SAR Image Regularization

with Fast Approximate Discrete Minimization

Loïc Denis, Florence Tupin, Jérôme Darbon and Marc Sigelle

Abstract

Synthetic aperture radar (SAR) images, like other coherent imaging modalities, suffer from speckle noise. The presence of this noise makes the automatic interpretation of images a challenging task and noise reduction is often a prerequisite for successful use of classical image processing algorithms.

Numerous approaches have been proposed to filter speckle noise. Markov Random Field (MRF) modelization provides a convenient way to express both data fidelity constraints and desirable properties of the filtered image. In this context, total variation minimization has been extensively used to constrain the oscillations in the regularized image while preserving its edges.

Speckle noise follows heavy-tailed distributions, and the MRF formulation leads to a minimization problem involving non-convex log-likelihood terms. Such a minimization can be performed efficiently by computing minimum cuts on weighted graphs. Due to memory constraints, exact minimization, although theoretically possible, is not achievable on large images required by remote sensing applications. The computational burden of the state-of-the-art algorithm for approximate minimization (namely the α-expansion) is too heavy specially when considering joint regularization of several images.

We show that a satisfying solution can be reached, in few iterations, by performing a graph-cut based combinatorial exploration of large trial moves. This algorithm is applied to joint regularization of the amplitude and interferometric phase in urban area SAR images.

I. INTRODUCTION

There are nowadays many SAR satellite sensors (EnviSat, Radarsat, ALOS ...) providing a huge amount of SAR images. The popularity of such sensors is linked to their all-weather and all-time capabilities, combined with their polarimetric and interferometric potential. The interferometric data, which are phase difference images, give either elevation or movement information. The launch of new sensors with improved resolution in 2007 (TerraSAR-X [47] and CosmoSkyMed [39]) opens new fields of applications. Particularly, the computation of Digital Elevation Models (DEM) becomes feasible with metric interferometric images, specially when tandem configurations will be available. These new data will contribute to urban monitoring which is an important issue for governmental agencies (risk analysis, disaster management, environmental protection, urban development

planning, ...). In this paper we are interested in filtering of SAR images for the the purpose of building delineation to perform 3D reconstruction.

However, SAR images are difficult to interpret not only with automatic image processing tools but also by human interpreters. This is mainly due to two specificities of the SAR system: first, SAR is coherent imagery and therefore subject to the speckle phenomenon; secondly, due to the microwave propagation, images are distance sampled leading to strong geometrical distortions.

Speckle is due to the interferences of waves reflected by many elementary reflectors inside a resolution cell. Although speckle has been extensively studied and is well modeled in some particular cases [23], [28], [30], speckle reduction remains one of the major issue in SAR image processing. Many filters have been proposed in the last twenty years and they can be classified in two categories: filters without explicit scene modeling based on Minimum Mean Square Error, and those with the explicit assumption of a scene distribution based on Maximum A Posteriori (MAP) or Maximum Likelihood criterion.

The first family contains the famous Lee [35], Kuan [34], and Frost [19] filters. More recent papers work in the wavelet domain [1]. In the second family, scene distribution hypothesis have lead to different filtering: Gaussian [33], Gamma [40], Fisher [42]. More elaborated models assuming that the scene is a Gaussian Markov random field [53] or establishing the probability density functions of the wavelet coefficients to do MAP filtering [18] have been developed.

Independently of the chosen filtering formula, parameter estimation is a crucial point. Indeed, the number of samples should be as big as possible, whereas the local stationarity should be verified inside the processing window. To solve this dilemma, many approaches have been proposed: gradient detection inside the analysis window [36], growing window strategy [43], [52], [54], feature (line, point, edge) detection [40]. Two excellent reviews with comparisons and improvements of many SAR filters can be found in [38] and [51].

One of the main interest of the Markovian framework is its ability to take into account both local non stationarity, specially the presence of edges, and a data acquisition model. Generally, the filtering corresponds to the computation of the MAP estimator. It consists of the minimization of an energy combining two types of information: a data driven term and a regularization term [22]. The first one is given by the physical mechanisms of radar processing. The second one reflects our knowledge about the reality of the imaged scene (also called "prior" term in the following). In the case of urban areas, many sharp discontinuities exist either in the amplitude image or in the interferometric one. Many models/priors have been investigated to cope with image discontinuities. There is a family of explicit edge processes [11], [22] and a family of well chosen functions which naturally preserve discontinuities [6], [21]. More recently, since the seminal work of [48], a great interest has been given to the minimization of total variation (TV) [12], [15], [41], [44], [45] due to its edge preserving behavior while still leading to a convex optimization problem. Various multiplicative noise models using Total Variation have been proposed [2], [17], [49]

Actually, the choice of the regularization function is closely linked to the optimization problem. Indeed, one of the main limitation to Markov Random Fields (MRF) in image processing was the optimization step. Although simulated annealing [22] has excellent theoretical performances, in practice, the computational burden might be very heavy. Deterministic approaches such as Iterated Conditional Modes [4] often converges toward a local minimum which can be far away from the exact solution. Thanks to graph-cut methods, i.e., computation

of a s-t minimum cut or by duality a maximum flow in a graph, exact *discrete* optimization schemes have been developed in some specific cases.

Such a combinatorial method has first been proposed in [46] for minimizing binary a class of energies. Then, Greig et al. [24] have used this approach to the study the behavior of the Ising model for binary image restoration. More recently, it has been shown in [32] that this approach works for any binary field whose prior is composed of pair-wise or triple-wise binary submodular functions. The case of non-binary fields has been addressed by few works. In [8] an excellent approximation result is presented where the prior corresponds to a semi-metric. Ishikawa has proposed a framework for exact optimization of convex regularization functions in the gray-level case [27]. With a different graph but the same size as the one of Ishikawa, exact optimization schemes for convex or levelable priors are also defined in [16]. In [14], it is shown that the approach of [8] converges toward a global minimizer for a subclass of non convex energies.

For convex energies, iterative approaches that allow to build much smaller graph are proposed in [5], [13], [31]. The particular case of the total variation minimization has been addressed in [10], [16]. Note that all of the above approaches are due to the efficient maximum flow/s-t minimum cut algorithm described in [7].

The contributions of the paper are the following: We propose a new fast algorithm for SAR scene reflectivity restoration and also for the joint regularization of amplitude and interferometric phase images. We have chosen to consider TV prior which is well adapted for urban areas. As will be seen in the next part, the data driven term is not convex. In this case, either [16] or [27] could provide exact optimization algorithms but at the price of a huge memory space due to the graph size. The α-expansion algorithm of [8] could also provide an approximate solution, but with a quite heavy computational burden. A new algorithm is presented providing a *fast* and *approximate* solution and able to deal with joint regularization of amplitude and phase image. The graph is of similar size to the one used to perform α-expansions, but based on a different principle. The obtained local minimum has been found satisfying in different practical cases. Empirical studies have shown that the minimum is very close to the global minimum computed by [16] with a great improvement of the needed memory space and of computation time.

The remainder of the article is organized as follows. In section II, the MRF model is presented, and particularly the data driven term in the case of SAR and InSAR images is detailed. Section III is dedicated to the presentation of our optimization algorithm after recalling other graph-cut based methods. In section IV, the model and minimization algorithm are compared to other methods. They are then applied to real InSAR images in the context of 3D reconstruction in urban areas. Section VI concludes about the proposed method.

II. MRF MODEL

A. MRF framework

It is assumed that an image u is defined on a finite discrete lattice S and takes values in a discrete integer set $\mathcal{L} = \{1, \ldots, L\}$. We denote by u_s the value of the image u at the site $s \in S$ and by (s, t) the related clique of order two. Given an observed image u, a Bayesian analysis using the MAP criterion consists of finding a restored image $\hat{u}$ that maximizes $p(\hat{u}|u) \propto p(u|\hat{u})p(\hat{u})$. It can be shown under the assumption of Markovianity of $\hat{u}$ and with some independence assumption on u conditionally to $\hat{u}$ ($P(u|\hat{u}) = \Pi_s P(u_s|\hat{u}_s)$) that the MAP
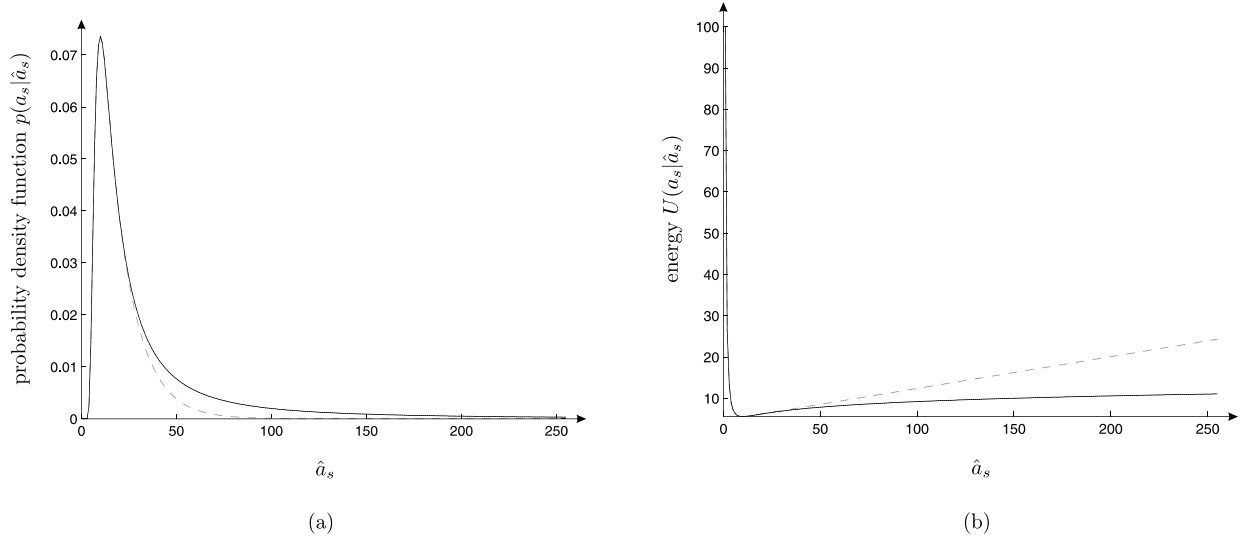
Fig. 1. Likelihood model for SAR amplitude. Continuous line: probability density function (a) and corresponding energy (b) followed by a single-look amplitude image ($a_s = 10$). Dashed-line: convex approximation. The convex approximation can not model the "heavy tail" that characterizes speckle noise.

problem becomes an energy minimization problem:

$$\hat{u}^{(MAP)} = \arg \min_{\hat{u}} E(\hat{u}|u) \ ,$$

with $E(\hat{u}|u) = \sum_s U(u_s|\hat{u}_s) + \beta \sum_{(s,t)} \psi(\hat{u}_s, \hat{u}_t)$, $U(u_s|\hat{u}_s) = -\log p(u_s|\hat{u}_s)$ the opposite of the log-likelihood and ψ a function modeling the prior chosen for the solution.

B. SAR and InSAR image formation

1) Distribution of the amplitude: The synthesized radar image z is complex-valued. Its amplitude $|z|$ is very noisy due to the interferences that occur inside a resolution cell. A classical model for speckle was developed by Goodman [23] and is valid for "rough" surfaces (the roughness being considered according to the wavelength of the sensor). Under this model, the amplitude a_s of a pixel s follows a Nakagami distribution depending on $\hat{a}_s$, the square root of the reflectivity [23]:

$$p(a_s|\hat{a}_s) = \frac{2M^M}{\Gamma(M)\hat{a}_s^{2M}} a_s^{(2M-1)} \exp\left(-\frac{Ma_s^2}{\hat{a}_s^2}\right). \tag{1}$$

with M the number of looks of the image (i.e. number of independent values averaged). For single-look images ($M = 1$), the density function simplifies to Rayleigh law.

This likelihood leads to the following energetic term:

$$U(a_s|\hat{a}_s) = M \cdot \left[\frac{a_s^2}{\hat{a}_s^2} + 2\log \hat{a}_s\right] \ ,$$

represented in figure 1 (continuous line).

This energy is not convex with respect to $\hat{a}_s$ (a_s is the fixed observed amplitude value), contrary to the quadratic energy that arises from a Gaussian likelihood assumption. A convex approximation is drawn with a dashed line in figure 1. For display purposes, the corresponding probability density function (pdf) has not been normalized so that it superimposes to the true pdf. It is clear from the figure that such an approximation fails to model the heavy tail (i.e. slowly decreasing pdf) which is typical of speckle noise.

2) Distribution of the interferometric phase: In the case of SAR interferometric data, the interferometric product is obtained by complex averaging of the hermitian product of the two SAR images z_1 and z_2 accurately registered:

$$\underline{\rho_s} = \frac{\sum_{i=1}^{W} z_{1_i} z_{2_i}^*}{\sqrt{\sum_{i=1}^{W} |z_{1_i}|^2 \sum_{i=1}^{W} |z_{2_i}|^2}}$$

with W the number of pixels of the averaging window centered on site s. The interferometric phase is given by the argument of $\underline{\rho_s}$. The coherence is given by $\rho_s = |\underline{\rho_s}|$ and measures the sample correlation between the two SAR images. It is an indicator of the interferometric phase reliability.

The pdf of the phase can be written as an expression implying hypergeometric functions [37]. A good approximation is given by a Gaussian model:

$$p(\phi_s|\hat{\phi}_s) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_{\phi_s}} \exp\left(-\frac{(\phi_s - \hat{\phi}_s)^2}{\hat{\sigma}_{\phi_s}^2}\right). \tag{2}$$

The standard deviation $\hat{\sigma}_{\phi_s}^2$ at site s is approximated by the Cramer-Rao bound:

$$\hat{\sigma}_{\phi_s}^2 = \frac{1 - \rho_s^2}{2L\rho_s^2}. \tag{3}$$

For low coherence areas (shadows or smooth surfaces, denoted *Shadows* in the following), this Gaussian approximation is less relevant and a uniform distribution model is better:

$$p(\phi_s|\hat{\phi}_s) = \frac{1}{2\pi}. \tag{4}$$

This leads to the following energy:

$$\begin{cases} U(\phi_s|\hat{\phi}_s) = 0 \text{ if } s \in Shadows, \\ U(\phi_s|\hat{\phi}_s) = \frac{(\phi_s - \hat{\phi}_s)^2}{\hat{\sigma}_{\phi_s}^2} \text{ otherwise.} \end{cases}$$

The energy $U(\phi|\hat{\phi})$ is convex. The variance dividing the quadratic term is a function of the coherence of the pixel. Although this coherence could also be considered as a random field to regularize, it will be considered as a fixed field in the following.

C. Prior model

As said in the introduction, the TV regularization prior is well adapted when dealing with strong discontinuities. Besides this prior has good properties for minimization since it is a convex function. The energetic term corresponding to the discretization of TV can be written [15] as follows:

$$E(\hat{u}) = \beta \sum_{(s,t)} w_{st}|\hat{u}_s - \hat{u}_t| \ ,$$

with $w_{st} = 1$ for the 4-nearest neighbors and $w_{st} = 1/\sqrt{2}$ for the 4 diagonal ones. We will not explicitly write the weights w_{st} in the following equations.

For the separate regularization of amplitude or phase images we have the following energies to minimize:

$$E(\hat{a}|a) = \sum_s M[\frac{a_s^2}{\hat{a}_s^2} + 2\log\hat{a}_s] + \beta_a \sum_{(s,t)} |\hat{a}_s - \hat{a}_t| \tag{5}$$

$$E(\hat{\phi}|\phi) = \sum_s \frac{(\phi_s - \hat{\phi}_s)^2}{\hat{\sigma}_\phi^2} + \beta_\phi \sum_{(s,t)} |\hat{\phi}_s - \hat{\phi}_t| \tag{6}$$

We consider in this paper the case of aerial high resolution images of urban areas. The elevation range is contained within a fringe so we do not have to handle the problem of phase unwrapping. Then, contrary to other SAR configurations, we do not have to take the wrapping into account in the regularization term which simplifies greatly the regularization problem. Joint phase regularization and unwrapping has been recently studied in [5] using a graph-cut approach.

The phase and amplitude information are hopefully linked since they reflect the same scene. Amplitude discontinuities are thus usually located at the same place as phase discontinuities and conversely. We propose in this paper to perform the joint regularization of phase and amplitude. To combine the discontinuities a disjunctive $\max$ operator is chosen.

Note that the MAP estimates are not modified if the energies of equations 5 and 6 are respectively divided by non-null terms β_a and β_ϕ/γ. Since the total variation of the amplitude and the phase are of the same order, this leads to a normalization of the likelihood energies. The joint prior model is defined by:

$$E(\hat{a}, \hat{\phi}) = \sum_{(s,t)} \max(|\hat{a}_s - \hat{a}_t|, \gamma|\hat{\phi}_s - \hat{\phi}_t|), \tag{7}$$

with γ a parameter that can be set to 1, and otherwise accounts for the relative importance given to the discontinuities of the phase ($\gamma > 1$) or of the amplitude ($\gamma < 1$).

The global joint energy term is then:

$$\begin{aligned}
E(\hat{a}, \hat{\phi}|a, \phi) = &\frac{1}{\beta_a} \sum_s M[\frac{a_s^2}{\hat{a}_s^2} + 2\log \hat{a}_s] \\
&+ \frac{\gamma}{\beta_\phi} \sum_s \frac{(\phi_s - \hat{\phi}_s)^2}{\hat{\sigma}_\phi^2} \\
&+ \sum_{(s,t)} \max(|\hat{a}_s - \hat{a}_t|, \gamma|\hat{\phi}_s - \hat{\phi}_t|).
\end{aligned} \tag{8}$$

Shadow areas: The regularized fields $\hat{a}$ and $\hat{\phi}$ at sites s located inside the detected shadow areas $Shadows$ are governed by the regularisation term. With the prior term defined in equation 8, the phase $\hat{\phi}_s$ for $s \in Shadows$ that minimizes the energy corresponds to an interpolation of the phase value at the surrounding sites. Shadow areas however are most of the time at ground level and not at an intermediate height between the top of the structure that creates the shadow and the ground at the shadow end. A modified regularization term that better describes this prior knowledge is therefore used for cliques involving one or both site(s) inside the shadow regions:

$$E(\hat{a}, \hat{\phi}) = \sum_{(s,t)} E(\hat{a}, \hat{\phi})_{(s,t)},$$

with $E(\hat{a}, \hat{\phi})_{(s,t)}$ defined as:

  (i) if $s \notin Shadows$ and $t \notin Shadows$,

    $E(\hat{a}, \hat{\phi})_{(s,t)} = \max(|\hat{a}_s - \hat{a}_t|, \gamma|\hat{\phi}_s - \hat{\phi}_t|),$

  (ii) if $s \in Shadows$ and $t \notin Shadows$ and $\hat{\phi}_s \leq \hat{\phi}_t$

    $E(\hat{a}, \hat{\phi})_{(s,t)} = |\hat{a}_s - \hat{a}_t| + \gamma|\hat{\phi}_s - \hat{\phi}_t|,$

  (iii) if $s \in Shadows$ and $t \notin Shadows$ and $\hat{\phi}_s > \hat{\phi}_t$

    $E(\hat{a}, \hat{\phi})_{(s,t)} = |\hat{a}_s - \hat{a}_t| + 2\gamma|\hat{\phi}_s - \hat{\phi}_t|,$

(iv) if $s \in Shadows$ and $t \in Shadows$

$$E(\hat{a}, \hat{\phi})_{(s,t)} = |\hat{a}_s - \hat{a}_t| + \gamma \left(\hat{\phi}_s - \hat{\phi}_t \right)^2 .$$

The cases where $s \notin Shadows$ and $t \in Shadows$ are treated in a symmetrical manner. Outside shadow areas (case i), the regularization term is the same as previously. To limit the effect of a given shadow area on the regularization of the amplitude, we independently regularize phase and amplitude in and at the limit of the shadows (cases ii to iv). To force the regularized phase inside a shadow to follow ground level, we penalize more heavily over-estimation (case iii) than under-estimation (case ii). Finally, a quadratic constraint (case iv) enforces a flat/smooth ground inside a shadow area. Note that in each case (i to iv) the prior term $E(\hat{a}, \hat{\phi})_{(s,t)}$ is convex and so is the prior energy $E(\hat{a}, \hat{\phi})$. The convexity of the prior energy is essential to apply the minimization algorithm described in section III.

D. Energy minimization problem

As said in the introduction, graph-cut based approaches are very efficient methods for MRF optimization. Nevertheless, only certain classes of energies can be exactly minimized. We briefly describe here the algorithms which can be used to minimize energies of equations 5, 6, and 8.

1) Exact minimization: First, concerning amplitude data, two graph-cut based algorithms have been proposed to minimize energies such as eq. 5.

The first one has been developed by Ishikawa [27]. It is able to handle any kind of data driven term and convex regularization. The graph is constituted by $N \times L$ nodes (a node for each pixel and grey level) plus two terminal nodes. In the case of TV regularization, there are $3N \times L$ in 4 connexity (resp. $5N \times L$ in 8 connexity) pairs of directed edges connecting nodes between successive levels for each pixel, and between neighboring pixels for a given level. For remote sensing application, the graph size is prohibitive since the full graph must be stored in memory.

The second method has been proposed in [16]. It is based on the notion of levelable energies, which means that the energy can be written as a sum on the level sets of $\hat{u}$. Since the convexity of the posterior energy is not guaranteed in our model (due to the non-convex log-likelihood of the amplitude), a fast algorithm based on a scaling search can not be applied [15]. In this case, a much wider graph linking the different level sets must be built whose size is similar to the one of Ishikawa and still prohibitive for remote sensing applications.

The problem is easier for phase images (equation 6) since the data driven term is convex. In that case, a fast algorithm is proposed in [15]. It consists in solving a set of binary problems associated to the level sets. A divide and conquer strategy permits to build a very fast algorithm.

Concerning the joint optimization of phase or amplitude, these algorithms have no straightforward extension to vectorial cases.

2) Approximate minimization: Since TV is a metric, α-expansion algorithm proposed in [8] can be applied. Starting from a current solution, this algorithm proposes to each pixel either to keep its current gray-level, or to take value α as new gray-level. The energy associated to this movement is minimized using a graph-cut. The succession of α-expansions other all possible values in $\mathcal{L}$ until convergence leads to a solution which is shown to be close to the global minimum. This approach has been shown in [14] to converge to a global minimizer when data fidelity is convex. If the set of all possible values can be large in the case of single

image regularization, its size becomes prohibitive when joint regularization is considered. We suggest in the next section a faster algorithm which is more suitable when large images or joint regularization are considered.

Table I summarizes the graph size and number of cuts required by α-expansion, exact minimization and the proposed algorithm. Existing algorithms can not satisfactorily handle (in terms of speed and general applicability) vectorial data.

TABLE I

COMPARISON OF GRAPH-CUTS BASED TECHNIQUES FOR MINIMIZATION OF MRF ENERGIES WITH A NON-CONVEX DATA TERM AND A CONVEX REGULARIZATION. EXPRESSIONS ARE GIVEN FOR A N PIXELS IMAGE WITH L POSSIBLE LABELS (8 CONNEXITY NEIGHBORHOODS).

	α-expansion [8]	exact minimization [16], [27]	**proposed algorithm**
minimum	local	global	local
graph size	N nodes	$N \times L$ nodes	N nodes
	$4 \times N$ arcs	$5 \times N \times L$ arcs	$4 \times N$ arcs
number of cuts	$\propto L$	1	$\log_2(L/2)$
extension to D-channel vectorial data	$\propto L^D$ cuts	hardly possible	$(3^D - 1)\log_2(L/2)$ cuts

III. PROPOSED ALGORITHM

Minimizing a non-convex energy is a difficult task as the algorithm may fall in a local minimum. Algorithms such as the Iterated Conditional Modes require a "good" initialization and then performs local changes to reduce the energy. Graph-cut approach provides a way to explore a combinatorial set of changes involving simultaneously all pixels. Following [8], we denote such changes *large moves*. Instead of allowing a pixel to either keep its previous value or change it to a given one (α-expansion), we suggest that a pixel could either remain unchanged or its value be increased (or decreased) by a fixed step. Such an approach has first been described independently in [5], [13], [31] and applied recently with unitary steps in [5]. We however use these large moves in a case of non-convex data term. The trial steps are chosen to perform a scaling sampling of the set of possible pixel values. We express the algorithm in the general case of joint regularization.

We describe in the following subsections the set of large moves considered, the associated graph construction and give the average complexity of the resulting algorithm.

A. Local minimization

First, let us introduce the set of images that lie within a single move in our algorithm. For the sake of generality, we denote by $\hat{\mathbf{u}}$ the vectorial field arising by associating to each component one of the images to jointly regularize. Then,

$$\mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)}) = \{\hat{\mathbf{u}} \, / \, \forall s \in S, \exists k_s \in \{0,1\}, \hat{\mathbf{u}}_s = \hat{\mathbf{u}}_s^{(n)} + k_s \mathbf{d}\}$$

is the set of images whose pixel value $\hat{\mathbf{u}}_s$ is either unchanged or increased by step $\mathbf{d}$. We define the "best" move $\hat{\mathbf{u}}^{(n)} \mapsto \hat{\mathbf{u}}^{(n+1)}$ as the one that minimizes the restriction of the energy to the set $\mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)})$:

$$\hat{\mathbf{u}}^{(n+1)} = \underset{\hat{\mathbf{u}}^{(n+1)} \in \mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)})}{\arg\min} E(\hat{\mathbf{u}}^{(n+1)}|\mathbf{u}).$$

The restriction of the energy to $\mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)})$ corresponds to an energy involving only the binary variables $(k_s)_{s \in S}$. The "best" move is then obtained by finding the optimal binary variables $(k_s)_{s \in S}$:

$$\underset{(k_s)_{s \in S}}{\arg\min} \sum_s U(\mathbf{u}_s|\hat{\mathbf{u}}_s^{(n)} + k_s\mathbf{d}) + \beta \sum_{(s,t)} \psi(\hat{\mathbf{u}}_s^{(n)} + k_s\mathbf{d}, \hat{\mathbf{u}}_t^{(n)} + k_t\mathbf{d}) \tag{9}$$

with U and ψ respectively the opposite of the log likelihood and the prior energy as defined in the general MRF formulation in section II-A.

Let us define $U_b(\mathbf{u}_s|k_s) \equiv U(\mathbf{u}_s|\hat{\mathbf{u}}_s^{(n)} + k_s\mathbf{d})$ and $\psi_b(k_s, k_t) \equiv \psi(\hat{\mathbf{u}}_s^{(n)} + k_s\mathbf{d}, \hat{\mathbf{u}}_t^{(n)} + k_t\mathbf{d})$ to emphasize that, for a given step $\mathbf{d}$, the move $\hat{\mathbf{u}}^{(n)} \mapsto \hat{\mathbf{u}}^{(n+1)}$ depends only on those binary variables. According to [32], an energy of binary variables arising from a first-order Markov model can be minimized by computing a minimum cut on a related graph provided its prior energy ψ_b satisfies the following submodular property:

$$\psi_b(0,1) + \psi_b(1,0) \geq \psi_b(0,0) + \psi_b(1,1).$$

To compute the "best" move using a s-t minimum-cut algorithm, the restriction $\psi_b(k_s, k_t)$ of the prior energy to $\mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)})$ must be submodular. The prior energy ψ must therefore verify for all $\mathbf{u}_s^{(n)}$ and all $\mathbf{u}_t^{(n)}$:

$$\psi(\hat{\mathbf{u}}_s^{(n)}, \hat{\mathbf{u}}_t^{(n)} + \mathbf{d}) + \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)}) \geq \\ \psi(\hat{\mathbf{u}}_s^{(n)}, \hat{\mathbf{u}}_t^{(n)}) + \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)} + \mathbf{d}). \tag{10}$$

Note that in most cases, the prior model $\psi(\hat{\mathbf{u}}_s, \hat{\mathbf{u}}_t)$ depends only on the difference $\hat{\mathbf{u}}_s - \hat{\mathbf{u}}_t$. This is the case in the model described in section II-C. For such prior models, condition 10 becomes:

$$\psi(\hat{\mathbf{u}}_s^{(n)} - \hat{\mathbf{u}}_t^{(n)} - \mathbf{d}) + \psi(\hat{\mathbf{u}}_s^{(n)} - \hat{\mathbf{u}}_t^{(n)} + \mathbf{d}) \geq 2\psi(\hat{\mathbf{u}}_s^{(n)} - \hat{\mathbf{u}}_t^{(n)})$$

which is verified by any convex prior ψ.

In conclusion, the *local* problem of finding the vectorial field $\hat{\mathbf{u}}^{(n+1)}$ located within a single move (i.e. $\hat{\mathbf{u}}^{(n+1)} \in \mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)})$) that minimizes the posterior energy $E(\hat{\mathbf{u}}^{(n+1)}|\mathbf{u})$ can be *exactly* solved by computing a minimum cut on a graph (described in next paragraph) provided that the regularization potential is convex and depends only on the difference $\hat{\mathbf{u}}_s - \hat{\mathbf{u}}_t$.

The model we described in section II consists of the sum of a non-convex likelihood term and a convex prior term. The above property therefore holds for this model and we give in the next paragraphs an algorithm for approximate global minimization based on exact local minimizations performed using graph-cuts.

B. Graph construction

We build a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, following the method of [32], to minimize the restriction of the energy to allowed moves of step $\mathbf{d}$.

The graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is directed, with nonnegative edge weights and two terminal vertices: the source $\mathscr{S}$ and the sink $\mathscr{P}$. The graph structure and the edge weights are chosen such that any cut[1] has a cost (i.e. sum of the

---

[1]a cut is a partition of the vertices into two disjoint sets $\mathcal{S}$ and $\mathcal{P}$ such that $\mathscr{S} \in \mathcal{S}$ and $\mathscr{P} \in \mathcal{P}$
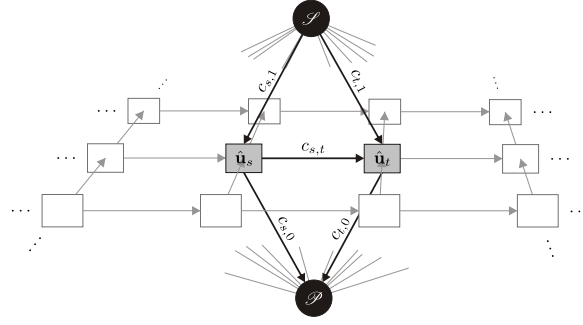
Fig. 2. Graph construction for local minimization.

cut edges capacities) corresponding to the energy to minimize. We create a vertice for each site s, all connected respectively to the source and the sink through two edges with capacity $c_{s,1}$ (resp. $c_{s,0}$). Finally, each clique (s,t) gives rise to an edge with capacity $c_{s,t}$ (fig. 2).

The capacities are set according to the additive method described in [32]. The first term in equation (9) is represented by the weights $c_{s,0}$ and $c_{s,1}$:

$$\begin{cases} c_{s,1} = \max(0, U(\mathbf{u}_s | \hat{\mathbf{u}}_s^{(n)} + \mathbf{d})) - U(\mathbf{u}_s | \hat{\mathbf{u}}_s^{(n)})) \\ c_{s,0} = \max(0, U(\mathbf{u}_s | \hat{\mathbf{u}}_s^{(n)}) - U(\mathbf{u}_s | \hat{\mathbf{u}}_s^{(n)} + \mathbf{d})). \end{cases}$$

To this weights are added the weights $c'_{.,.}$ (see notations in figure 2) representing each clique (second term of equation 9):

$$\begin{cases} c'_{s,1} & = & \beta \cdot \max\left(0, \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)}) - \psi(\hat{\mathbf{u}}_s^{(n)}, \hat{\mathbf{u}}_t^{(n)})\right) \\ c'_{s,0} & = & \beta \cdot \max\left(0, \psi(\hat{\mathbf{u}}_s^{(n)}, \hat{\mathbf{u}}_t^{(n)}) - \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)})\right) \\ c'_{t,1} & = & \beta \cdot \max(0, \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)} + \mathbf{d}) \\ & & -\psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)})) \\ c'_{t,0} & = & \beta \cdot \max(0, \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)}) \\ & & -\psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)} + \mathbf{d})) \\ c_{s,t} & = & \beta \cdot (\psi(\hat{\mathbf{u}}_s^{(n)}, \hat{\mathbf{u}}_t^{(n)} + \mathbf{d}) + \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)}) \\ & & -\psi(\hat{\mathbf{u}}_s^{(n)}, \hat{\mathbf{u}}_t^{(n)}) - \psi(\hat{\mathbf{u}}_s^{(n)} + \mathbf{d}, \hat{\mathbf{u}}_t^{(n)} + \mathbf{d})) \end{cases}$$

C. Approximate global minimization

When non-convex data terms such as Nakagami law described in section II-B1 are considered, the global minimization problem can not be exactly solved without considering each possible configuration (i.e. building a huge graph). On the other hand, when all terms are convex, it has been proven in [13] that a succession of local minimizations leads to the global minimum. An exploration based on different scalings of the step is then suggested to speed up convergence.

We follow here an heuristic method that combines the *exact* determination of the best moves, with no guarantee on how close to the global minimum we get. Section IV will illustrate on some simulated and real data that the obtained results are satisfying in practice with a speed and memory cost adequate for application use.

In one dimension, a scaling search is performed by looking for the best move with steps $d_i^+ = L/2^i$ and $d_i^- = -L/2^i$ for i from 1 to the desired precision (i.e. quantization level). In D dimensions, there are $3^D - 1$ vectorial steps $\mathbf{d}_i$ to consider for a given step size d_i:

$$\mathbf{d}_i \in \mathscr{S}(d_i) \stackrel{\text{def}}{=} \{0, -d_i, +d_i\}^D / \{0, \ldots, 0\}.$$

The joint-regularization algorithm is summarized here:

1: **for all** $s \in S$ **do**
2: $\quad \hat{\mathbf{u}}_s^{(0)} \leftarrow \{L/2, \ldots, L/2\}$
3: **end for**
4: $n \leftarrow 0$
5: **for** $i = 1$ to $precision$ **do**
6: $\quad d_i \leftarrow L/2^{i+1}$
7: $\quad$ **for all** $\mathbf{d}_i \in \mathscr{S}(d_i)$ **do**
8: $\qquad \hat{\mathbf{u}}^{(n+1)} \leftarrow \arg \min_{\hat{\mathbf{u}}^{(n+1)} \in \mathscr{S}_{\mathbf{d}}(\hat{\mathbf{u}}^{(n)})} E(\hat{\mathbf{u}}^{(n+1)} | \mathbf{u})$
9: $\qquad n \leftarrow n + 1$
10: $\quad$ **end for**
11: **end for**

Line 8 represents the exact binary energy minimization obtained by computing a minimum cut on a graph build according to section III-B. Note that if we perform unitary steps $\mathbf{d}_i \in \mathscr{S}(1)$ until convergence at the termination of our algorithm, exact minimization is then guaranteed for convex energies [13].

D. Complexity

The total number of cuts required by the algorithm depends on the precision chosen and on the number of jointly regularized images D. For a precision corresponding to the quantization level, the number of cuts is $(3^D - 1) \log_2(L/2)$. Joint regularization of the phase and the amplitude with 8 bit precision therefore requires 64 cuts, while the regularization of amplitude only or phase only is obtained after 16 cuts, to compare with respectively 65536 and 256 cuts with the α-expansion algorithm (see table I).

The algorithm we used to compute the cuts is Kolmogorov's freely available implementation of the augmenting path method described in [7]. For a $\mathcal{N}$ nodes and $\mathcal{E}$ edges graph, this algorithm has a high worst case complexity: $O(\mathcal{N}^2 \mathcal{E} |C|)$, with $|C|$ the cost of the cut. However, this algorithm performs well in practice for cuts arising from computer vision problems [7]. The complexity to expect on real data seems to be bounded by that of push-relabel algorithm ($O(\mathcal{N}^2 \sqrt{\mathcal{E}})$), which is the bound adopted in [5]. We will give the running times necessary for image regularization in our experiments conducted in section IV.

E. Hyper-parameter tuning

Hyper-parameter tuning is an essential issue as the regularized solution can be far from the true image if the hyper-parameters are incorrectly set. Depending on the target application (for example image enhancement prior to human photo-interpretation, or fully automatic image analysis), the optimal value of the hyper-parameter may be different. The range of possible values depends both on the log-likelihood term and on the prior term and is

very large. In the case of joint regularization, the hyper-parameters can differ by several orders of magnitude. An automatic method for adequate hyper-parameter estimation is therefore necessary.

Considerable effort has been devoted to hyper-parameter estimation [20], [29], [55]. One of the possible methods to perform hyper-parameter tuning is the analysis of the so-called L-curve [26]. This curve is the graphical representation of the regularization energy term with respect to the likelihood energy term. The corner of the curve corresponds to a good trade-off between under-regularization (steep part of the curve, where the regularization term can be largely improved with minor likelihood modification) and over-regularization (slowly varying part of the curve, where the regularization term can no longer be improved, whatever the likelihood price). Note however that the L-curve method is known to fail in some cases [25]. We successfully apply this method on simulated and real data in the next section.

IV. Experiments and algorithm comparison

A. Amplitude regularization

We evaluate here both the algorithm speed and the quality of the minimization on a synthetic image.

a) Algorithm comparison: Figure 3 compares the convergence of the ICM, α-expansion and the proposed algorithm on a simulated noisy image. The ground truth image consists of 4 regions denoted a, b, c and d on figure 3. Each region has a constant gray level (respectively 20, 40, 60, 80). The graph displays the energy decrease as a function of elapsed time computed on a laptop with a 2.4GHz Intel Core2 processor. The ICM is an example of an algorithm that involves local moves. It reaches convergence in about 30 seconds. Upon convergence, the obtained regularized image is far from the ground truth image and from the global minimum. Each iteration, depicted by a triangle on the graph, consists of a complete sweep of the image.

The α-expansion converges in 2 iterations for that example image (about 22s). The regularized image is close to the ground truth image, although a slight loss of contrast is visible. An iteration, also depicted by a triangle, consists of performing expansion moves for all possible labels. The energy of the image obtained at convergence is slightly less than that of the ground truth image. Ideally, the global minimum of the MRF model should correspond to the ground truth image. As illustrated with this example, this is not the case and the model minimization leads to a bias (further discussed in paragraph c).

The proposed algorithm gives an image almost identical to that obtained with α-expansion in less than 3 seconds. Its energy is also less than that of the ground truth image and slightly greater than that obtained with the α-expansion algorithm. Although the α-expansion and the proposed algorithm lead to a local minimum, the obtained regularized images are satisfactory. As the energy of these local minima is comparable to that of the ground truth image, these approximate minimization can be considered sufficient.

b) Automatic hyper-parameter estimation: The L-curve computed for β values in the range $[0, 1.2]$ is displayed in figure 4. As expected, the regularization term decreases as β is increased. As for the likelihood term, it increases with β. The two ends of the L-curve correspond to $\beta = 0$ (no regularization, null likelihood term) and $\beta \geq \beta_{lim}$ for which the regularized image is constant (null regularization term, maximum likelihood term). It has been empirically shown in [26] that the corner (i.e. maximum curvature point) of the L-curve gives a good regularization value β_{corner}. We have used the triangle method described in [9] to find automatically β_{corner} (depicted by a black triangle in figure 4). It seems that the use of a log-log scale for L-curve corner
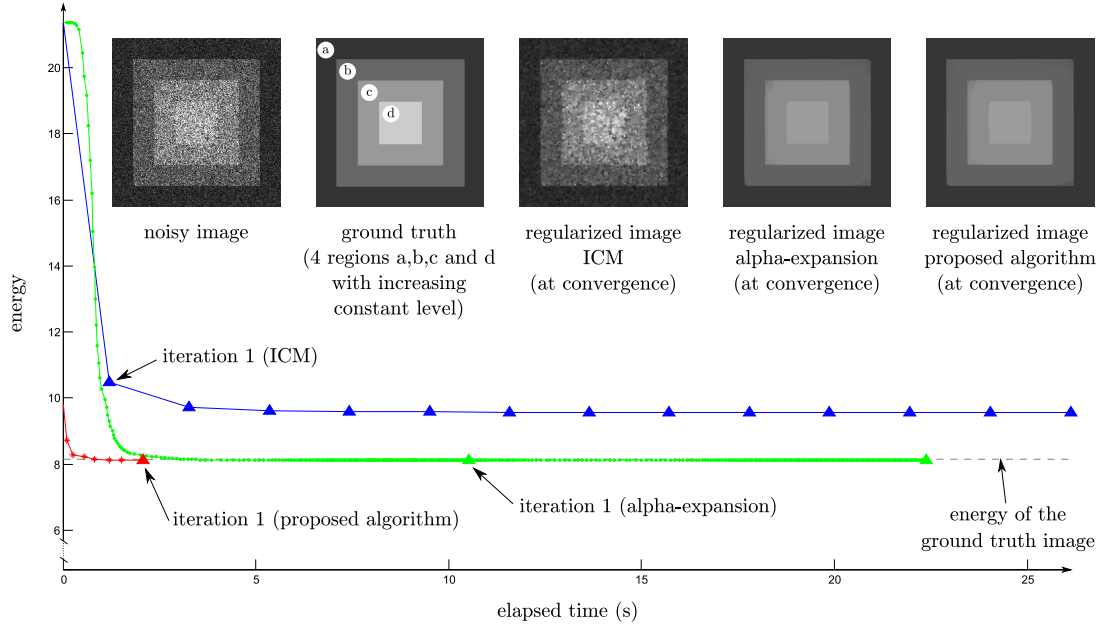
Fig. 3.   Convergence comparison of the Iterated Conditional Modes (ICM), α-expansion and the proposed algorithm.

TABLE II

BIAS/VARIANCE TRADEOFF: COMPARISON WITH SOME OTHER SPECKLE REDUCTION TECHNIQUES.

Region	Multi-look (11×11) (see text)				Lee (11×11) (ref [35])				Wu (ref [54])				**Proposed** ($\beta = 0.03$)				**Proposed** ($\beta = \beta_{corner}$)			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
Bias	-0.9	-1.1	-1.2	-3.5	-1.5	-0.8	-0.7	-2.9	-1.0	-1.4	-1.9	-4.9	-4.0	-2.7	-1.8	-4.8	1.0	-2.1	-5.3	-19.0
Standard deviation	4.7	3.7	4.5	5.0	5.6	4.3	5.1	6.4	3.8	7.1	9.1	12.2	5.7	4.9	3.7	3.7	0.02	0.8	1.0	0.5
MSE	23	15	22	37	33	19	26	49	15	52	86	173	49	32	17	37	1	5	29	363

detection as advised in [9] is less relevant when using TV regularization than it is for quadratic regularization . We therefore used linear scales as shown in figure 4. Three regularized images were computed for β values respectively less (sub-figure ①), equal (sub-figure ②) or greater (sub-figure ③) than β_{corner}. To enhance the details, we display the norm of the gradient of the regularized images (black means a high gradient norm, white is for null gradient) instead of the images themselves. Under- and over-regularization clearly correspond respectively to sub-figures ① and ③. The β_{corner} leads to a satisfying regularized image. The loss of contrast is visible in the change of gradient magnitude scale and increases with the value of the regularization hyper-parameter β.

   c) Bias/variance tradeoff: Table II compares denoised images obtained with different methods: multi-look filtering (averaging of the squared amplitude), Lee filtering [35], Wu's adaptive windows method [54] and the proposed regularization method. The multi-look and Lee filters are computed with a 11×11 window. All methods are applied to the noisy image shown in figure 3. The bias $< \hat{a} - \hat{a}_{true} >$, the mean square error
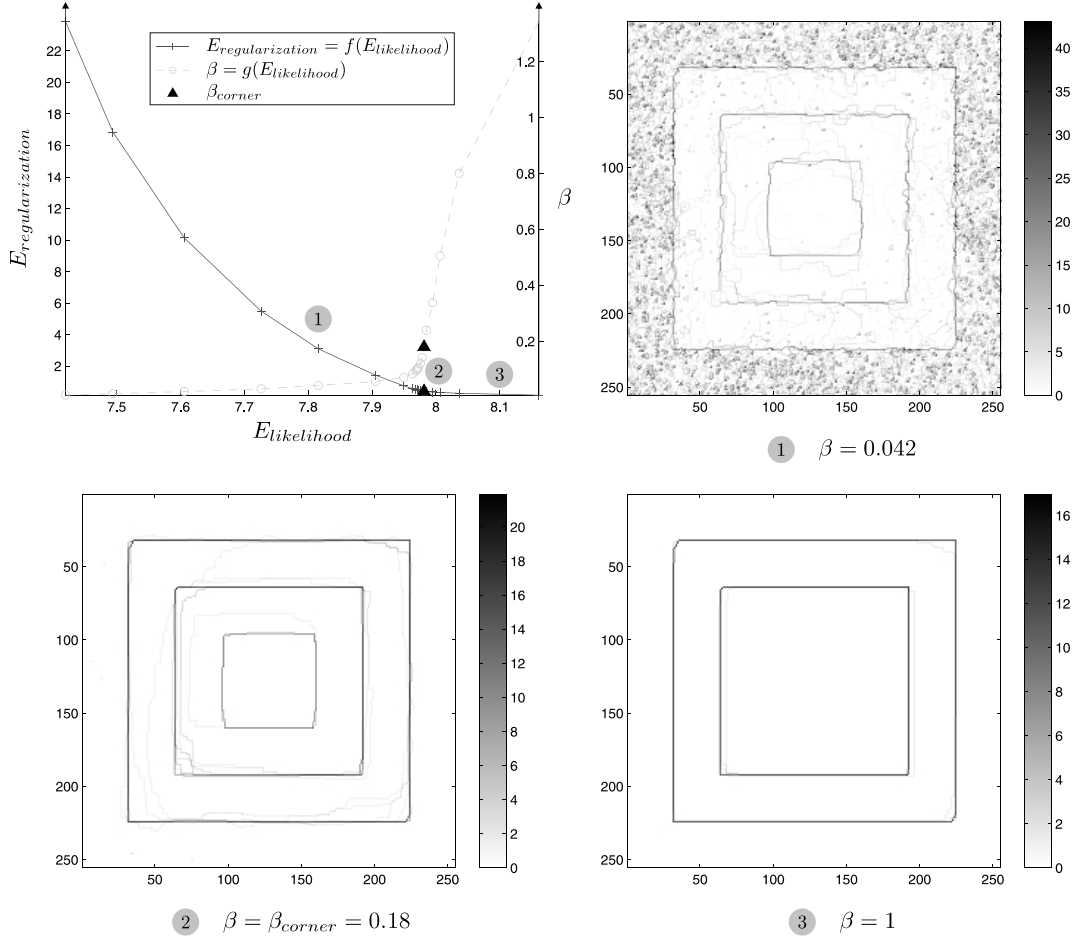
Fig. 4. Automatic hyper-parameter estimation: L-curve representation ($E_{regularization} = f(E_{likelihood})$) and corresponding β values. The detected β_{corner} value is displayed with a black triangle. (L-curve computation took less than 1 minute on this 256×256 image). Magnitude of the gradient of three images regularized with different β values are displayed to illustrate three different regions of the L-curve.

(MSE) $< (\hat{a} - \hat{a}_{true})^2 >$ and the standard deviation $\sqrt{MSE - bias^2}$ are computed[2] for the 4 homogeneous regions a,b,c and d. The proposed algorithm leads to images with very low variance (last column). Small regions with high reflectivity suffer however from a contrast loss (negative bias). This bias can be lowered to levels comparable to that of classical filters for under-regularized images at the cost of an increase of the variance. The MRF model considered penalizes variations and can therefore lead to images with uniform (flat) regions. This feature is of special interest in the context of urban areas. The bias that appears has been previously noticed and studied in the case of quadratic data terms (Gaussian likelihood) [41], [50]. We will show in section B and figure 7 that the contrast loss is limited when using joint regularization with the prior model designed in section II.

d) Level-dependent smoothing effect: It can be noticed from figure 4 that the image regions with high amplitude values tend to be smoothed first, while the noise in the low amplitude regions remains nearly unmodified for small values of the hyper-parameter ($\beta \ll \beta_{corner}$ as is the case for subfigure ①). This can

---

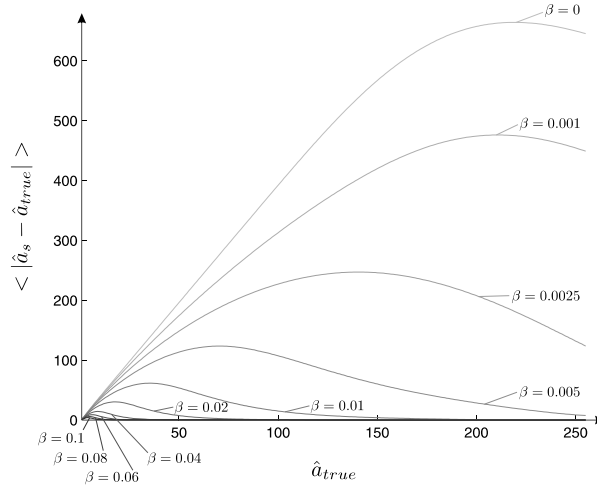[2]region "a" has been reduced to suppress boundary effects

Fig. 5. Expectation of the L1 error between the regularized value $\hat{a}_s$ and the (true) background value $\hat{a}_{true}$ as a function of $\hat{a}_{true}$. The curves were obtained for different regularization levels β. The limiting case $\beta = 0$ exhibits a linear part that illustrate the multiplicative nature of the noise. For high values of $\hat{a}_{true}$ truncation effects dominate the linear evolution (see text). When considering increasing β values, one can notice that the error is reduced more efficiently when the background level is high.

be intuitively understood by considering that speckle noise is multiplicative. Therefore, if we were to choose between two regularized values of equivalent likelihood in regions with different mean amplitude levels, the choice that would decrease most the global energy would be that which reduces the variations in the high amplitude region.

To study into more details this phenomenon, let us consider a constant region with amplitude $\hat{a}_{true}$. Due to the presence of noise, amplitude a is observed instead of $\hat{a}_{true}$ (a is considered to be a single-look image here: $M = 1$). The probability density function of a is given by equation 1. We are considering the filtered image $\hat{a}$ obtained by the MAP criterion using model of equation 5. Let us set the regularized values $\hat{a}_t$ of the neighbors of site s to the exact value $\hat{a}_{true}$ (i.e. $\forall t \in N_s, \hat{a}_t = \hat{a}_{true}$). We shall now consider the possible regularized values $\hat{a}_s$ at site s. In this specific case, $\hat{a}_s$ depends only on the noisy value a_s and the true value $\hat{a}_{true}$. The remaining L1 error after regularization is $\epsilon_s = |\hat{a}_s - \hat{a}_{true}|$. The expectation of this error is obtained by summing over all possible a_s values:

$$<\epsilon_s> \overset{\text{def}}{=} <|\hat{a}_s - \hat{a}_{true}|> = \sum_{a_s} |\hat{a}_s - \hat{a}_{true}| \cdot p(a_s|\hat{a}_{true}), \tag{11}$$

with $\hat{a}_s$ obtained by minimization

$$\hat{a}_s = \arg \min_{\hat{a}_s} E_s(\hat{a}_s|a_s, \hat{a}_{true})$$

of the local energy[3] E_s:

$$E_s(\hat{a}_s|a_s, \hat{a}_{true}) = U(a_s|\hat{a}_s) + 4\beta \cdot |\hat{a}_s - \hat{a}_{true}|.$$

Figure 5 represents the mean error $<\epsilon_s>$ as a function of the background level $\hat{a}_{true}$ for given regularization values β. These curves have been computed for integer values of $\hat{a}_{true}$ in $\mathcal{L}^* = [1, 255]$. Noisy amplitudes a_s

---

[3]defined here considering a 4 connexity neighborhood

have been sampled with 0.1 steps from 1 to 500 as the amplitude in SAR images is measured with a high dynamic (floating point values). For each triplet $(\beta,\ \hat{a}_{true}, a_s)$, $\hat{a}_s$ has been computed by searching for the minimum argument of E_s among integer values in range $\mathcal{L}^*$. By restricting the possible $\hat{a}_s$ values to the range $\mathcal{L}^*$, we introduce boundary effects. High values of $\hat{a}_{true}$ lead to noisy amplitudes a_s for which the energy E_s is minimized beyond the upper bound 255. Restricting $\hat{a}_s$ to lie within the range $\mathcal{L}^*$ moves the regularized values toward the true amplitude $\hat{a}_{true}$. The resulting error is therefore reduced, as can be noticed on the different curves for high values of $\hat{a}_{true}$ ($\hat{a}_{true} \gtrsim 200$).

The limiting case with no regularization ($\beta = 0$) corresponds to maximum likelihood estimation $\hat{a}_s = a_s$. The curve $\beta = 0$ is that of a linear function for $\hat{a}_{true}$ values where the boundary effects are negligible. This is the illustration of the multiplicative nature of speckle noise. As the regularization hyper-parameter β is increased, the linear dependency is not verified any more. The noise is then no more multiplicative and it can be observed, in agreement with our remark about figure 4 results, that the noise in high amplitude regions (i.e. high $\hat{a}_{true}$ values) is regularized more efficiently than that in low amplitude regions. This is related to the prior model we have chosen. For the application under consideration (recovery of urban structures), we find this model well adapted. For other purposes such as small targets detection in low signal-to-noise images, this model might be less suitable due to the risk of over-regularizing high-amplitude targets.

B. Joint regularization of InSAR images in urban area

We now consider joint regularization on high-resolution data acquired over the city of Toulouse, France. The images shown in figure 6(a) and 6(b) are 1200×1200 pixels extracts from single-pass interferometric SAR images acquired by RAMSES (ONERA SAR sensor) in X-band at sub-metric resolution.

The amplitude image is a 2-look image obtained after averaging the intensity of the two images of the interferometric pair. The interferogram has been computed on a 3×3 window and the coherence over detected shadow-areas set to 0.

We have set hyper-parameter γ to 1 and have estimated iteratively the values of β_a and β_ϕ using 1D L-curves: $\beta_a^{(0)}$ has been estimated considering an independent model. Then, $\beta_\phi^{(0)}$ has been estimated with β_a set to $\beta_a^{(0)}$ and $\hat{a}$ set to $\hat{a}^{(0)} = \hat{a}^{(MAP)}\big|_{\beta_a^{(0)}}$. The values $\beta_a^{(0)}$ and $\beta_\phi^{(0)}$ have then been refined respectively into $\beta_a^{(1)}$ and $\beta_\phi^{(1)}$ given $\hat{\phi}^{(0)} = \hat{\phi}^{(MAP)}\big|_{\beta_a^{(0)}, \beta_\phi^{(0)}, \hat{a}^{(0)}}$ (resp. $\hat{a}^{(1)} = \hat{\phi}^{(MAP)}\big|_{\beta_a^{(1)}, \beta_\phi^{(0)}, \hat{\phi}^{(0)}}$). Although this iterative refinement process could be carried on, the values $\beta_a^{(1)}$ and $\beta_\phi^{(1)}$ already provide satisfying regularization results. We obtained $\beta_a^{(1)} = 0.129$ and $\beta_\phi^{(1)} = 3.64 \times 10^3$ for the images shown in figure 6(a) and figure 6(b). The jointly regularized images are displayed in figure 6(c) and figure 6(d). The regularization process (given the hyper-parameter values) took less than 3 minutes with our implementation of the algorithm of section III. The hyper-parameters were determined using a 232×232 pixels sub-image as this step requires many regularization computations. Note that the hyper-parameters differ by 4 orders of magnitude, which would have made their manual tuning inconvenient. More subtle approaches have also been suggested to determine multiple regularization parameters (see reference [3]).
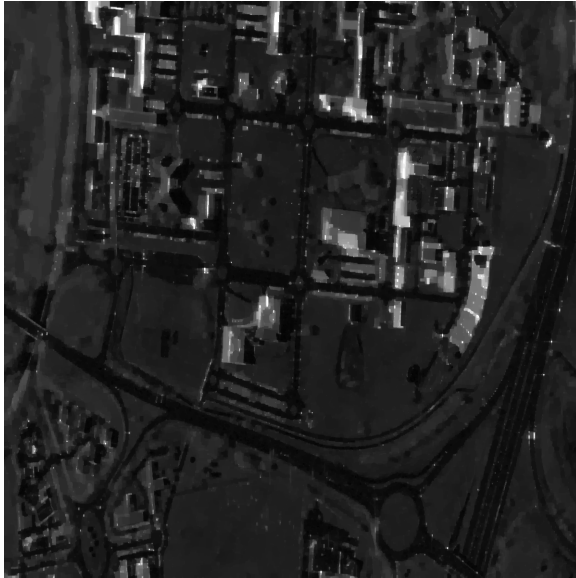
From the regularization results of figure 6 it can be noticed that the noise has been efficiently reduced both in amplitude and phase images. The sharp transitions in the phase image that correspond to man-made structures are well preserved.

Fig. 6. Joint regularization of InSAR images (1200×1200 pixels): (a) noisy amplitude, (b) noisy phase, (c) and (d) are respectively the jointly regularized amplitude and phase images for $\beta_a = 0.129$ and $\beta_\phi = 3.64 \times 10^3$ determined automatically with L-curves.

Joint regularization gives more precise contours than independent regularization as they are co-located from the phase and amplitude images (minimum cost images have transitions that occur between the same neighboring pixels). Small objects also tend to be better preserved by joint-regularization as illustrated in figure 7. In this figure, an excerpt showing a portion of streets is presented. Four dots (roughly vertically aligned) are visible in the noisy phase image and less clearly in the amplitude image. They correspond to the top of streetlights that is higher than the surrounding ground. In the independently regularized phase image $\hat{\phi}'$, some streetlights have nearly disappeared (see also the gradient image $|\nabla\hat{\phi}'|$ shown to ease visualization). In the jointly regularized
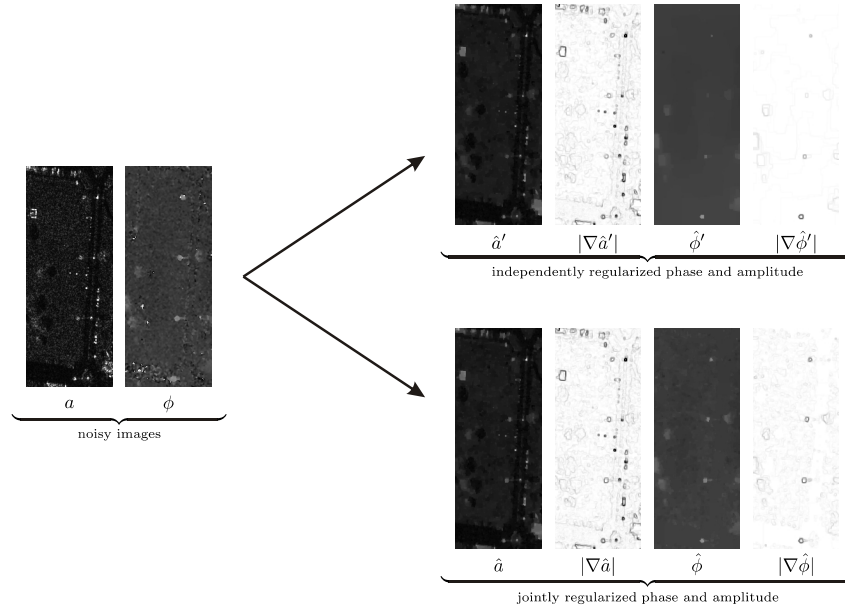
Fig. 7.   Illustration of the use of joint regularization to preserve small objects. A magnified portion of images displayed in figure 6, centred on a street, shows small round objects that corresponds to streetlights. They are noticeable both in amplitude (larger reflectivity than the ground) and phase (top of the streetlight is higher than surrounding ground). When independent regularization of phase and amplitude is performed, the true phase is lost for some of the streetlights that are merged with the ground by the regularization process. The streetlights are correctly preserved when regularization is jointly performed (bottom row of images).

phase image $\hat{\phi}$ the 4 streetlights are still visible, with comparable contrast from one another. The amplitude image, in which the streetlights are also present, has helped preserve these small objects. Note that the location of the contours in the jointly regularized images exactly coincide. As they are obtained in order to match both the amplitude and phase information, they are more precise than if independently set.

Note however that precise (and fair) comparison between joint and independent regularization is difficult to carry out as the values of the hyper-parameters are not directly related (since the models differ). The regularized images shown in figure 7 have been computed using hyper-parameter values obtained using the same L-curve procedure to reduce as much as possible this problem.

V. CONCLUSION

Speckle noise can be effectively reduced in SAR images with a Markov Random Field approach. TV minimization results in smoothed homogeneous regions while preserving sharp transitions. The Markovian formulation provides a convenient way to incorporate priors and to perform joint regularization. We have shown on real data that this can help to prevent over-regularization effects of objects that are visible in different images (such as amplitude and interferometric phase). Moreover, the contours of the jointly regularized images are more precise as all information is merged.

Heavy-tailed distributions such as Nakagami law that governs SAR amplitude lead to non-convex likelihood terms. The underlying minimization problem for MAP estimation is therefore difficult and many local minimum are present. Graph-cuts offer an efficient approach for these optimization problems. Although graph-cut-based algorithms that exactly minimize the target energy are known, they can hardly be applied in practice due to

computational and memory constraints. We derived a minimization algorithm suitable for (joint) regularization of large images.

The regularized images obtained both on synthetic and real data were satisfying. The algorithm is faster than existing graph-cut-based techniques. We have shown that joint regularization can be performed with little computation overload. It helps preventing loss of small objects (over-regularization) by merging all information.

The regularization prior chosen (TV) is responsible for the loss of contrast observed on the regularized images. Contrast loss is a known issue of MRF models with quadratic data fidelity and TV prior [41], [50]. We have observed that this contrast loss is also present in the case of speckle noise (Nakagami distribution). Under-regularization leads to better contrast preservation and a tradeoff between the bias and variance of the regularized image must be found depending on the target application. Moreover, the joint regularization scheme we have proposed better preserves the contrast compared to the channel independent prior scheme. Other MRF models could be considered to overcome this contrast issue provided the prior energy remains convex as required by our minimization algorithm. Another option would be to extend the iterative contrast restoration method of [45] to the case of speckle noise. This technique relies on solving a series of TV minimization problems. Its extension from the additive noise case with symmetrical log-likelihood to the case of multiplicative noise (asymmetrical log-likelihood) is however not straightforward. Another attempt has been proposed in [17] using the concept of levelable functions but it requires to perform a non-convex optimization and some deeper analysis still remains to be done.

The quality of the results could be improved for 3D urban modeling by introducing more elaborated prior knowledge in combination with contextual interpretation of the urban scene. The MRF model is flexible enough to incorporate higher level prior models. Including radar geometric deformations compensation in the regularization process could be an interesting step toward successful use of the regularized images.

REFERENCES

[1] F. Argenti and L. Alparone. Speckle removal from SAR images in the undecimated wavelet domain. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2363–2374, 2002.

[2] G. Aubert and J.F. Aujol. A Variational Approach to Removing Multiplicative Noise. *SIAM Journal on Applied Mathematics*, 68:925, 2008.

[3] M. Belge, M.E. Kilmer, and E.L. Miller. Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems*, 18(4):1161–1183, 2002.

[4] J. Besag. On the statistical analysis of dirty pictures. *J. R. Statist. Soc. B*, 48(3):259–302, 1986.

[5] J. M. Bioucas-Dias and G. Valad ao. Phase unwrapping via graph cuts. *IEEE Transactions on Image Processing*, 16(3):698–709, 2007.

[6] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[7] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, 2004.

[8] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2001.

[9] J.L. Castellanos, S. Gómez, and V. Guerra. The triangle method for finding the corner of the L-curve. *Applied Numerical Mathematics*, 43(4):359–373, 2002.

[10] A. Chambolle. Total variation minimization and a class of binary mrf models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume LNCS 3757, pages 136 – 152, St. Augustine, Florida, USA, 2005.

[11] P. Charbonnier. *Reconstruction d'image : régularisation avec prise en compte des discontinuités*. PhD thesis, Université de Nice Sophia Antipolis, 1994.

[12] P.L. Combettes and J.C. Pesquet. Image restoration subject to a total variation constraint. *IEEE Transactions on Image Processing*, 13(9):1213–1222, 2004.

[13] J. Darbon. *Composants logiciels et algorithmes de minimisation exacte d'énergies dédiées au traitement des images*. PhD thesis, Ecole Nationale Supérieure des Télécommunications (ENST E050), 2005.

[14] J. Darbon and S. Peyronnet. A vectorial self-dual morphological filter based on total variation minimization. In *Proceedings of the First International Conference on Visual Computing*, volume 3804 of *Lecture Notes in Computer Science Series*, pages 388–395, Lake Tahoe, Nevada, USA, December 2005. Springer-Verlag.

[15] J. Darbon and M. Sigelle. Image restoration with discrete constrained Total Variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276, December 2006.

[16] J. Darbon and M. Sigelle. Image restoration with discrete constrained Total Variation part II: Levelable functions, convex priors and non-convex cases. *Journal of Mathematical Imaging and Vision*, 26(3):277–291, December 2006.

[17] J. Darbon, M. Sigelle, and F. Tupin. The use of levelable regularization functions for MRF restoration of SAR images while preserving reflectivity. In SPIE, editor, *IS&T/SPIE 19th Annual Symposium Electronic Imaging*, volume E 112, 2007.

[18] S. Foucher, G. Bertin Bénié, and J.-M. Boucher. Multiscale MAP filtering of SAR images. *IEEE Transactions on Image Processing*, 10(1):49–60, 2001.

[19] V. S. Frost, J. Abbott Stiles, K. S. Shanmugan, and J. C. Holtzman. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4(2):157–166, March 1982.

[20] N.P. Galatsanos and A.K. Katsaggelos. Methods for choosing the regularization parameter and estimatingthe noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, 1(3):322–336, 1992.

[21] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-14(3):367–383, 1992.

[22] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[23] J.W Goodman. Statistical properties of laser speckle patterns. In *Laser Speckle and Related Phenomena*, volume 9, pages 9–75. J.C Dainty (Springer Verlag, Heidelberg, 1975), 1975.

[24] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *J. R. Statist. Soc. B*, 51(2):271–279, 1989.

[25] P.C. Hansen, T.K. Jensen, and G. Rodriguez. An adaptive pruning algorithm for the discrete L-curve criterion. *Journal of Computational and Applied Mathematics*, 198(2):483–492, 2007.

[26] P.C. Hansen and D.P. O'Leary. The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems. *SIAM Journal on Scientific Computing*, 14:1487, 1993.

[27] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, October 2003.

[28] E. Jakeman. On the statistics of K-distributed noise. *J. Phys. A: Math. Gen.*, 13:31–48, 1980.

[29] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia. Hyperparameter estimation for satellite image restoration using a MCMC maximum-likelihood method. *Pattern Recognition*, 35(2):341–352, 2002.

[30] J. K. Jao. Amplitude distribution of composite terrain radar clutter and the K-distribution. *IEEE Transactions on Antennas and Propagation*, AP-32(10):1049–1062, October 1984.

[31] V. Kolmogorov. Primal-dual algorithm for convex markov random fields. Technical report, Microsoft Research, 2005.

[32] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph-cuts ? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(2), 2004.

[33] Kuan, Sawchuk, Strand, and Chavel. Adaptive restauration of images with speckle. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(3):373–383, 1987.

[34] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive noise smoothing filter for images with signal dependant noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2):165–177, March 1985.

[35] J-S Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(2):165–168, March 1980.

[36] J.-S. Lee. Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17:24–32, 1981.

[37] J. S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller. Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1017–1028, 1994.

[38] J. S. Lee, I. Jurkevich, P. Dewaele, P. Wamback, and A. Oosterlinck. Speckle filtering of synthetic aperture radar images : A review. *Remote Sensing Reviews*, 8:313–340, 1994.

[39] F. Lombardini, F. Bordoni, and F. Gini. Feasibility study of along-track sar interferometry with the cosmo-skymed satellite system. In *IGARSS04*, volume 5, pages 337–3340, 2004.

[40] A. Lopes, E. Nezry, R. Touzi, and H. Laur. Structure detection, and statistical adaptive filtering in SAR images. *Int. J. Remote Sensing*, 14(9):1735–1758, 1993.

[41] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society, 2001.

[42] J.M. Nicolas. A Fisher-MAP filter for SAR image processing. In *IGARSS03*, volume 3, Toulouse, France, jul 2003.

[43] J.M. Nicolas, F. Tupin, and H. Maître. Smoothing speckle SAR images by using maximum homogeneous region filters: an improved approach. In *IGARSS01*, volume 3, pages 1503–1505, Sydney, Australie, jul 2001.

[44] M. Nikolova. A variational approach to remove outliers and impulse no ise. *Journal of Mathematical Imaging and Vision*, 20:99–120, 2004.

[45] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An Iterative Regularization Method for Total Variation Based Image Restoration. *SIAM journal on Multiscale modeling and Application*, 4:460–489, 2005.

[46] J.P. Picard and H.D. Ratlif. Minimum cuts and related problem. *Networks*, 5:357–370, 1975.

[47] R. Romeiser and H. Runge. Theoretical evaluation of several possible along-track InSAR modes of TerraSAR-X for ocean current measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 45:21–35, 2007.

[48] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D.*, 60:259–268, 1992.

[49] J. Shi and S. Osher. A nonlinear inverse scale space method for a convex multiplicative noise model. Technical Report 07-10, UCLA, 2007.

[50] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19(6), 2003.

[51] R. Touzi. A review of speckle filtering in the context of estimation theory. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2392–2404, 2002.

[52] G. Vasile, E. Trouvé, J.S. Lee, and V. Buzuloiu. Intensity-driven adaptive neighborhood technique for polarimetric and interferometric SAR parameters estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1609–1621, 2003.

[53] M. Walessa and M. Datcu. Model-based despeckling and information extraction of SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5), 2000.

[54] Y. Wu and H. Maître. Smoothing speckled synthetic aperture radar images by using maximum homogeneous region filters. *Optical Engineering*, 31(8):1785–1792, 1992.

[55] Z. Zhou, R.N. Leahy, and J. Qi. Approximate maximum likelihood hyperparameter estimation for Gibbs priors. *IEEE Transactions on Image Processing*, 6(6):844–861, 1997.

Loïc Denis is assistant professor at Ecole Supérieure de Chimie Physique Electronique de Lyon (CPE Lyon) since 2007. He was a postdoc at Télécom Paristech in 2006-2007 working on 3D reconstruction from interferometric synthetic aperture radar and optical data. His research interests include image denoising and reconstruction, radar image processing, and digital holography.

Florence Tupin received the engineering degree from Ecole Nationale Supérieure des Télécommunications (ENST) of Paris in 1994, and the Ph.D. degree from ENST in 1997. She is currently Professor at TELECOM ParisTech in the TSI (Image and Signal Processing) Department. Her main research interests are image analysis and interpretation, 3D reconstruction, Markov random field techniques, and synthetic aperture radar, especially for urban remote sensing applications.

Jérôme Darbon studied computer science at Ecole Pour l'Informatique et les Techniques Avanées (EPITA) in Paris and received the Ph.D. degree from Ecole Nationale des Télécommunications de Paris in 2005. He is currently a postdoc at the Department of Mathematics, University of California Los Angeles.
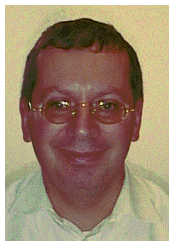
Marc Sigelle was born in Paris on 18th March 1954. He received an engineer diploma from Ecole Polytechnique Paris 1975 and from Ecole Nationale Supérieure des Télécommunications Paris in 1977. In 1993 he received a PhD from Ecole Nationale Supérieure des Télécommunications. He worked first at Centre National d'Etudes des Télécommunications in Physics and Computer algorithms. Since 1989 he has been working at Ecole Nationale Supérieure des Télécommunications in image and more recently in speech processing. His main subjects of interests are restoration and segmentation of signals and images with Markov Random Fields (MRF's), hyperparameter estimation methods and relationships with Statistical Physics. His interests concerned first reconstruction in angiographic medical imaging and processing of remote sensed satellital and synthetic aperture radar images, then speech and character recognition using MRF's and bayesian networks. His most recent interests concern a MRF approach to image restoration with Total Variation and its extensions.