Computational Statistics and Data Analysis 56 (2012) 543–558

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Locally adaptive image denoising by a statistical multiresolution criterion

Thomas Hotz [a,*], Philipp Marnitz [a], Rahel Stichtenoth [b,1], Laurie Davies [b], Zakhar Kabluchko [a,2], Axel Munk [a]

[a] Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstrasse 7, 37077 Göttingen, Germany

[b] Faculty of Mathematics, University of Duisburg-Essen, 45117 Essen, Germany

ARTICLE INFO

Article history:
Received 9 June 2010
Received in revised form 29 August 2011
Accepted 29 August 2011
Available online 6 September 2011

Keywords:
Image reconstruction
Statistical multiresolution criterion
Bandwidth selection

ABSTRACT

It is shown how to choose the smoothing parameter in image denoising by a statistical multiresolution criterion, both globally and locally. Using inhomogeneous diffusion and total variation regularization as examples for localized regularization schemes, an efficient method for locally adaptive image denoising is presented. As expected, the smoothing parameter serves as an edge detector in this framework. Numerical examples together with applications in confocal microscopy illustrate the usefulness of the approach.

1. Introduction

Image denoising is one of the main tasks in image analysis, as documented by numerous articles and books published on the subject, see for example Scherzer et al. (2009), Buades et al. (2005b) and Aubert and Kornprobst (2002). Statisticians have made many contributions to this area: using probabilistic models Bayesian methods were among the first with a statistical perspective on the subject, see Geman and Geman (1984), Besag (1986), Winkler (2003), and Thon et al. (2012). From a frequentist point of view image denoising becomes a smoothing or reconstruction problem, see for example Hall and Titterington (1986) and Polzehl and Spokoiny (2000, 2003) as well as Korostelev and Tsybakov (1993). The fact that many images feature sharp edges (see for example Chu et al. (1998) and Donoho (1999)) prompted a generalization of the well-established smoothing techniques in one-dimensional jump detection, see Ogden and Parzen (1996) and Qiu (2005, 2007). For a unifying framework for many popular numerical and statistical denoising techniques see Mrázek et al. (2006) and Charles and Rasson (2003) for the case of Poisson data.

Put simply, *statistical image denoising* amounts to reconstructing a noiseless image f given a noisy image y. Usually it is assumed that the noise ϵ is additive, $y = f + \epsilon$. In the following, it will be further assumed that the noise is generated at random: more specifically, for pixels (i, j)

$$y_{ij} = f_{ij} + \epsilon_{ij}, \tag{1}$$

* Corresponding author. Tel.: +49 551 39 13517; fax: +49 551 39 13505.
E-mail addresses: hotz@math.uni-goettingen.de (T. Hotz), marnitz@math.uni-goettingen.de (P. Marnitz), rahel.stichtenoth@uni-due.de (R. Stichtenoth), laurie.davies@uni-due.de (L. Davies), zakhar.kabluchko@uni-ulm.de (Z. Kabluchko), munk@math.uni-goettingen.de (A. Munk).

1 Present address: Sulfurcell Solartechnik GmbH, Groß-Berliner Damm 149, 12487 Berlin, Germany.

2 Present address: Institute of Stochastics, University of Ulm, Helmholtzstrasse 18, 89069 Ulm, Germany.

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.08.018

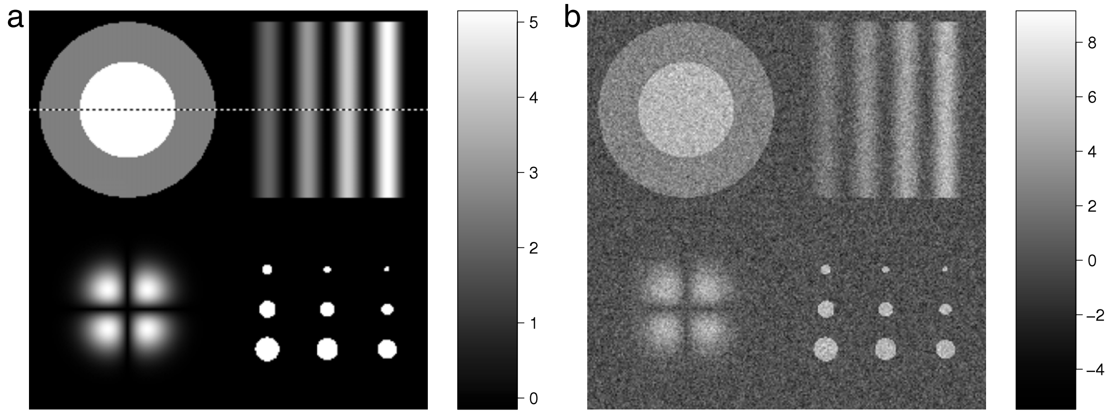

Fig. 1. (a) 256 × 256 pixel test image f taking values in [0, 5]; the dashed line indicates row 64 shown in detail in Fig. 8. (b) Simulated data y with noise level $\sigma = 1$.

with Gaussian white noise ϵ_{ij}, that is the ϵ_{ij} are independently and identically distributed Gaussian random variables with zero mean and variance σ^2,

$$\epsilon_{ij} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2). \tag{2}$$

Furthermore, it will be assumed that the value at each pixel is a real number, $y_{ij}, f_{ij}, \epsilon_{ij} \in \mathbf{R}$. This models a grey-scale image, although in practice grey levels are usually restricted to a finite number of discrete values, in particular to integers between 0 and 255. In many applications, a Gaussian assumption on the noise is therefore not very plausible and other noise processes would be more appropriate. Nonetheless, for the sake of simplicity the basic ideas will be expounded under the Gaussian assumption. Extensions to other models are briefly discussed at the end of Section 3. It may be noted that for a well-calibrated image y with a reasonable range of possible values, the Gaussian assumption is not very crucial as long as the errors are i.i.d., symmetric and do not have heavy tails.

Fig. 1 shows an artificial example where f exhibits varying degrees of smoothness (left), and Gaussian white noise has been added to obtain the data y (right). The assumption about the noise can then be exploited in order to distinguish between the 'true', noiseless image f and the noise ϵ as demonstrated in the following sections.

Clearly, this is impossible unless assumptions are made about f, for example that it varies slowly from pixel to pixel. In order to be able to formulate such 'smoothness' assumptions more precisely, let (with a slight abuse of notation) f_{ij} denote the values of a function f on a regular grid, that is $f : [0, 1]^2 \to \mathbf{R}$ with $f_{ij} = f(\frac{i}{n}, \frac{j}{n})$ where $i, j \in \{1, \ldots, n\}$ is a square grid chosen for ease of notation with $x_{ij} = (\frac{i}{n}, \frac{j}{n})$. The analysis can be extended all to rectangular grids, to higher dimensions, and to non-uniform sampling schemes through the use of finite elements (Ern and Guermond, 2004). Within this model f can be viewed as a function and 'smoothness' can be defined more rigorously to mean that f belongs to some function class, for example that f lies in a Sobolev or Besov ball, or that f has bounded variation, see Korostelev and Tsybakov (1993).

Image denoising techniques like the ones discussed in Section 2 generally require the choice of some *smoothing* or *regularization parameter* a which determines how much smoothing is to be applied. This parameter might be *localized*, allowing for different amounts of smoothing to be applied to different parts of the image; the regularization parameter then becomes a function $a : [0, 1]^2 \to \mathbf{R}$. This enables the user to adapt to differing levels of smoothness across the image. The reconstruction or denoised image $\hat{f}$ consequently depends on the values of the smoothness parameter. A purely data-driven and generally applicable way to choose a will be described in the following. It will be illustrated using two specific denoising techniques, namely linear diffusion and TV penalization. It is to be stressed, however, that the approach is in principle applicable to any regularization technique which depends on properly choosing a regularization parameter, the latter possibly being a function as described above.

The main idea can be summarized as follows. Consider the residuals $r_{ij} = y_{ij} - \hat{f}_{ij}$ which depend on the smoothness parameter a. If the image is oversmoothed, some structures which are present in f will be smoothed away, but these structures will then be visible in the residuals. In the case of a perfect reconstruction $\hat{f} = f$, however, the residuals form white noise, see (1). One possible way to decide whether the image was excessively smoothed is therefore to check whether the residuals look like white noise — if there is still some structure left in the residuals then the image must have been smoothed too much. This idea is at the heart of statistical methods for automatically selecting the regularization parameter.

The proposed key ingredient for choosing the smoothness parameter a is a *statistical multiresolution criterion* to be introduced in Section 3. An important feature of this criterion is that it not only detects whether the residuals deviate from white noise but also *localizes* the deviations. This then allows for a locally adaptive choice of a. In Section 4 an algorithm for a data-driven selection of a is given. Numerical details and results are given in Section 5, together with an example from