Total Variation with Automatic Hyper-Parameter Estimation

Jacinto Nascimento and João Sanches

***Abstract*— Medical diagnosis is often hampered by the quality of the images. This happens in a wide range of image modalities. Image noise reduction is a crucial step, however difficult to be accomplished. Bayesian algorithms have been commonly used with success, namely with *additive white Gaussian noise* (AWGN) model. In fact, the noise corrupting some of the most used medical imaging modalities is not additive neither Gaussian but multiplicative described by Poisson or Rayleigh distributions.**

This paper proposes a unified framework with automatic hyper parameters estimation. The proposed framework deals with AWGN but also with both Poisson and Rayleigh distributions. The algorithm proposed herein, is based on a *maximum a posteriori* (MAP) criterion with the edge preserving prior based on the *total variation* (TV), which avoids the distortion of relevant anatomical details. The denoising technique is performed via single parametric iterative scheme parameterized for each noise model considered. Tests with real data from several medical imaging modalities testify the performance of the algorithm.

***Index Terms*— Total variation, Despeckling, Denoising, Left Ventricle, Tracking, Ultrasound.**

I. INTRODUCTION

The extraction of the relevant features in medical images play an important role for medical diagnosis. However, such task is hampered by the presence of artifacts and distortions in the images introduced by the acquisition process. Additionally, the presence of multiplicative noise (speckle and Poisson) and the low signal to noise ratio (SNR) are typical in several medical image modalities and make this task even more difficult. This is typical in the Ultrasound (US) imagery, which is one of the most used for diagnosis purposes.

Given the above mentioned issues, noise reduction must be accomplished. The statistical characterization of the noise, introduced by the systems during the acquisition process, is strongly dependent on the acquisition system. Therefore, appropriated *denoising* algorithms must be used according to the image modality under consideration.

Two types of multiplicative noise assume particular importance: $i)$ speckle, usually modeled by a Rayleigh distribution and $ii)$ Poisson noise. They are multiplicative since the variance is not constant and depends on the parameters to be estimated. The speckle noise usually appears in acquisition processes involving coherent radiation like *LASER, Ultrasound* and *Synthetic Aperture Radar* (SAR) and the Poisson noise in systems involving counting procedures like PET/SPECT, functional MRI and fluorescence confocal microscopy.

The most common paradigm used to model the noise is the *adaptive white Gaussian noise* (AWGN). However, this approach can not deal with situations in which the noise can not be assumed additive neither space invariant across the image. This is the case of the multiplicative noise such as the *speckle* noise.

Several techniques and frameworks have been proposed to reduce the speckle noise without distorting the relevant clinical details. Among them the *Bayesian* [1] ones have been used with success by modeling the speckle noise by *Rayleigh* or *Rice* [13,19] distributions. Other methods comprise soft thresholding [2], wavelet based techniques [3,8]–[10], wavelet soft-shrinking [4], or multi-resolution based techniques [5], median filtering [6], anisotropic diffusion [7] or variational theory based [11].

In the Bayesian framework used in this paper the minimizer of an energy function is the desired denoised image. This energy function is the sum of two terms: i) the *Data Fidelity Term* that pulls the solution toward the data and ii) the *Prior Term* that regularizes the solution to remove the noise. A prior parameter allows to tune the degree of smoothness of the solution. Large values of the parameter makes it possible to remove large amounts of noise but also some anatomical details needed for the diagnosis. Small values of the parameter allows to keep these details but also keeps the noise. Therefore, a trade off is needed and a criterion to automatic select this parameter is difficult to establish.

Here, an unified framework with automatic model parameter estimation do deal with additive Gaussian and multiplicative noise described by Rayleigh and Poisson distributions is described.

Examples of synthetic and real data, as well as a tracker that tracks the *left ventricle* (LV) boundary of the heart during a cardiac cycle are used to illustrate the application of the method.

The paper is organized as follows: Section II formulates the estimation problem and the automatic prior parameter estimation. Section III describes experimental results and Section IV concludes the paper.

II. PROBLEM FORMULATION

Let $\mathbf{Y} = \{y(i,j)\}$ be a matrix representing the noisy image to be processed and $\mathbf{X} = \{x(i,j)\}$ the cleaned image to be estimated from $\mathbf{Y}$. In a Bayesian framework this is is done by solving the following equation

$$\hat{X} = \arg\min_{\mathbf{X}} E_Y(\mathbf{Y},\mathbf{X}) + E_X(\mathbf{X}) \quad (1)$$

where $E_Y(\mathbf{Y},\mathbf{X})$ is called the *data fidelity term* and $E_X(\mathbf{X})$ is called the *prior term*. The *data fidelity term* depends on the observations and on the noise generation model and the *prior term* is used to well pose the problem that is usually ill-posed and needs regularization.

Assuming independence of the observations, which is a current practice and a very convenient simplification from a mathematical point of view, the *data fidelity term* is

$$E_Y(\mathbf{Y},\mathbf{X}) = \sum_{i,j} d(i,j) \quad (2)$$

where $d(i,j) = -\log[p(y(i,j)|x(i,j)]$ are listed in Table I for the three observation models considered here: Gaussian, Poisson and Rayleigh.

The *prior term* is $E_X(\mathbf{X}) = -\log[p(\mathbf{X})]$ where

$$p(\mathbf{X}) = \frac{1}{Z} e^{-\alpha TV(\mathbf{X})} \quad (3)$$

is a Gibbs distribution, $TV(\mathbf{X}) = \sum_{i,j} g(i,j)$ is called the *Total Variation* of $\mathbf{X}$ and

$$g(i,j) = \sqrt{(x(i,j) - x(i-1,j))^2 + (x(i,j) - x(i,j-1))^2} \quad (4)$$

This work was partially supported by FCT, Portuguese Ministry of Science an Technology and Higher Education (which includes FEDER funds)

is the discrete estimation of the gradient magnitude computed at pixel (i,j). The prior term is thus

$$E_X(\mathbf{X}) = \alpha TV(\mathbf{X}) \tag{5}$$

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

TABLE I

GAUSSIAN, POISSON AND RAYLEIGH MODELS FOR THE OBSERVATIONS.

The minimization of $E(\mathbf{Y},\mathbf{X}) = E_Y(\mathbf{Y},\mathbf{X}) + E_X(\mathbf{X})$ (see equation (1)) is obtained by computing its stationary point, $\nabla E(\mathbf{Y},\mathbf{X}) = 0$ which leads to the following set of equations

$$\underbrace{\frac{d}{dx(i,j)}d(i,j)}_{\text{see table I}} + \frac{\partial}{\partial x(i,j)}\sum_{i,j} g(i,j) = 0, \quad 0 \le i,j \le N,M \tag{6}$$

where the first term is a total derivative because $d(i,j)$ only depends on $x(i,j)$ while the second term is a partial derivative because $g(i,j)$ depends also on the neighbors of $x(i,j)$. This set of equations is non linear and must be iteratively solved. Here an approximation is assumed for sake of simplicity, $g(i,j) \approx g(i+1,j) \approx g(i,j+1)$. By assuming this simplification the set of equations (6) may be written as follows

$$\frac{x(i,j) - x^{ML}(i,j)}{\xi(i,j)} - \alpha\frac{N_v}{g(i,j)}(x(i,j) - \bar{x}(i,j)) = 0 \tag{7}$$

where $N_v = 4$ is the number of neighbors of $x(i,j)$ and

$$\bar{x}(i,j) = \frac{1}{N_v}\sum_{k\in V(i,j)} x_k(i,j) \tag{8}$$

where $x_k(i,j)$ is the kth neighbor of $x(i,j)$. $x^{ML}(i,j)$ is obtained from table I, as well $\xi(i,j)$ that is

$$\xi(i,j) = \begin{cases} \sigma^2 & \text{Gaussian model} \\ x(i,j) & \text{Poisson model} \\ x^2(i,j) & \text{Rayleigh model} \end{cases} \tag{9}$$

The set of equations (7) can be rewritten as follows

$$x(i,j) = (1 - K(i,j,\mathbf{X}))x^{ML}(i,j) + K(i,j,\mathbf{X})\bar{x}(i,j) \tag{10}$$

with

$$K(i,j,\mathbf{X}) = \frac{1}{1 + \frac{g(i,j)}{\alpha N_v \xi(i,j)}} \tag{11}$$

where $K(i,j,\mathbf{X})$ depends on the $x(i,j)$ and on its neighbors, $x_k(i,j), \quad k \in V(i,j)$. By using the fixed point method the set of equations (10) may be iteratively solved with the following recursion

$$x^{t+1}(i,j) = (1 - K^t(i,j,\mathbf{X}))x^{ML}(i,j) + K^t(i,j,\mathbf{X})\bar{x}^t(i,j) \tag{12}$$

where ^{t} denotes the estimation at t-th iteration.

A. Automatic hyper-parameter

One of the most difficult problems when dealing with this Bayesian formulation is the choice of the prior parameter α. Most of the time this parameter is chosen manually in a trial and error basis. Several authors have proposed method to select it automatically [12] but it is still an open problem. Here, a new method is proposed.

Let us factorize the prior distribution (3)

$$p(\mathbf{X}) = \prod_{i,j=0}^{N,M} \frac{1}{Z(i,j)}e^{-\alpha g(i,j)} \tag{13}$$

which may be interpreted as the joint probability density function of NM *independent and identically distributed* (i.i.d) random variables with probability

$$p(g(i,j)) = \frac{1}{Z(i,j)}e^{-\alpha g(i,j)} \tag{14}$$

The independence of the gradient magnitude along the image may be not realistic but is an admissible interpretation for the equation (13) that results from the common accepted assumption that X is a *Markov random field* (MRF), being therefore, described by a Gibbs distribution [20].

This distribution is a true probability density function if

$$\int_0^\infty p(g)dg = 1 \tag{15}$$

where the integration is performed only for R^+ because $g(i,j) \ge 0$. From the equation (15) it is derived the following result

$$Z(i,j) = \frac{1}{\alpha} \tag{16}$$

which means

$$p(\mathbf{X}) = \alpha^{NM}e^{-\alpha TV(\mathbf{X})} \tag{17}$$

The energy function to be minimized is therefore

$$E(\mathbf{Y},\mathbf{X},\alpha) = E_Y(\mathbf{Y},\mathbf{X}) + \alpha TV(\mathbf{X}) - NM\log(\alpha) \tag{18}$$

The optimum value for α is the minimizer of $E(\mathbf{Y},\mathbf{X},\alpha)$

$$\alpha_{opt} = \arg\min_\alpha E(\mathbf{Y},\mathbf{X},\alpha) \tag{19}$$

which may be computed by solving the following equation

$$\frac{\partial E(\mathbf{Y},\mathbf{X},\alpha)}{\partial\alpha} = TV(\mathbf{X}) - \frac{NM}{\alpha} = 0 \tag{20}$$

or

$$\alpha_{opt} = \frac{NM}{TV} = \frac{1}{\bar{g}} \tag{21}$$

where $\bar{g}$ is the average value of the gradient magnitude computed over all pixels of the image.

The estimation of $\mathbf{X}$ is performed by using the recursion (12) where $K(i,j,\mathbf{X})$, defined in (11), is replaced by the following expression

$$K^t(i,j,\mathbf{X}) = \left(1 + \frac{\bar{g}^t g^t(i,j)}{N_v\xi(i,j)}\right)^{-1} \tag{22}$$

III. EXPERIMENTAL RESULTS

In this section several examples of application are presented. In the first part, real images from several biomedical imaging modalities are shown; in the second part two sequences of US images are used to track the LV boundary during several cardiac cycles.

A. Real Images

This section presents the results using real images. Fig. 1 (top) shows a functional MRI image corrupted by AWGN (left) and the resulting denoised image (right). At the bottom of each example, the main diagonal profile is extracted from the noisy and noiseless images and are superimposed for comparison purposes. In this case the final $\alpha_{opt} = 63.91$. Fig. 1 (bottom) also shows an MRI image of the knee corrupted by AWGN and the corresponding noiseless and profiles ($\alpha_{opt} = 58.39$). Fig. 2 shows a confocal fluorescent image of a cell ($\alpha_{opt} = 231.27$) (top) and an ultrasound images of a gall bladder ($\alpha_{opt} = 78.73$) (bottom) corrupted by Poisson and Rayleigh noise respectively.

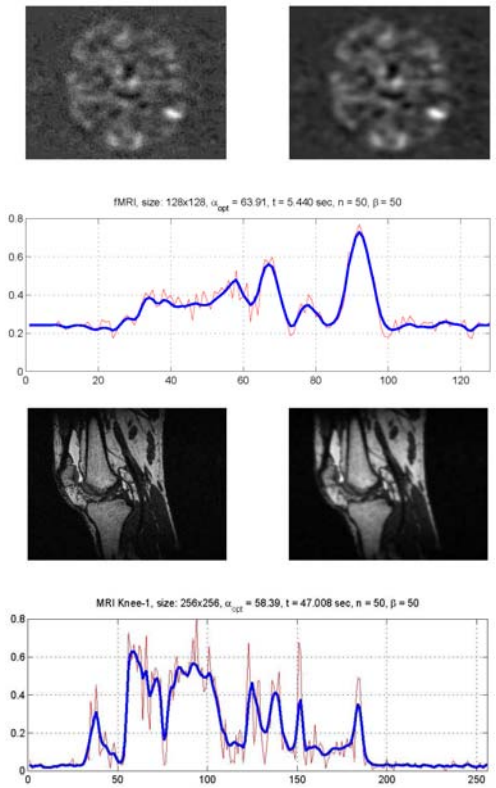

Fig. 1. Additive and Multiplicative noise reduction: (Top) Functional MRI (AWGN), (bottom) Knee (MRI - AWGN).

B. LV tracking

To assess the performance of the proposed method, the algorithm was tested using two ultrasound sequences of the LV. These sequences correspond to real medical exams performed on 2 different people. The sequences were obtained at a frame rate of $15\ \text{frames}\ \text{sec}^{-1}$ using an ultrasound probe operating at 1.7 Mhz. The size of the ultrasound sequences used in this study are: case $\#1$ - 376 frames (16 cycles); case $\#2$ - 470 frames (19 cycles).

In this study we provide a comparison of the following denoising methods: $i)$ median filtering; $ii)$ L_2 prior; $iii)$ TV prior; $iv)$ Benford prior and $v)$ optimal solution.

This study aims to track the LV during the sequences. (Fig. 3 shows a frame of one sequence and the corresponding denoised image).

The test is done as follows. The ultrasound sequences were first processed by each of the five denoising algorithms. All the methods were implemented using a non Gaussian observation model (Rayleigh distributed). A tracking algorithm was then applied to the denoised images to estimate the boundary of the LV. The

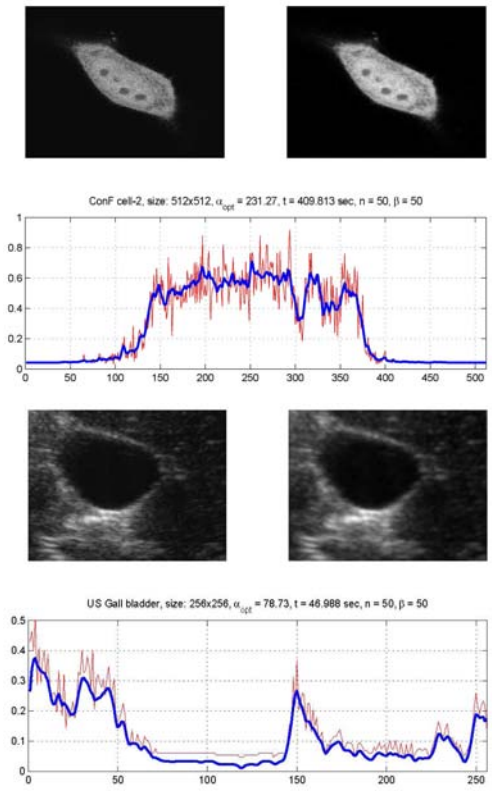

Fig. 2. Additive and Multiplicative noise reduction: (Top) Confocal (Poisson), (bottom) Gall Bladder (Ultrasound - Rayleigh).

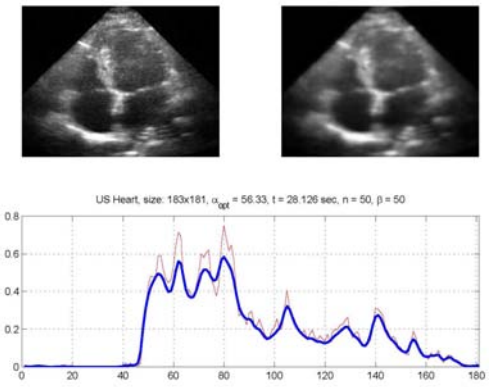

Fig. 3. Heart Ultrasound (Rayleigh).

tracker used in this study is the S-PDAF tracker proposed by the authors (see [14]). The output of the tracker for each of the denoising conditions was then compared with the ground truth 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.

To compare performance of different methods, we used five distance measures: $i)$ Hausdorff distance (*cf.* [15]); $ii)$ the average distance; $iii)$ Hammoude distance (*cf.* [16]); $iv)$ Mean Sum of

Squared Distance (MSSD) (*cf.* [17]) and Mean Absolute Distance (MAD) (*cf.* [18]). We next briefly describe them.

Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_x}\}$, and $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_y}\}$, 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 $\mathcal{Y}$

$$d(\mathbf{x}_i, \mathcal{Y}) = \min_j ||\mathbf{y}_j - \mathbf{x}_i|| \tag{23}$$

This is known as distance to the closest point (DCP). The average distance between the sets $\mathcal{X}$, $\mathcal{Y}$ is defined as

$$d_{AV} = \frac{1}{N_x} \sum_{i=1}^{N_x} d(\mathbf{x}_i, \mathcal{Y}) \tag{24}$$

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

$$d_{HDF}(\mathcal{X}, \mathcal{Y}) = \max \ \max_i\{d(\mathbf{x}_i, \mathcal{Y})\}, \max_j\{d(\mathbf{y}_j, \mathcal{X})\} \tag{25}$$

For the Hammoude metric, 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

$$d_{HMD} = \frac{\sharp\ (R_{\mathcal{X}} \cup R_{\mathcal{Y}}) - (R_{\mathcal{X}} \cap R_{\mathcal{Y}})}{\sharp(R_{\mathcal{X}} \cup R_{\mathcal{Y}})} \tag{26}$$

The MSSD and MAD distances are defined as follows

$$d_{MSSD}(\mathcal{X}, \mathcal{Y}) = \Big\{ \frac{1}{n} \sum_{i=1}^{n} d^2(\mathbf{x}_i, \mathcal{Y}) + \frac{1}{m} \sum_{i=1}^{m} d^2(\mathbf{y}_i, \mathcal{X}) \Big\} \tag{27}$$

$$d_{MAD}(\mathcal{X}, \mathcal{Y}) = \Big\{ \frac{1}{n} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathcal{Y}) + \frac{1}{m} \sum_{i=1}^{m} d(\mathbf{y}_i, \mathcal{X}) \Big\} \tag{28}$$

where d is the DCP given as in (23). In the two above distances, they do not give higher weights to longer sequences. In this paper, the distances in (27) and (28) do not have point correspondence as originally proposed in [17,18]. Thus, the nearest point from the other contour is taking as the corresponding point.

Table II show the fidelity in the representation of the LV contour obtained in the two US sequences, which is an important feature when tracking the contour of the object of interest. These values correspond to the mean values of the metrics described herein. However they appear to be close to each other, these small differences can reflect significant differences in long sequences, such as the ones used in this paper. From Table II we conclude that, in the second sequence the best results are shared between the TV prior and the optimal solution. In the first sequence, the best scores belong to the method proposed herein. These results testify that the proposed method exhibit attractive properties, providing valuable results and capable to compete with other methods recently proposed.

TABLE II

METRICS MEAN VALUES FOR THE FIRST AND SECOND SEQUENCES.

		Median	L_2	TV	Bfd	Opt
d_{Hmd}	case #1	0.20	0.18	0.19	0.17	**0.14**
	case #2	0.25	0.22	0.19	0.20	**0.18**
d_{Av}	case #1	4.66	4.20	4.19	3.81	**3.35**
	case #2	4.58	4.33	**3.57**	3.74	3.75
d_{Hdf}	case #1	13.74	12.83	12.78	11.33	**10.04**
	case #2	12.17	11.16	**9.48**	9.95	10.56
d_{MSSD}	case #1	2.13	1.61	1.57	1.26	**1.00**
	case #2	2.14	1.56	1.46	1.79	**1.29**
d_{MAD}	case #1	0.25	0.23	0.23	0.21	**0.19**
	case #2	0.26	0.23	**0.20**	0.21	**0.20**

IV. CONCLUSIONS

This paper proposes an unified framework with automatic model parameter estimation do deal with additive Gaussian and multiplicative noise, described by Rayleigh and Poisson distributions. A Bayesian unified framework is described, where the denoised image comes from a minimization of an energy function containing two terms: *data fidelity term* and *prior term*. It is shown how the prior term can be estimated automatically. The method proposed herein exhibits a very good compromise between smoothness-vs-anatomical details. Experimental results testify an improvement over the methods previously proposed.

REFERENCES

[1] Z. Zeng and I. Cumming, "Bayesian speckle noise reduction using the discrete wavelet transform", *Int. geoscience and remote sensing symp.* pp. 6-10, 1998.

[2] S. Gupta, L. Kaur, R. C. Chauhan and S. C. Saxena, "A wavelet based statistical approach for speckle reduction in medical ultrasound images", *Med. and Bio. Eng. and Computing*, vol. 42, no. pp. 189-192, 2004.

[3] I. Duskunovic, A. Pizurica, G. Stippel, W. Philips and I. Lemahieu, "Wavelet based denoising techniques for ultrasound images", *Proc. IEEE Eng. in Med. and Bio. Soc. Conf.*, vol. 4, pp. 2662-2665, 2000.

[4] Y. Yue, M. M. Croitoru, A. Bidani, J. B. Zwischenberger and J. W. Clark, "Ultrasonic speckle suppression using robust nonlinear wavelet diffusion for lv volume quantification", *Proc. of the Int. Conf. of the EMBS*, pp. 1609-1612, 2004.

[5] A. Achim, A. Bezerianos and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images", *IEEE Trans. Med. Imag.*, vol. 20, pp. 772-783, 2001.

[6] T. Loupas and W. McDicken and P. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images", IEEE Trans. Circuits Syst., vol. 36, 129-135, 1989.

[7] J. Montagnat and H. Delingette, "Space and Time Shape Constrained Deformable Surfaces for 4D Medical Image Segmentation", *Proc. of the 3rd Int. Conf. on Med. Imag. Comp. and Computer-Assisted Intervention*, pp. 195-205, 2000.

[8] Hui Cheng and Qiuze Yu, Jinwen Tian and Jian Liu, Image denoising using wavelet and support vector regression", *Proc. of the 3rd Int. Conf. on Image and Graphics*, pp. 43-46, 2004.

[9] Alle M. Wink and Jos B.T.M. Roerdink, "Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing", *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 374-387, 2004.

[10] E. J. Balster, Y. F. Zheng and R.L. Ewing, "Feature-based wavelet shrinkage algorithm for image denoising", *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2024-2039, 2005.

[11] R. Pan and S. J. Reeves, "Efficient Hubber-Markov Edge-Preserving Image Restoration", *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 3728-3735, 2006.

[12] A. Bouhamidi and K. Jbilou, "Sylvester Tikhonov-regularization methods in image restoration", J. Comput. Appl. Math.", pp. 86-98, vol. 206, no. 1, 2007.

[13] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images", *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 229-240, 2006.

[14] 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, 2004.

[15] D. P. Huttenlocher and G. A. Klanderman and W. J. Rucklidge, "Comparing images using Hausdorff distance", *IEEE Trans. Pattern Anal. Machine Intell.* vol. 15, pp. 850-863, 1993.

[16] A. Hammoude, *Computer-assited endocardial border identification from a sequence of two-dimensional echocardiographic images*, Ph.D. thesis, Univ. Washington, Seattle, WA, 1988.

[17] Y. Akgul and C. Kambhamettu, "A coarse-to-fine deformable contour optimization framework", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 2, pp. 174-186, 2003.

[18] I. Mikić and S. Krucinki and J. D. Thomas, "Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates", *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 274-284, 1998.

[19] J. Sijbers, A. den Dekker, P. Scheunders, and D. V. Dyck, "Maximum likelihood estimation of Rician distribution parameters", *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 357-361, 1998.

[20] J. Besag, On the Statistical Analysis of Dirty Pictures, J. R. Statist. Soc. B, vol.48, no 3,pp.259-302,1986.