

Medical Image Noise Reduction Using the Sylvester–Lyapunov Equation

João M. Sanches, *Member, IEEE*, Jacinto C. Nascimento, *Member, IEEE*, and Jorge S. Marques

Abstract—Multiplicative noise is often present in medical and biological imaging, such as *magnetic resonance imaging* (MRI), *ultrasound*, *positron emission tomography* (PET), *single photon emission computed tomography* (SPECT), and *fluorescence microscopy*. Noise reduction in medical images is a difficult task in which linear filtering algorithms usually fail. Bayesian algorithms have been used with success but they are time consuming and computationally demanding. In addition, the increasing importance of the 3-D and 4-D medical image analysis in medical diagnosis procedures increases the amount of data that must be efficiently processed. This paper presents a Bayesian denoising algorithm which copes with additive white Gaussian and multiplicative noise described by Poisson and Rayleigh distributions. The algorithm is based on the *maximum a posteriori* (MAP) criterion, and edge preserving priors which avoid the distortion of relevant anatomical details. The main contribution of the paper is the unification of a set of Bayesian denoising algorithms for additive and multiplicative noise using a well-known mathematical framework, the Sylvester–Lyapunov equation, developed in the context of the Control theory.

Index Terms—Despeckling, image denoising, medical imaging.

I. INTRODUCTION

MEDICAL imaging acquisition technologies and systems introduce noise and artifacts in the images that should be attenuated by denoising algorithms. The denoising process, however, should not destroy anatomical details relevant from a clinical point of view. For instance, the noise corrupting the ultrasound images often contains relevant medical information useful for diagnosis proposes [1]–[3]. A significant effort has been made in the last decades to develop reliable and fast algorithms to remove noise without distorting or destroying relevant clinical information [4], [5]. Despite the huge advances in the computer technology, the computational efficiency is still a key issue, since the amount of information grows as we extend the analysis from 2-D images to 3-D and 4-D data [6]–[9].

The *additive white Gaussian noise* (AWGN) model is the simplest and the most used model to describe noisy signals [10] and images [11]. However, this model is not acceptable when the image is corrupted by multiplicative or impulsive noise [11], [12].

Several techniques have been proposed to transform the multiplicative noise in order to use the AWGN paradigm. Among

which the log conversion [11] for speckle and the Anscombe transform [13] for Poisson have been widely used. However, these approaches leads to rough approximations of Gaussian distribution as it will be discussed in Section II.

Two types of multiplicative noise often arise in several imaging modalities: speckle and Poisson noise. Both types are called multiplicative in the sense that their variance is not constant but depends on the parameters to be estimated [14]. The speckle noise usually appears in acquisition processes involving coherent radiation like high intensity *LASER* [15], *ultrasound* [16] and *synthetic aperture radar* (SAR) [17] and the Poisson noise in systems involving counting procedures like PET/SPECT [18], functional MRI [19], fluorescence confocal microscopy [20] and very low intensity *LASER* [21]. The additive Gaussian noise model is used in *computed tomography* (CT) [22] and sometimes in low intensity MRI [23]. In fact, the noise corrupting the MRI images is usually modeled by a Rice distribution which can be approximated by a Gaussian distribution for low image intensities and by a Rayleigh distribution for high intensity regions [23]. Common distributions describing speckle noise are: Rayleigh [23], Poisson [20], K-distribution [24], Nakagami [25], Fisher-Tippet [26], and generalized gamma (GG) [27].

In this paper, a framework is presented to deal with additive white Gaussian and multiplicative noise described by Rayleigh and Poisson distributions. The goal is not to describe a new denoising algorithm, neither a new and faster computational procedure but to describe a new formalism for denoising images corrupted by multiplicative noise.

The paper shows that the MAP estimate of the image is the solution of a Sylvester–Lyapunov equation

$$\Phi \mathbf{X} + \mathbf{X} \Phi + \mathbf{Q} = 0 \quad (1)$$

where \mathbf{Q} is a known matrix in the case of the Gaussian noise and it depends on \mathbf{X} in the case of the Rayleigh and Poisson distributions.

The main contribution of this paper is the unification, in a single framework, of a set of Bayesian denoising algorithms for additive white Gaussian noise (AWGN) and multiplicative noise by using a well known mathematical framework, the Sylvester–Lyapunov equation, for which there are fast and computationally efficient algorithms described in the literature [28], [29] and implemented in scientific processing packages, e.g., Matlab.

The proposed algorithm is compared with several despeckling algorithms [30].

II. STATE OF THE ART

Linear filtering, e.g., Wiener filter [10], was the first approach used to deal with multiplicative noise [14]. Although linear filtering reduces the amount of noise in the image it over-smooths

Manuscript received September 13, 2007; revised May 1, 2008. Published August 13, 2008 (projected). This work was supported in part by FCT, Portuguese Ministry of Science and Technology and Higher Education (which includes FEDER funds). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stanley J. Reeves.

The authors are with the Instituto Superior Tecnico, Instituto de Sistemas e Robotica (IST/ISR), 1049-001 Lisboa, Portugal (e-mail: jmrs@alfa.ist.utl.pt).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2008.2001398

the transitions and anatomical details that should be preserved [27]. This difficulty arises because images corrupted with this type of noise often present low *signal-to-noise ratio* (SNR) and the nonlinear and nonadditive nature of the corrupting process makes the linear approach not appropriate [22]. This difficulty can be overcome by adopting specific multiplicative models to describe the speckle noise [11], [31]–[33].

In the last two decades, nonlinear filtering methods have been successfully used to deal with the multiplicative noise using several approaches [34]. The median filter has been extensively used to process multiplicative and impulsive noise [11], [35] because it is a very simple technique and the results are usually impressive. Several variants of the basic median filter were proposed, e.g., adaptive median filters. An example is the *weighted median filter* (WMF) proposed in [36] for despeckling of ultrasound images. The filter considers a small window centered at each pixel as usual but each element of the input image is considered several times according to the local statistics of the image.

More recently, [30] proposes a novel stochastically driven filtering method with constant and variable size windows. A simpler version of this filter with constant window, is proposed in [37]. In [30], a novel despeckling method iteratively removes outliers by determining the local mean and standard deviation from an adaptively varying window. By removing outliers (local extrema) at each iteration, this method produces a convergent sequence of images by squeezing the stochastically distributed pixels values to a limiting value. It was experimentally shown by the authors that the proposed filter outperforms all the median filters considered in the experiments.

Wavelet based despeckling algorithm have also been extensively used in medical imaging since the seminal works of Donoho on soft-thresholding were published [38]. This method is based on the multi scale decomposition of the noisy image and in the processing of the image coefficients at coarser scales [23], [34], [39]–[42]. For example, in [40], the authors compute the multiscale decomposition of the logarithmic of the ultrasound image and model the coefficients at each scale by the alpha-stable heavy-tailed distribution. A Bayesian based nonlinear operator is then used to remove the speckle noise at each scale. A similar approach is described in [41] using *generalized Gaussian distribution* (GGD).

In [23], the same multiscale decomposition is used to identify wavelet coefficients at different scales with high correlation. It is assumed that these coefficients describe anatomical details of the image which should be preserved. This allows the elimination of the noise without removing anatomical information. Comprehensive surveys on wavelets in medical imaging can be found in [43] and [44].

An usual approach to deal with multiplicative noise is to apply a logarithmic transformation to the observations. If the observation model is $y(n) = x(n)\eta(n)$ [11], then $\log y(n) = \log x(n) + \log \eta(n)$ where it is assumed that $\log \eta(n) \sim \mathcal{N}(\mu, \sigma^2)$. However, this assumption is not true, as discussed in [26] and [27]. In fact, this observation model does not accurately describe the *speckle* [16] and Poisson [45] noise generation process dealt in the paper. Besides, even if the referred model is correct the distribution of $r(n) = \log(\eta(n))$ would not be Gaussian as can easily be demonstrated. Nevertheless, this approximation, called *homomorphic wavelet despeckling* (HWDS), is used by several authors (e.g., [46])

producing some of the most efficient and fast denoising state-of-the-art algorithms to deal with multiplicative noise.

Bayesian approaches have been used to deal with this type of noise and they have produced some of the most advanced algorithms to remove multiplicative noise. Bayesian methods formulates the denoising task as an estimation problem, where the likelihood function and a prior distribution are jointly maximized. Three issues must be chosen in this framework: *i*) the statistical observation model, *ii*) the prior distribution and, *iii*) the optimization method [47]. Bayesian methods are often combined with the wavelet decomposition leading to efficient and fast algorithms [48]–[50].

In this paper, we present a unifying Bayesian algorithm which is able to deal with AWGN and multiplicative (Rayleigh and Poisson) noise and with several prior distributions.

This paper is organized as follows. Section III formulates the denoising problem in a Bayesian framework. Section IV describes the proposed method based on the Sylvester–Lyapunov equation, the simpler linear case of the AWGN and Gaussian prior. Section V uses a majorize/minimize (MM) optimization algorithm to denoise the images with multiplicative noise. Section VI presents examples of applications and comparisons with other methods and Section VII concludes the paper.

Appendix VIII-A discusses edge preserving priors, emphasizing the ones using *adaptive potential functions* (APF). Appendix VIII-B contains the pseudo-code of the algorithm.

III. PROBLEM FORMULATION

Let \mathbf{X} be a $N \times M$ unknown image to be estimated/reconstructed from a noisy image, \mathbf{Y} . The *maximum a posteriori* (MAP) estimate of \mathbf{X} is the solution of the following optimization problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} E(\mathbf{X}, \mathbf{Y}) \quad (2)$$

where

$$E(\mathbf{X}, \mathbf{Y}) = \underbrace{E_Y(\mathbf{X}, \mathbf{Y})}_{\text{Data fidelity term}} + \underbrace{E_X(\mathbf{X})}_{\text{Prior term}}. \quad (3)$$

$E_Y(\mathbf{X}, \mathbf{Y}) = -\log p(\mathbf{Y}|\mathbf{X})$, is called the *data fidelity term* and attracts the solution toward the data while $E_X(\mathbf{X}) = -\log p(\mathbf{X})$, is called the *prior* or *internal energy* and regularizes the solution removing the noise.

The computation of \mathbf{X} , based on the minimization of the data fidelity term $E_Y(\mathbf{X}, \mathbf{Y})$, is the *maximum likelihood* (ML) estimation problem and it is usually an ill-posed problem in the Hadamard sense because the solution is not unique and it may not depend continuously on the data [47], [51], [52]. To overcome this difficulty, a regularization term is added, turning the problem into a well-posed problem. The distribution $p(\mathbf{X})$ introduces prior knowledge about the image to be estimated, and it usually favors smooth solutions. The MAP energy function, $E(\mathbf{X}, \mathbf{Y})$ has a global minimum, called MAP solution, which is, very often, difficult to find because the MAP optimization function (3) may not be convex.

Insight on the *a priori* distributions of images are difficult to attain. This is especially true in medical applications where assumptions about the prior distribution may lead to wrong diagnosis. The common assumption about these images is that they are band-limited, changing slowly in space except near the

organs boundaries where abrupt transitions are expected. This is a difficult assumption because the location of the transitions are unknown and must be estimated. \mathbf{X} can be modeled as a *Markov random field* (MRF) under the assumption that neighboring pixels have similar intensities, except if they are located at a transition.

The joint probability density function of a MRF, given by the Hammersley-Clifford theorem [53], [54] is a Gibbs distribution

$$p(\mathbf{X}) = \frac{1}{Z} e^{-U(\mathbf{X})} \quad (4)$$

where $U(\mathbf{X})$ is the Gibbs energy and Z is the partition function, which is very difficult to compute. A typical choice for the Gibbs energy function is [55]

$$U(\mathbf{X}) = \alpha \sum_{i,j} \rho(\delta_v(i,j), \delta_h(i,j)) \quad (5)$$

where $\delta_v(i,j) = x(i,j) - x(i-1,j)$ and $\delta_h(i,j) = x(i,j) - x(i,j-1)$ are the first-order vertical and horizontal differences respectively, computed at pixel $x(i,j)$. $\rho(\delta_v, \delta_h)$ is a potential function and α is the prior hyper-parameter that may be automatically estimated [41], [56]. Here, it is chosen in a trial and error basis.

Assuming conditional independence on the observations, $p(\mathbf{Y}|\mathbf{X}) = \prod_{i,j} p(y(i,j)|x(i,j))$ and defining

$$d(x,y) = -x \log p(y|x) \quad (6)$$

the data *fidelity term*, $E_Y(\mathbf{X}, \mathbf{Y}) = -\log p(\mathbf{Y}|\mathbf{X})$, is

$$E_Y(\mathbf{X}, \mathbf{Y}) = \sum_{i,j} d(x(i,j), y(i,j)) \quad (7)$$

and the energy is given by

$$E(\mathbf{X}, \mathbf{Y}) = \sum_{i,j} d(x(i,j), y(i,j)) + \alpha \sum_{i,j} \rho(\delta_v(i,j), \delta_h(i,j)). \quad (8)$$

The conditional independence of the observations may not be realistic in some cases. However, it is a convenient hypothesis because it separates spatial dependence and data dependence effects making the problem much easier. Furthermore, the inclusion of the statistical dependence in the observation model may not lead to relevant improvement of the final solution, as noted in [57].

The minimization of the energy function (3) is graphically illustrated in Fig. 1. In this 1-D example the MAP solution minimizes the total potential energy which is the sum of the energies $d(x,y)$ associated with the data springs which pull the solution towards the data and the regularization energies $\rho(\delta_v, \delta_h)$ associated with the other springs which smooth the solution and avoid fast changes.

The MAP solution can be obtained by computing the stationary points of (3) by solving the equation

$$\nabla_X E_Y(\mathbf{X}, \mathbf{Y}) + \nabla_X E_X(\mathbf{X}) = 0 \quad (9)$$

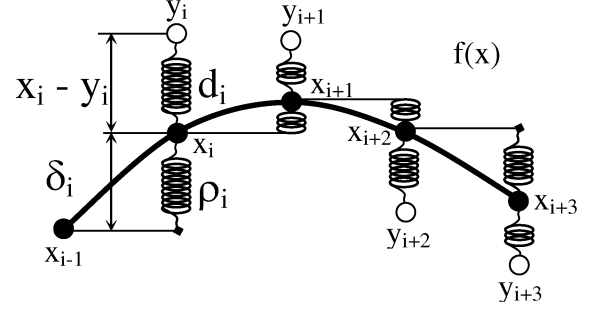


Fig. 1. Mechanical interpretation of the optimization task for the 1-D case.

where ∇_X denotes the gradient operator with respect to (w.r.t.) \mathbf{X} .

In the following sections, three observation models are considered based on the Gaussian, Rayleigh and Poisson distributions and three regularization models are also discussed: quadratic (L_2), *total variation* (TV), and Benford (Bfd).

IV. AWGN AND GAUSSIAN PRIOR MODELS

Let us first consider the simpler case of *additive white Gaussian noise* (AWGN) model. We will show that the denoising problem can be solved by a Sylvester-Lyapunov equation, independent of the observation model. Later on, multiplicative noise will be considered and included in this formalism.

Let \mathbf{Y} be the observed image obtained by corrupting the original image \mathbf{X} with additive white Gaussian noise (AWGN)

$$\mathbf{Y} = \mathbf{X} + \mathbf{\Gamma} \quad (10)$$

$\mathbf{\Gamma} = \{\gamma_{ij}\}$ is a random image with $\gamma_{ij} \sim \mathcal{N}(0, \sigma^2)$. The prior distribution is also assumed to be Gaussian

$$-\log(p(\mathbf{X})) = \alpha \sum_{i,j=0}^{N-1, M-1} [\delta_v^2(i,j) + \delta_h^2(i,j)] + C \quad (11)$$

where C is a constant and the potential function $\rho(\delta_v, \delta_h)$ is quadratic.

The energy function in this case is

$$E(\mathbf{X}, \mathbf{Y}) = \frac{1}{2\sigma^2} \sum_{i,j=0}^{N-1, M-1} [x(i,j) - y(i,j)]^2 + \alpha \sum_{i,j=0}^{N-1, M-1} [\delta_v^2(i,j) + \delta_h^2(i,j)] \quad (12)$$

and the partial derivative w.r.t. a given pixel, $x(k,l)$, is

$$\frac{\partial E(\mathbf{X}, \mathbf{Y})}{\partial x(k,l)} = \frac{1}{\sigma^2} (x(k,l) - y(k,l)) + 2\alpha \left[N_v x(k,l) - \sum_{s \in S(x(k,l))} x_s(k,l) \right] = 0 \quad (13)$$

where $S(x(k, l))$ is the 4-neighborhood of $x(k, l)$ and $N_v = 4$. This equation can be rewritten as follows:

$$x(k, l) = (1 - \mathcal{K})y(k, l) + \mathcal{K}\bar{x}(k, l) \quad 0 \leq k, l \leq N - 1, M - 1 \quad (14)$$

where $\mathcal{K} = 2N_v\beta/(1 + 2N_v\beta)$ is a constant, $\beta = \alpha\sigma^2$ and

$$\bar{x}(k, l) = \frac{1}{N_v} \sum_{s \in S(x(k, l))} x_s(k, l) \quad (15)$$

is the average intensity in the 4-neighborhood of $x(k, l)$. The parameter β depends on the noise variance σ^2 and the parameter α , which is manually chosen in a trial and error basis.

In the case of Gaussian observation and prior distributions, the set of (15) is linear and a closed form solution may be computed by using matrix manipulation. However, for the multiplicative observation models (Poisson and Rayleigh) and non Gaussian priors (TV and Benford), considered later, this set of equation is no longer linear and must be iteratively solved, e.g., by using the following recursion expression based on the fixed point algorithm [58]

$$x^t(k, l) = (1 - \mathcal{K})y(k, l) + \mathcal{K}\bar{x}^{t-1}(k, l) \quad (16)$$

where $\bar{x}^{t-1}(k, l)$ is the average value of the neighbors of $x(k, l)$ computed in the previous iteration, $t - 1$.

The iterative algorithm sweeps all elements of \mathbf{X} until convergence is achieved. The final estimate depends on several factors such as the initialization, the updating order of the image pixels of \mathbf{X} and on the prior parameter α . The convergence conditions depend on the α parameter, i.e., the higher is α , more stable is the iterative algorithm. In fact, for small values of α the algorithm may diverge. This behavior is expected, since the denoising process based on the ML criterion is an *ill posed* inverse problem. The goal of the prior distribution, is to regularize the solution and to stabilize the estimation algorithm. Therefore, the smaller is the value of α the more *ill-posed* the problem is.

Let us now formalize the estimation process using matrix notation, defining $\mathbf{x} = \text{vect}(\mathbf{X})$, $\mathbf{y} = \text{vect}(\mathbf{Y})$ as two column vectors corresponding to lexicographical [10] column ordering of the images \mathbf{X} and \mathbf{Y} , where the columns of each image matrix are stacked in vectors. Using this notation, the energy function (3) may be written as follows:

$$E(\mathbf{x}, \mathbf{y}) = \frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{y}\|^2 + \alpha(\|D_v \mathbf{x}\|^2 + \|D_h \mathbf{x}\|^2) \quad (17)$$

where $\|\mathbf{v}\|^2 = \mathbf{v}^T \mathbf{v}$ is the squared Euclidean norm of \mathbf{v} and D_v, D_h are $NM \times NM$ difference matrix operators. The minimization of (17) leads to the following set of equations:

$$[\mathbf{I} + 2\sigma^2\alpha(D_v^T D_v + D_h^T D_h)] \mathbf{x} = \mathbf{y}. \quad (18)$$

This is a system of NM equations and NM unknowns whose solution is the MAP estimate of \mathbf{x} . It is not possible to solve the system by matrix inversion since the matrix $\mathbf{A} = [\mathbf{I} + 2\sigma^2\alpha(D_v^T D_v + D_h^T D_h)]$ has huge dimensions. For instance, if \mathbf{X} is a 128×128 image, \mathbf{A} is a square matrix of dimensions 16384×16384 . Another difficulty is the computational burden and memory associated with the vectorial tasks.

These difficulties are avoided by locally processing the image data, in a pixel-wise or block-wise way.

An alternative method is now proposed to avoid dealing with such high-dimensional matrices keeping the objects within reasonable dimensions, e.g., equal to the dimension of the original image.

The energy function (12) can be written as

$$\begin{aligned} E(\mathbf{X}, \mathbf{Y}) &= \text{tr} \left[\frac{1}{2\sigma^2} (\mathbf{X} - \mathbf{Y})^T (\mathbf{X} - \mathbf{Y}) \right] \\ &\quad + \alpha \text{tr} [(\phi_v \mathbf{X})^T (\phi_v \mathbf{X}) \\ &\quad + (\phi_h \mathbf{X}^T)^T (\phi_h \mathbf{X}^T)] \\ &= \text{tr} \left[\frac{1}{2\sigma^2} (\mathbf{X} - \mathbf{Y})^T (\mathbf{X} - \mathbf{Y}) \right] \\ &\quad + \alpha \text{tr} \left[\underbrace{\mathbf{X}^T (\phi_v^T \phi_v) \mathbf{X}}_{\Theta_N} + \underbrace{\mathbf{X} (\phi_h^T \phi_h) \mathbf{X}^T}_{\Theta_M} \right] \end{aligned} \quad (19)$$

where \mathbf{X}, \mathbf{Y} are $N \times M$ matrices, ϕ_v , and ϕ_h are $N \times N$, $M \times M$ matrices used to compute the vertical and horizontal first order differences; tr denotes the trace of a matrix. Both matrices, ϕ_v and ϕ_h , have the following structure (but different dimensions)

$$\phi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & -1 \\ -1 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & -1 & 1 & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & -1 & 1 \end{pmatrix}. \quad (20)$$

Computing the derivative of E with respect to the unknown matrix \mathbf{X} , we obtain a stationary condition

$$\frac{dE}{d\mathbf{X}} = \frac{1}{2\sigma^2} (\mathbf{X} - \mathbf{Y}) + 2\alpha(\Theta_N \mathbf{X} + \mathbf{X} \Theta_M) = 0 \quad (21)$$

where $[(dE/d\mathbf{X})]_{k,l} = \partial E(\mathbf{X}, \mathbf{Y}) / \partial x(k, l)$ and

$$\begin{aligned} \Theta_N &= \phi_N^T \phi_N \\ \Theta_M &= \phi_M^T \phi_M \end{aligned} \quad (22)$$

are two symmetric circulant matrices, with the following structure [see (23) shown at the bottom of the next page].

After straightforward manipulations, (21) can be written as

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} + \mathbf{C} = 0 \quad (24)$$

where

$$\mathbf{A} = \frac{1}{2} \mathbf{I}_N + 2\alpha\sigma^2 \Theta_N \quad (25)$$

$$\mathbf{B} = \frac{1}{2} \mathbf{I}_M + 2\alpha\sigma^2 \Theta_M \quad (26)$$

$$\mathbf{C} = -\mathbf{Y} \quad (27)$$

and $\mathbf{I}_N, \mathbf{I}_M$ are the identity matrices of order N, M , respectively. Equation (24) is the well-known Sylvester equation [59] (the Lyapunov equation is a particular case of this equation, when $\mathbf{B} = \mathbf{A}^H$). In this equation, $\mathbf{X}, \mathbf{Y}, \mathbf{A}$ and \mathbf{B} are $N \times M$, $N \times M$, $N \times N$ and $M \times M$ -dimensional matrices, respectively. Therefore, all the matrices involved in this equation have the size of the original data. Note that, if \mathbf{X} is a square images, \mathbf{A} and \mathbf{B} are equal, symmetric, and real matrices, and (21) becomes the Lyapunov equation.

TABLE I
GAUSSIAN, POISSON, AND RAYLEIGH MODELS FOR THE OBSERVATIONS

Obs. Model	$p(y x)$	$d(x, y)$	$\frac{d}{dx}d(x, y)$	x^{ML}
Gaussian	$Ke^{-\frac{1}{2\sigma^2}(x-y)^2}$	$\frac{1}{2\sigma^2}(x-y)^2$	$\frac{x-x^{ML}}{\sigma^2}$	y
Poisson	$\frac{x^y e^{-x}}{y!}$	$x - y \log(x)$	$\frac{x-x^{ML}}{x}$	y
Rayleigh	$\frac{y}{x}e^{-y^2/(2x)}$	$\frac{y^2}{2x} - \log(\frac{y}{x})$	$\frac{x-x^{ML}}{x^2}$	$y^2/2$

There are efficient algorithms to solve the Sylvester–Lyapunov equation [28], [29], [60] available in several mathematical packages, e.g., Matlab and Mathematica, which avoid the need of huge matrices and the solution of huge systems of equations.

The complexity of these algorithms is about $MN^2 + NM^2$ floating point operations, for $N \times M$ images [61].

This section showed that the denoising problem with Gaussian observation and prior models can be solved by a Sylvester–Lyapunov equation (24) where \mathbf{A} and \mathbf{B} are constant matrices and \mathbf{C} is also a constant matrix depending on \mathbf{Y} . In the other cases (non Gaussian noise and prior) considered in this paper, the solution is also obtained by the Sylvester–Lyapunov equation but the matrix \mathbf{C} is no longer constant.

V. DENOISING WITH EDGE PRESERVING PRIORS

This section extends the previous algorithm to other data models and prior distributions. We consider three data distributions which are often used in medical imaging: Gaussian, Rayleigh, and Poisson. These distributions can be used to deal with CT, ultrasound, PET/SPECT and MRI. We also consider in this section edge preserving priors that will smooth the image in slowly varying regions but preserve the transitions. This type of priors allow a better representation of anatomical details. The priors considered in this paper are: quadratic (L_2), total variation (TV), and Benford (Bfd).

The minimization of $E(\mathbf{X}, \mathbf{Y})$ is performed by solving the equation

$$\nabla_X E(\mathbf{X}, \mathbf{Y}) = 0 \quad (28)$$

which leads to the following set of equations:

$$\sum_{i,j=0}^{N-1,M-1} \frac{d}{dx(k,l)} [d(x(i,j), y(i,j))]$$

TABLE II

GAUSSIAN (L_2), TOTAL VARIATION (TV), AND BENFORD (BFD) PRIORS

Prior Model	$\rho(\delta_v, \delta_h)$	$\frac{\partial}{\partial x(k,l)} E_X(\mathbf{X})$
Gaussian	$g^2(k, l)$	$2N_v[x(k, l) - \bar{x}(k, l)]$
TV	$g(k, l)$	$\frac{N_v}{g(k, l)}[x(k, l) - \bar{x}(k, l)]$
Benford	$\log g(k, l)$	$\frac{N_v}{g^2(k, l)}[x(k, l) - \bar{x}(k, l)]$

$$+ \alpha \underbrace{\sum_{i,j=0}^{N-1,M-1} \frac{\partial}{\partial x(k,l)} [\rho(\delta_v(i,j), \delta_h(i,j))]}_{\frac{\partial}{\partial x(k,l)} E_X(\mathbf{X})} = 0 \quad 0 \leq k, l \leq N-1, M-1 \quad (29)$$

where $(d/dx)d(x, y)$ is defined in Table I for three observation models.

The potential functions, $\rho(\delta_v, \delta_h)$ and the partial derivatives of $E_X(\mathbf{X})$ with respect $x(k, l)$ are listed in Table II where $\bar{x}(k, l)$ is defined in (15) and $g(k, l) = \sqrt{\delta_v^2(k, l) + \delta_h^2(k, l)}$ is the gradient magnitude computed of \mathbf{X} at (k, l) th pixel. In the computation of $(\partial/\partial x(k, l))E_X(\mathbf{X})$, in Table II, it is assumed that $g(k, l) \approx g(k+1, l) \approx g(k, l+1)$.

From Tables I and II, it is concluded that (29) are linear only in the case of Gaussian observation and prior models, as shown in Section IV. In the other cases, the minimization of the energy function $E(\mathbf{X}, \mathbf{Y})$ leads to a huge set of non linear equations, which must be iteratively solved.

This problem will be tackled by using the MM algorithm [62], [63]. The MM algorithm is a powerful tool to deal with non

$$\Theta = \begin{pmatrix} a2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & \cdots & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & \cdots & \cdots & 0 & 0 & -1 & 2 \end{pmatrix} \quad (23)$$

TABLE III

WEIGHTS FOR THE L_2 TV AND BFD PRIORS, WHERE g_{t-1} DENOTES THE GRADIENT MAGNITUDE VALUE COMPUTED IN THE PREVIOUS ITERATION. SINCE WE ARE DEALING WITH ISOTROPIC FIELDS, $\omega_h = \omega_v$

Weights	L_2	TV	Bfd
w	α	α/g_{t-1}	α/g_{t-1}^2

quadratic optimization problems. It amounts to iteratively minimizing a non quadratic energy function by iteratively solving a set of simpler quadratic energy functions, $E_2(\mathbf{X}, \hat{\mathbf{X}}_{t-1})$, depending on the previous estimates, $\hat{\mathbf{X}}_{t-1}$ such that the following relations hold:¹

$$E_2(\hat{\mathbf{X}}_{t-1}, \hat{\mathbf{X}}_{t-1}) = E(\hat{\mathbf{X}}_{t-1}) \quad (30)$$

$$E_2(\mathbf{X}, \hat{\mathbf{X}}_{t-1}) \geq E(\mathbf{X}) \quad (31)$$

for every $\mathbf{X} \in V(\hat{\mathbf{X}}_{t-1})$ where $V(\hat{\mathbf{X}}_{t-1})$ is a neighborhood of $\hat{\mathbf{X}}_{t-1}$. This means that the auxiliary function E_2 must be greater or equal than the true function E in the vicinity of $\hat{\mathbf{X}}_{t-1}$ and both functions must be equal at $\hat{\mathbf{X}}_{t-1}$.

One of the most used MM algorithm is the *reweighted least squares* (RWLS) algorithm [10] that is used in the minimization of functions based on the L_p norm.

Here, we consider a specific MM algorithm proposed in [64], where the quadratic energy function to be optimized in each iteration is

$$\begin{aligned} E(\mathbf{X}, \mathbf{Y}) = & \sum_{i,j=0}^{N-1,M-1} d(x(i,j), y(i,j)) \\ & + \frac{1}{2} \sum_{i,j=0}^{N-1,M-1} [\omega_v(i,j) \delta_v^2(i,j) \\ & + \omega_h(i,j) \delta_h^2(i,j)] \end{aligned} \quad (32)$$

where $d(x, y)$ is defined in (6), $\omega_v(i, j)$ and $\omega_h(i, j)$ are weights updated in each iteration and computed as follows:

$$\omega(\delta) = -\frac{1}{\delta p(\delta)} \frac{dp(\delta)}{d\delta} \quad (33)$$

where ω stands for ω_v and ω_h , and

$$p(\delta) = Ke^{-\alpha\rho(\delta)} \quad (34)$$

is the distribution of δ . Table III shows the weights computed with (33) for the three priors considered, L_2 , TV and Bfd.

The partial derivative of $E(\mathbf{X}, \mathbf{Y})$ w.r.t. $x(k, l)$ computed from (32) lead to

$$\begin{aligned} \frac{\partial E(\mathbf{X}, \mathbf{Y})}{\partial x(k, l)} = & \frac{d}{dx(k, l)} d(x(k, l), y(k, l)) \\ & + [\omega_v(k, l) \delta_v(k, l) + \omega_h(k, l) \delta_h(k, l)] \\ & - [\omega_v(k+1, l) \delta_v(k+1, l) \\ & + \omega_h(k, l+1) \delta_h(k, l+1)] \end{aligned} \quad (35)$$

where $\omega_v(k, l) = \omega_h(k, l) = \omega(k, l)$ because we are using an isotropic MRF, as can be seen from Table III. Assuming the following approximations:

$$\omega_v(k+1, l) \approx \omega_v(k, l) = \omega(k, l) \quad (36)$$

$$\omega_h(k, l+1) \approx \omega_h(k, l) = \omega(k, l) \quad (37)$$

¹The dependence on \mathbf{Y} was dropped for the sake of simplicity

TABLE IV

GRADIENT OF THE DATA ENERGY $\nabla_{\mathbf{X}} E_Y(\mathbf{X}, \mathbf{Y})$ FOR THE GAUSSIAN, POISSON, AND RAYLEIGH MODELS (SEE TABLE I). THE \odot OPERATOR REPRESENTS A ELEMENT WISE OR HADAMMARD MULTIPLICATION

Observation Model	$\nabla_{\mathbf{X}} E_Y(\mathbf{X}, \mathbf{Y})$
Gaussian	$(\mathbf{X} - \mathbf{X}^{ML}) \sigma^{-2}$
Poisson	$(\mathbf{X} - \mathbf{X}^{ML}) \odot \mathbf{X}^{-\odot 1}$
Rayleigh	$(\mathbf{X} - \mathbf{X}^{ML}) \odot \mathbf{X}^{-\odot 2}$

TABLE V

MATRIX Σ FOR THE GAUSSIAN, POISSON, AND RAYLEIGH OBSERVATION MODELS AND FOR THE L_2 TV AND BFD PRIORS. THE \odot OPERATOR REPRESENTS A ELEMENT WISE OR HADAMMARD OPERATION

Σ	Gaussian	Poisson	Rayleigh
L_2	$1/(\alpha\sigma^2)$	$1 \odot /(\alpha\mathbf{X})$	$1 \odot /(\alpha\mathbf{X}^{\odot 2})$
TV	$\mathbf{G}/(\alpha\sigma^2)$	$\mathbf{G} \odot /(\alpha\mathbf{X})$	$\mathbf{G} \odot /(\alpha\mathbf{X}^{\odot 2})$
Bfd	$\mathbf{G}^{\odot 2}/(\alpha\sigma^2)$	$\mathbf{G}^{\odot 2} \odot /(\alpha\mathbf{X})$	$\mathbf{G}^{\odot 2} \odot /(\alpha\mathbf{X}^{\odot 2})$

equation (35) may be rewritten as follows:

$$\begin{aligned} & \frac{\partial E(\mathbf{X}, \mathbf{Y})}{\partial x(k, l)} \\ & = \frac{d}{dx(k, l)} d(x(k, l), y(k, l)) + \omega(k, l) \\ & \quad \times \left[\underbrace{\delta_v(k, l) - \delta_v(k+1, l) + \delta_h(k, l) - \delta_h(k, l+1)}_{4x(k, l) - x(k-1, l) - x(k+1, l) - x(k, l-1) - x(k, l+1)} \right] \end{aligned} \quad (38)$$

and, therefore, the stationary point of $E(X, Y)$ can be computed by solving the following set of equations:

$$\begin{aligned} & \frac{1}{\omega(k, l)} \frac{d}{dx(k, l)} d(x(k, l), y(k, l)) \\ & + \left[\underbrace{2x(k, l) - x(k-1, l) - x(k+1, l)}_{\text{Vertical differences}} \right. \\ & \quad \left. + \underbrace{2x(k, l) - x(k, l-1) - x(k, l+1)}_{\text{Horizontal differences}} \right] = 0 \end{aligned} \quad 0 \leq k, l \leq N, M.$$

The set of (39) can be written in a compact way [see (21) and Table IV]

$$\Sigma \odot (\mathbf{X} - \mathbf{X}^{ML}) + (\Theta_N \mathbf{X} + \mathbf{X} \Theta_M) = 0 \quad (39)$$

where \odot is the Hadamard operator (component wise multiplication), the elements of \mathbf{X}^{ML} are defined in Table I, Θ_N, Θ_M are given by (23) and Σ is a $N \times M$ matrix depending on the observation and prior models, defined in Table V.

Equation (39), is an extension of (21). However, it is a non-linear equation since Σ depends on \mathbf{X} which must be iteratively solved. By using the fixed point method the following recursion is obtained:

$$\Sigma_{t-1} \odot (\mathbf{X}_{t-1} - \mathbf{X}^{ML}) + (\Theta_N \mathbf{X} + \mathbf{X} \Theta_M) = 0 \quad (40)$$

where Σ_{t-1} is computed using the estimate \mathbf{X}_{t-1} obtained in the previous iteration. Equation (40), solved in each iteration, is the well-known Sylvester equation

$$\Theta_N \mathbf{X} + \mathbf{X} \Theta_M + \mathbf{C}_{t-1} = 0 \quad (41)$$

where $\mathbf{C}_{t-1} = \Sigma_{t-1} \odot (\mathbf{X}_{t-1} - \mathbf{X}^{\text{ML}})$.

Let us now discuss the number of solution of this equation. The Sylvester equation, for an arbitrary matrix \mathbf{C}_{t-1} , has a unique solution if and only if $(\lambda_N)_i \neq -(\lambda_M)_j$ for $0 \leq i \leq N-1$ and $0 \leq j \leq M-1$, where $(\lambda_N)_i$ and $(\lambda_M)_j$ are the eigenvalues of Θ_N and Θ_M respectively [65], [66]. These are circulant matrices and the eigenvalues of a $N \times N$ circulant matrix, $\text{circulant}(c_0, c_1, \dots, c_{N-1})$, are given by the *discrete Fourier transform* (DFT) of the first column [10]

$$\lambda_i = \sum_{k=0}^{N-1} c_k e^{-j2\pi i k / N}. \quad (42)$$

Using (23), we obtain

$$(\lambda_L)_i = 2[1 - \cos(2\pi i / L)] \quad (43)$$

with $0 \leq i \leq L-1$ and $L = \{N, M\}$. Therefore, for $i = j = 0$ there are two equal and symmetrical eigenvalues of Θ_N and Θ_M , $(\lambda_N)_0 = (\lambda_M)_0 = 0$, which means that (41) does not have a unique solution. To overcome this difficulty let us rewrite (40) as follows:

$$\Phi_N \mathbf{X} + \mathbf{X} \Phi_M + \mathbf{Q}_{t-1} = 0 \quad (44)$$

where

$$\Phi_N = \frac{\beta}{2} \mathbf{I}_N + \Theta_N \quad (45)$$

$$\Phi_M = \frac{\beta}{2} \mathbf{I}_M + \Theta_M \quad (46)$$

$$\mathbf{Q}_{t-1} = \Sigma_{t-1} \odot (\mathbf{X}_{t-1} - \mathbf{X}^{\text{ML}}) - \beta \mathbf{X}_{t-1} \quad (47)$$

where \mathbf{I}_N and \mathbf{I}_M are $N \times N$ and $M \times M$ identity matrices respectively; β is condition parameter used to improve the stability of the algorithm. In fact, the new matrix Φ_L is also circulant

$$\Phi_L = \text{circulant} \left[\underbrace{\left[\frac{\beta}{2} + 2, -1, \dots, 0, \dots, -1 \right]}_{L \text{ elements}} \right]. \quad (48)$$

The eigenvalues are now

$$(\lambda_L)_i = \frac{\beta}{2} + 2[1 - \cos(2\pi i / L)] \quad (49)$$

with $0 \leq i \leq L-1$. These new eigenvalues are all strictly positive and, therefore, $(\lambda_N)_i \neq -(\lambda_M)_j$, which means that there is a unique solution for (44) in each iteration [65]. The condition number

$$\kappa = \frac{|\lambda_{\max}|}{|\lambda_{\min}|} = \frac{\beta/2 + 4}{\beta/2} = 1 + \frac{8}{\beta} \quad (50)$$

is a measure of the numerical stability of the algorithm.

The proposed iterative algorithm described by equations (44)–(47) has two key features: 1) it does not require the vectorization of images and 2) it is based on the solution of a Sylvester equation in each iteration which can be done by fast algorithms described in the literature.

VI. EXPERIMENTAL RESULTS

This section presents three experiments comparing the performance of the proposed algorithm with other methods. The methods considered in this study are: i) median filter, ii) Wiener

filter, iii) *squeeze box filter* (SBF) [30] iv) platelets [67], the method proposed in this paper using v) L_2 , vi) TV, and vii) Bfd priors. The first set of experiments (Section VI-A) comprises a Monte Carlo (MC) test in two scenarios, using synthetic image corrupted with Gaussian noise, and with Rayleigh noise distributions. In the case of synthetic images corrupted with Rayleigh noise (where we try to simulate the noise present in medical images), each noisy pixel is obtained by using a random number generator, $y(i, j) = R(x(i, j))$, where x is the original noiseless pixel and $p(y) = (y/x)e^{-y^2/2x}$. The algorithms are evaluated by MC tests using the SNR criterion. The second set of experiments (Section VI-B), presents a comparison of the methods in two scenarios: *ultrasound* (US) images (Rayleigh distributed) and *fluorescence confocal microscopy* (FCM) images (Poisson distributed). We also present the image profiles in three image modalities: US, MRI, and FCM. Finally, in the third set of experiments (Section VI-C), we evaluate the performance of the methods in a tracking task using two ultrasound sequences of the LV. The first sequence contains 490 frames and the second one contains 470 frames (corresponding to 26 and 19 cardiac cycles, respectively). The tracking is done using the *shape-probabilistic data association filter* (S-PDAF) recently proposed in [68]. To evaluate the performance of the algorithms, we compare the contour estimates provided by the tracker with the ground truth (reference contours). This comparison is done as follows: we selected four images from each cardiac cycle (two images in the systole phase and two images in the diastole phase) and asked to the user to manually define the LV contour for each of these images. For the first sequence, the user segmented 72 images: 36 images extracted during the systole phase and other 36 images during the diastole phase. For the second sequence, the user segmented 78 images: 39 images in systole phase, and 39 images in diastole phase. The selection of the frames was random within each cycle. Three metrics are used in these tests: i) Hausdorff distance, ii) average distance, and iii) Hamoude distance [69]. The Hausdorff and Hamoude distances are metrics since they verify all the axioms of a metric. However, the average distance is not a metric since is not symmetric. The Hamoude distance is a normalized version of the Hamming distance (see Section IV-C) which is widely used to compare binary strings and images in information theory. The formulated methodology allows us to obtain an objective evaluation of the contours accuracy provided by the methods considered herein.

A. Synthetic Images

To assess the performance of the above mentioned methods, we performed MC tests in two scenarios using synthetic image of a gray square in a black background. In the first scenario the image is corrupted with i) Gaussian noise, in the second with ii) Rayleigh noise distribution. In the first set of experiment, the σ parameter was varied in the set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$, and we carried out 20 experiments for each σ value. In the case of Rayleigh noise we consider $p(y|x) = (y/\eta x) \exp((-y^2/2\eta x))$ the η parameter varied in the interval $\{10, 100, 500, 1000, 5000\}$. Fig. 3 shows the mean of SNR for the different values of σ and η . The SNR is computed along the main diagonal of the image.

It can be seen for Gaussian noise [Fig. 3(a)] that for small values of σ , the Wiener method exhibits the best SNR. However, it strongly degrades as σ increases. The same can be said about the median filtering. It also has a poor performance for higher

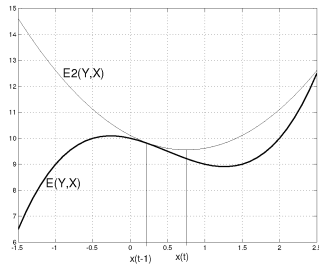
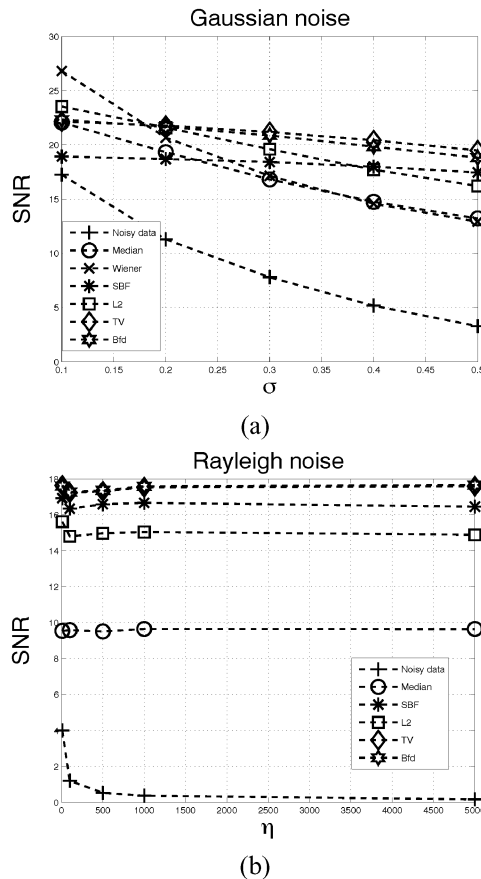


Fig. 2. Majorize/minimize (MM) optimization.

Fig. 3. Evolution of SNR as a function of parameter σ and η : (a) Gaussian, (b) Rayleigh noise distributions.

values of σ . The SBF has a good and almost constant SNR, and the best SNR is achieved by the Sylvester algorithm with TV and Benford priors. Fig. 3(b) reports the performance of the methods with Rayleigh noise. The SBF and TV, Benford priors exhibit remarkable results. As in the previous experiment, the TV and Bfd prior exhibit the best results.

A second experiment is shown in Fig. 4. In this study, it was created a synthetic phantom of an ultrasound image with the Field II simulation program (see [70] for details) that simulates the ultrasound image generation process where the noise is multiplicative, in the sense that the variance is space dependent. The goal is to compare the techniques in a image corrupted with Rayleigh noise.

Fig. 4 shows the simulation of an US image and the output of the various despeckling algorithms.² The phantom consists of a background region class with a “one” pixel value, and other two classes with “zero” (dark region) and “ten” pixel values (bright region). Each foreground object consists of a two sets of five bright disks with increasing (left) and decreasing radius (right), respectively [from the top to the bottom image, see left image in Fig. 4(a)].

The original size of the image is 500×390 . However, a resize of this image to 512×256 was needed, due to the constraints imposed by the method proposed in [67].

We compute the SNR along a vertical line passing through the centers of the left bright circles. In this study the comparison comprises the following methods: i) median filter, ii) Wiener, iii) SBF [30] and the proposed techniques, iv) L_2 , v) TV, and vi) Bfd. In this comparison, the platelets method is not presented since it is not tailored to deal with Rayleigh noise.

Fig. 4 shows the output, as well as the intensity profile image for each method. The red line is the true profile and the blue line is the image profile after denoising.

In this experiment, five iterations were used for the proposed techniques (L_2). The average time required to denoise the image was:

- SBF : 6.05 s;
- L_2 : 7.26 s;
- TV: 7.60 s;
- Bfd: 7.55 s.

From Fig. 4, it should be stressed that, although the TV and Bfd priors provide competitive results, the SBF method provides the best SNR, and an excellent output with sharp transitions [see Fig. 4(d)].

Fig. 5 show the results when the image is corrupted with Poisson noise. In this experiment, we integrate the Platelets method proposed in [67], which can be a basis for comparison for Poisson denoising (see [71] for a comprehensive description). Thus, the comparison comprises the following methods: i) Platelets [67] and the proposed techniques, ii) L_2 , iii) TV, iv) Bfd. Here, we do not include the Wiener as well as median filters, since the results are very poor. In this comparison the SBF method is not presented since it is not tailored to deal with Poisson noise.

We used 20 iterations in the platelets method [67] and 5 iterations in the proposed methods as in the previous experiment. iterations were used, in the proposed methods we maintained five iterations as in the previous experiment. The average time per iteration was the following:

- Platelets : 23.90 s;
- L_2 : 7.77 s;
- TV: 8.03 s;
- Bfd: 7.94 s.

Fig. 5 shows that the proposed techniques outperform the platelets, exhibiting attractive SNR as well as the processing time to denoise the image.

B. Still Images

This section evaluates the denoising algorithms in US, MRI and Fluorescence Confocal Microscopy (FCM) images. Three examples are presented (see Fig. 6): an US image of the carotid, a MRI image of a knee and a FCM of a cell. For comparison purposes, we selected the appropriate method (which depends

²This test is similar to the one presented in [30].

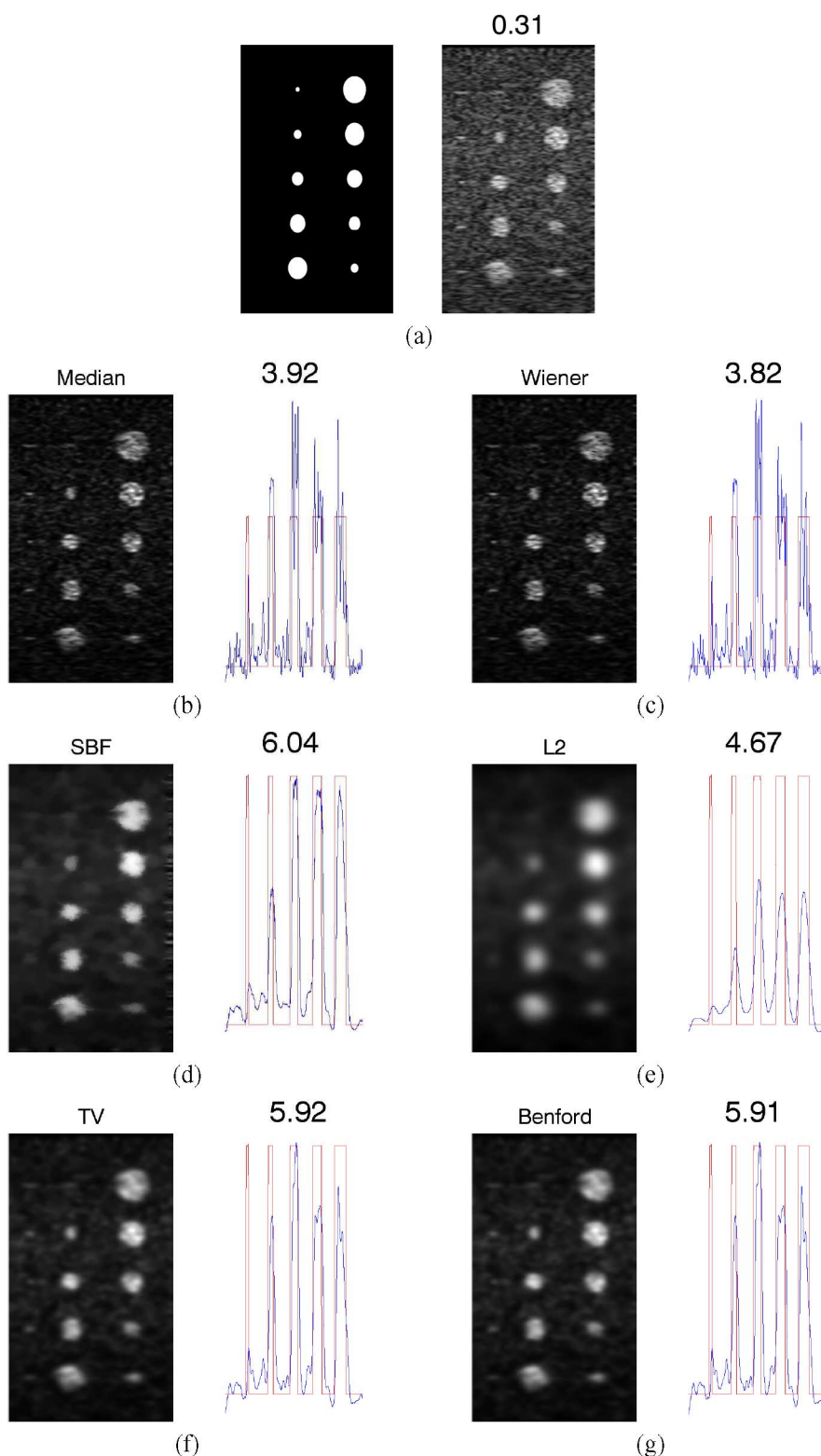


Fig. 4. Ideal profile (square wave in red) and estimated profile obtained with various methods (in blue): (a) original image with size 512×256 , (b) median—SNR = 3.93, (c) Wiener—SNR = 3.83, (d) SBF—SNR = 6.05, (e) L_2 —SNR = 4.67, (f) TV—SNR = 5.92, (g) Bfd—SNR = 5.92.

on the image modality) and the framework herein proposed (L_2 , TV and Bfd priors). Thus, we compare the framework with: i) the SBF method (carotid image); ii) Wiener filter (knee image); iii) Platelets method (cell image).

Fig. 7 shows the denoising results for a US image of carotid where the noise was modeled by a Rayleigh distribution (fourth

column in Table V). The Bayesian estimates presented in this study were obtained using a Rayleigh data distribution which accounts for multiplicative noise and the three priors. The comparison of the results obtained by several methods shows that the best results are obtained by the TV and Bfd priors as well as the SBF, which is well suited to deal with US images. This is

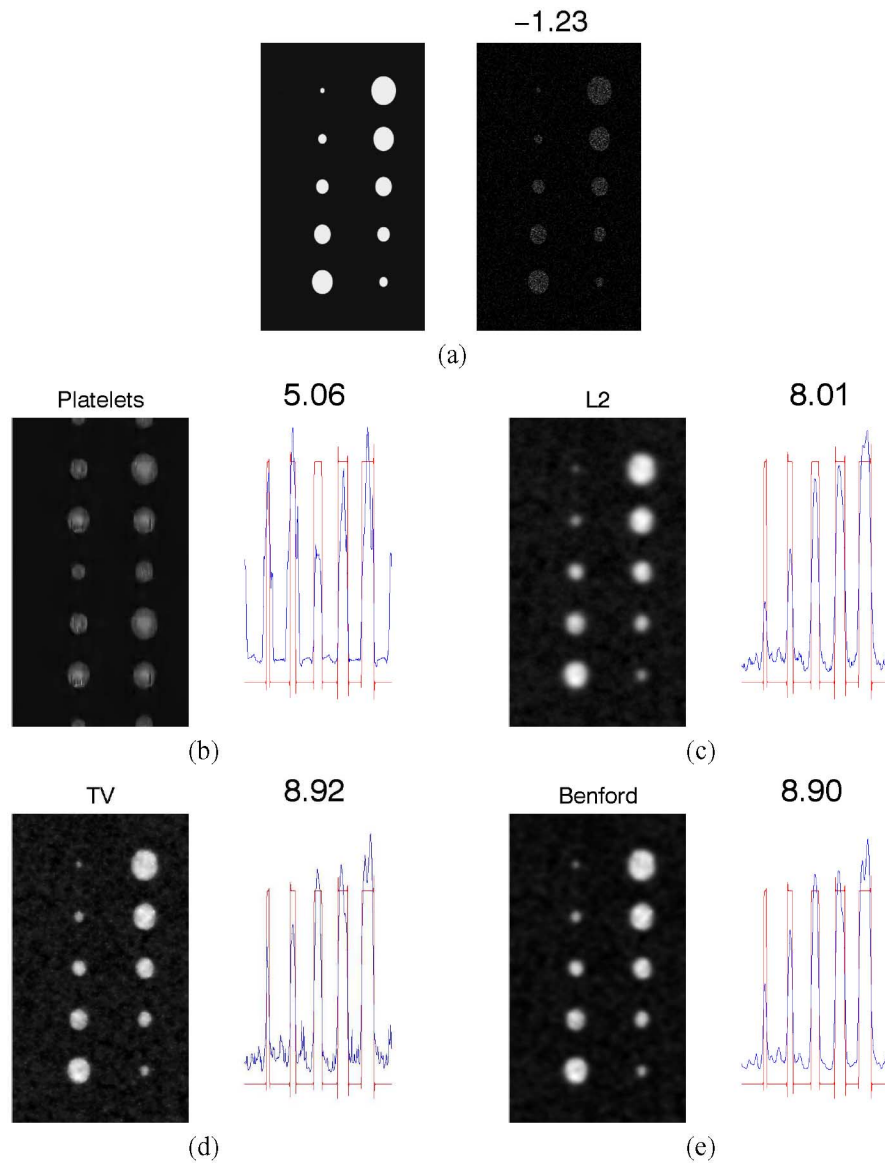


Fig. 5. Ideal profile (square wave in red) and estimated profile obtained with various methods (in blue): (a) original image with size 512×256 , (b) platelets—SNR = 5.06, (c) L_2 —SNR = 8.01, (d) TV—SNR = 8.93, (e) Bfd—SNR = 8.90.

illustrated in Fig. 7, which shows the image profiles produced by each method (blue line).

Fig. 8 shows the results obtained with MRI image. In this case, the noise is modeled as being white, additive and Gaussian (second column in Table V). It is desirable to compare the results with the Wiener method since it is a benchmark when dealing with Gaussian noise. It can be seen from the figure that the framework methods perform much better than the Wiener filtering.

Fig. 9 shows the results obtained with FCM image. The three proposed methods were implemented using a non Gaussian observation model (Poisson distributed). Once again, the methods proposed in this paper outperforms the platelets, exhibiting remarkable results concerning the obtained image profiles.

C. Heart Sequences

The denoising algorithm proposed in this paper was evaluated in the context of heart tracking in ultrasound images. The ultrasound sequences were obtained using an ultrasound probe

operating at 1.7 MHz and were sampled at 15 frames/s^{-1} . Two sequences of the heart cavities were used in this study with 490 and 470 frames, corresponding to 26 and 19 cardiac cycles, respectively. The size of each frame is 200×248 pixels for both sequences. This study aims to track the left ventricle during the whole sequences. Fig. 10 shows a frame from both sequences. These images present a poor quality (low SNR) and the multiplicative *speckle* noise which corrupts these images is modeled by a *Rayleigh* distribution. Furthermore, the boundaries of the LV can hardly be seen when they suffer a sudden motion. This happens mainly in diastole phase.

The test was done as follows. The ultrasound sequences were first processed by each of the denoising algorithms: i) median filtering, ii) squeeze box filter [30], and the proposed method with iii) L_2 norm, iv) total variation, and v) Benford priors. The last three methods were implemented using a non Gaussian observation model (Rayleigh). A tracking algorithm was then applied to the denoised images to estimate the boundary of the left ventricle during the whole sequences. The tracker used in this

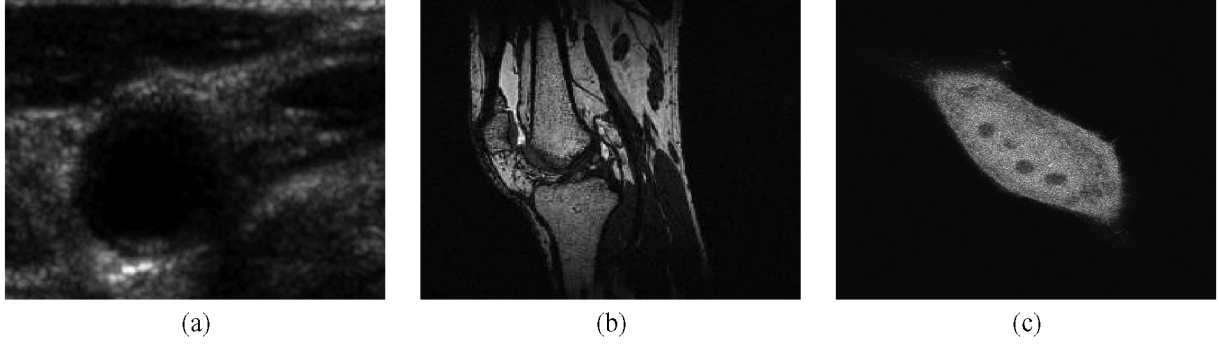


Fig. 6. Sizes and corresponding images used in this study: (a) 440×346 US carotid image, (b) 256×256 MRI knee image, (c) 512×512 cell FCM image.

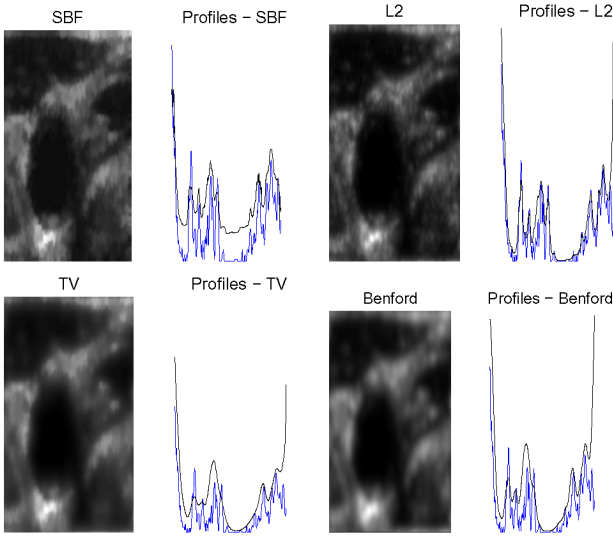


Fig. 7. Image output and resulting image profiles (along diagonal line) of the methods: (blue) profile of the original image, (black) profile of the estimated image. Top row (from left to right): SBF, L2 norm. Bottom row (from left to right): TV, Bfd priors.

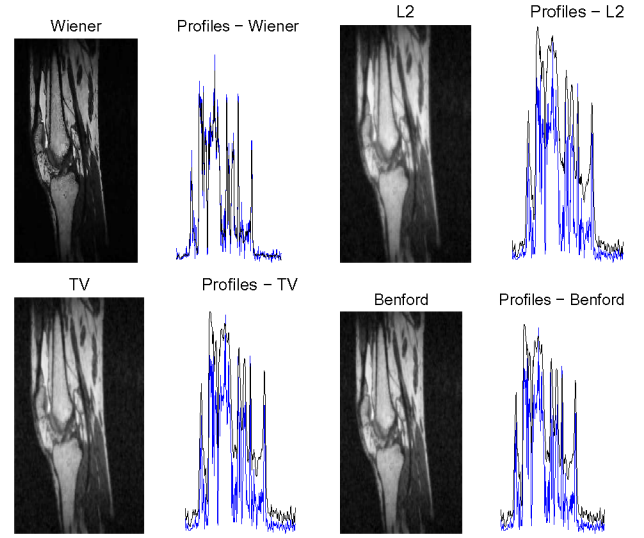


Fig. 8. Image output and resulting image profiles (along diagonal line) of the methods: (blue) profile of the original image, (black) profile of the estimated image. Top row (from left to right): Wiener filtering, L2 norm. Bottom row (from left to right): TV, Bfd priors.

study is the S-PDAF tracker proposed by the authors in [68]. This tracker is able to discard spurious edges since it relies on a robust association of the detected features (intensity transitions) with the contour. This is specially important since multiple detections are expected. The output of the tracker for each of the denoising conditions was then compared with the ground truth contours of the left ventricle.

Three metrics were used in this test to compare the output of the tracker with the reference contours defined by the user: the Hausdorff distance, the average distance and the Hamoude metric. We will briefly define each of them.

Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_x}\}$, and $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_y}\}$, be two sets of points obtained by sampling the estimated contour and the reference contour. We define the distance of \mathbf{x}_i to the curve \mathcal{Y} as the distance from \mathbf{x}_i to the closest point of \mathcal{Y}

$$d(\mathbf{x}_i, \mathcal{Y}) = \min_j \|\mathbf{y}_j - \mathbf{x}_i\|. \quad (51)$$

This is denoted as distance to the closest point (DCP).

The average distance between the sets \mathcal{X} , \mathcal{Y} is defined as

$$d_{av} = \frac{1}{N_x} \sum_{i=1}^{N_x} d(\mathbf{x}_i, \mathcal{Y}) \quad (52)$$

where N_x is the length of the \mathcal{X} and the Hausdorff distance between both sets is defined as the maximum of the DCP's between the two curves [72]

$$d_{\max}(\mathcal{X}, \mathcal{Y}) = \max(\max_i \{d(\mathbf{x}_i, \mathcal{Y})\}, \max_j \{d(\mathbf{y}_j, \mathcal{X})\}). \quad (53)$$

We have also Hamoude metric [69]. Let $R_{\mathcal{X}}$, $R_{\mathcal{Y}}$ be the image regions inside the two contours. We compute the number of points which belongs to only one of these regions (e.g., obtained by pixel-wise XOR operation) and normalize it by the number of points of the union of both regions

$$d_H = \frac{\#((R_{\mathcal{X}} \cup R_{\mathcal{Y}}) - (R_{\mathcal{X}} \cap R_{\mathcal{Y}}))}{\#(R_{\mathcal{X}} \cup R_{\mathcal{Y}})}. \quad (54)$$

Fig. 11 shows the features detected in the ultrasound image after the denoising operation. These features are obtained by detecting intensity transitions along directions orthogonal to the best available estimate of the contour. Details can be found in [68].

Fig. 12 shows a frame processed by each of the denoising techniques, as well as the intensity profile on the image line

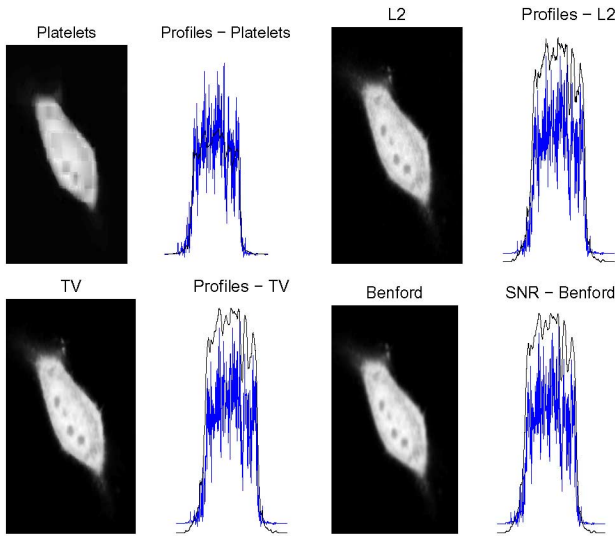


Fig. 9. Image output and resulting image profiles (along diagonal line) of the methods: (blue) profile of the original image, (black) profile of the estimated image. Top row (from left to right): platelets, L_2 norm. Bottom row (from left to right): TV, Bfd priors.

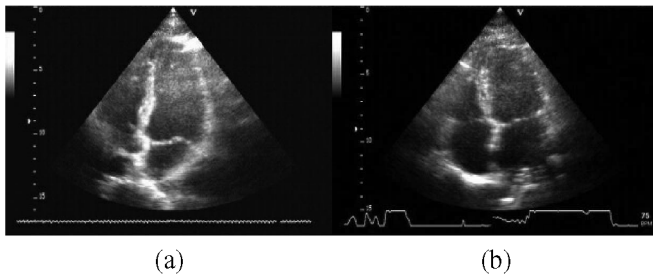


Fig. 10. Ultrasound images of the left ventricle extracted from two test sequences: (a) first sequence, (b) second sequence.

[see the white line in Fig. 12(a)] locations used to perform the comparison.

Table VI shows the performance of the S-PDAF tracker using the: Average distance, Hausdorff distance, and the Hamoude metric. It is concluded that good results are obtained with all the denoising algorithms. The best results in sequence 1 are achieved by the proposed algorithm with the Benford prior. The best results in sequence 2 are obtained by the proposed algorithm with the total variation prior and by the SBF method.

From this paper, we conclude that the TV and Bfd prior proposed in this framework compare well with all the other denoising techniques.

Table VII shows the average time per frame needed to denoise the images from both sequences in a Matlab implementation on a Pentium 2-GHz CPU. It is seen from this table that the median filtering is the fastest method, whilst the TV and Bfd prior are slower since they are the output of an iterative algorithm (six iterations were used).

Table VIII displays the percentage outlier features (i.e., features that do not belong to the left ventricle boundary) detected by the tracker (the results listed in *first sequence* row of this table are obtained with the data sequence displayed in Fig. 11). From Table VIII, we see that the first sequence is noisier than the second sequence (all the methods provide higher number of outliers in the first sequence). The L_2 norm provides less outliers

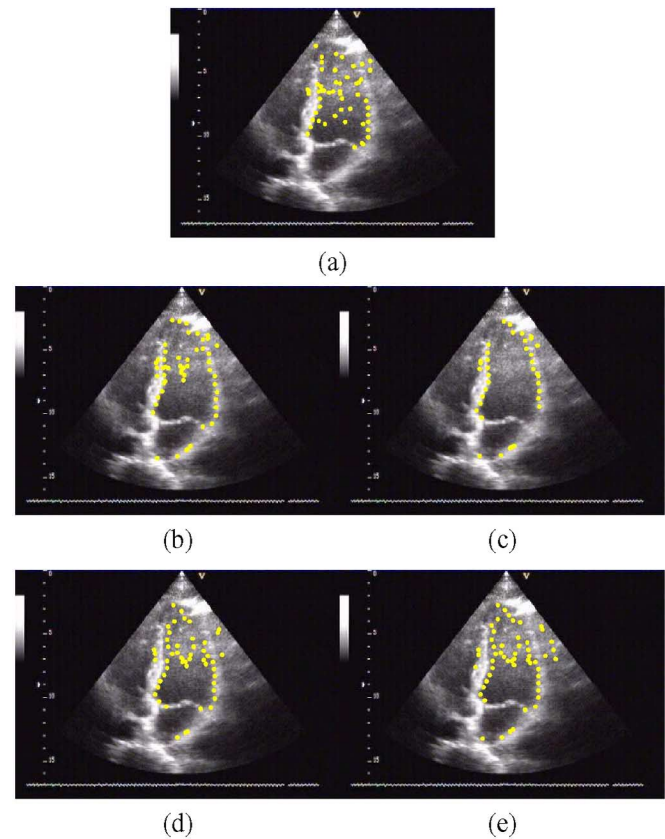


Fig. 11. Outliers detected by the tracker in one frame of the first sequence: (a) median filtering, (b) SBF, (c) L_2 norm, (d) TV, (e) Bfd.

for both sequences than the remaining methods. The Median filter provides the worst result, the TV and Bfd are best ones and seems to be equivalent in the first sequence. This is ascertained in Fig. 11 where one frame of the first sequence is displayed, the yellow dots are the features detected by the tracker, most of them located outside (i.e., inner/outer regions of the contour) from the left ventricle boundary. The median filtering provides the higher number of outliers. The L_2 prior is the best method not providing inner features. The TV and Bfd priors, appear to be, once again, similar to each other.

Table IX, shows the total and average time (in seconds) needed to track the whole sequences. It is seen that the tracker with the L_2 norm provides the fastest processing time. This happens due to the reasons already mentioned (i.e., small number of outlier features). From this table, we stress that all the methods allow a much faster tracking process when compared with the median filtering. Recall that, with the denoising methods herein proposed, we gain 3 to 4 min to track the LV contour, which is an expressive difference.

VII. CONCLUSION

Bayesian methods rely on the minimization of an energy function with two terms: a data dependent term and a regularization term. If the image is corrupted by Gaussian additive noise and the prior is Gaussian, the MAP estimate is obtained by solving a system of linear equations with MN equations. This system has a huge dimension which prevents the computational direct solution by matrix inversion and leads to the use of numeric algorithms.

TABLE VI
METRICS MEAN VALUES FOR THE FIRST AND SECOND SEQUENCES

		Median	SBF	L_2	TV	Bfd
Hammoude	1st sequence	0.20	0.18	0.18	0.19	0.17
	2nd sequence	0.25	0.18	0.22	0.19	0.20
Average	1st sequence	4.66	4.32	4.20	4.19	3.81
	2nd sequence	4.58	3.64	4.33	3.57	3.74
Hausdorff	1st sequence	13.74	13.25	12.83	12.78	11.33
	2nd sequence	12.17	9.70	11.16	9.47	9.95

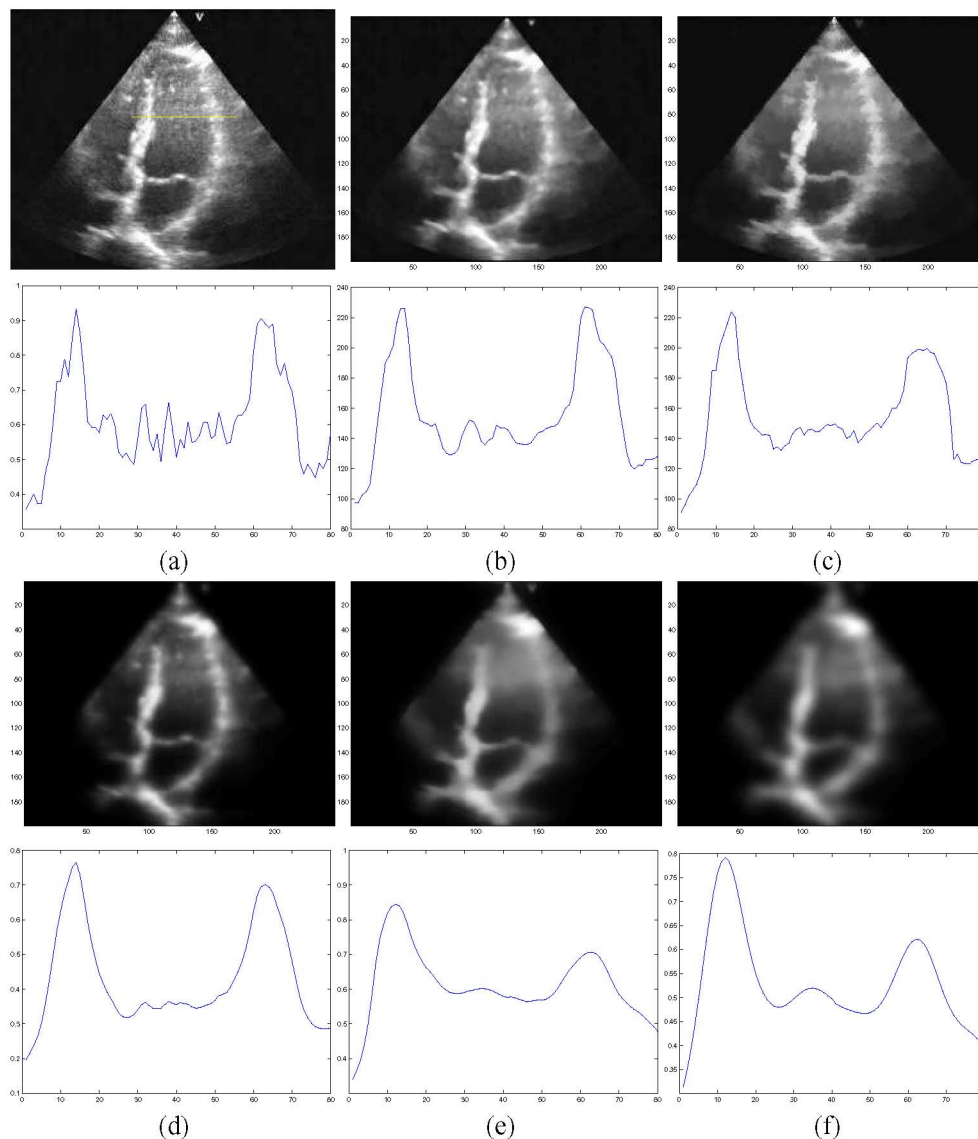


Fig. 12. Denoising images with the corresponding intensity profile. The yellow line is the location in which the intensity profile is computed for each of the methods: (a) real image, (b) median filtering, (c) SBF (d) L_2 norm, (e) TV , (f) Bfd .

This paper shows that the solution of the denoising problem with Gaussian noise and prior can be obtained by solving the Sylvester–Lyapunov matrix equation. This equation avoids the use of matrices with huge dimensions and can be efficiently solved using standard numerical algorithms. In the case of non Gaussian noise (e.g., multiplicative noise) and non Gaussian priors (e.g., total variation or Benford) the same approach can still be used. However, the coefficients of the Sylvester–Lya-

punov equation are no longer constant and depend on the denoised image. This difficulty is solved by recursively solving the Sylvester–Lyapunov equation using a fixed point algorithm.

Experimental results using several priors and noise models are presented to evaluate the performance of the proposed algorithm. A comparison with state of the art denoising methods is also included. All the methods were applied to ultrasound, CT and MRI images which have different types of noise statistics.

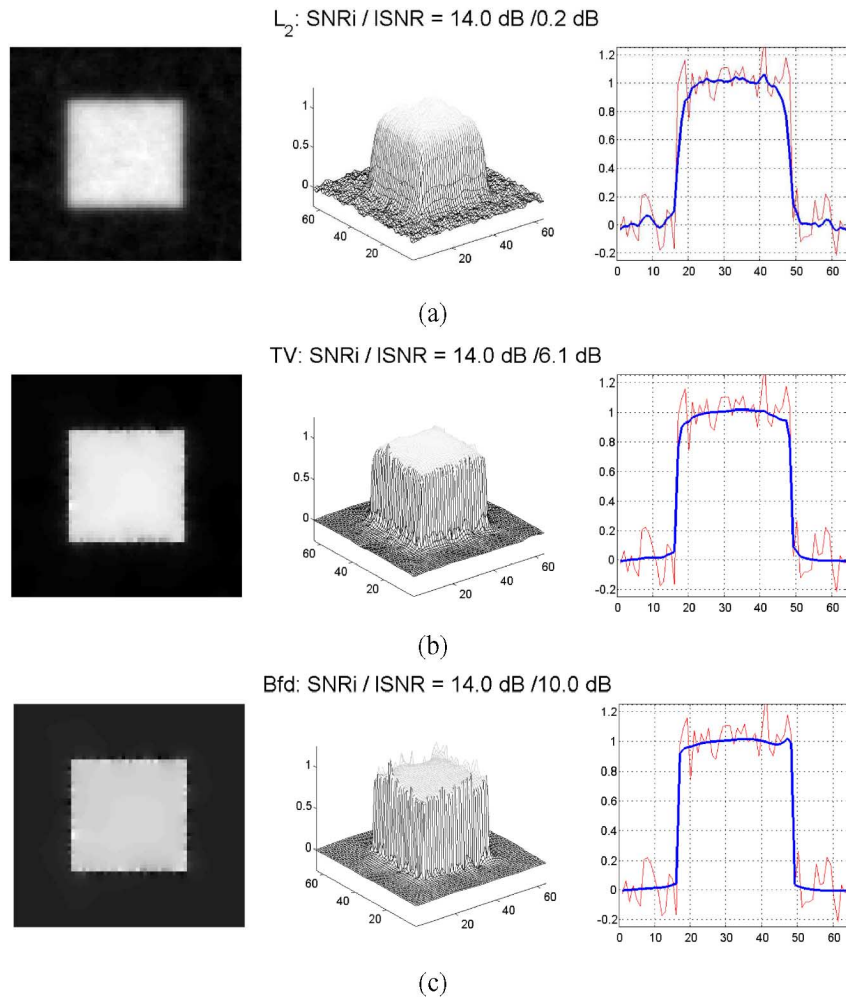


Fig. 13. Filtered images and the corresponding profiles taken at diagonal line of the image. (a) L_2 norm, (b) TV, and (c) Bfd prior. Also shown is the initial SNR (SNRi) and the corresponding improvement that was achieved (ISNR).

TABLE VII
AVERAGE TIME PER FRAME (VALUES IN SECONDS) TAKEN TO
DENOISE THE SEQUENCES FOR ALL METHODS

	Median	SBF	L_2	TV	Bfd
1st sequence	0.10	2.49	2.25	5.93	6.25
2nd sequence	0.10	2.56	2.38	5.90	6.22

TABLE VIII
PERCENTAGE OF OUTLIER FEATURES DETECTED BY
THE TRACKER FOR BOTH SEQUENCES

	Median	L_2	TV	Bfd	SBF
1st sequence	56	8	15	16	19
2nd sequence	11	5	11	6	12

TABLE IX
AVERAGE TIME SPENT BY THE S-PDAF TRACKER (VALUES IN SECONDS):
TOTAL TIME TAKEN TO TRACK THE CONTOUR AND AVERAGE TIME PER FRAME

	Median	SBF	L_2	TV	Bfd
1st sequence	1.19	0.76	0.69	0.76	0.79
2nd sequence	0.75	0.69	0.66	0.66	0.72

The Sylvester–Lyapunov algorithm proposed in the paper manages to achieve competitive results with the best tested methods in a wide variety of experimental conditions.

APPENDIX

A. Edge Preserving Priors

The Gaussian prior used in the Section IV, with quadratic potential functions, allows an efficient elimination of the noise assigning a quadratic cost to the differences between neighboring pixels. However, it also assigns a quadratic cost to transitions. Therefore, the MAP images, obtained with quadratic potential have smooth transitions which lead to a loss of anatomical details. This effect is shown in Fig. 13 where a synthetic square image is created with a zero background, and unit square foreground. This image is corrupted with additive white Gaussian noise (AWGN) with distribution $\mathcal{N}(0, 0.25^2)$. The edges of the filtered image appear smoothed as shown in Fig. 13(a), while in the case of edge preserving priors, TV and Bfd, shown in Fig. 13(b) and (c), respectively, the edges are preserved (less distorted).

To overcome this difficulty, other prior functions were proposed in the literature [73]. These priors present a quadratic behavior for small differences and a sub-quadratic behavior for

Algorithm 1 Unified Framework

Initialization:

$$\Theta_N \leftarrow \phi_N^T \phi_N;$$

$$\Theta_M \leftarrow \phi_M^T \phi_M;$$

$$\alpha_{\text{end}} \leftarrow \alpha;$$

$$\alpha \leftarrow \alpha_{\text{begin}};$$

Selection of the matrix Σ :

Σ	Gaussian	Poisson	Rayleigh
L_2	$1/(\alpha\sigma^2)$	$1 \odot /(\alpha\mathbf{X})$	$1 \odot /(\alpha\mathbf{X}^{\odot 2})$
TV	$\mathbf{G}/(\alpha\sigma^2)$	$\mathbf{G} \odot /(\alpha\mathbf{X})$	$\mathbf{G} \odot /(\alpha\mathbf{X}^{\odot 2})$
Bfd	$\mathbf{G}^{\odot 2}/(\alpha\sigma^2)$	$\mathbf{G}^{\odot 2} \odot /(\alpha\mathbf{X})$	$\mathbf{G}^{\odot 2} \odot /(\alpha\mathbf{X}^{\odot 2})$

Select one line and one column from the above Table according to the *Prior* and *Observation Model* respectively to obtain Σ .

for $t = 2$ to N **do**

Compute:

$$\Phi_N \leftarrow \frac{\beta}{2} \mathbf{I}_N + \Theta_N;$$

$$\Phi_M \leftarrow \frac{\beta}{2} \mathbf{I}_M + \Theta_M;$$

$$\mathbf{Q}^{t-1} \leftarrow \Sigma^{t-1} \odot (\mathbf{X}^{t-1} - \mathbf{X}^{\text{ML}}) - \beta \mathbf{X}^{t-1};$$

Solve:

$$\Phi_N \mathbf{X} + \mathbf{X} \Phi_M + \mathbf{Q}^{t-1} = 0;$$

$$\alpha \leftarrow \alpha + \frac{1}{2}(\alpha_{\text{end}} - \alpha_{\text{begin}});$$

end for

higher differences. It is assumed that higher differences are due to transitions and should not receive a strong penalty while small differences that are associated to noise should be attenuated.

Table X shows several edge preserving potential functions, $\rho(x)$, and the corresponding influence functions, $\psi(x) = d\rho(x)/dx$ [56] and Fig. 14 shows their graphical representations as well as the quadratic potential function.

The displayed potential functions are the L_1 norm, the L_p norm, also known as *Generalized Gaussian Markov Random Fields* (GGMRF) [74], [75], the Geman and McClure [55], the Lorentzian, the Huber [73], and the Benford [76] potential functions. Almost all of these potential functions present a quadratic behavior for small values of x and a sub-quadratic behavior for higher values of x . In the case of L_1 and Huber potential functions, the asymptotic behavior is linear, for the case of the Lorentzian and Benford the asymptotic behavior is logarithmic, in the case of Geman and McClure the asymptotic behavior is a constant value and finally in the case of L_p (GGMRF) prior the asymptotic behavior is something between the L_1 and L_2 priors. For all priors, the asymptotic behavior is sub-quadratic which means that large differences among neighboring nodes are not so penalized as in the case of the quadratic prior L_2 .

Each potential function, described above, depends only on one first order difference. When applied to image processing, all first order vertical and horizontal differences are independently considered which means that the cliques are pairs of neighboring pixels [54]. More complex edge preserving poten-

tial functions, called here *adaptive potential functions* (APF), have been recently used to cope with higher order cliques. One example is the TV of the discrete image \mathbf{X} [27], [75], [77], defined by

$$U(\mathbf{X}) = \alpha \sum_{i,j=0}^{N-1,M-1} g(i,j) \quad (55)$$

where $g(i,j) = \sqrt{\delta_v^2(i,j) + \delta_h^2(i,j)}$ is a discrete approximation of the gradient magnitude at pixel (i,j) .

$U(\mathbf{X})$ is a discrete version of the TV of a continuous scalar function, $\mathcal{X}(x) : \Omega \rightarrow \mathbb{R}$ with $\Omega \subset \mathbb{R}^2$, defined as follows:

$$TV(\mathcal{X}(x)) = \int_{\Omega \subset \mathbb{R}^2} |\nabla \mathcal{X}(x)| dx. \quad (56)$$

Each potential function in (55) depends on both horizontal, $\delta_h(i,j)$ and vertical, $\delta_v(i,j)$ differences.

These potential functions [see (57), shown at the bottom of the page], have higher order than the previous ones because they involve two interactions with the neighbors: one with the nearest neighbor in the horizontal direction and the other with the nearest neighbor in the vertical direction.

To compare the total variation prior with the previous potential functions, listed in Table X, let us consider that one of the differences is constant. The potential function associated to the other can be written as follows:

$$\rho(x) = \sqrt{x^2 + \delta} \quad (58)$$

$$g(i,j) = \sqrt{\underbrace{(x(i,j) - x(i-1,j))^2}_{\delta_h} + \underbrace{(x(i,j) - x(i,j-1))^2}_{\delta_v}} \quad (57)$$

TABLE X
COMMON EDGE PRESERVING PRIORS (EXCLUDING L_2)

	Potential - $\rho(x)$	Influence - $\psi(x) = d\rho(x)/dx$
L_2	x^2	$2x$
Edge preserving priors		
L_1	$ x $	$\text{sign}(x)$
L_p (GGMRF)	$ x ^p$	$\text{sign}(x)p x ^{p-1}$
Geman and McClure	$\frac{x^2}{1+x^2}$	$\frac{2x}{(1+x^2)^2}$
Lorentzian	$\log(1 + \frac{1}{2}(\frac{x}{\sigma})^2)$	$\frac{x/\sigma}{1 + \frac{1}{2}(\frac{x}{\sigma})^2}$
Huber	$x^2/2\epsilon + \epsilon/2, x \leq \epsilon$ $ x , x > \epsilon.$	$x/\epsilon, x \leq \epsilon$ $\text{sign}(x), x > \epsilon.$
Benford	$\log(x + \epsilon)$	$\text{sign}(x)/(x + \epsilon)$

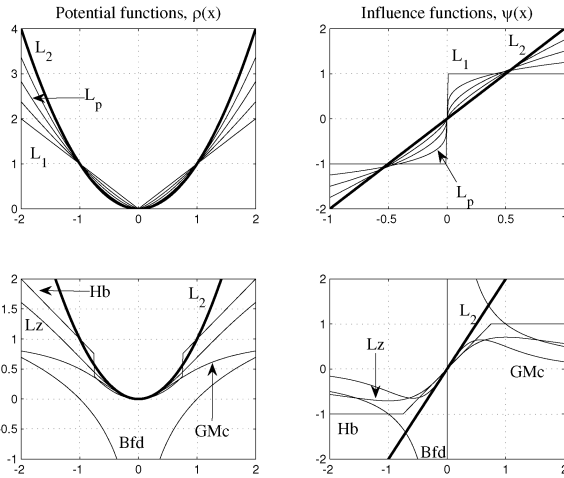


Fig. 14. Edge preserving potential functions.

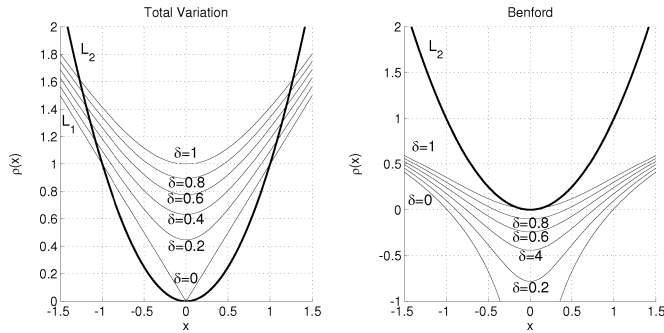


Fig. 15. TV—Adaptive priors, (left) TV, and (right) Benford.

which is a function depending on the space varying parameter δ , the other interaction involved in the clique. Fig. 15(a) displays this function for several values of δ , as well as the quadratic potential function. As before, this function presents a sub-quadratic behavior, typical in edge preserving potential functions, depending on the other difference considered as constant. This means that a penalization in one direction (for instance vertical direction, δ_v) depends on the first order difference on the other direction (for instance horizontal direction, δ_h) and vice versa.

In [76], a new adaptive potential function is proposed based on the Benford law, with the following energy function:

$$U(\mathbf{X}) = \sum_{i,j=0}^{N-1, M-1} \log[g(i, j)] \quad (59)$$

where $g(i, j)$, as previously, is the gradient magnitude at pixel (i, j) . This new prior is derived by assuming the existence of a natural prior for real life images [78] and specially for medical images [76].³ Fig. 15(b) displays this APF for several values of δ and compare the respective curves with the quadratic potential function. Once again these potential functions present a desirable sub-quadratic behavior for edge preserving priors.

B. Pseudo-Code Implementation

ACKNOWLEDGMENT

The authors would like to thank Prof. P. C. Tay, Prof. S. T. Acton, and Prof. J. A. Hossack for providing the source code of the SBF method proposed in [30]. They would also like to thank Prof. R. Willett and Prof. R. Nowak for providing the source code of the platelets proposed in [67], which was helpful to extend the comparison of the techniques presented in the paper. The ultrasound images were kindly provided by Prof. F. Pinto from the *Instituto Cardiovascular de Lisboa* (ICVL).

REFERENCES

- [1] J. Stoitsis, S. Golemati, K. S. Nikita, and A. N. Nicolaides, "Characterization of carotid atherosclerosis based on motion and texture features and clustering using fuzzy c-means," presented at the 26th Int. Conf. IEEE EMBS, 2004.
- [2] C. I. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 902–912, Jul. 2003.
- [3] J. Bleck, U. Ranft, M. Gebel, H. Hecker, M. Westhoff-Beck, C. Thiesemann, S. Wagner, and M. Manns, "Random field models in textural analysis of ultrasonic images of the liver," *IEEE Trans. Med. Imag.*, vol. 15, no. 6, pp. 796–801, Dec. 1996.
- [4] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.
- [5] J. L. P. Dzung, L. Pham, and C. Xu, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, 2000.
- [6] A. Hua and L. Yezzi, "Vessels as 4D curves: Global minimal 4D paths to extract 3D tubular surfaces," presented at the Int. Conf. Computer Vision and Pattern Recognition Workshop, Jun. 2006.
- [7] T.-S. Lee, W. Segars, and B. Tsui, "Study of parameters characterizing space-time gibbs priors for 4D map-rbi-em in gated myocardial perfusion spect," in *Proc. IEEE Nuclear Science Symp. Conf. Record*, 2005, vol. 4, pp. 2124–2128.
- [8] T. Deschamps, R. Malladi, and I. Ravve, "Fast evolution of image manifolds and application to filtering and segmentation in 3D medical images," *IEEE Trans. Vis. Comput. Graph.*, vol. 10, pp. 525–535, 2004.
- [9] J. Montagnat and H. Delingette, "Space and time shape constrained deformable surfaces for 4D medical image segmentation," in *Proc. 3rd Int. Conf. Med. Imag. Comput. and Computer-Assisted Intervention*, 2000.
- [10] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, 2000.

³In practice, we use the potential function $\rho(\delta_h, \delta_v) = \log[\sqrt{\delta_h^2 + \delta_v^2} + \epsilon]$ where ϵ is a small constant to guarantee that the argument is strictly positive.

- [11] A. K. Jain, *Fundamentals of Digital Image Processing*, 1989.
- [12] A. Toprak and I. Gler, "Suppression of impulse noise in medical images with the use of fuzzy adaptive median filter," *J. Med. Syst.*, vol. 30, no. 6, 2006.
- [13] F. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 15, pp. 246–254, 1948.
- [14] J. Abbot and F. Thurstone, "Acoustic speckle: Theory and experimental analysis," *Ultrasound Imag.*, vol. 1, pp. 303–324, 1979.
- [15] B. Zagar and H. Weiss, "Processing of laser speckle interferometer signals for high precision strain measurement," in *Proc. 8th Int. Symp. Artificial Intelligence Based Measurement and Control*, 1991, vol. 1, pp. 99–105.
- [16] C. Burckhardt, "Speckle in ultrasound B-mode scans," *IEEE Trans. Son. Ultrason.*, vol. SU-25, no. 1, pp. 1–6, Jan. 1978.
- [17] P. Xu, "Despeckling sar-type multiplicative noise," *Int. J. Remote Sens.*, vol. 20, no. 13, pp. 2577–2596, Sep. 1999.
- [18] J. M. Ollinger and J. A. Fessler, "Positron emission tomography," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 43–55, Jan. 1997.
- [19] G. E. Hagberg, G. Zito, F. Patria, and J. N. Sanes, "Improved detection of event-related functional MRI signals using probability functions," *NeuroImage*, vol. 2, pp. 1193–1205, 2001.
- [20] G. van Kempen, L. van Vliet, and P. Verveer, "Application of image restoration methods for confocal fluorescence microscopy," in *Proc. SPIE 3-D Microscopy: Image Acquisition and Processing IV*, 1997, vol. 2984.
- [21] E. Saleh and M. Teich, N. Y. Wiley, Ed., in *Proc. Fundamentals of Photonics*, 1991.
- [22] P. Gravel, G. Beaudoin, and J. A. D. Guise, "A method for modeling noise in medical images," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1221–1232, Oct. 2004.
- [23] P. Bao and L. Zhang, "Noise reduction for magnetic resonance images via adaptative multiscale products thresholding," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, Sep. 2003.
- [24] T. K. Keyes and W. T. Tucker, "The k-distribution for modeling the envelope amplitude of a backscattered signal," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 46, no. 4, pp. 883–887, Jul. 1999.
- [25] P. Shankar, "A general statistical model for ultrasonic backscattering from tissues," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 47, no. 5, pp. 727–736, May 2000.
- [26] J. M. Sanches and J. S. Marques, "Compensation of log-compressed images for 3d ultrasound," *Ultrasound Med. Biol.*, vol. 29, no. 2, pp. 239–253, 2003.
- [27] O. V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 53, no. 1, pp. 64–78, Jan. 2006.
- [28] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C[F4]$," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [29] A. Barraud, "A numerical algorithm to solve $A^T X A - X = Q$," *IEEE Trans. Autom. Control*, vol. AC-22, 1977.
- [30] P. C. Tay, S. T. Acton, and J. A. Hossack, "Ultrasound despeckling using an adaptive window stochastic approach," in *Proc. Int. Conf. Image Processing*, 2006, pp. 2549–2552.
- [31] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound b-scans," *IEEE Trans. Son. Ultrason.*, vol. 30, no. 3, pp. 156–163, May 1983.
- [32] J. Dainty, "Laser speckle and related phenomena," in *Laser Speckle and Related Phenomena*. New York: Springer-Verlag, 1984, vol. 9, p. 146, XVII, 342 figs., Also Topics in Applied Physics.
- [33] R. F. Wagner, M. F. Insana, and S. W. Smith, "Fundamental correlation lengths of coherent speckle in medical ultrasonic images," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 35, no. 1, pp. 34–44, Jan. 1988.
- [34] C. Loizou, C. Pattichis, C. Christodoulou, R. H. Istepanian, M. Pantziaris, and A. Nicolaides, "Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery," *IEEE Trans. Visual. Comput. Graph.*, vol. 52, no. 10, pp. 1653–1669, Oct. 2005.
- [35] T. Huang, G. Yang, and G. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 1, pp. 13–18, Jan. 1979.
- [36] T. Loupas, W. McDicken, and P. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images," *IEEE Trans. Circuits Syst.*, vol. 36, no. 1, pp. 129–135, Jan. 1989.
- [37] P. C. Tay, S. T. Acton, and J. A. Hossack, "A stochastic approach to ultrasound despeckling," in *Proc. 3rd IEEE Int. Symp. Biomedical Imaging: Macro to Nano*, 2006, pp. 221–224.
- [38] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, Mar. 1994.
- [39] S. Zhong and V. Cherkassky, "Image denoising using wavelet thresholding and model selection," in *Proc. Int. Conf. Image Processing*, 2000, vol. 3, pp. 262–265.
- [40] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 772–783, Aug. 2001.
- [41] S. Gupta, L. Kaur, R. C. Chauhan, and S. C. Saxena, "A wavelet based statistical approach for speckle reduction in medical ultrasound images," *Med. Biol. Eng. Comput.*, vol. 42, pp. 534–537, 2004.
- [42] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," presented at the IEEE Int. Conf. Image Processing, Sep. 2005.
- [43] A. F. Laine, "Wavelets in temporal and spatial processing of biomedical images," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 511–550, 2000.
- [44] Y. Jin, E. Angelini, and A. Laine, *Wavelets in Medical Image Processing: Denoising, Segmentation, and Registration, Handbook of Medical Image Analysis: Advanced Segmentation and Registration Models*. New York: Springer, 2004.
- [45] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Melbourne, FL: Krieger, 1992.
- [46] H. Guo, J. E. Odegaard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. Burrus, "Wavelet based speckle reduction with application to sar based atd/r," in *Proc. Int. Conf. Image Processing*, 1994, pp. 75–79.
- [47] C. R. Vogel, *Computational Methods for Inverse Problems*, 2002.
- [48] X. Huang, A. Madoc, and A. Cheetham, "Wavelet-based bayesian estimator for poisson noise removal from images," in *Proc. Int. Conf. Multimedia and Expo*, Jul. 2003, vol. 1, pp. I-593–I-596.
- [49] D. Gleich and M. Datcu, "Gauss-markov model for wavelet-based sar image despeckling," *IEEE Signal Process. Lett.*, vol. 13, no. 6, pp. 365–368, Jun. 2006.
- [50] J. Xu and S. Osher, "Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 534–544, Feb. 2007.
- [51] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton Univ. Bull.*, pp. 49–52, 1902.
- [52] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction in positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, Feb. 1982.
- [53] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [54] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images," *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, pp. 564–584, 1987.
- [55] S. Geman and D. McClure, *Bayesian Image Analysis: An Application to Single Photon Emission Tomography*, pp. 12–18, 1985.
- [56] S. Z. Li, "Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 916–932, Sep. 1998.
- [57] E. Rignot and R. Chelappa, "Segmentation of polarimetric synthetic aperture radar data," *IEEE Trans. Image Process.*, vol. 1, no. 1, pp. 281–300, Jan. 1992.
- [58] S. D. Conte and C. de Boor, *Elementary Numerical Analysis*. New York: McGraw-Hill, 1981.
- [59] V. Balakrishnan and E. F., Eds., "Linear matrix inequalities in control theory and applications," *Int. J. Robust and Nonlinear Control*, vol. 6, no. 9/10, pp. 896–1099, Nov.–Dec. 1996.
- [60] D. Calvetti and L. Reichel, "Application of ad iterative methods to the restoration of noisy images," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 1, pp. 165–186, 1996.
- [61] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, "Solution of the sylvester matrix equation $axbt + cxd = e$," *ACM Trans. Math. Softw.*, vol. 18, no. 2, pp. 223–231, 1992.
- [62] K. Lange, *Optimization*, ser. Springer Texts in Statistics.. New York: Springer, Jun. 2004.
- [63] R. Pan and S. J. Reeves, "Efficient huber-markov edge-preserving image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 3728–3735, Dec. 2006.
- [64] J. Dias, "Fast gem wavelet-based image deconvolution algorithm," presented at the Int. Conf. Image Processing, 2003.
- [65] P. Benner, *Factorized Solution of Sylvester Equations With Applications in Control*, May 2004.
- [66] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, "Linear matrix inequalities in system and control theory," presented at the Annu. Allerton Conf. Communication, Control and Computing, Oct. 1993.
- [67] R. M. Willett and R. D. Nowak, "Platelets: A multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 332–350, Mar. 2003.

- [68] J. Nascimento and J. S. Marques, "Robust shape tracking in the presence of cluttered background," *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 852–861, Dec. 2004.
- [69] A. Hammoude, "Computer-assited endocardial border identification from a sequence of two-dimensional echocardiographic images," Ph.D. dissertation, Univ. Washington, Seattle, WA, 1988.
- [70] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped apodized and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 39, no. 2, pp. 262–267, Feb. 1992.
- [71] K. Timmermann and R. Nowak, "Multiscale bayesian estimation of poisson intensities," in *Proc. 31st Asilomar Conf. Signals, Systems, Comput.*, 1997, pp. 85–90.
- [72] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, pp. 850–863, 1993.
- [73] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, 1996.
- [74] C. Bouman and K. Sauer, "A generalized gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.
- [75] C. Vogel and M. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 7, no. 7, pp. 813–824.
- [76] J. Sanches and J. Marques, "Image reconstruction using the Benford law," presented at the IEEE Proc. Int. Conf. Image Processing, 2006.
- [77] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1-2, pp. 89–97, 2004.
- [78] J. Jolion, "Image and the benford's law," *J. Math. Imag. Vis.*, vol. 14, pp. 73–81, 2001.



João M. Sanches (M'06) received the E.E., M.Sc., and Ph.D. degrees from the Technical University of Lisbon, Portugal, in 1991, 1996, and 2003, respectively.

Currently, he is an Assistant Professor in the Electrical and Computer Engineering Department, Instituto Superior Técnico, Lisbon, and a Researcher at the Institute for Systems and Robotics. He has taught in the area of signal processing and control. His Ph.D. work was in medical imaging reconstruction, and, since then, his main research interests

have been in biomedical engineering with several publications in this field. Today, his research work is mainly in functional MRI, 3-D ultrasound, confocal imaging microscopy, and neurophysiology, mostly in collaboration with the Molecular Medicine Institute at the Hospital de Santa Maria, Lisbon.



Jacinto C. Nascimento (M'06) received the E.E. degree from the Instituto Superior de Engenharia de Lisboa, and the M.Sc. and Ph.D. degrees from the Instituto Superior Técnico, Lisbon, in 1995, 1998, and 2003, respectively.

Presently, he is a Posdoctorate Researcher at IST affiliated with Institute for Systems and Robotics (ISR). His research interests are image processing, shape tracking, robust estimation, medical imaging, and video surveillance.



Jorge S. Marques received the E.E., M.Sc., and Ph.D. degrees, and the aggregation title from the Technical University of Lisbon, Lisbon, Portugal, in 1981, 1984, 1990, and 2002, respectively.

Currently, he is an Associate Professor with the Electrical and Computer Engineering Department, Instituto Superior Técnico, Lisbon, and a Researcher at the Institute for Systems and Robotics. He has published over 140 papers in international journals and conferences and he is the author of the book *Pattern Recognition: Statistical and Neural Methods*

(IST Press, 2005, 2nd ed., in Portuguese).

Dr. Marques was the Co-Chairman of the IAPR Conference IbPRIA 2005 and President of the Portuguese Association for Pattern Recognition from 2001 to 2003. His research interests are in the areas of statistical image processing, shape analysis, and pattern recognition.