Weighted maximum likelihood denoising with iterative and probabilistic patch-based weights

Débruitage par maximum de vraisemblance pondérée par une méthode non-locale itérative et probabiliste

---

Charles-Alban Deledalle
Loïc Denis
Florence Tupin

2009D004

Janvier 2009

Département Traitement du Signal et des Images
Groupe TII : Traitement et Interprétation des Images

Débruitage par Maximum de Vraisemblance Pondérée par une Méthode Non-Locale Iterative et Probabiliste

Charles-Alban Deledalle, Loïc Denis et Florence Tupin

Résumé

Le débruitage d'images est un problème important en traitement d'images puisque le bruit rend difficile l'interprétation visuelle et automatique des images. Ainsi, une étape de prétraitements est souvent requise pour débruiter les image avant de procéder à leur analyse. L'objectif est de présenter une nouvelle approche pour le débruitage d'images quand un modèle de bruit décorrélé est disponible. Le filtre proposé est une généralisation de l'algorithme des moyennes non locales introduit par Buades *et al.* (2005), qui réalise une moyenne des valeurs des pixels similaires (les pixels ayant des voisinages similaires). La définition du filtre est plus générale puisqu'elle se base sur une estimation au sens du maximum de vraisemblance pondéré dont les poids sont définis via des patches et adaptés au modèle de bruit (et non pas seulement aux bruits blancs additifs Gaussiens). Nous proposons aussi d'étendre l'algorithme des moyennes non locales à un schéma itératif similaire à l'approche "expectation-maximization" de Dempster (1977) qui semble plus pertinent pour des images ayant un faible rapport signal sur bruit. Les expériences ont été menées sur l'image "Lena", dégradée pas un fort bruit blanc additif Gaussien, et sur des images radar à ouverture synthétique, corrompues par un bruit de chatoiement multiplicatif.

Mots-clés

Débruitage d'images, moyennes non-locales, maximum de vraisemblance pondérée, imagerie radar à synthèse d'ouverture

C. Deledalle et F. Tupin sont à l'Institut TELECOM, TELECOM ParisTech, CNRS LTCI, Paris, FRANCE, e-mail : charles-alban.deledalle@telecom-paristech.fr et florence.tupin@telecom-paristech.fr.

L. Denis est à l'École Supérieure de Chimie Physique Électronique de Lyon et au Laboratoire Hubert Curien, CNRS UMR 5516, ST-ÉTIENNE, e-mail : loic.denis@cpe.fr.

12 janvier 2009

Weighted Maximum Likelihood Denoising with Iterative and Probabilistic Patch-Based Weights

Charles-Alban Deledalle, Loïc Denis and Florence Tupin,

Abstract

Image denoising is an important problem in image processing since the noise makes difficult their visual or automatic interpretation. Hence, a well-adapted preprocessing step is required to denoise the images before analyzing them. The paper presents a new approach for image denoising when an uncorrelated noise model is provided. The proposed filter is a generalization of the non local means algorithm introduced by Buades *et al.* [1], which performs an average of the values of similar pixels (pixels with similar neighborhoods). The definition of the filter is more general since it is based on a weighted maximum likelihood estimation with patch-based weights adapted to a noise model (without restriction to additive white Gaussian noise). We also propose to extend the non local means to an iterative scheme similar to the expectation-maximization approach [2] which seems especially relevant for low signal to noise ratio images. The experiments are realized on the "Lena" image, corrupted by a strong additive white Gaussian noise, and on Synthetic Aperture Radar (SAR) images, corrupted by a multiplicative speckle noise.

Index Terms

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

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.

I. INTRODUCTION

This paper presents a denoising method using a weighted maximum likelihood estimator with probabilistic patch-based weights. An iterative procedure is used to refine the estimations of the weights and seems to be well adapted for low signal to noise ratio images. The filter performances are given on images damaged by a strong additive White Gaussian Noise (WGN) and Synthetic Aperture Radar (SAR) images damaged by a multiplicative speckle noise.

Image denoising consists of retrieving the "true" image from a noisy image. Numerous denoising methods have been proposed in the literature. Their definitions can be very different, as the anisotropic diffusion [3], the total variation minimization [4], [5] and the wavelet thresholding [6]. However, most of them can be traced back to a weighted averaging of the noisy image pixel values [7] in order to retrieve the true image. Their difference lies in the definition of the weights. According to different assumptions and models, lots of weight definitions have been proposed. Recently, Buades *et al.* presented the Non-Local (NL) means filter [1]. This one performs a weighted averaging based on patches which relies on image redundancy. The weights are set in function of a weighted Euclidean distance between patches centered around the pixels of the image. The NL means filter gives very good results and has attracted lot of attention these last years. Unfortunately, this method suffers from a lack of flexibility to be applied for any noise distribution since the Euclidean distance is adapted for additive WGN.

This paper presents the Probabilistic Patch-Based (PPB) filter which is a generalization of the NL means filter. Indeed, PPB corresponds to the NL means filter in the case of additive WGN model. However, PPB is not necessarily a weighted averaging filter since it is derived from a patch-based weighted maximum likelihood which can be adapted to different noise models (i.e., not only restricted to additive WGN models). PPB can process any image corrupted by an uncorrelated noise model when the probability density function, which models the noise component, is known. Note that is different from the unsupervised information-theoretic adaptive filter [8] which is general since it does not make assumptions neither about image nor noise properties. Moreover, PPB is well-adapted for low signal to noise ratio images thanks to an iterative procedure which refines the patch-based weights. Some generalizations and iterative frameworks of the NL means filter have already been proposed in [9]–[13] and will be considered in the following.

Section II introduces the PPB filter as a weighted likelihood estimator where weights are patch-based and defined by a probabilistic approach. Section III presents the iterative procedure based on a Bayesian framework with a *prior* model defined by a Kullback-Leibler criterion in the gray levels of a previously

estimated image. Section IV presents two applications of PPB adapted respectively to additive WGN and to the multiplicative speckle noise present in SAR images. Finally, Section V presents the experiments carried out on optical and SAR images.

II. Patch-based Weighted Maximum Likelihood

This section presents the probabilistic patch-based filter as a weighted maximum likelihood estimation with a probabilistic patch-based weight definition.

A. Weighted Maximum Likelihood

The image denoising problem is considered as the estimation $\hat{u}$ of a "true" image u^* from a noisy observation u. The images are considered to be defined over a discrete regular grid Ω and we denote by u_s a pixel value at site $s \in \Omega$. We assume the noise is uncorrelated and that a model of the noise is available at site s (namely the likelihood). We consider this noise model at each site s as a random process governed by an unknown parameter θ_s^*, and described by the probability density function $p(u_s|\theta_s^*)$. The parameter θ^* can be different from u^*. However, we assume the noise-free image u^* is directly related to this unknown parameter θ^*. For instance, in the following, the denoised amplitude of the SAR images will be different from the estimated parameter (the reflectivity) but will be equal to its square root. Then, to denoise an image is equivalent to find the best estimation $\hat{\theta}$ of θ^*.

For each site s, the Maximum Likelihood Estimation (MLE) is defined as the estimation $\hat{\theta}_s$ of the underlying parameter θ_s^* from a set of independent and identically distributed random variables:

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

where $\mathcal{S}_{\theta_s^*}$ is the set of the sites governed by the same distribution θ_s^* and $\delta_{\mathcal{S}_{\theta_s^*}}$ is 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 all $s \in \Omega$ are unknown, since the knowledge of $\mathcal{S}_{\theta_s^*}$ is equivalent to that of the noise-free image θ^*. Hence, a good trade-off is to approximate the indicator $\delta_{\mathcal{S}_{\theta_s^*}}(t)$ by a weight $w(s,t)$. The Weighted Maximum Likelihood Estimation (WMLE) is given by

$$\hat{\theta}_s^{(WMLE)} = \arg\max_{\theta_s} \sum_{t} w(s,t) \log p(u_t|\theta_s) \tag{1}$$

where the weights $w(s,t) \geq 0$ are defined according to the interaction of the sites s and t. As shown in Section IV, in the particular case of additive White Gaussian Noise (WGN) model, the best estimation

is defined by a Weighted Averaging (WA):

$$\hat{\theta}_s^{(WA)} = \frac{\sum_t w(s,t)u_t}{\sum_t w(s,t)}. \tag{2}$$

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 $\mathcal{S}_{\theta_s^*}$ (as soon as the weights verify $0 \leq w(s,t) \leq 1$). 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. A well-done definition of the weights constitutes the key to the success of the WMLE filter. 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) 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 and textures for which the ergodicity assumption is invalid. Instead of defining $w(s,t)$ in spatial domain, Yaroslavsky proposed a weight definition based on gray level scale [14] also known as sigma-filter [15]. The weight $w(s,t)$ increases when the values u_s and u_t are more similar. Such a filter was then refined by the SUSAN filter [16] and the bilateral filter [17] 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 [18]. The weight $w(s,t)$ is defined by comparing two patches Δ_s and Δ_t centered respectively around the sites s and t:

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

where $u_{s,k}$ and $u_{t,k}$ is relatively the k-th neighbor in the patch Δ_s and Δ_t, the weights α_k define a centered symmetric Gaussian kernel and h^2 controls the decay of the exponential function. The setting

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

of the parameter h^2 is a difficult task addressed in [10], [19]. 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 weights used in the NL means algorithm. The idea is to generalize the NL means algorithm to non additive WGN models. According to the previous comments, 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 u, that two patches Δ_s and Δ_t have same parameters. That holds by assuming pixels with similar neighborhoods come from same distribution, which is the NL means assumption. The patch-based similarity probability involves the following weight definition

$$w(s,t)^{(PPB)} := \lambda\ p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t}|u)^{1/h^2} \quad (4)$$

where $\theta^*_{\Delta_s}$ and $\theta^*_{\Delta_t}$ denote respectively the parameters sub-images in the windows Δ_s and Δ_t, and $\lambda > 0$ and $h > 0$ are real values. The λ value is an arbitrary proportionality value which does not affect the WMLE. The h parameter comes from the NL means algorithm and acts on the width of the fuzzy set to control the amount of filtering. Kervrann *et al.* explain the parameter h probably counterbalances the invalidity of the patch independency assumption [10]. Indeed, we assume the similarity probability can be decomposed as follow $\prod_k p(\theta^*_{s,k}\ =\ \theta^*_{t,k}|u_{s,k}, u_{t,k})$. In a Bayesian framework, without knowledge on $p(\theta^*_{s,k} = \theta^*_{p,k})$ and $p(u_{s,k}, u_{t,k})$, it is natural to consider the probability $p(\theta^*_{s,k}\ =\ \theta^*_{t,k}\mid u_{s,k}, u_{t,k})$ proportional to the likelihood $p(u_{s,k},\ u_{t,k}\mid \theta^*_{s,k}\ =\ \theta^*_{t,k})$. According to Equation 5 (at the top of the page) and under the independent assumption on $u_{s,k}$ and $u_{t,k}$, the similarity probability is given by

$$\frac{1}{Z}\int_D p(u_{s,k}|\theta^*_{s,k} = \theta) p(u_{t,k}|\theta^*_{t,k} = \theta)\ d\theta \quad (6)$$

where Z is a normalization constant equal to $\int p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)\ d\theta$, D is the definition domain of the parameter θ, and $p(\theta^*_{s,k}\ =\ \theta, \theta^*_{t,k}\ =\ \theta)$ is assumed to be independent of s, t, k and θ.

C. Related Works and Motivations

The PPB filter is different from the solution in [9] proposed for noise reduction in magnetic resonance images. To avoid the Euclidean distance, 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 most similar sites t with respect to the ℓ_1 distance between u_{Δ_s} and u_{Δ_t}.

Note that PPB differs also from the Bayesian NL means filter proposed in [10] and used for ultrasound speckle reduction in [20]. On the one hand, Bayesian NL means filter minimizes a weighted mean square error function. WMLE seems to be better suited to cope with non-additive WGN, since the loss function considers the noise property. In the particular case of additive WGN model, the same estimation is obtained by weighted mean square error minimization, which is the weighted averaging proposed in Equation 2. On the other hand, to cope with non additive WGN, the Euclidean based weights are substituted by the conditional probability $p\left(u_{\Delta_s}|\theta^*_{\Delta_s} = u_{\Delta_t}\right)$. This approach assumes that u_{Δ_t} provides a good approximation on the true parameter $\theta^*_{\Delta_t}$. We suggest that the similarity probability $p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t}|u)$ 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 refined 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 thanks to two terms. The first term is the data fidelity used in the previous section and the second term is based on the comparison between patches from the image obtained 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 probability (see Equation 4). The idea is to refine iteratively this probability by refining an estimation of the image parameter. Let us consider a given previous estimation $\hat{\theta}^{i-1}$ of θ^*. Then the patch-based similarity probability involves this new weight definition:

$$w(s,t)^{(it.\ PPB)} := \lambda\ p(\theta^*_{\Delta_s} = \theta^*_{\Delta_t}|u, \hat{\theta}^{i-1})^{1/h^2}.$$

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

$$p(\theta^*_{s,k} \ = \ \theta^*_{t,k} \mid u_{s,k}, u_{t,k}, \hat{\theta}^{i-1}) \propto$$

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

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 estimation $\hat{\theta}^{i-1}$. We assume the equality $\theta^*_{s,k} \ = \ \theta^*_{t,k}$ is more likely to hold as the distributions $\hat{\theta}^{i-1}_{s,k}$ and $\hat{\theta}^{i-1}_{t,k}$ get closer. The *prior* term is a function of a similarity between these two data distributions. We suggest to use a symmetrical version of the Kullback-Leibler divergence over an exponential decay function

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

where D is the domain of pixel values and $T > 0$ is a positive real value. That corresponds to the Kullback-Leibler divergence based kernel used in [21]. The parameters T and h^2 act as dual parameters to balance the trade-off between the noise reduction and the estimation fidelity.

The refining procedure can be performed iteratively. Indeed, at iteration $i-1$, the whole estimations $\hat{\theta}_s$ provide the estimation $\hat{\theta}^{i-1}$ used at iteration i. Note that $\hat{\theta}^{i-1}$ is updated once $\hat{\theta}_s$ is evaluated for all sites $s \in \Omega$. That is called a synchronous local iterative method [22]. This kind of algorithms converges to a local optimum depending on the initial parameter $\hat{\theta}^1$. For best performances, the initial estimation has to be a high signal to noise ratio image which preserves the thin structures existing of the noisy image. A way to construct such an initialization is given in Section IV.

B. Related Works

In PPB, 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 previous estimation 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 [23] and the gradient descent proposed in [11], where only previous parameter estimations are used for similarity criteria. 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. The same idea is explored in

[12], [13] where the total variation is eventually minimized iteratively on a non-local graph defined by the original noisy image and the previous estimation.

The iterative PPB is related to the Expectation-Maximization (EM) procedure [2]. As in PPB, the EM algorithm is a two steps iterative algorithm which converges to a local optimum depending on the initial estimation. The first step (E-Step) evaluates a complete-data likelihood expectation by computing sufficient parameters using a previous estimation. The second step (M-Step) maximizes this likelihood. In our filter, the first step requires to evaluate the weighted likelihood by computing the similarity probability using the previous estimation $\hat{\theta}^{i-1}$. The second step evaluates the WMLE to produce the new refined estimation. As in the EM procedure, PPB considers also the previous estimation as "true" parameters. According to our experiments, this consideration involves the model stability over the different iterations and provides the convergence of our method. The algorithm differs from the EM procedure in the definition of the objective function since our function is not related to a complete-data likelihood expectation. Our latent variable is the set of the indicator functions $\delta_{\mathcal{S}_{\theta_s^*}}(t)$. Under the NL means assumption, a model of $\delta_{\mathcal{S}_{\theta_s^*}}(t) = 1$ is provided by the similarity of the neighborhoods of s and t. However, these neighborhoods cannot be used to create a model of $\delta_{\mathcal{S}_{\theta_s^*}}(t) = 0$ (having different neighborhoods does not imply following different distributions). Then, our objective function considers only the event $\delta_{\mathcal{S}_{\theta_s^*}}(t) = 1$, whereas a complete-data likelihood expectation would be done over all possible values of the latent variable: $\delta_{\mathcal{S}_{\theta_s^*}}(t) = 1$ and $\delta_{\mathcal{S}_{\theta_s^*}}(t) = 0$. Finally, our latent variable definition makes the algorithm locally defined for all sites s. Then, PPB 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 two different noise distributions. On the one hand, we will consider images corrupted by an additive White Gaussian Noise (WGN). On the other hand, we will consider multiplicative speckle noise present in SAR images. Finally, the setting of the parameters and the complexity are discussed.

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:

$$p(I_{s,k}, I_{t,k} | \mu^*_{s,k} = \mu^*_{t,k}) = \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{|I_{s,k} - I_{t,k}|^2}{4\sigma^2}\right),$$

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

Finally the weights $w(s,t)^{(it.\ PPB)}$ at iteration i can be defined as

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

In a non-iterative version, $T \to +\infty$, the filter is exactly the NL means filter. Then PPB can be considered as an iterative generalization of the NL means filter. Note that the Gaussian kernel, defined by the weights α_k in Equation 3, is not use in PPB, but can be introduced easily in the model.

In SAR images, the information sought (the reflectivity) is generally considered to be corrupted by the multiplicative Goodman's speckle noise model [24]. Thus, the pixel amplitudes A_s are modeled as independent and identically distributed according to the following Rayleigh distribution

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

where R^* is the underlying reflectivity image. Then, from the first order optimality condition, the following estimation

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

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

$$p(A_{s,k}, A_{t,k} | R^*_{s,k} = R^*_{t,k}) = \frac{4A_{s,k}A_{t,k}}{A^2_{s,k} + A^2_{t,k}},$$

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

Finally the weights $w(s,t)^{(it.\ PPB)}$ at iteration i can be defined as

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

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

$$\log\left(\frac{A_1}{A_2} + \frac{A_2}{A_1}\right) \tag{7}$$

where A_1 and A_2 are two observed amplitude values.

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 [25], the fast non-local means [26], and the solution implemented here and proposed by Darbon *et al.* in [27] with a time complexity given by $O(4|\Omega||W|)$. For best performances, Buades *et al.* suggest to use a search window of size $|W| = 21 \times 21$ and a similarity window of size $|\Delta| = 7 \times 7$ [1]. Kervrann and Boulanger showed that the size of $|W|$ acts as a bias-variance trade-off on the estimation [23]. When the window size increases, the variance decreases but the estimation is more biased because there are more values coming from different distributions. According to this trade-off, the initial image parameter θ^1 is computed by the iterative PPB 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.

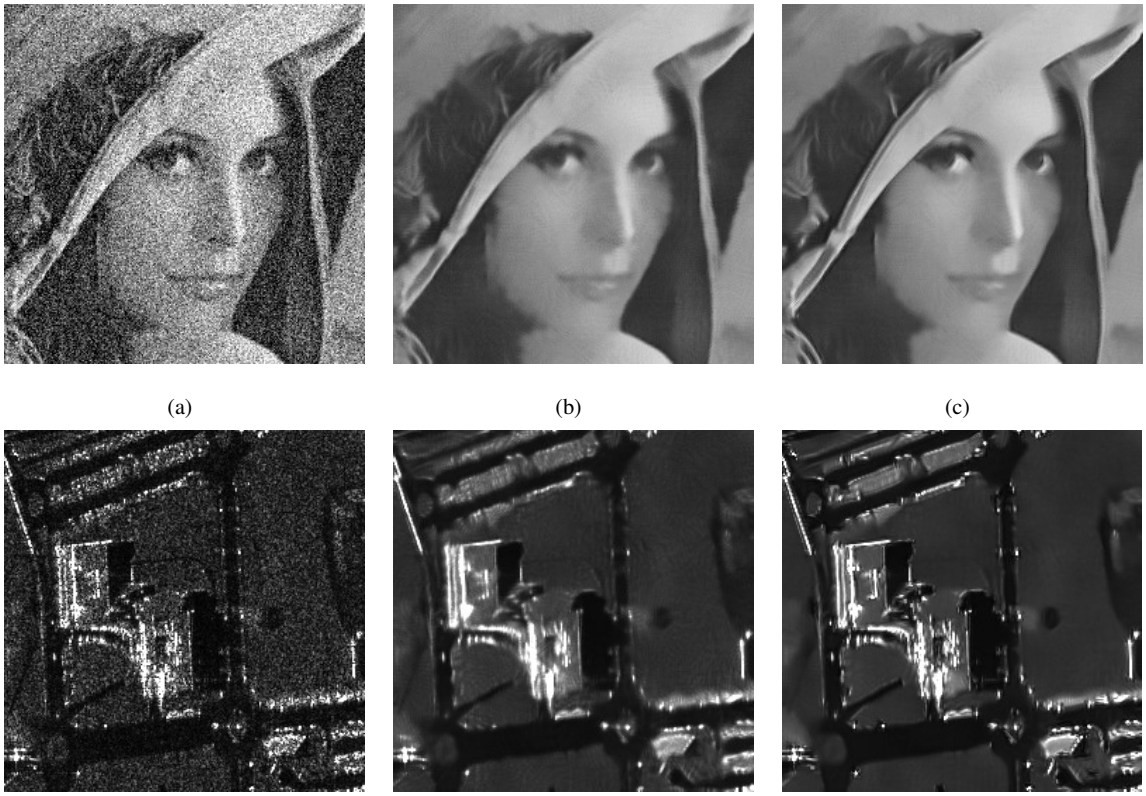
V. Experiments and Results

A. Qualitative Evaluation of the Denoising Algorithms

This section presents the results obtained on two images. The first image is the 512×512 *Lena* image corrupted by an additive WGN with a standard deviation $\sigma = 40$. The second image is a 512×512 SAR image of the CNES in Toulouse (France) sensed by RAMSES (ONERA) and provided by the French spatial agency (CNES). This image is assumed to follow the multiplicative Goodman's Speckle Noise (GSN) model.

On both images, the non-iterative and the iterative Probabilistic Patch-Based (PPB) filters are applied. Note that in the case of the non-iterative filter applied on the "Lena" image, our filter provides the same result as the NL means filter. For all experiments, we use a search window of size $|W| = 21 \times 21$ and a similarity window of size $|\Delta| = 7 \times 7$. In case of the non-iterative procedure, the parameter h^2 has been tuned to $29.0\sigma^2$ for additive WGN, and $h^2 = 2.65$ for multiplicative GSN. For the iterative procedure, the parameters have been set to $h^2 = 37.2\sigma^2$ and $T = 0.33$ for additive WGN, and $h^2 = 5.54$ and $T = 2.39$ for multiplicative GSN

The denoised images are displayed on Figure 1. For reasons of visibility, only 256×256 sub-images are shown here, the full size images are available at http://www.tsi.enst.fr/~deledall/ppb.php. Note that the squared root of the estimated reflectivity images are displayed in order to be homogeneous and comparable to the original amplitude image. The images obtained with the iterative methods seem to be



(a) (b) (c)

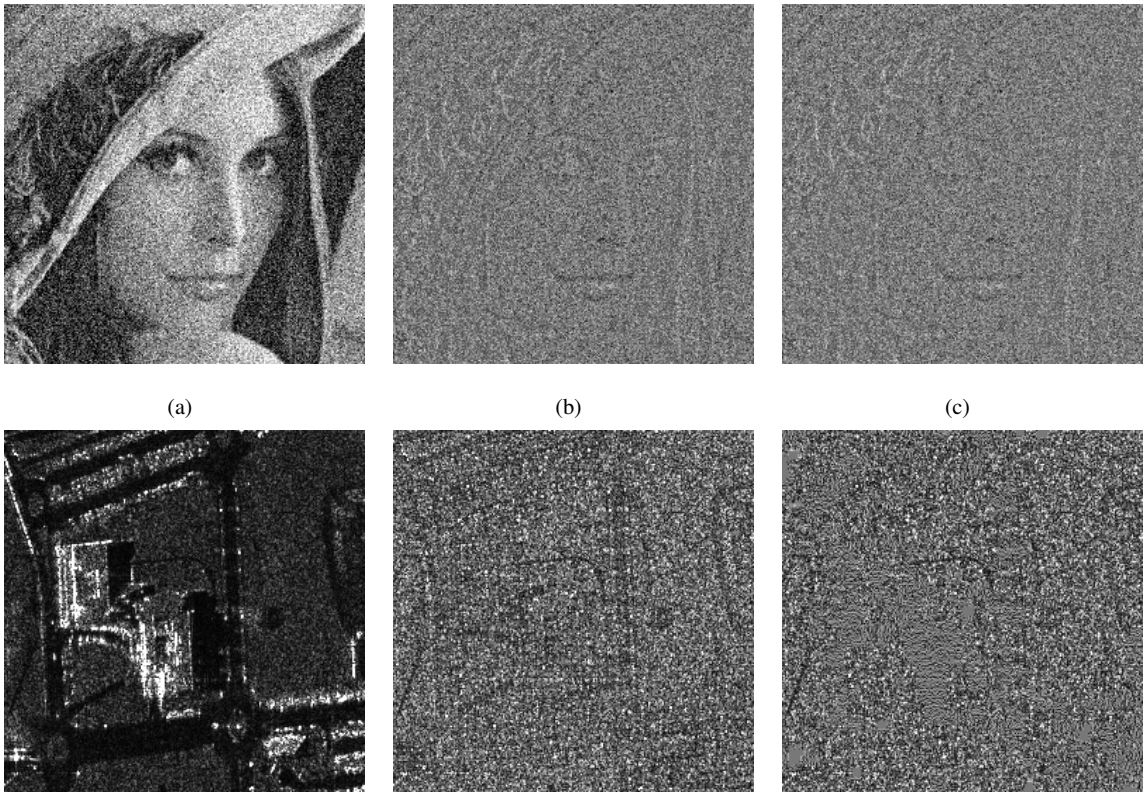
(d) (e) (f)

Fig. 1. (a-c) "Lena" images and (d-f) SAR images of Toulouse ©DGA ©ONERA. (a) Image corrupted by an additive white Gaussian noise with $\sigma = 40$. (b,c) Denoised image obtained respectively with our non-iterative (NL means) and iterative filter derived for additive white Gaussian noise. (d) SAR image originally corrupted by a multiplicative Goodman's speckle noise. (e,f) Denoised image obtained respectively with our non-iterative and iterative filter derived for multiplicative Goodman's speckle noise.

well smoothed with a better edge and shape preservation. In the case of SAR images, the bright scatterers are restored with a high precision (as the three bright lines on the building left side which constitute relevant targets). Unfortunately, all methods seem to attenuate thin and dark structures as the mouth of *Lena* and the two thin streets existing in the SAR image.

B. Analysis of the "Method Noise"

In the case of additive WGN and multiplicative GSN model, the information removed by a filter is obtained respectively by the difference $u - \hat{\mu}$ and the ratio $A/\hat{R}^{1/2}$. That is referred in [1] as "method noise". An ideal denoising procedure would give a method noise without any structure (i.e., uncorrelated),



(a) (b) (c)

(d) (e) (f)

Fig. 2. (a-c) "Lena" images and (d-f) SAR images of Toulouse ©DGA ©ONERA (a) Image corrupted by an additive white Gaussian noise with $\sigma = 40$. (b,c) "Method noise" obtained respectively with our non-iterative (NL means) and iterative filter derived for additive white Gaussian noise. (d) SAR Image originally corrupted by a multiplicative Goodman's speckle noise. (e,f) "Method noise" obtained respectively with our non-iterative and iterative filter derived for multiplicative Goodman's speckle noise. The method noise should ideally correspond to a noise image free of any structure.

and following the noise distribution.

If object structures are present in method noise, that means that the related objects are not well restored in the denoised image. Figure 2 displays the method noise for the 4 denoised images presented above. It can be noticed that structures are less present when the iterative approach is used and the noise seems more decorrelated. The remaining structures correspond actually to the damaged ones in the denoised images such as the mouth of *Lena* and the two thin streets existing in the SAR image.

To validate our filter, we also compute numerical statistics over the obtained method noises. In case of additive WGN, the removed information should be Gaussian distributed with zero mean $\hat{\mu} = 0$ and standard deviation $\hat{\sigma} = 40$. For multiplicative GSN, it should be Rayleigh distributed with parameter

$\hat{R} = 1$ and standard deviation $\sqrt{1 - \pi/4} \approx 0.463$. In all cases, the correlation between the method noise and its one pixel left-translated version is computed and should be close to 0. Table I shows the obtained statistics for the fourth images. All results are improved by the iterative approach. Moreover, the statistics are very close to the underlying distribution with a little correlation.

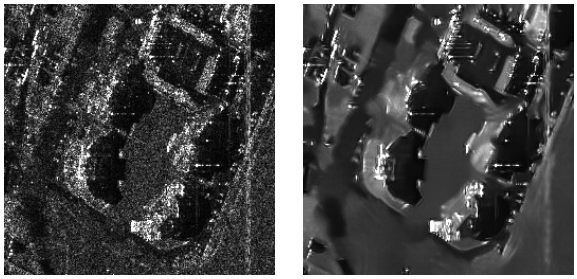
TABLE I

STATISTICS OF THE "METHOD NOISE"

Lena image	$\hat{\mu}$	$\hat{\sigma}$	Correlation
non-it. PPB (NL means)	0.086	36.88	0.015
it. PPB	$\mathbf{0.018}$	$\mathbf{37.06}$	$\mathbf{0.009}$

SAR image of Toulouse	$\hat{R}$	$\hat{\sigma}$	Correlation
non-it. PPB	0.826	0.422	0.045
it. PPB	$\mathbf{0.863}$	$\mathbf{0.429}$	$\mathbf{0.027}$

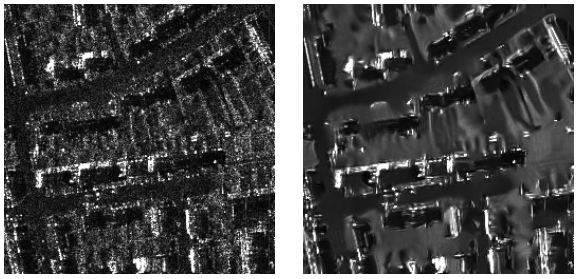
C. Overview



(a) (b)

Fig. 3. SAR image of Bayard (France). (a) Original image ©DGA ©ONERA. (b) Denoised image obtained by our iterative filter.

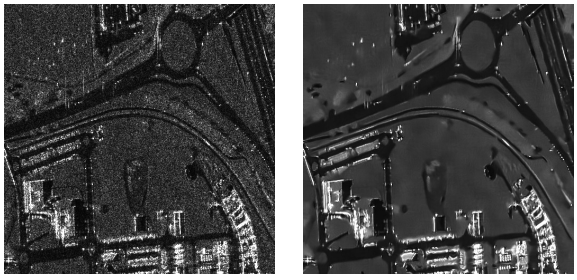
This section presents an overview of different results obtained with PPB on real SAR images. In all experiments, the algorithm is executed with a search window of size 21×21, a similarity window of size 7×7 and the parameters $h^2 = 5.54$ and $T = 2.39$. Figure 3 and 4 are two SAR acquisitions of Bayard and Cheminot from Saint-Pol-sur-Mer (France), sensed in 1996 by RAMSES of ONERA. Figure 5 is a SAR acquisition of the CNES in Toulouse (France) sensed also by RAMSES and provided by the CNES. Figure 6 displays a SAR image of an agriculture region in Lelystadt (Netherlands), sensed



(a) (b)

Fig. 4. SAR image of Cheminot (France). (a) Original image ©DGA ©ONERA. (b) Denoised image obtained by our iterative filter.

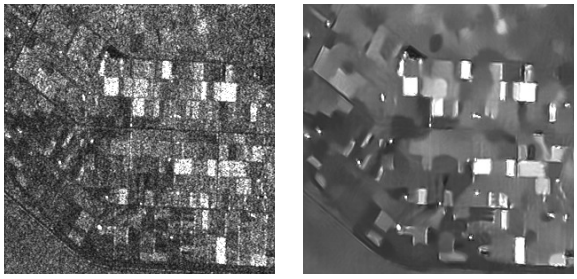
by ERS-1 with 3-looks (PRI data) and provided by the European Space Agency (ESA). These 4 images provide a testing set which presents a good diversity: different sensors (RAMSES/ERS), different scenes (urban/agricultural), different data (mono-look/multi-looks).



(a) (b)

Fig. 5. SAR image of Toulouse (France). (a) Original image ©DGA ©ONERA. (b) Denoised image obtained by our iterative filter.

On these 4 different images, the speckle effect is strongly reduced. The spatial resolution seems to be very well preserved: buildings, sidewalks, streets, fields seems to be well restored. Note that for the 3-looks image, we adjust the parameters $h^2 = 4.41$ and $T = 0.90$ to produce an optimal result. Moreover, bright scatterers (numerous in urban area) are very well conserved. The fields, streets and buildings shapes are perfectly restored. Only few thin and dark details are lost during the filtering, such the dirt tracks between the fields of Lelystadt.



(a) (b)

Fig. 6. 3-looks SAR image of Lelystadt (France). (a) Original image ©ESA. (b) Denoised image obtained by our iterative filter.

D. Algorithm Convergence

In practice, the algorithm converges and the solution depends on the initial estimation $\hat{\theta}^1$. Figures 7 and 8 show the evolution of the similarity between the successive estimations $\hat{\theta}^t$ and $\hat{\theta}^{t+1}$ for images respectively corrupted by additive WGN and multiplicative GSN. According to the above comments, the similarity is expressed respectively as an Euclidean distance and with the criterion defined in Equation 7 (normalized according to the image size $|\Omega|$). In case of additive WGN corrupted images, the curves

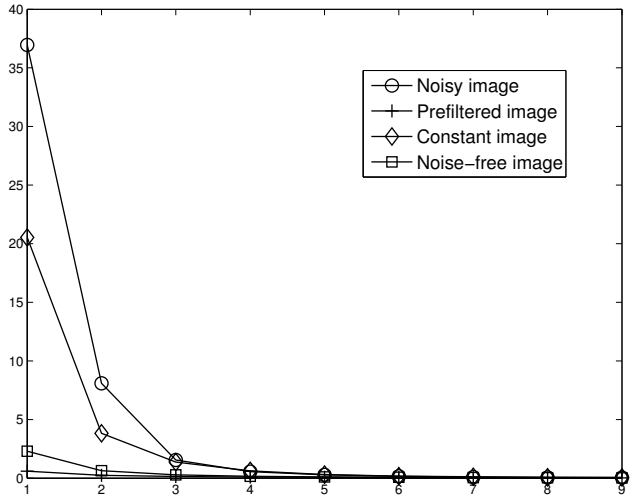


Fig. 7. Euclidean distance evolution, in function of the iterations, between two successive estimations of an image corrupted by additive white Gaussian noise. The evolution is measured for different initial estimations: the noisy image, a constant image, the noise-free image and our prefiltered estimation procedure (see Section IV).

converge to 0 whatever the initialization. In case of multiplicative GSN corrupted images, the curves converge to $\log(2)$ whatever the initialization. That means that, after enough iterations, they are no more change between two successive iterations: the iterative procedure has reached convergence. Moreover, using a prefiltered image (built as explained in Section IV) seems to accelerate the speed of convergence. Finally, we can notice that the noise-free image is not necessarily a local optimum. Indeed, when using the noise-free image as initialization, some changes are applied before reaching convergence.

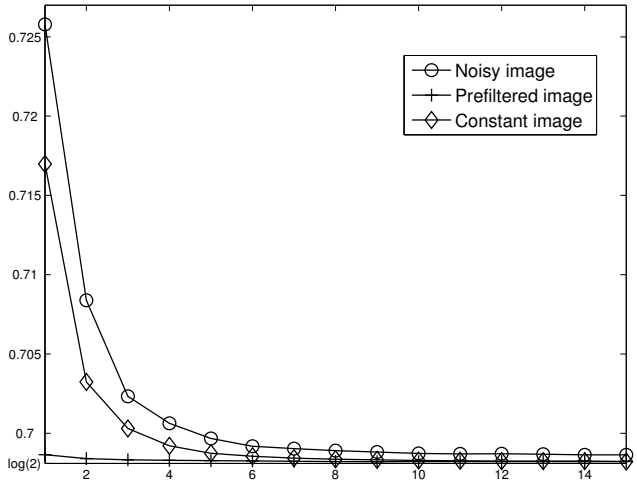


Fig. 8. Evolution of the criterion defined in Equation 7, in function of the iterations, between two successive estimations of an image corrupted by multiplicative Goodman's speckle noise. The evolution is measured for different initial estimations: the noisy image, a constant image and our prefiltered estimation procedure (see Section IV).

VI. Conclusion

A new method was proposed for image denoising which can be adapted to different noise distributions. This method is based on the non local means filter [1] and is related to the iterative Expectation Maximization procedure [2]. The proposed filter gives promising results for optical images, damaged by an additive White Gaussian Noise (WGN), as well as for Synthetic Aperture Radar (SAR) images, damaged by a multiplicative speckle noise. The noise, present in the input images, is perfectly smoothed in the homogeneous regions and the object contours are very well restored (conservation of the resolution). Moreover we can consider from our numerical experiments, that the reflectivity in SAR images is very well recovered, with a good restoration of bright scatterers. A drawback of the filter is the suppression of thin and dark details in the regularized images. In a future work, the local adaptive window selection,

presented in [23], will be studied and generalized to recover such thin and dark details in case of non additive WGN. We have seen that the filter elaboration, based on the statistics of the processed images, has led to define a suitable similarity criteria for SAR images. This similarity criteria will be studied in other applications such as displacement estimation.

ACKNOWLEDGMENTS

The authors would like to thank Marc Sigelle for his interesting comments and constructive criticisms, the Office National d'Etudes et de Recherches Aérospatiales and the Délégation Générale pour l'Armement for providing the data.

APPENDIX A
PROOF OF THE ALGORITHM FOR ADDITIVE WGN

The similarity probability is given by

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

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

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

Respectively the following holds

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

$$= \quad -\frac{\mu_1^2 - \mu_2^2}{2\sigma^2} + \frac{\mu_1\mu_2 - \mu_2^2}{\sigma^2}$$

Then, the symmetric Kullback-Leibler divergence is given by

$$-\frac{\mu_1^2 - \mu_2^2}{2\sigma^2} + \frac{\mu_1^2 - \mu_1\mu_2}{\sigma^2} + \frac{\mu_1^2 - \mu_2^2}{2\sigma^2} - \frac{\mu_1\mu_2 - \mu_2^2}{\sigma^2}$$
$$= \quad \frac{\mu_1^2 - 2\mu_1\mu_2 + \mu_2^2}{\sigma^2} = \frac{|\mu_1 - \mu_2|^2}{\sigma^2}$$

APPENDIX B

PROOF OF THE ALGORITHM FOR MULTIPLICATIVE GSN

First, note the following equality

$$\int_0^{\infty} \frac{A}{x^2} \exp\left(-\frac{B}{x}\right)\, dx$$
$$= \quad A \int_{\infty}^{0} \frac{x^2}{B^2} \exp(-x) \times -\frac{B}{x^2}\, dx$$
$$= \quad \frac{A}{B} \underbrace{\int_0^{\infty} \exp(-x)\, dx}_{=1}$$
$$= \quad \frac{A}{B}$$

Then, the similarity probability is given by

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

The Kullback-Leibler divergence is given by

$$\int_0^{+\infty} \frac{2t}{R_1} \exp\left(-\frac{t^2}{R_1}\right) \times \left(\log\frac{R_2}{R_1} + \frac{t^2}{R_2} - \frac{t^2}{R_1}\right)\, dt$$

$$= \quad \log\frac{R_2}{R_1} \underbrace{\int_0^{+\infty} \frac{2t}{R_1} \exp\left(-\frac{t^2}{R_1}\right)\, dt}_{=1}$$
$$+ \left(\frac{1}{R_2} - \frac{1}{R_1}\right) \underbrace{\int_0^{+\infty} \frac{2t^3}{R_1} \exp\left(-\frac{t^2}{R_1}\right)\, dt}_{=R_1}$$
$$= \quad \log\frac{R_2}{R_1} + \frac{R_1}{R_2} - 1$$

Respectively the following holds

$$\int_0^{+\infty} \frac{2t}{R_2} \exp\left(-\frac{t^2}{R_2}\right) \times \left(\log \frac{R_2}{R_1} + \frac{t^2}{R_2} - \frac{t^2}{R_1}\right) dt$$

$$= \log \frac{R_2}{R_1} + 1 - \frac{R_2}{R_1}$$

Then, the symmetric Kullback-Leibler divergence is given by

$$\log \frac{R_2}{R_1} + \frac{R_1}{R_2} - 1 - \log \frac{R_2}{R_1} - 1 + \frac{R_2}{R_1}$$

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

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] 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.

[3] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[4] 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.

[5] T. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 231–241, 2001.

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

[7] D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image and Vision Computing*, vol. 22, no. 1, pp. 73–81, 2004.

[8] 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.

[9] 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.

[10] 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.

[11] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient Nonlocal Means for Denoising of Textural Patterns," *IEEE Transactions on Image Processing*, 2007.

[12] G. Peyre, S. Bougleux, and L. Cohen, "Non-local Regularization of Inverse Problems," in *Proc. of ECCV*, vol. 2008, 2008.

[13] M. Mignotte, "A non-local regularization strategy for image deconvolution," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2206–2212, 2008.

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

[15] J. Lee, "Digital noise smoothing and the sigma filter," *Comput. Graph. Image Process*, vol. 24, pp. 255–269, 1983.

[16] 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.

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

[18] 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.

[19] M. Sambora, "A Non-local Maximum-Likelihood Denoising Algorithm," in *Aerospace Conference, 2007 IEEE*, 2007, pp. 1–7.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] J. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Am*, vol. 66, no. 11, pp. 1145–1150, 1976.

[25] 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.

[26] 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.

[27] 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.

Dépôt légal : 2009 – 1er trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

TELECOM ParisTech

Institut TELECOM - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.telecom-paristech.frfr

Département TSI