMS$_2(x, y)$ has, for n tending to infinity, probability zero to occur under some regularity conditions. It is included just for completeness. The explicit formulation of the estimator (5) is helpful when investigating theoretical properties of the estimator. For a one-dimensional version of the estimator (when recovering discontinuous curves) this can be seen from Gijbels et al. [45].

C. Selection of the bandwidth B and the threshold u

Different choices of u would lead to different types of surface estimation. For example, if $u = 0$, then the estimator $\widehat{m}(x, y)$ privileges edge preservation since it only uses $\widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$ in such a case, whereas if $u = \max_{(x,y)\in[0,1]\times[0,1]}$ diff(x, y), then $\widehat{m}(x, y)$ becomes the conventional estimator $\widehat{a}_c(x, y)$ which is ideal for removing noise. Any choice of u between these two extreme values would lead to an estimator having both edge preserving and noise removing properties. The trade-off between these two properties depends on the value of u. Theoretically, the best choice of u is the one minimizing the Mean Integrated Squared Error (MISE) of the surface estimator, defined by:

$$\text{MISE} = \text{E}\left\{\int_0^1 \int_0^1 \left(\widehat{m}(x, y) - m(x, y)\right)^2 f_{X,Y}\; dxdy\right\} \qquad (7)$$

This definition can be used for both the fixed design and random design cases. In the case of fixed design with equispaced design points, we can simply set $f_{X,Y} = 1$. Obviously, the

MISE value defined in (7) depends on the true surface m which is unknown. So a data-driven criterion is still required to choose u properly.

Similarly, the bandwidth matrix B used in (2) and (3) also needs to be chosen properly. For simplicity, we assume that B is a diagonal matrix with the same diagonal elements, which implies that we impose the same amount of smoothing in the x and y directions. This is a common practice in the image processing literature (e.g., [4]). Then, choosing B is equivalent to choosing one of its diagonal elements, denoted by h.

In this article, we suggest choosing both h and u by the following cross-validation (CV) procedure:

$$(h_n, u_n) = \arg\min_{h,u} \sum_{i=1}^{n} \left(Z_i - \widehat{m}^{(-i)}(X_i, Y_i) \right)^2, \tag{8}$$

where $\widehat{m}^{(-i)}(X_i, Y_i)$ is the estimator of $m(X_i, Y_i)$ using all the data points except the point (X_i, Y_i, Z_i).

One might think that choosing both parameters h and u by the CV procedure (8) would require a big computational demand, but this is not true. In view of (5), once we have the three estimators ($\widehat{a}_c,\ \widehat{a}_1$ and $\widehat{a}_2$) (for fixed h), we do not have to re-compute these estimators for different values of u, which saves a great amount of computation.

In Fig. 1 (d), we show the proposed estimator of the *step* surface. Here, the edge is preserved quite well compared to results shown in Fig 1 (c).

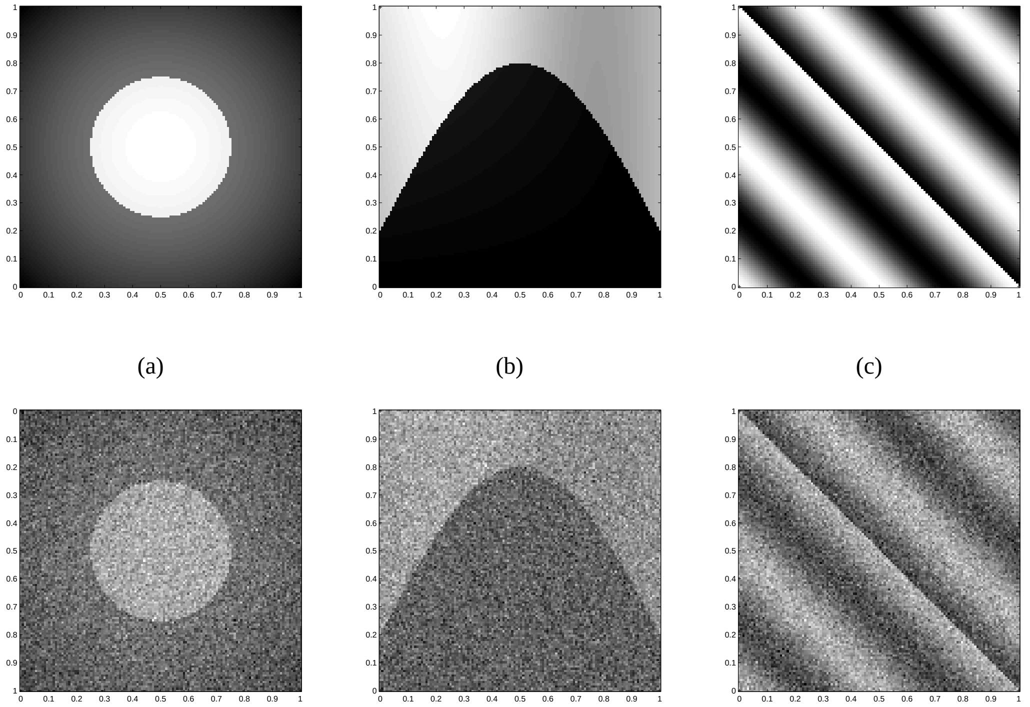
III. Comparison with other methods and heteroscedastic case

In this section, we compare via simulations our method with four existing ones which are mainly used in the image denoising literature: wavelets, Markov random field (MRF), median filter, and bilateral filter [23]. For the wavelet method, we use the R package wavethresh by Nason and Silverman [33]. The following parameters should be chosen properly before this method can be used: the wavelet family, the threshold method (hard or soft), the levels of wavelet coefficients to be thresholded, and the number of vanishing moments of the wavelet family. We also compare with the wavelet decomposition method from Portilla *et al.* [32] using the BLS-GSM Image Denoising Matlab Toolbox 1.0.3. For the MRF method, the Fortran package developped by P. Qiu is used, which is based on the method from Godtliebsen and Sebastiani [9] and Geman and Geman [5] which has some similarities with regularization techniques. This procedure has three parameters to choose which describe the joint prior distribution of the true image and the line process (representing edges in an image). The median filter has only one parameter involved, which is the size of the neighbourhood to smooth. The bilateral filtering involves taking a weighted average of image intensities at nearby pixels. The procedure has two smoothing parameters: a scaling factor for each of the two weight functions (univariate and bivariate Gaussian density functions for the image intensities and the pixel locations respectively).

We consider three different test models. Model 1 consists of a smooth quadratic surface which has a circular edge with constant jump size (see Fig. 3 (a)). Model 2 is depicted in Fig. 3 (b); here, the jump size varies along an edge in a sinusoidal form. Model 3 is shown in Fig. 3 (c). In this case, continuity parts of the image consist of regions with large first-order derivatives. These three models can be described by the following functions: for $(x, y) \in [0, 1] \times [0, 1]$,

$$\begin{aligned} m_1(x, y) &= -2(x-0.5)^2 - 2(y-0.5)^2 + I_{[(x-0.5)^2+(x-0.5)^2<0.25^2]}, \\ m_2(x, y) &= 0.25(1-x)y + (1+0.2\sin(2\pi x))I_{[y>0.6\sin(\pi x)+0.2]}, \\ m_3(x, y) &= \cos(4\pi(1-x-y)) - 2\cos(4\pi(1-x-y))I_{[x+y-1>0]}, \end{aligned} \tag{9}$$

where $I_{[A]}$ equals to 1 if A is true, and 0 otherwise. For Model 1, the jump size is one along the circular edge. For Model 2, the jump size varies from 0.8 to 1.2. For Model 3, the jump size is 2 along the edge line. Figs. 3 (d), (e) and (f) show noisy versions of the three models when $n = 128^2$ and $\varepsilon \sim N(0; 0.5^2)$. Various surface estimation methods are evaluated in



(a) (b) (c)

(d) (e) (f)

Fig. 3. (a) Model 1; (b) Model 2; (c) Model 3; (d) Noisy version of (a); (e) Noisy version of (b); (f) Noisy version of (c). In (d), (e) and (f), $\varepsilon \sim N(0, 0.5^2)$.

two different sample sizes: $n = 128^2$ and $n = 256^2$, and two values of the error variance: $\sigma^2 = 0.2^2$ and $\sigma^2 = 0.5^2$, where it is assumed that $\varepsilon \sim N(0; \sigma^2)$. For each method, its procedure parameters are selected such that the approximated Mean Integrated Squared Error (MISE), computed based on $N = 100$ samples and denoted by $\widehat{\text{MISE}}$, reaches the minimum except for the BLS-GSM method where the default parameters (optimized) are used. For comparison purposes, both $\widehat{\text{MISE}}$ and $\widehat{\text{MISE}}_e$ are reported, where $\widehat{\text{MISE}}_e$ is the local $\widehat{\text{MISE}}$ value computed in a band, denoted as $B(m, h)$, around edges of the surface m with radius equal to the selected h-value. More specifically, $\widehat{\text{MISE}}_e$ is defined by

$$\widehat{\text{MISE}}_e = \frac{\sum_{i=1}^{n} (\widehat{m}(X_i, Y_i) - m(X_i, Y_i))^2 I[(X_i, Y_i) \in B(m, h)]}{\sum_{i=1}^{n} I[(X_i, Y_i) \in B(m, h)]} \tag{10}$$

Results for Models 1, 2, and 3 are shown in Tables I, II, and III, respectively.

TABLE I

$\widehat{\text{MISE}}$ (FIRST COLUMNS) AND $\widehat{\text{MISE}}_e$ (SECOND COLUMNS) VALUES FOR MODEL 1 WHEN $N = 100$ AND $n = 128^2$ OR 256^2.

	$n = 128^2$				$n = 256^2$			
Method	σ=0.2 $(h = 0.047)$		σ=0.5 $(h = 0.074)$		σ=0.2 $(h = 0.029)$		σ=0.5 $(h = 0.051)$	
$\widehat{m}$ (Proposed procedure)	0.0012	0.0044	0.0055	0.0172	0.0006	0.0033	0.0027	0.0121
Wavethresh	0.0058	0.0337	0.0125	0.0389	0.0032	0.0277	0.0071	0.037
BLS-GSM	0.0020	0.0097	0.0061	0.0179	0.0012	0.0090	0.0032	0.0140
MRF	0.0011	0.0029	0.0109	0.0327	0.0005	0.0017	0.0059	0.0216
Median Filter	0.0032	0.0165	0.0115	0.0364	0.0019	0.0153	0.0072	0.0306
Bilateral Filter	0.0023	0.0082	0.0104	0.0313	0.0016	0.0092	0.0067	0.0281

TABLE II

$\widehat{\text{MISE}}$ (FIRST COLUMNS) AND $\widehat{\text{MISE}}_e$ (SECOND COLUMNS) VALUES FOR MODEL 2 WHEN $N = 100$ AND $n = 128^2$ OR 256^2.

	$n = 128^2$				$n = 256^2$			
Method	σ=0.2 $(h = 0.055)$		σ=0.5 $(h = 0.090)$		σ=0.2 $(h = 0.035)$		σ=0.5 $(h = 0.059)$	
$\widehat{m}$ (Proposed procedure)	0.0010	0.0053	0.0052	0.0214	0.0005	0.0035	0.0024	0.0150
Wavethresh	0.0064	0.0475	0.0130	0.0504	0.0035	0.0379	0.0079	0.0529
BLS-GSM	0.0019	0.0135	0.0064	0.0260	0.0011	0.0114	0.0034	0.0206
MRF	0.0007	0.0033	0.0107	0.0401	0.0004	0.0026	0.0065	0.0323
Median Filter	0.0036	0.0218	0.0120	0.0497	0.0021	0.0196	0.0075	0.0433
Bilateral Filter	0.0024	0.0111	0.0107	0.0410	0.0017	0.0116	0.0069	0.0387

TABLE III

$\widehat{\text{MISE}}$ (FIRST COLUMNS) AND $\widehat{\text{MISE}}_e$ (SECOND COLUMNS) VALUES FOR MODEL 3 WHEN $N = 100$ AND $n = 128^2$ OR 256^2.

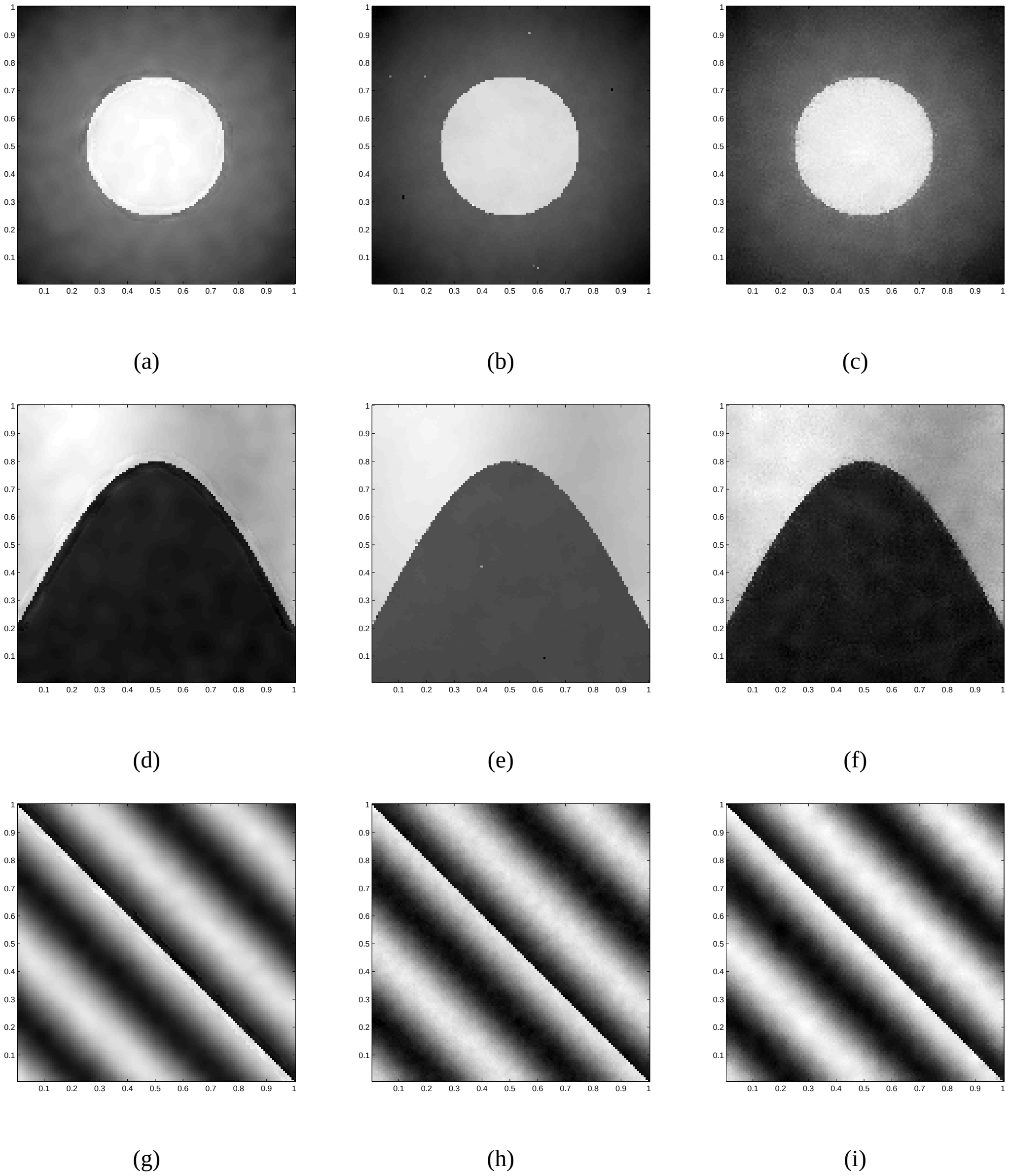
	$n = 128^2$				$n = 256^2$			
Method	σ=0.2 $(h = 0.039)$		σ=0.5 $(h = 0.051)$		σ=0.2 $(h = 0.029)$		σ=0.5 $(h = 0.039)$	
$\widehat{m}$ (Proposed procedure)	0.0018	0.0055	0.0071	0.0244	0.0007	0.0025	0.0026	0.0099
Wavethresh	0.0076	0.0579	0.0250	0.1696	0.0047	0.0396	0.0124	0.1004
BLS-GSM	0.0039	0.0125	0.0118	0.0481	0.0018	0.0088	0.0066	0.0326
MRF	0.0033	0.0040	0.0126	0.0326	0.0013	0.0015	0.0054	0.0160
Median Filter	0.0044	0.0185	0.0211	0.1045	0.0025	0.0250	0.0116	0.0848
Bilateral Filter	0.0025	0.0062	0.0184	0.0955	0.0012	0.0042	0.0113	0.0740

For the three models, we can see that the selected h increases when noise level σ increases and sample size n decreases, which is intuitively reasonable. We can also see that the proposed estimator behaves well in both $\widehat{\text{MISE}}$ and $\widehat{\text{MISE}}_e$ values in all cases. The MRF method performs very well when σ is small; whereas the proposed method outperforms the MRF method when σ is relatively large. For Models 1 and 2 the BLS-GSM wavelet method performs somewhat comparable, but (slightly) worse, than the proposed method. Fig. 4 shows the estimators by $\widehat{m}$, the MRF method and the bilateral filtering, corresponding to a median $\widehat{\text{MISE}}$ performance, for the three models when $n = 128^2$ and $\sigma = 0.2$.

Heteroscedastic error variance

In model (1), we assume the noise level σ to be constant on the entire design space. In some applications this assumption may not hold. Then a more appropriate model would be

$$Z_i = m(X_i, Y_i) + \sigma(X_i, Y_i)\varepsilon_i,\ \ i = 1, \ldots, n. \tag{11}$$



(a) (b) (c)

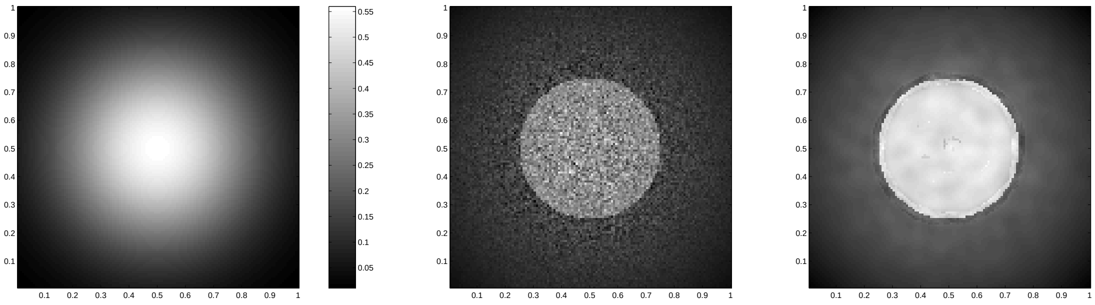
(d) (e) (f)

(g) (h) (i)

Fig. 4. Surface estimators corresponding to a median performance of the $\widehat{\text{MISE}}$ when $n = 128^2$ and $\sigma = 0.2$. Plots (a), (d), and (g): $\widehat{m}$; plots (b), (e), and (h): MRF; plots (c), (f), and (i): Bilateral filter.

where ε_i are n i.i.d. random variables with zero mean and unit variance and $\sigma(x, y)$ is a smooth function describing the noise level in the design space. In this heteroscedasticity setting the proposed procedure can still be used. For example, we simulate the Model 1 ($\varepsilon \sim N(0, 1)$) with a σ function with a bell shape such that more noise is present in the foreground of the image and less in the background (see Figs 5 (a) and (b)) which is more realistic in applications. In the example the function $\sigma(x, y)$ is taken such that $\int_0^1 \int_0^1 \sigma(x, y)\, dxdy = 0.2$.

The estimation, based on a simulation with sample size 128^2, is depicted in Fig. 5 (c). The $\widehat{\text{MISE}}$ is equal to 0.0027 and $\widehat{\text{MISE}}_e = 0.0148$ which seems to be reasonable compared to the corresponding results in the homoscedastic case (see Table I for $\sigma = 0.2$ and $\sigma = 0.5$).



(a) (b) (c)

Fig. 5. (a) Function $\sigma(x, y)$; (b) Noisy version of Model 1 and (c) Estimation by the proposed procedure.

IV. Corner preserving

A corner can be defined as an intersection point of two edge segments. Clearly, the discussed surface estimator cannot preserve a corner well because we divide the kernel support into two half circles and hence $\widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$ are biased at a corner.

Here, we suggest an improvement of the proposed procedure for corner preservation, which is based on the following measure of the cornerness originally proposed by Yang *et al.* [46] and Chabat *et al.* [47]:

$$c(x, y) = (1 - g(x, y))^2 ||(m_x, m_y)|| \tag{12}$$

where $m_x = \frac{\partial m}{\partial x}$, $m_y = \frac{\partial m}{\partial y}$, $||.||$ is the Euclidian norm, and

$$g(x, y) = \frac{(\iint_V (m_x^2 - m_y^2)\, dxdy)^2 + (\iint_V (2m_x m_y)\, dxdy)^2}{(\iint_V (m_x^2 + m_y^2)\, dxdy)^2}, \tag{13}$$

with V a neighbourhood around the point (x, y).

It can be shown that $c(x, y)$ is large in the neighbourhood of a corner and small elsewhere. In what follows, $c(x, y)$ is normalized by $c(x, y)/\max(c(x, y))$ such that it belongs to $[0, 1]$.

All quantities involved in the estimator of c, which is denoted as $\widehat{c}$, have been previously computed in (2). The neighbourhood V is taken to be a square centered at (x, y) and with length $2h$.

We now define a new estimator $\widetilde{m}$ of m based on $\widehat{c}$ and $\widehat{m}$ as follows:

$$\text{If } \widehat{c}(x,y) > C,\ \widetilde{m}(x,y) = \begin{cases} \widetilde{a}_1(x,y) & \text{if} \quad \text{WRMS}_{\tilde{1}}(x,y) < \text{WRMS}_{\tilde{2}}(x,y) \\ \widetilde{a}_2(x,y) & \text{if} \quad \text{WRMS}_{\tilde{1}}(x,y) > \text{WRMS}_{\tilde{2}}(x,y) \\ \frac{\widetilde{a}_1(x,y)+\widetilde{a}_2(x,y)}{2} & \text{if} \quad \text{WRMS}_{\tilde{1}}(x,y) = \text{WRMS}_{\tilde{2}}(x,y), \end{cases} \tag{14}$$

where $\widetilde{a}_1(x, y)$ and $\widetilde{a}_2(x, y)$ are modified versions of $\widehat{a}_1(x, y)$ and $\widehat{a}_2(x, y)$ defined below, $\text{WRMS}_{\tilde{1}}(x, y)$ and $\text{WRMS}_{\tilde{2}}(x, y)$ are the corresponding WRMS values, and $C \geq 0$ is a threshold. If $\widehat{c}(x, y) \leq C$, we let $\widetilde{m}(x, y) = \widehat{m}(x, y)$. From this definition, we can see that when C is chosen 0, $\widetilde{m}$ depends only on $\widetilde{a}_1(x, y)$ and $\widetilde{a}_2(x, y)$, which means that $\widetilde{m}$ privileges corner preservation in such a case. On the other hand, when C is chosen 1, $\widetilde{m} = \widehat{m}$, and consequently the procedure (14) can only preserve flat edges while removing noise.

The estimators $\widetilde{a}_1(x, y)$ and $\widetilde{a}_2(x, y)$ follow the same definitions as $\widehat{a}_1$ and $\widehat{a}_2$ (see equation (3)), except that the quantity $K_B((X_i - x), (Y_i - y))$ is replaced by $L_H((X_i - x), (Y_i - y))$, i.e., a new kernel function L and a new bandwidth matrix H are used here. The matrix H is defined by

$$H = \begin{pmatrix} \cos(\alpha), & -\sin(\alpha) \\ \sin(\alpha), & \cos(\alpha) \end{pmatrix} \begin{pmatrix} h/k_1, & 0 \\ 0, & hk_2 \end{pmatrix},$$

where $k_j \in [0, 1]$, for $j = 1, 2$, $\alpha = \tan^{-1}(\widehat{m}_y(x, y)/\widehat{m}_x(x, y))$, and $L_H((X_i - x), (Y_i - y)) = |H|^{-1}K(H^{-1}(X_i - x, Y_i - y)^t)2\cos^2(\beta_i)$. In the simulation examples below, we take $k_1 = 1$ and $k_2 = 0.5$. The angles β_i are chosen as shown in Fig. 6 (b). Using this setup, the support of $K(H^{-1}(X_i - x, Y_i - y)^t)$ becomes an ellipse centered at (x, y), with its longer axis of length h/k_1 parallel to the gradient direction, and its shorter axis of length hk_2 perpendicular to the gradient direction. If we are close to a corner point, then the direction of the gradient

$(\widehat{m}_x(x,y), \widehat{m}_y(x,y))$ approximates the direction of the bisectrix of the corner (see Fig 6). Thus, more data points are used in a corner region in this case, compared to the case when the quantity K_B is used as in equation (3) (cf. Fig. 6 (a)). The factor $2\cos^2(\beta_i)$ guarantees that more weights are given to observations near the bisectrix and less weights elsewhere.

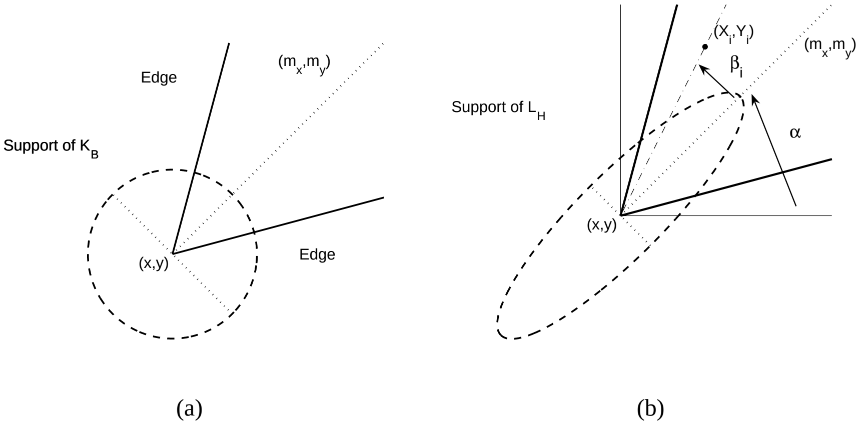


Fig. 6. (a) Support of K_B at a corner point; (b) Support of L_H at a corner point with a nearby point (X_i, Y_i).

For illustration, we apply this procedure to a noisy version of Fig. 7 (a), shown in Fig. 7 (b), in which $n = 100^2$ and $\varepsilon \sim N(0; 0.1^2)$. In the true image, there are three different corners. Fig. 7(c) shows the surface estimator without corner preserving (i.e., $C = 1$ in (14)). It can be seen that the three corners are blurred, as expected, and the blurring is more noticeable around a sharper corner. In Figs. 7 (d) and (e), we plot respectively the new estimator $\widetilde{m}$ with $C = 0.2$ and the corresponding standardized cornerness measure $\widehat{c}(x,y)$. It can be seen that the corners are preserved quite well. For more discussion on this corner preserving improvement of the method, see [48].

V. OTHER ILLUSTRATIONS

In this section, we illustrate the proposed method on several different greyscale images whose grey levels are in $[0, 255]$. Fig. 8 (a) shows a noisy version of an image of three circles, in which $n = 256^2$ and $\sigma = 75^2$. The estimator $\widetilde{m}$ with $C = 1$ and u and h chosen by CV is presented in Fig. 8 (b). We can see that the circular edges are preserved quite well

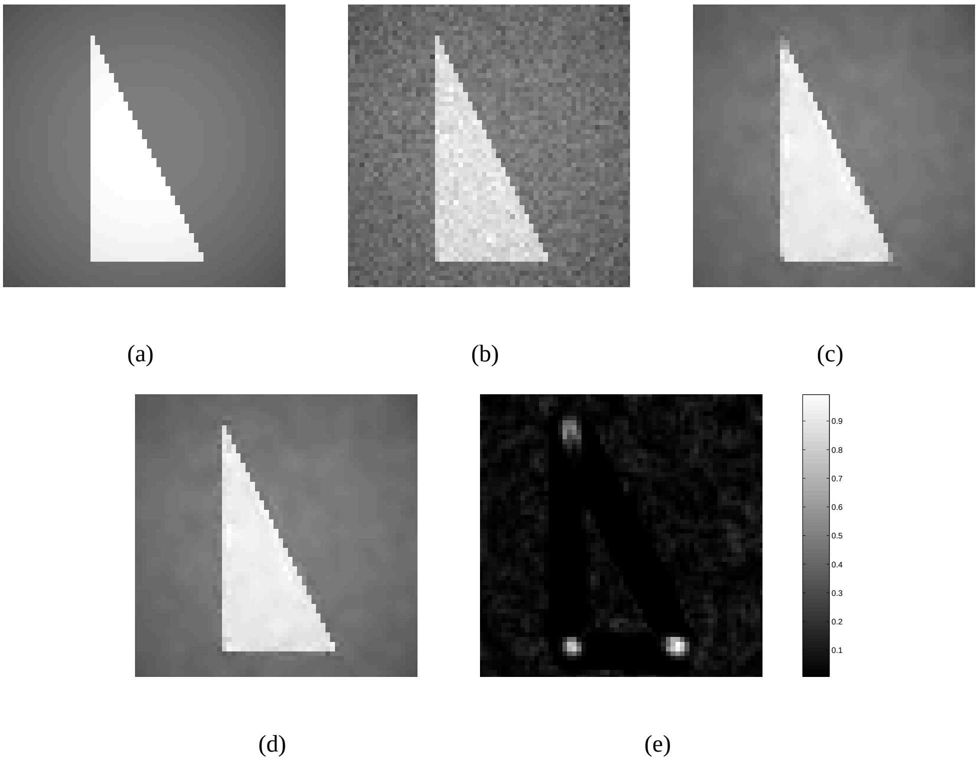


Fig. 7. (a) True image with three selected points (little squares); (b) Noisy version of (a) with $\varepsilon \sim N(0; 0.1^2)$; (c) Surface estimator without corner preserving; (d) Corner-preserving surface estimator; (e) Normalized, estimated cornerness measure $\widehat{c}(x, y)$;

in both the interior region and places near the boundaries.

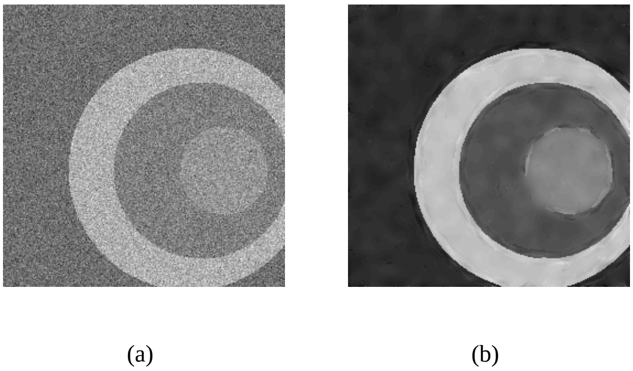


Fig. 8. (a) Noisy image; (b) Denoised image.

Figs. 9 (a) and (b) depict respectively a noisy image of a rod with $n = 256^2$ and $\sigma^2 = 25^2$ and the denoised image $\widetilde{m}$ with $C = 1$ and u and h chosen by CV. Again, edges are preserved well by $\widetilde{m}$, whereas noise is removed efficiently as well. Fig. 10 (a) shows a log-transformed C-band, HH-polarization, synthetic aperture radar (SAR) image recorded by Dr. E. Attema at

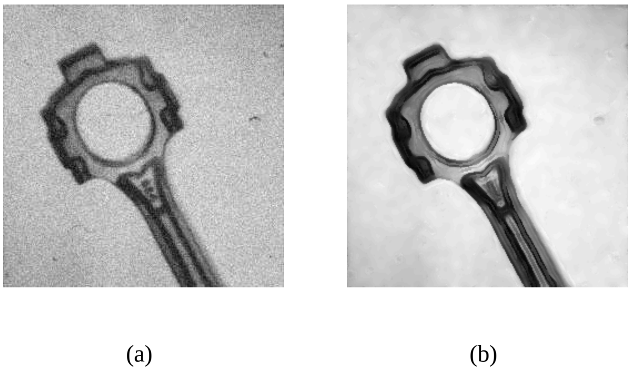


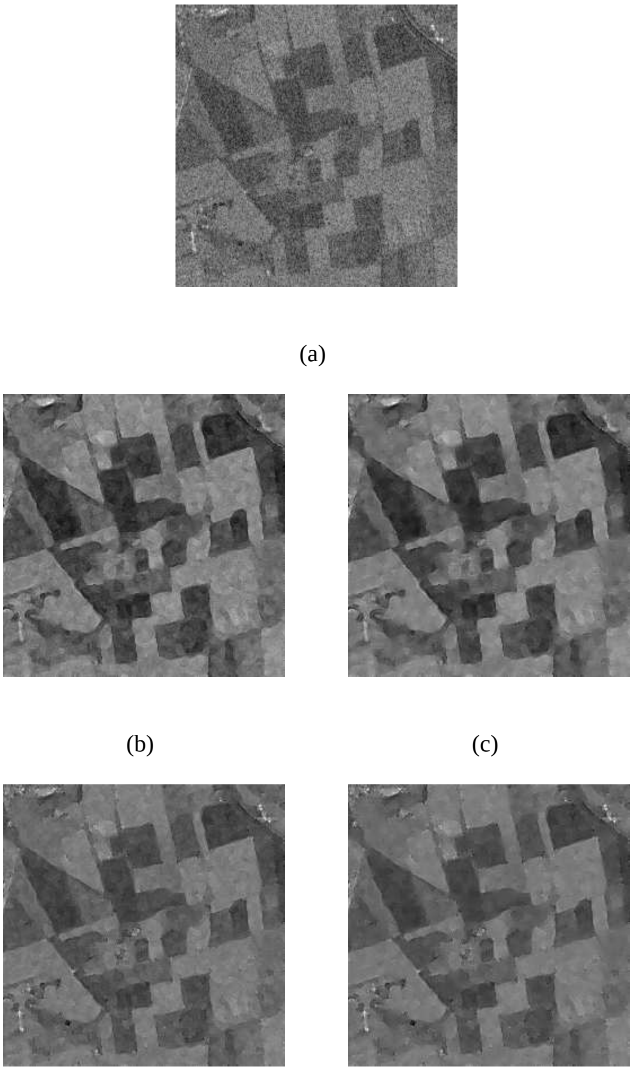
Fig. 9. (a) Noisy image; (b) Denoised image.

the European Space Research and Technology Centre in Noordwijk, The Netherlands. It can be seen that this noisy image has some corner points. Figs. 10 (b)-(e) present the estimator $\widetilde{m}$ in various cases: $C = 1$ (i.e., without corner preservation) and $u = 0$ in plot (b); $C = 1$ and $u = 51$ in plot (c); $C = 0.25$ and $u = 0$ in plot (d); and finally $C = 0.25$ and $u = 51$ in plot (e). The bandwidth h is chosen via CV in all cases. It can be seen that corners are better preserved when $C = 0.25$.

VI. Random Design

As mentioned in Section II-A, the proposed method can also be used in the case of non-equispaced design. This problem has been studied by Strohmer [49] and Arigovindan *et al.* [50], among some others, in applications of digital image processing.

For illustration, we randomly select 10% of the original pixels of Fig. 3 (a) (shown in Fig. 11 (a)) and then add noise from $N(0; 0.2^2)$ distribution to the selected pixels. The noisy version of Fig. 11 (a) is shown in Fig. 11 (b). In both images, black regions do not include any pixels. Our goal is to estimate the whole image from the randomly selected true pixels (i.e., from those in plot (a)) and from these pixels plus noise as well (i.e., from those in plot (b)). Figs. 11 (c) and (d) present the two sets of results, respectively. It can be seen that the proposed method works well in general. Fig. 11 (e) shows the $\widehat{\text{MISE}}$ values of the estimated



(a)

(b)

(c)

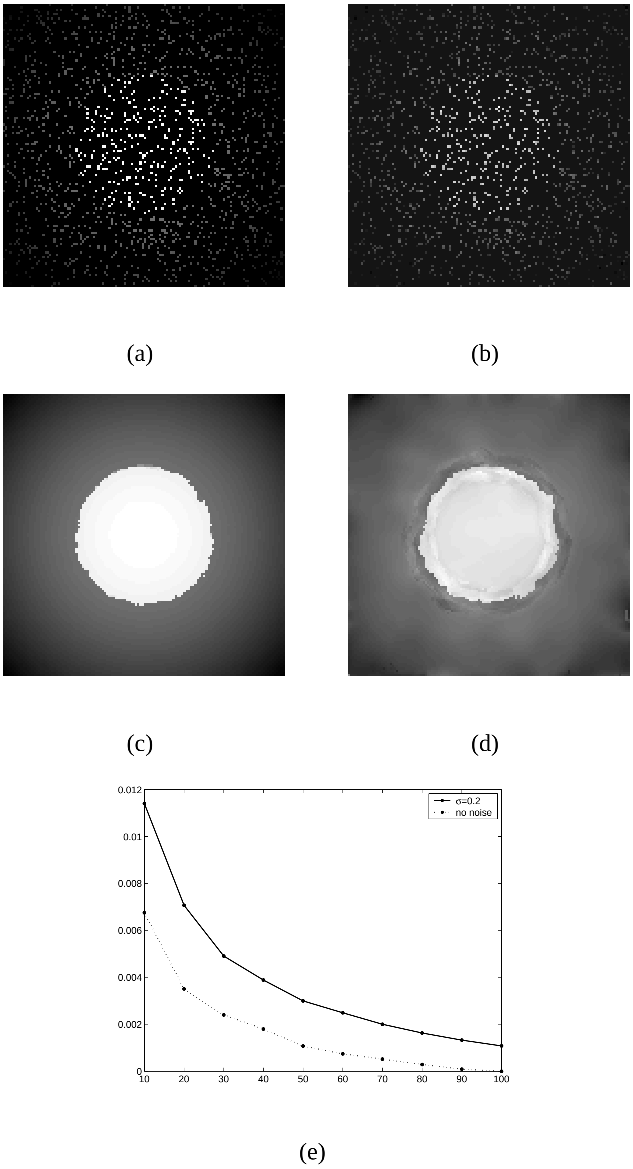
(d)

(e)

Fig. 10. (a) Noisy image; (b) Denoised image when $C = 1$ (i.e., without corner preservation) and $u = 0$; (c) Denoised image when $C = 1$ and $u = 51$; (d) Denoised image when $C = 0.25$ and $u = 0$; (e) Denoised image when $C = 0.25$ and $u = 51$.

images for several different percentages of the selected pixels. The dotted line denotes results from the randomly selected true pixels, whereas the solid line denotes results from these pixels plus noise. In the proposed procedure, the values of h and u are selected by CV and C is fixed at 1.

We also apply the proposed method to the *pepper* image shown in Fig. 12(a) with $n = 512^2$. In this case, 30% original pixels were randomly selected. The selected pixels and the


(a)

(b)

(c)

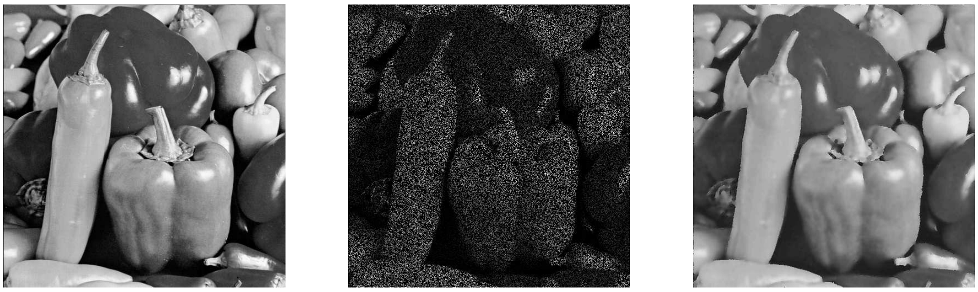
(d)

(e)

Fig. 11. (a) 10 $\%$ randomly selected true pixels from Fig. 3 (a); (b) Noisy version of (a); Plots (c) and (d) show the estimators from (a) and (b), respectively; (e) $\widehat{\text{MISE}}$ values for several different percentages from randomly selected pixels with (solid line) and without (dotted line) noise.

reconstructed image from them are shown in Figs. 12 (b) and (c), respectively. Again, in the proposed method, h and u are selected by CV and C is fixed at 1.

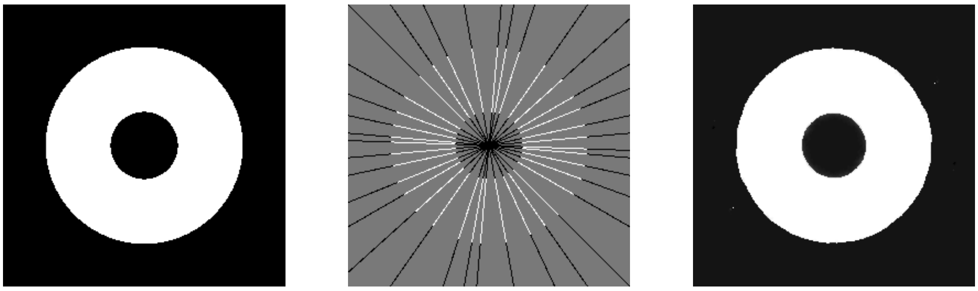
Finally, we test the proposed method with the *ring* image of size 256×256 pixels, which is similar to the image used in Arigovindan *et al.* [50]. For this circularly symmetric image, we randomly select 18 radial directions and pixels in these directions constitute the random design. The true image, selected pixels, and reconstructed image are presented in Figs. 13



(a) (b) (c)

Fig. 12. (a) True *pepper* image; (b) Sampled pixels; (c) Reconstructed image.

(a)-(c). The ring appears to be reconstructed well.



(a) (b) (c)

Fig. 13. (a) True image; (b) Sampled pixels; (c) Reconstructed image.

VII. Conclusion

This paper presents a new procedure for edge-preserving image denoising. It is based on local linear kernel surface estimation in the statistical framework. Advantages of the proposed procedure are its non-iterative feature, its explicit formulation (local smoothing method) and consequently its numerical simplicity. Numerical studies show that it performs well in applications. We also demonstrated that this method can be used for reconstructing images or surfaces from sparse noisy data. The proposed method could also possibly be used in conjunction with other available methods, such as the mean shift method by nonlinear filtering or more general methods (see e.g. Comaniciu [51] and Barash and Comaniciu [52]), by applying the proposed surface estimation procedure to their iteration steps. Note that the proposed method already improves upon the bilateral filtering, by using neighborhoods that are based upon the gradient estimation.

A limitation of the proposed procedure is the use of a global bandwidth parameter. Local bandwidth selection methods should be developed in the future, which might lead to a possible improvement of denoising in some cases.

VIII. Acknowledgements

The authors are grateful to the anonymous referees for their helpful comments and suggestions.

References

[1] P. Qiu, *Image Processing and Jump Regression Analysis*, New York: John Wiley & Sons, 2005.

[2] T. Poggio, "Early vision: From computational structure to algorithms and parallel hardware," *Compt. Vision Graphics Image Processing*, vol. 31, pp. 139-155, 1985.

[3] M.A. Garcı́a, "Efficient surface reconstruction from scattered points through geometric data fusion," *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 559-566, 1994.

[4] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, New York: Addison-Wesley Publishing Company, 1992.

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.

[6] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *Journal of the Royal Statistical Society (Series B)*, vol. 48, pp. 259-302, 1986.

[7] J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussions)," *Journal of the Royal Statistical Society (Series B)*, vol. 36, pp. 192-236, 1974.

[8] J. Moussouris, "Gibbs and Markov systems with constraints," *Journal of Statistical Physics*, vol. 10, pp. 11-33, 1974.

[9] F. Godtliebsen and G. Sebastiani, "Statistical methods for noisy images with discontinuities," *Journal of Applied Statistics*, vol. 21, pp. 459-477, 1994.

[10] A. Blake and A. Zisserman, *Visual Reconstruction*, Cambridge, MA: MIT Press, 1987.

[11] D. Geiger and F. Girosi, "Parallel and deterministic algorithms for MRFs: surface reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 401-412, 1991.

[12] D. Geman and G. Reynolds “Constrained restoration and the recovery of distributions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 367-383, 1992.

[13] J.A. Fessler, H. Erdogan and W.B. Wu, “Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise,” *IEEE Transactions on image processing*, vol. 9, pp. 1049-1055, 2000.

[14] J.L. Marroquin, F.A. Velasco, M. Rivera and M. Nakamura, “Gauss-Markov measure field models for low-level vision,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 337-347, 2001.

[15] J.W. Tukey, *Exploratory Data Analysis*, Reading Massachusetts: Addison-Wesley, 1977.

[16] N.C. Gallagher, Jr. and G.L. Wise, “A theoretical analysis of the properties of median filtering,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 1136-1141, 1981.

[17] T.S. Huang, ed. *Two-Dimensional Digital Signal Processing*, New York: Springer-Verlag, 1981.

[18] A.C. Bovik, T.S. Huang and D.C. Munson, “The effect of median filtering on edge estimation and detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 181-194, 1987.

[19] D.R.K. Brownrigg, “The weighted median filtering,” *Communications of the ACM*, vol. 27, pp. 807-818, 1984.

[20] P. Haavisto, M. Gabbouj and Y. Neuvo, “Median based idempotent filters,” *Journal of Circuits, Systems, and Computers*, vol. 1, pp. 125-148, 1991.

[21] T. Sun, M. Gabbouj and Y. Neuvo, “Center weighted median filters: some properties and their applications in image processing,” *Signal Processing*, vol. 35, pp. 213-229, 1994.

[22] P. Saint-Marc, J. Chen and G. Medioni, “Adaptive smoothing: a general tool for early vision,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 514-529, 1991.

[23] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images,” *Proceedings of the 1998 IEEE International Conference on Computer Vision*, pp. 839-846, Bombay, India, 1998.

[24] J. Koenderink, “The structure of images,” *Biological Cybernetics*, vol. 50, pp. 363-370, 1984.

[25] A. Hummel, “Representations based on zero-crossings in scale-space,” in *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pp. 204-209, 1987.

[26] P. Perona and J. Malik, “Scale space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629-639, 1990.

[27] D. Barash, “A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 844-847, 2002.

[28] D.L. Donoho and I.M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.

[29] D.L. Donoho and X. Huo, “Beamlets and multiscale image analysis,” *Lecture Notes in Computational Science and Engineering: Multiscale and Multiresolution Methods*, New York: Springer, pp. 149-196, 2001.

[30] S.G. Chan, B. Yu and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” *IEEE Transactions on Image Processing*, vol. 9, pp. 1522-1531, 2000.

[31] A.A.T. Figueiredo and R.D. Nowak, “Wavelet-based image estimation: an empirical Bayes approach using Jeffrey’s noninformative prior,” *IEEE Transactions on Image Processing*, vol. 10, pp. 1322-1331, 2001.

[32] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, “Image denoising using scale mixtures of gaussians in the wavelet domain,” *IEEE Transactions on Image Processing*, vol. 12, pp. 1338-1351, 2003.

[33] G. Nason and B. Silverman, “The discrete wavelet transform in S,” *Journal of Computational and Graphical Statistics*, vol. 3, pp. 163-191, 1994.

[34] S.Z. Li, “On discontinuity-adaptative smoothness prior in computer vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 576-586, 1995.

[35] M. Rivera and J.L. Marroquin, “Adaptive rest condition potentials: first and second order edge-preserving regularization, *Journal of Computer Vision and Image Understanding*, vol. 88, pp.76-93, 2002.

[36] M. J. Black and A. Rangarajan, “On the unification of line processes, outlier rejection, and robust statistics with applications in early vision, *International Journal of Computer Vision*, vol. 19, pp 57-91, 1996.

[37] S.S. Sinha and B.G. Schunck, “A two-stage algorithm for discontinuity-preserving surface reconstruction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 36-55, 1992.

[38] R.L. Stevenson and E.J. Delp, “Viewpoint invariant recovery of visual surfaces from sparse data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 897-909, 1992.

[39] J.H. Yi and D.M. Chelberg, “Discontinuity-preserving and viewpoint invariant reconstruction of visible surfaces using a first order regularization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 624-629, 1995.

[40] P. Qiu, “Discontinuous regression surfaces fitting,” *The Annals of Statistics*, vol. 26, pp. 2218-2245, 1998.

[41] C.-K. Chu, I.K. Glad, F. Godtliebsen and J.S. Marron, “Edge preserving smoothers for image processing (with discussion),” *Journal of the American Statistical Association*, vol. 93, pp. 526-556, 1998.

[42] J. Polzehl and V.G. Spokoiny, “Adaptative weights smoothing with applications to image restoration,” *Journal of the Royal Statistical Society*, Vol. B-62, pp. 335-354, 2000.

[43] P. Qiu, “The local piecewisely linear kernel smoothing procedure for fitting jump regression surfaces,” *Technometrics*, vol. 46, pp. 87-98, 2004.

[44] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, New York: Chapman and Hall, 1996.

[45] I. Gijbels, A. Lambert and P. Qiu, “Jump-preserving regression and smoothing using local linear fitting: a compromise,” *Under revision*.

[46] G.Z. Yang, P. Burger, D.N. Firmin and S.R. Underwood, “Structure adaptative anisotropic image filtering,” *Image and*

Vision Computing, vol. 14, pp. 135-145, 1996.

[47] F. Chabat, G.Z. Yang and D.M. Hansell, "A corner orientation detector," *Image and Vision Computing*, vol. 17, pp. 761-769, 1999.

[48] I. Gijbels, A. Lambert and P. Qiu, "Edge-preserving image denoising and estimation of discontinuous surfaces," *Technical Report*. Website http://www.stat.ucl.ac.be/∼alambert.

[49] T. Strohmer, "Computationally attractive reconstruction of bandlimited images from irregular samples," *IEEE Transactions on Image Processing*, vol. 6, pp. 540-548, 1997.

[50] M. Arigovindan, M. Sülhing, P. Hunziker and M. Unser, "Variational Image Reconstruction from Arbitrarily Spaced Samples: A Fast Multiresolution Spline Solution," *IEEE Transactions on Image Processing*, vol. 14, pp 450-460, 2005.

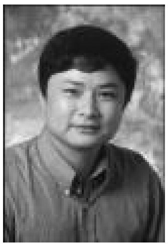
[51] D. Comaniciu, "An algorithm for data-driven bandwidth selection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp 281-283, 2003.

[52] D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift", *IEEE Image and Vision Computing*, vol. 22, pp 73-81, 2004.

Irène Gijbels was born in Belgium. She received a university degree in mathematics from the Catholic University of Leuven, Belgium, and a PhD degree in Sciences, specialization mathematical statistics from the Central Jury, Belgium.

After her PhD study she was a Senior Research Assistant at the National Science Foundation Belgium. From 1992 till 2004 she was with the Institute of Statistics, Université catholique de Louvain, Louvain-la-Neuve, Belgium. She joined the Catholic University of Leuven, Belgium, in October 2004. Her current research interests include non- and semiparametric curve and surface estimation, deconvolution problems, semi-and nonparametric inference and dimension reduction.

Alexandre Lambert received a university degree in civil engineering in applied mathematics from Catholic University of Louvain, Belgium and Ph.D. degree in Sciences, specialization in statistics from the Institute of Statistics, Catholic University of Louvain, Belgium. His current research interests include nonparametric estimation of discontinuous curves and surfaces and jump-edge detection in the statistical framework.



Peihua Qiu was born in Shanghai, P.R. China. He received B.S. degree in mathematics and M.S. degree in statistics from Fudan University, Shanghai, China, M.S. degree in statistics from University of Georgia, Athens, GA, and Ph.D. degree in statistics from University of Wisconsin, Madison, WI.

From 1996 to 1998, he was with the Biostatistics Center at the Ohio State University, where he worked on biostatistical research and consulting. He was an assistant professor during 1998-2002, and has been an associate professor since 2002, in the School of Statistics at the University of Minnesota, Minneapolis, MN. His current research interests include curve and surface estimation from noisy data, image restoration, image segmentation, statistical process control, and survival analysis.