Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights

Charles-Alban Deledalle, *Student Member, IEEE,* Loïc Denis and Florence Tupin, *Senior Member, IEEE*

***Abstract*—Image denoising is an important problem in image processing since noise may interfere with visual or automatic interpretation. This paper presents a new approach for image denoising in the case of a known uncorrelated noise model. The proposed filter is an extension of the Non Local means (NL means) algorithm introduced by Buades *et al.* [1], which performs a weighted average of the values of similar pixels. Pixel similarity is defined in NL means as the Euclidean distance between patches (rectangular windows centered on each two pixels). In this paper a more general and statistically grounded similarity criterion is proposed which depends on the noise distribution model. The denoising process is expressed as a weighted maximum likelihood estimation problem where the weights are derived in a data-driven way. These weights can be iteratively refined based on both the similarity between noisy patches and the similarity of patches extracted from the previous estimate. We show that this iterative process noticeably improves the denoising performance, especially in the case of low signal-to-noise ratio images such as Synthetic Aperture Radar (SAR) images. Numerical experiments illustrate that the technique can be successfully applied to the classical case of additive Gaussian noise but also to cases such as multiplicative speckle noise. The proposed denoising technique seems to improve on the state of the art performance in that latter case.**

***Index Terms*—Image denoising, Non Local means (NL means), Weighted Maximum Likelihood Estimation (WMLE), Patch-Based methods, Synthetic Aperture Radar (SAR)**

I. Introduction

IMAGE denoising is a key pre-processing step in many cases, e.g., low-light or high-speed imaging, low-cost sensor usage in embedded systems, or also echographic, sonar and radar coherent imaging techniques. Noise can generally be well modeled with parametric distributions, either grounded on physical or empirical considerations. Denoising then amounts to estimate the space-varying value of these parameters. Numerous denoising techniques have been proposed in the image processing literature. The majority of them consider an additive white Gaussian noise model. Others have been specifically designed for non-Gaussian and/or non-additive noise models. Few denoising techniques provide general methodology that apply to different noise models[1].

Manuscript received January 14, 2009; revised May 14, 2009.

C. Deledalle and F. Tupin are with Institut Telecom, Telecom ParisTech, CNRS LTCI, Paris, FRANCE, e-mail: charles-alban.deledalle@telecom-paristech.fr and florence.tupin@telecom-paristech.fr.

L. Denis is with École Supérieure de Chimie Physique Électronique de Lyon and Laboratoire Hubert Curien, CNRS UMR 5516, ST-ÉTIENNE, e-mail: loic.denis@cpe.fr.

[1]Markov random fields are representative of such approaches.

The crudest denoising approach is the direct application of the spatial coherence principle. It considers noisy samples in a window centered on a given pixel as all following the distribution of that pixel. This leads to the moving average for Gaussian noise distributions, or more generally to maximum likelihood estimates such as the multi-look averaging classically used in SAR image community [2]. The isotropy of the filter can be improved by considering a weighted average with a circular symmetric kernel such as a Gaussian kernel of given standard deviation. The fundamental limitation of those techniques comes from the loss of resolution in the filtered image. The same smoothing effect is applied equally to homogeneous regions, and to edges or textured zones. More elaborate denoising techniques all aim at better preserving image structures (edges, texture) while suppressing the noise. Very different strategies have been proposed to achieve this goal:

Markov Random Field (MRF) approaches introduce a prior model of the noise-free image and search for a compromise between that prior and noisy data. It is crucial for these techniques to define a suitable prior that guarantees both the smoothness of the denoised image and the preservation of its structure. Total variation [3] is an example of a prior that enforces smoothness while preserving edges. Priors however tend to bias the denoised image, especially when high noise levels are considered. Markovian priors are *local* in essence, and lead to stronger attenuation of several small disconnected regions than that of a single region [4]. In practice, edge-preserving MRF models generally lead to minimization problems with non-smoothness and/or non-convexity issues.

It has been shown by Donoho and Johnstone [5] that spatial adaptation of the smoothing can be obtained by wavelet soft-thresholding. Wavelet shrinkage can be interpreted from a Bayesian perspective as resulting from a ℓ_1 sparsity-promoting prior. Wavelet and cosine bases are indeed well known to be able to capture most of a signal or image in a few coefficients. This property is exploited by compression techniques such as JPEG and JPEG2000. For denoising applications, orthogonal transforms like the wavelet or discrete cosine transforms lead to a separation of signal and noise. Noise can then be strongly suppressed by zeroing the least significant coefficients. Such approaches can be applied to additive Gaussian noise [6], [7] and have been extended to multiplicative speckle noise [8]–[13]. They can be improved by using shape-adaptive domains [14], and sparse decompositions with over-complete or learned dictionaries [15]–[17].

Spatial adaptivity can also be reached by considering shape-adaptive windows [18]–[20] or spatially variable bandwidth selection [21]. Windows can naturally be generalized by weights. In the framework of weighted averaging, it has been proposed in the variants of the bilateral filter [22]–[25] to use the gray level difference between the central pixel and each neighboring pixel to define the corresponding weights. More recently, Buades *et al.* proposed a more selective data-driven way to set the weights [1], based on the Euclidean distance between the patch that surrounds the central pixel and the one that surrounds a given neighboring pixel. This patch-based filter can be considered *non-local* as pixel values far apart can be averaged together, depending on the weight values (i.e., surrounding patch similarity). The Non Local means (NL means) principle has lead to several extensions and modifications that will be considered in the following paragraphs. The best denoising techniques so far for additive Gaussian noise [26]–[28] consist of a combination of the ideas of similar patches selection with that of sparsification by transforms or learned dictionaries. We refer the reader to the very recent survey by Katkovnik *et al.* [29] for a deeper analysis of the connections and evolutions of all denoising approaches we mentioned here.

We describe in this paper a general methodology for patch-based denoising and its application to additive Gaussian and multiplicative speckle noises. The NL means algorithm [1] appears as a special case. We give in Section II an interpretation of our denoising technique in the framework of weighted maximum likelihood. The similarity between noisy patches is defined from the noise distribution. We suggest in Section III to refine iteratively the obtained weights by also including the similarity between restored patches. This leads to an iterative algorithm (Section IV) which is then compared in Section V with the state of the art techniques for Gaussian or speckle noise suppression.

II. Patch-based Weighted Maximum Likelihood

This section introduces the proposed denoising method in the framework of Weighted Maximum Likelihood Estimation (WMLE) investigated in [30]. Contrary to Polzehl and Spokoiny in [30], we define the weights following a statistically grounded patch-based approach.

A. Weighted Maximum Likelihood Estimator

We consider image denoising as an estimation $\hat{u}$ of the "true" image u^* from noisy data v. The images are considered to be defined over a discrete regular grid Ω and we denote by v_s a pixel value at site $s \in \Omega$. We consider an uncorrelated noise model defined by a parametric noise distribution $p(v_s|\theta_s^*)$ (namely the likelihood), with θ_s^* a space-varying unknown parameter[2]. Then, denoising an image is assumed to be equivalent to find the best estimate $\hat{\theta}$ of θ^*.

At each site s, the Maximum Likelihood Estimator (MLE) defines an estimate $\hat{\theta}_s$ of the underlying parameter θ_s^* from a set $\mathcal{S}_{\theta_s^*}$ of independent and identically distributed random variables by:

$$\begin{aligned}\hat{\theta}_s^{(MLE)} &\triangleq \arg\max_{\theta_s} \sum_{t \in \mathcal{S}_{\theta_s^*}} \log p(v_t|\theta_s) \\ &= \arg\max_{\theta_s} \sum_t \delta_{\mathcal{S}_{\theta_s^*}}(t) \log p(v_t|\theta_s),\end{aligned}$$

with $\delta_{\mathcal{S}_{\theta_s^*}}$ the indicator function of $\mathcal{S}_{\theta_s^*}$ (i.e., $\delta_{\mathcal{S}_{\theta_s^*}}(t) = 1$ if $t \in \mathcal{S}_{\theta_s^*}$, 0 otherwise). The MLE is unbiased and asymptotically efficient. In practice, the sets $\mathcal{S}_{\theta_s^*}$ for each $s \in \Omega$ are unknown. Hence, we only approximate $\delta_{\mathcal{S}_{\theta_s^*}}(t)$ by data-driven weights $w(s,t) \geq 0$. This leads to the Weighted Maximum Likelihood Estimation (WMLE) given by

$$\hat{\theta}_s^{(WMLE)} \triangleq \arg\max_{\theta_s} \sum_t w(s,t) \log p(v_t|\theta_s). \quad (1)$$

WMLE is known to reduce the mean squared error by reducing the variance of the estimate at the cost of a bias introduced by samples that follow a distribution with a parameter θ_t^* different to θ_s^* [31]. The WMLE framework has first been applied to image denoising by Polzehl and Spokoiny [30].

As shown in Section IV, in the particular case of additive white Gaussian noise (WGN) model[3], the corresponding WMLE estimate is defined by a Weighted Averaging (WA):

$$\hat{\theta}_s^{(WA)} \triangleq \frac{\sum_t w(s,t) v_t}{\sum_t w(s,t)}. \quad (2)$$

This is consistent with the numerous denoising methods existing in image processing and based on a weighted averaging. The weights used to approximate the indicator function can be seen as membership values over a fuzzy set version of $\mathcal{S}_{\theta_s^*}$ (with proper weight normalization). This fuzzy set introduces a bias in the estimation since similar noisy values coming from different distributions are incorporated. However, this drawback is counterbalanced by decreasing the variance of the estimation. Actually, more pixel values are included in the fuzzy set which decreases the variance of the estimation (note that for pixel values defined on a continuum, the probability measure $P(\theta_s = \theta_p)$ is zero, which means that we almost never find two pixels following the same distribution, thus we do not average pixel values therefore leaving the noisy image unchanged). According to this bias-variance trade-off, WMLE can outperform MLE for well-chosen weights. That is the purpose of the next section.

B. Setting the Weights between Noisy Patches

The definition of the weights $w(s,t)$ is the main problem addressed here. As noted in [30], a well-chosen definition of the weights constitutes the key to the success of WMLE filters. Under ergodic process assumption, $w(s,t)$ can be defined locally in the neighborhood of the site s. That is the case of the Box filter (also known as multi-look filter in the context of SAR images processing [2]) and the Gaussian filter. The local neighborhood is fixed by the weights $w(s,t)$ which increase when the sites s and t are closer. Unfortunately, this kind of filter is inappropriate to denoise singular features such as edges

[2]parameter θ^* may differ from u^* but depends deterministically on u^*, as does SAR reflectivity with respect to SAR amplitude in Section IV-B

[3]and more generally, for noise distributions in the exponential family [30]

$$\frac{p(v_{s,k},\ v_{t,k},\ \theta^*_{s,k}\ =\ \theta^*_{t,k})}{p(\theta^*_{s,k}\ =\ \theta^*_{t,k})} = \frac{\int p(\ v_{s,k},\ v_{t,k} \mid \theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta\ ) p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)\ d\theta}{\int p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)\ d\theta} \tag{5}$$

and textures for which the ergodicity assumption is invalid. Instead of defining $w(s,t)$ in spatial domain, Yaroslavsky proposed a data-driven weight definition based on gray level scale [22] also known as sigma-filter [24]. The weight $w(s,t)$ increases when the values v_s and v_t are more similar. Such a filter was then refined by the SUSAN filter [23] and the bilateral filter [25] which is defined both in spatial and gray level scales.

More recently, Buades *et al.* proposed the Non-Local (NL) means filter which relies on image redundancy [1]. It takes inspiration on the patch-based approach proposed for texture synthesis by Efros and Leung [32]. The weight $w(s,t)$ is defined by comparing two patches Δ_s and Δ_t centered respectively around the sites s and t:

$$w(s,t)^{(NL)} \triangleq \exp\left(-\frac{1}{h}\sum_k \alpha_k |v_{s,k} - v_{t,k}|^2\right) \tag{3}$$

where $v_{s,k}$ and $v_{t,k}$ are respectively the k-th neighbor in the patch Δ_s and Δ_t, the weights α_k define a centered symmetric Gaussian kernel and h controls the decay of the exponential function. The similarity is expressed by a weighted Euclidean distance over the two windows. This is well-adapted for additive WGN models.

The Probabilistic Patch-Based (PPB) filter aims to define a suitable patch-based weight to generalize the Euclidean distance based weight used in the NL means algorithm. The idea is to extend the NL means algorithm to non additive WGN models. According to the previous comments, the weights can be seen as a membership value over the fuzzy set version of $\mathcal{S}_{\theta^*_s} = \{t | \theta^*_t = \theta^*_s\}$. In a probabilistic patch-based approach, we express this weight by the probability, given the noisy image v, that the two patches Δ_s and Δ_t have the same parameters. We follow the same idea as that of the NL-means and assume equal values for the central pixel of two statistically close image patches. The patch-based similarity probability involves the following weight definition:

$$w(s,t)^{(PPB)} \triangleq\ p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t} | v)^{1/h} \tag{4}$$

where $\theta^*_{\Delta_s}$ and $\theta^*_{\Delta_t}$ denote the sub-image extracts from the parameter image θ^* in the respective windows Δ_s and Δ_t, and $h > 0$ is a scalar parameter. The h parameter is similar to that of the NL means algorithm and acts on the size of the fuzzy set (i.e., the number significant weights) to control the amount of filtering. Kervrann *et al.* justify the introduction of parameter h by saying that it probably counterbalances the invalidity of the patch independence assumption [33]. Under an independence assumption on the pixel of the patches, the similarity probability may be decomposed into a product $\prod_k p(\theta^*_{s,k}\ =\ \theta^*_{t,k} | v_{s,k}, v_{t,k})$. In a Bayesian framework, without knowledge on $p(\theta^*_{s,k} = \theta^*_{p,k})$ and $p(v_{s,k}, v_{t,k})$, the probability $p(\theta^*_{s,k}\ =\ \theta^*_{t,k} \mid v_{s,k}, v_{t,k})$ can be considered proportional to the likelihood $p(v_{s,k},\ v_{t,k} \mid \theta^*_{s,k}\ =\ \theta^*_{t,k})$. According to Equation 5 (at the top of the page) and under independence assumption on $v_{s,k}$ and $v_{t,k}$, the similarity likelihood is given by

$$p(v_{s,k},\ v_{t,k} \mid \theta^*_{s,k}\ =\ \theta^*_{t,k}) \propto \int_D p(v_{s,k} | \theta^*_{s,k} = \theta) p(v_{t,k} | \theta^*_{t,k} = \theta)\ d\theta \tag{6}$$

where D is the definition domain of the parameter θ. Note that the denominator $\int p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)\ d\theta$ in Equation 5 acts as a proportionality constant, as well as the quantity $p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)$ which is assumed to be independent of s, t, k and θ. When D is a continuum as $\mathbb{R}$, $\mathbb{R}^+$ and $\mathbb{C}$ then $p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)$ corresponds to an uniform improper prior density. After acceptance of this paper, we have become aware of the probabilistic similarity measure in [34] expressed as in Equation 5 with an uniform prior density $p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)$ defined on a bounded interval D.

C. Related Works and Motivations

In the non local approaches for density estimation introduced in [35]–[37], the authors do not make any assumptions on the noise distribution. The parameters are estimated by minimizing a variational energy related to the joint probability $p(v_{\Delta_s})$. Having no prior on this joint probability, the energy is expressed by a multivariate isotropic Gaussian kernel. The Probabilistic Patch Based (PPB) filter does not requires such a simplification since each quantity is defined directly from the noise model.

The robust M-estimator in [38] consists on replacing the exponential decay function in Equation 3 by a more suitable function of the Euclidean distance. Their work is restricted on additive WGN but could be extended here easily to replace the power function in Equation 4 by a more suitable function of the similarity probability. The use of the robust M-estimator in a probabilistic patch-based approach could be the purpose of a future work.

The PPB filter is also different from the non local solution in [39] proposed for noise reduction in magnetic resonance images. To avoid the Euclidean distance used in Equation 3, the authors propose to use a MLE where the similarity is used to select the suitable pixel values. The MLE is then performed over the set of the most similar sites t with respect to the ℓ_1 distance between v_{Δ_s} and v_{Δ_t}.

The Bayesian NL means filter, proposed in [33] and used for ultra-sound speckle reduction in [40], minimizes the Bayesian risk instead of maximizing the weighted likelihood. In the particular case of additive WGN model, the same estimation is obtained by WMLE and Bayesian risk minimization, which is the weighted averaging given in Equation 2. To cope with non additive WGN, the Euclidean based weights are substituted by the conditional probability $p\left(v_{\Delta_s} | \theta^*_{\Delta_s} = v_{\Delta_t}\right)$. This approach assumes that v_{Δ_t} provides a good approximation on the true

parameter $\theta^*_{\Delta_t}$. We suggest that the similarity probability $p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t}|v)$ is more suitable since it does not make such a strong assumption. In the case of additive WGN, this conditional probability and our similarity probability involve the same weight definition. Since the Bayesian NL means filter makes this strong assumption, the authors proposed a two steps algorithm to refine the weights. Based on a similar idea, we describe in the next section an iterative procedure to enhance the estimation.

III. Iterative Denoising

This section presents the iterative procedure used to refine the patch-based weights estimation. The weights are defined at each iteration as the product of two terms. The first term corresponds to the similarity between noisy patches as defined in section II-B. The second evaluates the similarity between the restored patches extracted from the denoised image at the previous iteration.

A. Refining the Weights with Denoised Patches

The Probabilistic Patch-Based (PPB) filter is a WMLE filter where the weights are defined by the similarity probabilities (see Equation 4). The idea is to refine iteratively these weights by including patches from the estimate of the image parameters. Let us consider at iteration i the previous estimate $\hat{\theta}^{i-1}$ of θ^*. Then, the patch-based similarity probability can be extended by introducing the knowledge of $\hat{\theta}^{i-1}$:

$$w(s,t)^{(it.\ PPB)} \triangleq p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t}|v, \hat{\theta}^{i-1})^{1/h}. \quad (7)$$

With the same considerations as in Section II-B, the similarity probability can be decomposed as a product of the probabilities $p(\theta^*_{s,k} = \theta^*_{t,k} \mid v_{s,k}, v_{t,k}, \hat{\theta}^{i-1})$. In a Bayesian framework, without knowledge on $p(v_{s,k}, v_{t,k})$, and assuming the event $v_{s,k},\ v_{t,k} \mid \theta^*_{s,k} = \theta^*_{t,k}$ is independent of $\hat{\theta}^{i-1}$, the following relation holds:

$$p(\theta^*_{s,k} = \theta^*_{t,k} \mid v_{s,k}, v_{t,k}, \hat{\theta}^{i-1}) \propto \underbrace{p(v_{s,k},\ v_{t,k} \mid \theta^*_{s,k} = \theta^*_{t,k})}_{\text{likelihood}} \times \underbrace{p(\theta^*_{s,k} = \theta^*_{t,k} \mid \hat{\theta}^{i-1})}_{\textit{a priori}}. \quad (8)$$

The likelihood term corresponds to the data fidelity and was defined in the previous section (see Equation 6). The *prior* term measures the validity of $\theta^*_{s,k} = \theta^*_{t,k}$ given the estimate $\hat{\theta}^{i-1}$. We assume the equality $\theta^*_{s,k} = \theta^*_{t,k}$ is more likely to hold as the data distributions with parameters $\hat{\theta}^{i-1}_{s,k}$ and $\hat{\theta}^{i-1}_{t,k}$ get closer. The *prior* term is a function of a similarity between these two data distributions. Polzehl and Spokoiny used the Kullback-Leibler divergence between the estimates[4] as a test statistic of hypotheses $\theta^*_{s,k} = \theta^*_{t,k}$ [30]. We also suggest using a symmetrical version of the Kullback-Leibler divergence over an exponential decay function

$$p(\theta^*_{s,k} = \theta^*_{t,k} \mid \hat{\theta}^{i-1}) \propto \exp\left(-\frac{1}{T}\int_D \left(p(t|\hat{\theta}^{i-1}_{s,k}) - p(t|\hat{\theta}^{i-1}_{t,k})\right) \log \frac{p(t|\hat{\theta}^{i-1}_{s,k})}{p(t|\hat{\theta}^{i-1}_{t,k})}\, dt\right) \quad (9)$$

[4]they considered in their work noise distributions from the exponential family

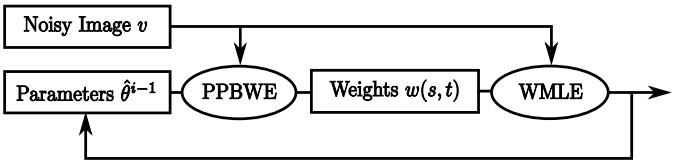


Fig. 1. Scheme of the iterative Probabilistic Patch-Based (PPB) filter. The PPB Weights Estimator (PPBWE) computes the weights $w(s,t)$ by using the noisy image v and the estimate $\hat{\theta}^{i-1}$. The WMLE computes the new parameters $\hat{\theta}^i$ by using the estimated PPB weights $w(s,t)$ and the noisy image v. The procedure is repeated until there is no more change between two consecutive estimates.

where D is the domain of pixel values and $T > 0$ is a positive real value. This corresponds to the Kullback-Leibler divergence based kernel used in [41]. The parameters T and h act as dual parameters to balance the trade-off between the noise reduction and the fidelity of the estimate. The reader can find a detailed discussion about the influence of such parameters in [30].

This refining procedure is performed iteratively. Indeed, at iteration $i-1$, estimates $\hat{\theta}_s$ provide the estimate $\hat{\theta}^{i-1}$ used at iteration i. Note that $\hat{\theta}^{i-1}$ is updated only after $\hat{\theta}_s$ is evaluated for all sites $s \in \Omega$. This corresponds to a synchronous local iterative method [42]. This kind of algorithms converges to a solution depending on the initial parameter $\hat{\theta}^1$. For best performances, the initial estimate should preferably have a good signal to noise ratio with preservation of the thin structures existing in the noisy image. A way to construct such an initialization is given in Section IV-C.

Figure 1 illustrates the whole procedure:

1 first, the PPB Weights Estimator computes the weights $w(s,t)$ by using the noisy image v for the likelihood term (Equation 6) and by using the estimate $\hat{\theta}^{i-1}$ for the prior term (Equation 9);
2 then, the WMLE computes the new parameters $\hat{\theta}^i$ by using the estimated PPB weights $w(s,t)$ and the noisy image v (Equation 1);
3 steps 1 and 2 are repeated until there is no more change between two consecutive estimates.

B. Related Works

In the iterative PPB filter, the weights are defined by two terms. The first term, the data fidelity, depends on the original noisy image and considers its pixel values as a realization of the noise generative model. The second term is calculated from the previously estimated image and considers its pixel values as the "true" parameters of the noise generative model. This idea is different from the iterative NL means versions defined in [43] and the gradient descents proposed in [35], [36], [38], where only previously estimated parameters are used to compute the similarity criterion. In [35], [38] a weighted averaging is performed on the previously estimated image instead of the noisy image. Our approach seems to converge to a solution closer to the noise-free image since the solution remains guided by the noisy image over the different iterations.

In [44], the parameters are estimated iteratively by the Maximum A Posteriori (MAP) estimator under assumptions

of interactions between pixels of the "true" image parameters. The similarity with our method is the use of a non-local approach to define the graph of interactions which is computed and re-estimated using the original noisy image and the previous estimate. In [45], the parameters are also estimated by the MAP estimator but under the assumption that the "true" image is close to the non-local means result. Such a method is used in [46] in the case of speckle reduction. The authors proposed to eventually reestimate the non-local weights by using the previous estimate. The first difference with our filter is that their non-local estimation corresponds to a weighted average instead of a WMLE. Another difference is that they compare the model parameters with a Minkowski distance instead of the symmetric Kullback-Leibler divergence.

The iterative PPB filter is related to the Expectation-Maximization (EM) procedure [47]. The EM algorithm is a two steps iterative algorithm which converges to a local optimum depending on the initial estimate. The first step (E-Step) evaluates a complete-data likelihood expectation by computing sufficient parameters using a previous estimate, while we evaluate a weighted likelihood by computing similarity probabilities using the previous estimate $\hat{\theta}^{i-1}$. The second step (M-Step) maximizes the complete-data likelihood expectation, while we maximize the weighted likelihood. As in the EM procedure, the PPB filter considers also the previous estimate as "true" parameters. According to our experiments, this consideration involves the model stability over the different iterations and provides the convergence of our method. Nevertheless, our function is not related to a complete-data likelihood expectation over our latent variable $\delta_{\mathcal{S}_{\theta^*_s}}(t)$. The similarity between two patches is a good indication that their central values are close (as demonstrated by the performance of NL-means). Dissimilar patches however do not provide any clue on the difference or closeness between the central values. The complete-data likelihood expectation that should be computed in a normal E-Step is therefore less relevant in our context. Finally, our latent variable definition makes the algorithm locally defined for all sites s. Then, the PPB filter is a synchronous local iterative method while an EM algorithm would try to resolve iteratively the problem directly on the global image.

IV. Algorithm Derivation in the cases of Gaussian and Speckle Noises

This section presents the derived algorithms from the iterative Probabilistic Patch-Based (PPB) filter for additive White Gaussian Noises (WGN) and multiplicative speckle noises present in SAR images. Finally, the automatic setting of the parameters and the algorithm complexity are discussed.

A. Derivation in the case of Gaussian Noise

Under the additive WGN model assumption, the pixel values I_s are independent and identically distributed according to the normal distribution $\mathcal{N}(\mu^*_s, \sigma^2)$ where μ^* is the underlying noise-free image and σ^2 the noise variance. Then, it is straightforward to show from the first order optimality condition that

$$\hat{\mu}_s^{(WMLE)} = \frac{\sum_t w(s,t) I_t}{\sum_t w(s,t)}$$

must hold to maximize the WMLE defined in Equation 1, and according to Appendix A:

$$p(I_{s,k}, I_{t,k} | \mu^*_{s,k} = \mu^*_{t,k}) \quad \propto \quad \exp\left(-\frac{|I_{s,k} - I_{t,k}|^2}{4\sigma^2}\right),$$

$$p(\mu^*_{s,k} \; = \; \mu^*_{t,k} \mid \hat{\mu}^{i-1}) \quad \propto \quad \exp\left(-\frac{1}{T}\frac{|\hat{\mu}^{i-1}_{s,k} - \hat{\mu}^{i-1}_{t,k}|^2}{\sigma^2}\right).$$

Finally the weights at iteration i can be defined as

$$w(s,t)^{(it.\ PPB)} = \exp\left[-\sum_k \left(\frac{1}{h}\frac{|I_{s,k} - I_{t,k}|^2}{4\sigma^2} + \frac{1}{T}\frac{|\hat{\mu}^{i-1}_{s,k} - \hat{\mu}^{i-1}_{t,k}|^2}{\sigma^2}\right)\right].$$

In a non-iterative version, $T \to +\infty$, the filter is exactly the NL means filter. Then the PPB filter can be considered as an iterative extension of the NL means filter. Note that the Gaussian kernel, defined by the weights α_k in Equation 3, is not used in the PPB filter (i.e., PPB is purely non-local), but can be re-introduced easily.

B. Derivation in the case of Speckle Noise

In SAR images, the information sought (the reflectivity) is considered to be corrupted by the multiplicative Goodman's speckle noise model [48]. The pixel amplitudes A_s are modeled as independent and identically distributed according to the following Nakagami-Rayleigh distribution

$$p(A_s | R^*_s) = \frac{2L^L}{\Gamma(L) R^{*L}_s} A_s^{2L-1} \exp\left(-\frac{L A_s^2}{R^*_s}\right)$$

where R^* is the underlying reflectivity image and L the equivalent number of looks. Note that the underlying noise-free amplitude image A^* is the square root of the reflecivity image R^*. From the first order optimality condition, the following estimation

$$\hat{R}_s^{(WMLE)} = \frac{\sum_t w(s,t) A_t^2}{\sum_t w(s,t)}$$

must hold to maximize the WMLE defined in Equation 1, and according to Appendix B:

$$p(A_{s,k}, A_{t,k} | R^*_{s,k} = R^*_{t,k}) \quad \propto \quad \left(\frac{A_{s,k} A_{t,k}}{A^2_{s,k} + A^2_{t,k}}\right)^{2L-1},$$

$$p(R^*_{s,k} \; = \; R^*_{t,k} \mid \hat{R}^{i-1}) \quad \propto \quad \exp\left(-\frac{L}{T}\frac{|\hat{R}^{i-1}_{s,k} - \hat{R}^{i-1}_{t,k}|^2}{\hat{R}^{i-1}_{s,k}\hat{R}^{i-1}_{t,k}}\right).$$

Finally the weights at iteration i can be defined as

$$w(s,t)^{(it.\ PPB)} = \exp\left[-\sum_k \left(\frac{1}{\tilde{h}} \log\left(\frac{A_{s,k}}{A_{t,k}} + \frac{A_{t,k}}{A_{s,k}}\right) + \frac{L}{T}\frac{|\hat{R}^{i-1}_{s,k} - \hat{R}^{i-1}_{t,k}|^2}{\hat{R}^{i-1}_{s,k}\hat{R}^{i-1}_{t,k}}\right)\right].$$

where $\tilde{h} = h/(2L-1)$. In a non-iterative version, $T \to +\infty$, the filter is based on the same scheme as the NL means filter by

substituting the Euclidean distance with a similarity criterion adapted to speckle noise and given by

$$\log\left(\frac{A_1}{A_2}+\frac{A_2}{A_1}\right) \tag{10}$$

where A_1 and A_2 are two observed amplitude values.

C. Choice of the Window Sizes and the Initialization

For complexity reasons, the pixels t are restricted to a window W_s centered around s. Then, the algorithm complexity is given by $O(|\Omega||W||\Delta|)$ where $|\Omega|$, $|W|$ and $|\Delta|$ are respectively the image size, the search window size and the similarity patch size. Several optimizations have been proposed as the block-based approach [49], the fast non-local means [50], the improved NL means [38] and the solution implemented here and proposed by Darbon *et al.* in [51] with a time complexity given by $O(4|\Omega||W|)$.

As suggested by Buades *et al.* a search window of size $|W| = 21 \times 21$ is used and a similarity window of size $|\Delta| = 7 \times 7$ [1]. Finally, the computational time of our method is of 22 seconds and 35 seconds per iteration for the additive WGN implementation and the multiplicative speckle noise respectively (image of size $|\Omega| = 512 \times 512$, using an Intel Pentium D 3.20GHz).

Kervrann and Boulanger showed that the size of $|W|$ acts as a bias-variance trade-off on the estimation [43]. When the window size increases, the variance decreases but the estimation is more biased because there are more values coming from different distributions. Then, we suggest to compute the initial image parameters θ^1 by the iterative PPB filter with a smaller search window size. Thus, small features will be preserved and noise reduced before proceeding to a stronger denoising in the following steps.

D. Choice of the Filtering Parameters

The purpose is to find automatically a value of h. Since $w(s,t)^{(PPB)} \propto \exp\left(\frac{1}{h}\log p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t}|v)\right)$, the parameter h can be seen as a normalization of a similarity criterion defined by $c(v_{\Delta_s}, v_{\Delta_t}) = -\sum_k \log p(v_{s,k},\ v_{t,k}|\ \theta^*_{\Delta_s} = \theta^*_{\Delta_t})$. The weights range is directly related to the range of variation of this criterion. In order to normalize the criterion values, we suggest to subtract from the criterion its mathematical expectation. Once the criterion values are centered around zero, the value of h can be chosen as an α-quantile of the distribution of the new centered criterion between two random noisy patches v_{Δ_s}, v_{Δ_t} with identical "true" parameters $\theta^*_{\Delta_s} = \theta^*_{\Delta_t}$. That is different from [30], [43], where h was chosen accordingly to the distribution of the original non-centered criterion. In our case, the minimum of the similarity criterion is not necessarily zero which requires to center the data first. Note that centering the data corresponds to a division of the weights after applying the exponential decay function. Since WMLE is independent to multiplicative constants, the definition of the PPB filter remains unchanged. Finally the parameter h, defined as the α-quantile of the new centered criterion, is given by:

$$h \triangleq q - \mathbb{E}[c(v_{\Delta_s}, v_{\Delta_t})] \tag{11}$$
$$\text{with} \qquad q = \mathbb{F}^{-1}_{c(v_{\Delta_s}, v_{\Delta_t})}(\alpha)$$

where q, $\mathbb{E}$ and $\mathbb{F}$ denote respectively the α-quantile, the expectation and the cumulative distribution function.

Unlike h, the parameter T cannot be defined from the distribution of the similarity criterion between the estimated patches. Indeed, this distribution depends on the statistical model of the previous estimate $\hat{\theta}^{i-1}$ which is non-stationary and depends on the unknown image θ^*. An adaptive estimation of these distribution could be considered in a future work to choose a local adaptative parameter T as done in [37], [38], [43], [52]. We choose to tune manually T after a value of h is fixed. Our results indicate that this value ranges around $0.20|\Delta|$ for any image and noise model when h is set according to the Equation 11 with $\alpha = 0.92$.

E. Convergence and Choice of the Number of Iterations

To study the convergence of the iterative PPB filter, the Signal to Noise Ratio (SNR) is computed and given by

$$SNR(\hat{u}^i, u^*) = 10\log_{10}\frac{Var[u^*]}{\frac{1}{|\Omega|}\sum_{s\in\Omega}(u^*_s - \hat{u}^i_s)^2}. \tag{12}$$

Figure 2 shows the evolution of the SNR during the iterations for three initializations: our preestimated image proposed in Section IV-C, the noise-free image and a constant image. The iterative PPB filter is studied on the 512×512 *Lena* image corrupted by additive WGN with standard deviation $\sigma = 40$ and multiplicative Goodman's speckle noise with equivalent number of looks $L = 3$. The algorithm converges for all initializations and similar solutions are reached, which illustrates that the algorithm is not very sensitive to the initialization. Note that the noise-free image is not a fixed-point of our algorithm since some changes are applied before reaching convergence. According to Section III-A, the similarities between the successive estimates $\hat{\theta}^i$ and $\hat{\theta}^{i+1}$ can be measured by the

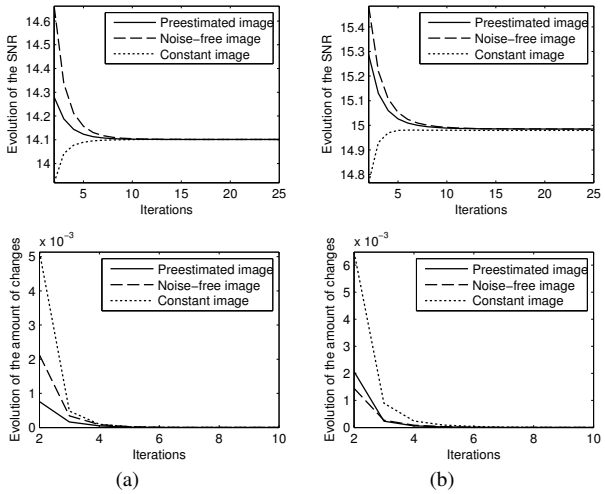


Fig. 2. Evolution of (top) the SNR and (bottom) the symmetric Kullback-Leibler divergence between two successive estimates. The evolution is measured for three different initial estimates: a preestimated image, the noise-free image and a constant image. The iterative PPB filter is studied on the 512×512 *Lena* image corrupted by (a) an additive WGN with standard deviation $\sigma = 40$ and (b) a multiplicative Goodman's speckle noise with equivalent number of looks $L = 3$.

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

Fig. 3. (a) From top to bottom, corrupted images of *Barbara*, *Boat*, *House* and *Lena* by an additive WGN with standard deviation $\sigma = 40$. Denoised images using (b) the K-SVD filter, (c) the BM3D filter, (d) our non-iterative PPB filter (i.e the NL-Means filter) and (e) 25 iterations of our PPB filter.

symmetrical Kullback-Leibler divergence. Then, we suggest to use this criterion to measure the amount of changes between two successive estimated images. Figure 2 shows the evolution of this criterion during the iterations (normalized according to the image size $|\Omega|$). The curves converge to 0 whatever the initialization. This means that, after enough iterations, there are no more changes between two successive iterations: the iterative procedure has reached convergence. Thus, we propose to use this measure as a stopping criterion. In practice, the procedure converges in about 15 iterations. Note that the speed of convergence depends also on the parameter T.

V. Experiments and Results

A. Results on Synthetic Images

This section presents visual and numerical results obtained on four synthetic images corrupted by additive WGN and multiplicative Goodman's Speckle Noises (GSN). The corrupted images are obtained from four classical noise-free images: *Barbara*, *Boat*, *House* and *Lena*. On all noisy images, the non-iterative and the iterative Probabilistic Patch-Based (PPB) filters are applied. A search window of size $|W| = 21 \times 21$ and a similarity window of size $|\Delta| = 7 \times 7$ are used. For the non-iterative procedure, the parameter h has been set with $\alpha = 0.88$. For the iterative procedure, the parameters have been set to $\alpha = 0.92$ and $T = 0.20|\Delta|$ for all experiments. We used 25 iterations of the iterative PPB filter to ensure to reach convergence. Some comparison with the latest state-of-the-art filters are provided. For additive WGN, the comparisons have been performed with the NL-means [1], the K-Singular Value Decomposition (K-SVD) [53] and the Block-Matching and 3D collaborative filtering (BM3D) [26]. Note that the NL-means filter corresponds here to the non-iterative PPB filter. For multiplicative GSN, the comparisons have been performed with the Wavelet-based Image-denoising Nonlinear SAR (WIN-SAR) filter [11] and the MAP filter based on Undecimated Wavelet Decomposition and image Segmentation (MAP-UWD-S) [13].

Figures 3 and 4 present the obtained denoised images for the images corrupted respectively by additive WGN with a standard deviation $\sigma = 40$ and by multiplicative GSN with an equivalent number of look $L = 3$. Note that these two noise levels have been chosen because they provide comparable levels of SNR. For reasons of space and visibility, only small sub-images are shown here. To assess the quality of the denoising methods, the reader can compare the full size images at http://www.tsi.enst.fr/~deledall/ppb.php. Some complementary comparisons are provided on this webpage.

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

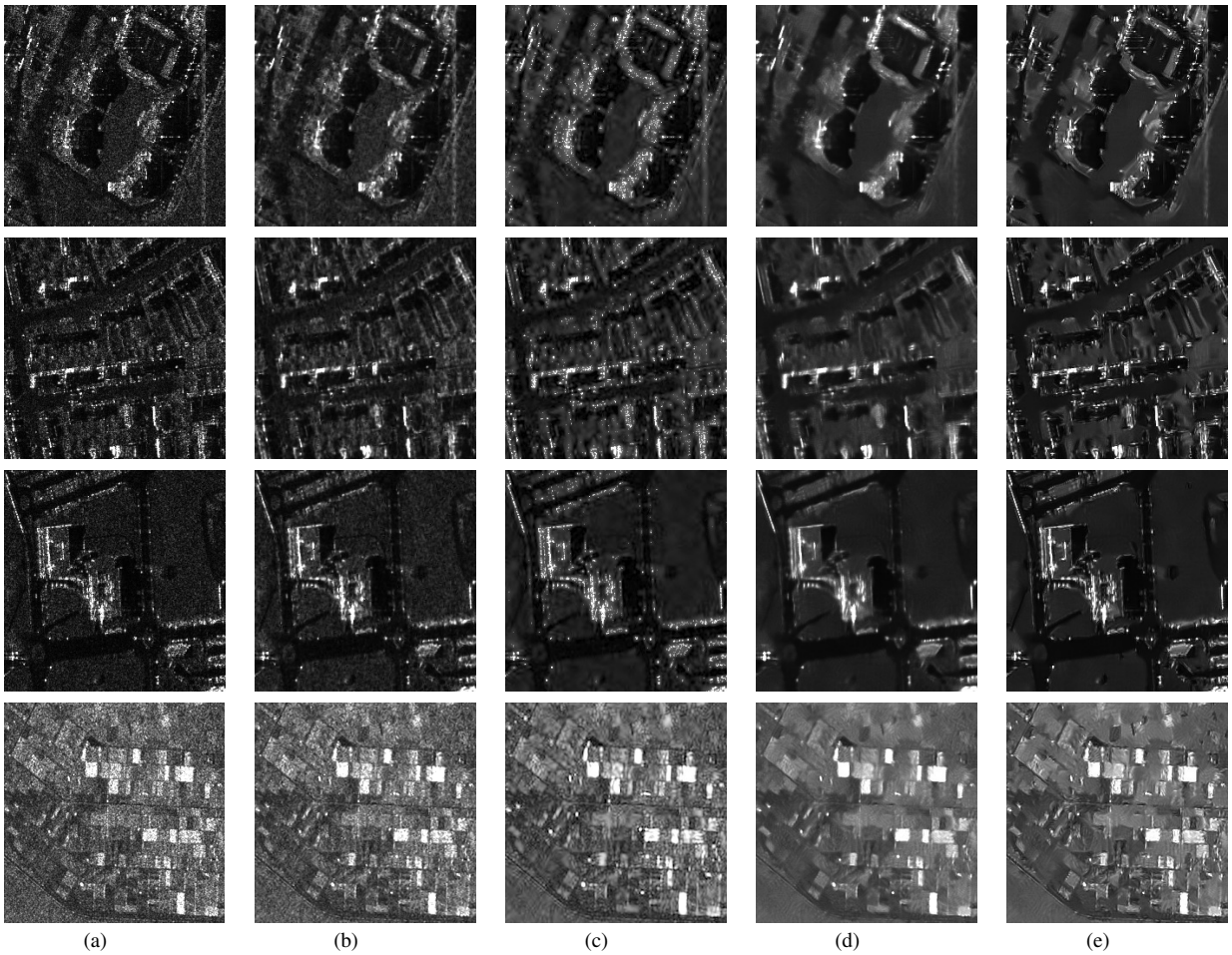
Fig. 4. (a) From top to bottom, corrupted images of *Barbara*, *Boat*, *House* and *Lena* by a multiplicative speckle noise with equivalent number of looks $L = 3$. Denoised images using (b) the WIN-SAR filter, (c) the MAP-UWD-S filter, (d) our non-iterative PPB filter and (e) 25 iterations of our PPB filter.

TABLE I

SNR VALUES OF ESTIMATED IMAGES USING DIFFERENT DENOISING METHODS FOR IMAGES CORRUPTED BY (LEFT) AN ADDITIVE WGN WITH DIFFERENT STANDARD DEVIATIONS AND BY (RIGHT) A MULTIPLICATIVE SPECKLE NOISE WITH DIFFERENT EQUIVALENT NUMBERS OF LOOKS

	$\sigma = 10$	$\sigma = 20$	$\sigma = 40$	$\sigma = 60$
Barbara				
Noisy image	14.73	8.80	3.09	0.04
K-SVD	21.02	17.43	13.01	09.29
BM3D	**21.48**	**18.38**	**14.59**	**12.14**
PPB non-it. (NL means)	19.85	16.97	12.85	10.24
PPB 25 it.	18.69	15.96	13.49	10.99
Boat				
Noisy image	13.41	7.42	1.63	-1.49
K-SVD	18.87	15.62	11.78	9.04
BM3D	**19.09**	**16.09**	**12.83**	**10.55**
PPB non-it. (NL means)	17.59	14.63	11.06	8.96
PPB 25 it.	17.19	14.51	11.63	9.50
House				
Noisy image	13.27	7.26	1.45	-1.62
K-SVD	21.15	18.31	14.36	10.22
BM3D	**21.77**	**18.94**	**15.78**	**13.28**
PPB non-it. (NL means)	20.25	17.55	13.33	10.40
PPB 25 it.	19.59	17.03	14.20	11.57
Lena				
Noisy image	13.59	7.60	1.81	-1.25
K-SVD	20.93	17.81	14.18	11.09
BM3D	**21.27**	**18.42**	**15.33**	**13.05**
PPB non-it. (NL means)	20.12	17.10	13.66	11.33
PPB 25 it.	19.50	16.90	14.20	11.99

	$L = 1$	$L = 2$	$L = 4$	$L = 16$
Barbara				
Noisy image	-1.09	1.69	4.61	10.57
WIN-SAR	8.82	10.48	12.04	15.82
MAP-UWD-S	9.65	11.44	13.28	16.93
PPB non-it.	9.79	11.88	**14.05**	**17.83**
PPB 25 it.	**10.58**	**12.51**	13.98	16.59
Boat				
Noisy image	-2.99	-0.18	2.70	8.67
WIN-SAR	8.57	10.65	12.14	15.17
MAP-UWD-S	9.26	10.68	**12.31**	**15.71**
PPB non-it.	8.71	10.49	12.22	15.33
PPB 25 it.	**9.43**	**10.91**	12.25	15.10
House				
Noisy image	-3.55	-0.76	2.11	8.10
WIN-SAR	8.69	11.42	13.15	16.24
MAP-UWD-S	10.34	11.97	13.72	17.24
PPB non-it.	9.06	11.61	14.29	**18.27**
PPB 25 it.	**10.46**	**12.98**	**14.50**	17.42
Lena				
Noisy image	-2.45	0.34	3.25	9.19
WIN-SAR	10.35	13.00	14.72	17.90
MAP-UWD-S	11.87	13.53	15.14	**18.65**
PPB non-it.	11.05	13.20	15.18	18.61
PPB 25 it.	**12.16**	**13.95**	**15.25**	18.10



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

Fig. 5. (a) From top to bottom, SAR images of *Bayard* (France) ©DGA ©ONERA, *Cheminot* (France) ©DGA ©ONERA, *Toulouse* (France) ©DGA ©ONERA and *Lelystadt* (Netherlands) ©ESA. Denoised images using (b) the WIN-SAR filter, (c) the MAP-UWD-S filter, (d) our non-iterative PPB filter and (e) 25 iterations of our PPB filter.

Recall that for synthetic SAR images the noise-free image is given by the squared root of the estimated reflectivity images (Section IV-B). The images obtained with the iterative PPB filter seem to be well smoothed with a better edge and shape preservation than the non-iterative PPB filter (and than NL means filter in case of additive WGN). The images denoised by the K-SVD and the BM3D filters present some artifacts while the PPB filter provides smoother regions with comparable edge preservation. However, our PPB filter seems to attenuate the image contrast and thin and dark structures as the mouth of *Lena*, the eyes of *Barbara* and the ropes of the *Boat*, while the BM3D filter preserves these structures. This phenomenon can in part be explained by the high value of α-quantile chosen to get a (qualitatively satisfying) low variance in homogeneous regions. It could also be reduced by considering smaller search windows, at the cost of a larger remaining noise variance (see Section IV-C).

The images denoised by the WIN-SAR and the MAP-UWD-S filters are less smoothed than the images obtained by the PPB filter. Moreover, the WIN-SAR filter blurs the edges and the MAP-UWD-S filter introduces some artifacts in the neighborhood of the edges. Finally, our PPB filter seems to be working equally well for additive WGN and multiplicative GSN when the SNR is similar. Thus, the PPB filter seems to be an efficient extension of the NL means filter to take into account different noise degradation models.

To quantify the denoising qualities, Table I presents numerical results for images corrupted by additive WGN with standard deviations $\sigma = 10, 20, 40$ and 60 and by multiplicative GSN with equivalent number of looks $L = 1, 2, 4$ and 16. The performance criterion used is the Signal to Noise Ratio (SNR) presented in Equation 12. We observe that the iterative PPB filter improves on the non-iterative PPB filter for low SNR images. High SNR images (standard deviation $\sigma \leq 30$ or equivalent number of looks $L \geq 4$) do not require iterative refinement of the weights. In case of additive WGN, the PPB filter is better than the K-SVD filter for low SNR images, but is out-performed by the BM3D filter at all SNR values. In the case of multiplicative GSN, the PPB filter out-performed all the state-of-the-art filters considered for low SNR images. Nevertheless, the PPB filter provides comparative results to the MAP-UWD-S filter for high SNR images (i.e., with a large equivalent number of looks L). The iterative PPB filter is then more relevant since SAR images are generaly provided for a low equivalent number of look as $L = 1, 2, 3$ or 4.

B. Results on Real SAR Images

This section presents an overview of different results obtained on four real SAR images with the same state-of-the-art speckle filters as the ones used above and our PPB filter. There are two single-look SAR acquisitions identified as *Bayard* and *Cheminot* from Saint-Pol-sur-Mer (France), sensed in 1996 by RAMSES of ONERA. There is a single-look SAR acquisition identified as *Toulouse* of the CNES in Toulouse (France) sensed also by RAMSES and provided by the CNES. And there is a multi-look SAR image identified as *Lelystadt* of an agriculture region in Lelystadt (Netherlands), sensed by ERS-1 with 3 equivalent number of looks (PRI data) and provided by the European Space Agency (ESA). All these images are assumed to follow the multiplicative Goodman's speckle noise model. These 4 images provide a testing set which presents a good diversity: different sensors (RAMSES/ERS), different scenes (urban/agricultural), different noises (single-look/multi-looks). In all experiments, the algorithms are executed with the same parameters described in Section V-A.

Figure 5 presents the obtained denoised images for the different real SAR images and the different denoising filters. The results obtained with our iterative PPB filter seem to be well smoothed with a better edge and shape preservation than other filters. The speckle effect is strongly reduced and the spatial resolution seems to be well preserved: buildings, sidewalks, streets, fields are well restored. Moreover, the bright scatterers (numerous in urban area) are well restored. Unfortunately, the PPB filter seems to attenuate thin and dark structures existing in the SAR image, such as the thin streets in *Cheminot* and *Toulouse* and the channels between the fields of *Lelystadt*.

VI. Conclusion

A general methodology was proposed for image denoising which can be adapted to different noise distributions. This method is based on the Non-Local means filter (NL means) [1] and can be adapted to non-additive and/or non Gaussian noises. An iterative procedure is proposed to enhance the denoising quality in case of low signal to noise ratio images and its efficiency has been shown on additive Gaussian noise and multiplicative speckle noise. The proposed filter is outperformed by the BM3D filter [26] in the case of additive Gaussian noise. However, our filter provides interesting and promising results for Synthetic Aperture Radar (SAR) images, damaged by a multiplicative speckle noise. The noise, present in the input images, is well smoothed in the homogeneous regions and the object contours are well restored (preservation of the resolution). Moreover we can consider from our numerical experiments, that the reflectivity in SAR images is well recovered, without introducing undesired artifacts, with a good restoration of bright scatterers. More generally, it provides a new framework for the estimation of image parameters when the uncorrelated noise model is known. This framework could be used for multi-channel image denoising or joint denoising of several images as done for joint interferometric SAR data denoising in [54]. A drawback of the filter is the suppression of thin and dark details in the regularized images. In a future work, we will try to better preserve thin and dark details, by using adaptive filtering as done in [37], [38], [43], [52]. The filter elaboration, based on the statistics of the processed images, has led to define a suitable patch-similarity criterion for SAR images. This similarity criterion will be applied to other applications such as pattern tracking and displacement estimation.

Acknowledgments

The authors would like to thank Marc Sigelle and the anonymous reviewers for their interesting comments and constructive criticisms, Fabrizio Argenti and Tiziano Bianchi for helping us by processing our data with the MAP-UWD-S filter, Alin Achim for providing us his WIN-SAR filter and the Office National d'Etudes et de Recherches Aérospatiales and the Délégation Générale pour l'Armement for providing the RAMSES data.

Appendix A
Weights derivation for additive WGN

The similarity probability is given by

$$\begin{aligned}&\int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|I_1-\mu|^2+|I_2-\mu|^2}{2\sigma^2}\right)\, d\mu \\ =\;& \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{|I_1-I_2|^2}{4\sigma^2}\right)\end{aligned}$$

according to the convolution of two Gaussian functions.

Now, note the following statement

$$|t-\mu_2|^2 - |t-\mu_1|^2 = -(\mu_1^2-\mu_2^2) + 2t(\mu_1-\mu_2)$$

Then, the Kullback-Leibler divergence is given by

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|t-\mu_1|^2}{2\sigma^2}\right) \times \frac{|t-\mu_2|^2-|t-\mu_1|^2}{2\sigma^2}\, dt$$

$$\begin{aligned}=\;& -\frac{\mu_1^2-\mu_2^2}{2\sigma^2} \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|t-\mu_1|^2}{2\sigma^2}\right)\, dt}_{=1} \\ &+\frac{\mu_1-\mu_2}{\sigma^2} \underbrace{\int_{-\infty}^{+\infty} \frac{t}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|t-\mu_1|^2}{2\sigma^2}\right)\, dt}_{=\mu_1} \\ =\;& -\frac{\mu_1^2-\mu_2^2}{2\sigma^2} + \frac{\mu_1^2-\mu_1\mu_2}{\sigma^2}\end{aligned}$$

Finally, the symmetric Kullback-Leibler divergence is

$$\begin{aligned}&-\frac{\mu_1^2-\mu_2^2}{2\sigma^2} + \frac{\mu_1^2-\mu_1\mu_2}{\sigma^2} + \frac{\mu_1^2-\mu_2^2}{2\sigma^2} - \frac{\mu_1\mu_2-\mu_2^2}{\sigma^2} \\ =\;& \frac{\mu_1^2-2\mu_1\mu_2+\mu_2^2}{\sigma^2} = \frac{|\mu_1-\mu_2|^2}{\sigma^2}\end{aligned}$$

APPENDIX B
WEIGHTS DERIVATION FOR MULTIPLICATIVE GSN

First, note the following equality

$$\int_0^\infty \frac{A}{x^n} \exp\left(-\frac{B}{x}\right) dx \quad = \quad AB^{1-n}\Gamma(n-1)$$

Then, the similarity probability is given by

$$\int_0^\infty \frac{4L^{2L}A_1^{2L-1}A_2^{2L-1}}{\Gamma(L)^2R^{2L}} \exp\left(-\frac{L(A_1^2+A_2^2)}{R}\right) dR$$
$$= 4L\frac{\Gamma(2L-1)}{\Gamma(L)^2}\left(\frac{A_1A_2}{A_1^2+A_2^2}\right)^{2L-1}$$

The Kullback-Leibler divergence is given by

$$\int_0^\infty \frac{2L^Lt^{2L-1}}{\Gamma(L)R_1^L} \exp\left(-\frac{Lt^2}{R_1}\right)\left(L\log\frac{R_2}{R_1}+\frac{Lt^2}{R_2}-\frac{Lt^2}{R_1}\right) dt$$

$$= L\log\frac{R_2}{R_1}\underbrace{\int_0^\infty \frac{2L^L}{\Gamma(L)R_1^L}t^{2L-1}\exp\left(-\frac{Lt^2}{R_1}\right) dt}_{=1}$$
$$+\left(\frac{L}{R_2}-\frac{L}{R_1}\right)\underbrace{\int_0^\infty \frac{2L^L}{\Gamma(L)R_1^L}t^{2L+1}\exp\left(-\frac{Lt^2}{R_1}\right) dt}_{=R_1}$$

$$= L\log\frac{R_2}{R_1}+L\frac{R_1}{R_2}-L$$

Finally, the symmetric Kullback-Leibler divergence is

$$L\log\frac{R_2}{R_1}+L\frac{R_1}{R_2}-L-L\log\frac{R_2}{R_1}-L+L\frac{R_2}{R_1}$$
$$= L\frac{R_1}{R_2}+L\frac{R_2}{R_1}-2L = L\frac{|R_1-R_2|^2}{R_1R_2}$$

REFERENCES

[1] A. Buades, B. Coll, and J. Morel, "A Non-Local Algorithm for Image Denoising," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, 2005.

[2] J. Bruniquel and A. Lopes, "On the true multilook intensity distribution in SAR imagery," in *1998 IEEE International Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS'98*, vol. 1, 1998.

[3] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[4] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse problems*, vol. 19, no. 6, pp. 165–187, 2003.

[5] D. Donoho and J. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[6] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage." *Journal of the American Statistical Association*, vol. 90, no. 432, 1995.

[7] S. Chang, B. Yu, M. Vetterli *et al.*, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[8] M. Bhuiyan, M. Ahmad, and M. Swamy, "Spatially adaptive wavelet-based method using the cauchy prior for denoising the SAR images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 4, pp. 500–507, 2007.

[9] H. Xie, L. Pierce, and F. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2196–2212, 2002.

[10] F. Argenti and L. Alparone, "Speckle removal from SAR images in the undecimated wavelet domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2363–2374, 2002.

[11] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, 2003.

[12] F. Argenti, T. Bianchi, and L. Alparone, "Multiresolution MAP despeckling of SAR images based on locally adaptive generalized Gaussian pdf modeling," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3385–3399, 2006.

[13] T. Bianchi, F. Argenti, and L. Alparone, "Segmentation-Based MAP Despeckling of SAR Images in the Undecimated Wavelet Domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 9, pp. 2728–2742, 2008.

[14] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1395–1411, 2007.

[15] J. Starck, E. Candes, and D. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on image processing*, vol. 11, no. 6, pp. 670–684, 2002.

[16] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[17] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Modeling and Simulation*, vol. 7, no. 1, pp. 214–241, 2008.

[18] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing," in *Proc. XII Eur. Signal Process. Conf., EUSIPCO 2004*, pp. 101–104.

[19] J. Park, W. Song, and W. Pearlman, "Speckle filtering of SAR images based on adaptive windowing," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 146, no. 4, pp. 191–197, 1999.

[20] G. Vasile, E. Trouve, J. Lee, and V. Buzuloiu, "Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric SAR parameters estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1609–1621, 2006.

[21] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *Image Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 349–366, Feb. 2007.

[22] L. Yaroslavsky, *Digital Picture Processing*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1985.

[23] S. Smith and J. Brady, "SUSAN – A new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[24] J. Lee, "Digital image smoothing and the sigma filter," *Computer Vision, Graphics, and Image Processing*, vol. 24, pp. 255–269, 1983.

[25] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*, 1998, pp. 839–846.

[26] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, p. 2080, 2007.

[27] ——, "A Nonlocal and Shape-Adaptive Transform-Domain Collaborative Filtering," in *Proc. 2008 Int. Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2008*.

[28] ——, "BM3D Image Denoising with Shape-Adaptive Principal Component Analysis," in *SPARS'09 – Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[29] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," http://www.cs.tut.fi/~foi/papers/KFKA-LocalNonLocalDenoising-IJCV-Preprint2009.pdf, Tech. Rep., 2009.

[30] J. Polzehl and V. Spokoiny, "Propagation-separation approach for local likelihood estimation," *Probability Theory and Related Fields*, vol. 135, no. 3, pp. 335–362, 2006.

[31] J. Fan, M. Farmen, and I. Gijbels, "Local maximum likelihood estimation and inference," *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pp. 591–608, 1998.

[32] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999.

[33] C. Kervrann, P. Pérez, and J. Boulanger, "Bayesian non-local means, image redundancy and adaptive estimation for image representation and applications," in *SIAM Conf. on Imaging Science*, San Diego, CA, July 2008.

[34] Y. Matsushita and S. Lin, "A Probabilistic Intensity Similarity Measure based on Noise Distributions," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*, 2007, pp. 1–8.

[35] S. P. Awate and R. T. Whitaker, "Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering," in *CVPR (2)*, 2005, pp. 44–51.
[36] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient Nonlocal Means for Denoising of Textural Patterns," *IEEE Transactions on Image Processing*, 2007.
[37] N. Azzabou, N. Paragios, and F. Guichard, "Uniform and textured regions separation in natural images towards MPM adaptive denoising," *Lecture Notes in Computer Science*, vol. 4485, p. 418, 2007.
[38] B. Goossens, H. Luong, A. Pižurica, and W. Philips, "An improved non-local denoising algorithm," in *Proc. Int. Workshop on Local and Non-Local Approximation in Image Processing (LNLA'2008), Lausanne, Switzerland*, 2008.
[39] L. He and I. Greenshields, "A Non-Local Maximum Likelihood Estimation Method for Rician Noise Reduction in MR Images," *Medical Imaging, IEEE Transactions on : Accepted for future publication*, 2008.
[40] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot, "Bayesian non-local means-based speckle filtering," in *Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'08)*, Paris, France, May 2008.
[41] P. Moreno, P. Ho, and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," *Advances in Neural Information Processing Systems*, vol. 16, 2004.
[42] E. Bratsolis and M. Sigelle, "Fast SAR image restoration, segmentation, and detection of high-reflectance regions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2890–2899, 2003.
[43] C. Kervrann and J. Boulanger, "Local Adaptivity to Variable Smoothness for Exemplar-Based Image Regularization and Representation," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 45–69, 2008.
[44] G. Peyre, S. Bougleux, and L. Cohen, "Non-local Regularization of Inverse Problems," in *Proc. of ECCV*, vol. 2008, 2008.
[45] M. Mignotte, "A non-local regularization strategy for image deconvolution," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2206–2212, 2008.
[46] N. Azzabou and N. Paragios, "Spatio-temporal Speckle Reduction in Ultrasound Sequences," in *Proceedings of the 11th international conference on Medical Image Computing and Computer-Assisted Intervention-Part I*. Springer, 2008, pp. 951–958.
[47] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
[48] J. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Am*, vol. 66, no. 11, pp. 1145–1150, 1976.
[49] A. Buades, B. Coll, and J. Morel, "A Review of Image Denoising Algorithms, with a New One," *Multiscale Modeling and Simulation*, vol. 4, no. 2, p. 490, 2005.
[50] P. Coupe, P. Yger, and C. Barillot, "Fast Non Local Means Denoising for 3D MR Images," *Lecture Notes In Computer Science*, vol. 4191, pp. 33–40, 2006.
[51] J. Darbon, A. Cunha, T. Chan, S. Osher, and G. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 1331–1334, 2008.
[52] G. Gilboa, N. Sochen, and Y. Zeevi, "Estimation of optimal PDE-based denoising in the SNR sense," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2269–2280, 2006.
[53] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, p. 4311, 2006.
[54] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "SAR image regularization with fast approximate discrete minimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1588–1600, 2009.

Charles-Alban Deledalle received the engineering degree from Ecole Pour l'Informatique et les Techniques Avancées (EPITA) and the Science & Technology master's degree from the University Pierre et Marie Currie (Paris 6), both in Paris in 2008. He is currently pursuing the Ph.D. degree at Telecom ParisTech. His main interests are image denoising, analysis and interpretation, especially in multi-modal synthetic aperture radar imagery.

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.