MAP entropy estimation: applications in robust image filtering

J. I. de la Rosa
ismaelrv@ieee.org

Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Antiguo Camino a la Bufa No. 1, Col. Centro. C. P. 98000, Zacatecas, Zac., México

J. Villa-Hernandez

Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Antiguo Camino a la Bufa No. 1, Col. Centro. C. P. 98000, Zacatecas, Zac., México

E. de la Rosa M.

Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Antiguo Camino a la Bufa No. 1, Col. Centro. C. P. 98000, Zacatecas, Zac., México

E. Gonzalez-Ramirez

Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Antiguo Camino a la Bufa No. 1, Col. Centro. C. P. 98000, Zacatecas, Zac., México

O. Gutierrez

Instituto Tecnológico Superior de Fresnillo, Av. Tecnológico No. 2000 Col. Solidaridad, C.P. 99010 Fresnillo, Zac., México

N. Escalante

Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Antiguo Camino a la Bufa No. 1, Col. Centro. C. P. 98000, Zacatecas, Zac., México

R. Ivanov

Facultad de Física, Universidad Autónoma de Zacatecas, Calz. Solidaridad esquina paseo de la Bufa S/N, C. P. 98060, Zacatecas, Zac., México

G. Fleury

École Supérieure d'Électricité (SUPELEC), Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette Cedex, France

We introduce a new approach for image filtering in a Bayesian framework. In this case the probability density function (pdf) of the likelihood function is approximated using the concept of non-parametric or kernel estimation. The method is based on the generalized Gaussian Markov random fields (GGMRF), a class of Markov random fields which are used as prior information into the Bayesian rule, which principal objective is to eliminate those effects caused by the excessive smoothness on the reconstruction process of images which are rich in contours or edges. Accordingly to the hypothesis made for the present work, it is assumed a limited knowledge of the noise pdf, so the idea is to use a non-parametric estimator to estimate such a pdf and then apply the entropy to construct the cost function for the likelihood term. The previous idea leads to the construction of Maximum a posteriori (MAP) robust estimators, since the real systems are always exposed to continuous perturbations of unknown nature. Some promising results of three new MAP entropy estimators (MAPEE) for image filtering are presented, together with some concluding remarks.
[DOI: http://dx.doi.org/10.2971/jeos.2013.13047]

Keywords: Digital image processing, image recognition, algorithms and filters, image reconstruction-restoration, inverse problems

1 INTRODUCTION

The image restoration approaches or recuperation of an image to its original condition given a degraded image, passes by reverting the effects caused by a distortion functional which in some cases is known or must be estimated. The basic idea in Bayesian estimation is to construct a Maximum a posteriori (MAP) of the modes or so called estimator of true images by using Markov Random Fields (MRF's). The evolution of the basic idea has caused the development of new algorithms which consider new models of contextual information which is led by the MRF's and the final aim is the restoration of real images. The idea is based on a robust scheme which could be adapted to reject outliers, tackling situations where noise is present in different forms during the acquisition process. In the case of classical MAP filters, usually the additive Gaussian noise is considered, however in some applications this noise is non-Gaussian [1] or unknown (with some partial knowledge). This is a source of information which imposes a key rule in

the image processing context (the contextual or spatial information), that represents the likelihood function or correlation between the intensity values of a well specified neighborhood of pixels. Also, the modelling when using MRF's takes into account such spatial interaction and it was introduced and formalized in [2] where it is shown the powerfulness of these statistical tools [3]–[7]. The image modeling in the context of the present paper lead us to assume a limited knowledge about the image noise pdf (see Eq. (3), where $p(e) = p(y|x)$), so we propose to use the data itself to obtain a non-parametric Entropy Estimate (EE) of the log-likelihood pdf ($\widehat{p}_{n,h}(e)$) [8]–[10], similar methodologies have been proposed in recent works by Milanfar et al, [11, 12]. Then the log-likelihood will be optimized together with a log-MRF to obtain the MAP image estimation.

A variety of applications in signal processing and instrumen-

J. Europ. Opt. Soc. Rap. Public. **8**, 13047 (2013)

J. I. de la Rosa, *et al.*

tation are based in statistical modelling analysis. One of the most used is the linear regression model (simple or multivariable in the case of images)

$$y_{i,j} = x_{i,j}^{\top}\theta_{i,j} + e_{i,j}, \text{ with } e \sim p(e), \qquad (1)$$

where y represents the response (observed data, acquired data, or explained variables), to x explicative variables for $i = 1, \ldots, N$ and $j = 1, \ldots, M$, and a system response parameterized by θ which is associated to data (y, x). In some applications θ are functional parameters which will be estimated by an identification procedure if x are known, but if θ are known and x are unknown, the estimation is made about x, or the estimation can be made for both cases (e.g. blind deconvolution). The e variables are independent random processes identically distributed accordingly to $p(e)$. A natural extension of the linear regression model, is the non-linear regression model, but now it is based on a parameterized function $f(\cdot)$ (see Eq. (2)). This function is nonlinear with respect to the parameters, and its use is also considered because it has been shown in a large variety of signal processing and control applications that the modelling when using nonlinear functions could be more realistic. The perturbations affecting the analyzed system are also modelled as stochastic processes. Then, for this case:

$$y_{i,j} = f(x, \theta)_{i,j} + e_{i,j}, \text{ with } e \sim p(e). \qquad (2)$$

The proposed nonparametric procedure is led by the classical kernel estimators, it will be presented in Section 3. The principal contribution of this work is explained in Sections 4 and 5, where a comparison of the different proposed MAP-Entropy estimators (MAPEE) is presented. This comparison shows the performance and the improvement of estimation results when one takes into account the influence of the bandwidth parameter h used by the classical kernel estimators, and when it is discarded by using two other nonparametric procedures. Thus, a more general criterion is proposed and it is constructed on the basis of three nonparametric estimators of the residuals distribution, and some new efficient estimators could be built (efficiency in the sense of calculus time) [13, 14] and [15]. Finally, some concluding remarks about the MAPEE proposed estimators are given in Section 6.

2 BAYESIAN FILTERING AND LOG-LIKELIHOOD APPROXIMATED BY EE

The problem of image estimation (e.g. restoration) into a Bayesian framework deals with the solution of an inverse problem, where the estimation process is carried out in a whole stochastic environment. All variables presented along this paper are, x: which represent a Markov random field (or image to be estimated), y: represents the observed image with noise and distorted, and $\hat{x}$: is the estimator of x with respect to data y, and $p(\cdot)$ is a probability density function.

Maximum A Posteriori (MAP) estimator:

$$\begin{aligned} \hat{x}_{\mathbf{MAP}} &= \arg\max_{x \in \mathbf{X}} p(x|y) \\ &= \arg\max_{x \in \mathbf{X}} \left(\log p(y|x) + \log g(x)\right) \\ &= \arg\min_{x \in \mathbf{X}} \left(-\log p(y|x) - \log g(x)\right), \end{aligned} \qquad (3)$$

in this case, the estimator is regularized by using a Markov random field function $g(x)$ which model all prior information as a whole probability distribution, where $\mathbf{X}$ is the set of pixels x capable to maximize or minimize $p(x|y)$, and $p(y|x)$ is the likelihood function from y given x.

2.1 Markov random fields

The Markov random fields (MRF) can be represented in a general way by using the following cost function:

$$g(x) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathbf{C}} V_c(x)\right), \qquad (4)$$

where Z is a normalization constant, $\mathbf{C}$ is a set of "cliques" c or local neighborhoods of pixels, and $V_c(x)$ is a weighting function given over the local group of points c. Generally, the "cliques" correspond to the sets of neighborhoods of pixels if $\forall s, r \in c$, s and r are neighbors, and one can construct a neighborhood system called ∂s; for the 8 closest neighbors $\partial s = \{r : |s - r| < 2\}$. The Markov random fields have the capacity to represent various image sources.

There is a variety of MRF models which depend on the cost functions also known as potential functions that can be used. Each potential function characterizes the interactions between pixels in the same local group [4, 6].

2.2 Likelihood pdf Entropy estimators (EE)

A classical procedure to estimate x when θ is known (from Eq. (1) and (2)), is based on a cost function or criterion $\mathcal{J}(x)$ which varies in function $\psi(\cdot)$ of residuals or noise $e(x)$, where:

$$e_{i,j}(x) = y_{i,j} - x_{i,j}^{\top}\theta_{i,j} \quad \text{or} \quad e_{i,j}(x) = y_{i,j} - f(x, \theta)_{i,j}, \qquad (5)$$

and so

$$\mathcal{J}(x) = \sum_{i=1}^{N}\sum_{j=1}^{M} \psi(e_{i,j}(x)). \qquad (6)$$

This is, for example, the case of the maximum likelihood (ML) estimator:

$$\hat{x}_{\mathbf{ML}} = \arg\min_{x \in \mathbf{X}} \left[-\sum_{i=1}^{N}\sum_{j=1}^{M} \log p(e_{i,j}(x))\right]. \qquad (7)$$

The ML estimator is optimal when all information about the distribution $p(e)$ is accessible. When the knowledge about $p(e)$ is imprecise or wrong, the estimator $\hat{x}_{\mathbf{ML}}$ is possibly suboptimal. Moreover, under certain circumstances, in image processing restoration, it results in an ill-posed problem or produces excessive noise and also causes smooth of edges. The regularization of the ML estimator gives a more effective approach called Maximum A Posteriori (MAP) estimator which reduces noise and smoothness at the same time.

Our proposition for a new MAP scheme is to use a Generalized Gaussian MRF introduced by Bouman and Sauer in [4, 7], together with three kernel estimators used in [8, 9], and [10] to obtain cost functionals or criterions based on the entropy of the approximated likelihood function (first term of Eq. (3)) $\hat{p}_{n,h}(e)$. Thus, $-\log p(y|x)$ is built on the basis of the entropy

J. Europ. Opt. Soc. Rap. Public. **8**, 13047 (2013)

J. I. de la Rosa, *et al.*

of an estimate (EE) version $\widehat{p}_{n,h}(e)$ of the distribution $p(e)$. A first proposition is due to Pronzato and Thierry [16, 17], the approximation is obtained using the classical kernel estimators which uses the empirical distribution of the random vector $e_1(x), \ldots, e_n(x)$, the next expression denote such estimators:

$$\widehat{p}_{n,h}(e) = \widehat{p}_{n,h}(e|e_1(x), \ldots, e_n(x)) = \frac{1}{n} \sum_{i=1}^{n} K_h(e - e_i). \quad (8)$$

This expression assumes the hypothesis that $p(e)$ is symmetric, two times differentiable and positive, indeed, it is assumed that $K(\cdot)$ is a kernel weighted function which satisfies some imposed conditions treated in the work of Masry [18] and subsequently taken back by Devroye [19]–[22], Berlinet [23], and Loader [24] in some of their research work. The bandwidth $h = h_n$ is given in function of the sample size n, this parameter could be considered as a sequence of positive numbers that must satisfy: $h_n \to 0$ and $nh_n \to \infty$ when $n \to \infty$. The strong uniform consistency of $\widehat{p}_{n,h}(e)$ and its convergence toward $p(e)$, depend on a convenable procedure of bandwidth selection. A simple and faster procedure which has been retained in this work is the technique proposed and developed by Terrell [14, 15]. In the two dimensional kernel cases the previous idea has been extended in this work according to the equation:

$$\begin{aligned} \widehat{p}_{n,h}(e) &= \widehat{p}_{n,h}(e|e_{1,1}(x), \ldots, e_{n,n}(x)) \\ &= \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} K_h(e - e_{k,l}). \end{aligned} \quad (9)$$

for a sample size of $n \times n$. If the convergence and consistence of $\widehat{p}_{n,h}(e)$ is assumed, such that $\widehat{p}_{n,h}(e) \to p(e)$, then the entropy criterion over $\widehat{p}_{n,h}(e)$ can be approximated to $-\log p(y|x)$. The fact that the entropy of any probability density function is invariant by translation, leads to consider one practical artifact to build a suitable criterion. An extended criterion $\widehat{p}_{n,h}(e_E)$ is based on the residuals or noise extended vector which is given by: $e_E = \{e_1(x), \ldots, e_n(x), -e_1(x), \ldots, -e_n(x)\}$ and on a suitable choice of h:

$$\mathcal{J}_e(x) = \mathrm{H}_A(\widehat{p}_{n,h}(e_E)) \approx -\log p(y|x), \quad (10)$$

where $\mathrm{H}_A(f) = -\int_{-A_n}^{A_n} f(x) \log f(x) dx$. This last idea is also applied for the two dimensional case. Finally, if we assume that the EE is a version of the log-likelihood function into the MAP estimator, then a first version of the MAP-Entropy Estimator (MAPEE) which assumes unknown noise pdf can be constructed from the fact that $-\log p(y|x)$ can be approximated by the entropy of an estimate version $\widehat{p}_{n,h}(e)$ of the distribution $p(e)$, that is $\mathrm{H}_A(\widehat{p}_{n,h}(e_E))$, thus:

$$\widehat{x}_{\mathbf{MAPEE}} = \arg \min_{x \in \mathbf{X}} (\mathrm{H}_A(\widehat{p}_{n,h}(e_E)) - \log g(x)). \quad (11)$$

3 KERNEL STRUCTURES

A function of the form $K(z)$ is assumed as a fixed kernel $K_h(z) = 1/(h^d)K(z/h)$, where $h > 0$, this parameter is called the kernel bandwidth. The fundamental problem in kernel density estimation lies in both the selection of an appropriate value for h and the selection of the kernel structure. The choice of $K(z)$ could depend on the smoothness of $p(e)$. Three different kernels or nonparametric schemes are reviewed in this section to approximate $\widehat{p}_{n,h}(e_E)$. The gaussian kernel, which has proved to give good performance when h is selected by using the over-smoothed principle introduced by Terrell [14, 15] when the errors vector is generally of finite length n. The second kernel is the cosine based weights functions, where h is viewed in a different way. Finally, the third kernel is obtained from a recent class of Hilbert kernels [21]. It avoids the bandwidth h selection and its performance depends on other parameters, which selection is very easy (parameters d and k are defined in Section 3.3).

3.1 Normal or gaussian kernel

Among the different classical kernels [23], the gaussian kernel has been chosen since it is one of the most popular and easy to implement estimator. The next expression resumes this estimator by a sum of exponentials:

$$\widehat{p}_{n,h}(z) = \frac{1}{n^2 h \sqrt{2\pi}} \sum_{k=1}^{n} \sum_{l=1}^{n} \exp\left(-\frac{(z - z_{k,l})^2}{2h^2}\right). \quad (12)$$

In such a case, and considering that a fixed kernel structure has been chosen, Terrell [14] proposes to use an over-smoothed bandwidth h_n that corresponds to:

$$h = 3\left(\frac{1}{2\sqrt{\pi}(35)}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}}, \quad (13)$$

this bandwidth value guarantees the minimization of the Mean Integrated Squared Error (MISE), σ is the standard deviation of the sample z, and $\int K(z)^2 dz = \frac{1}{(2\sqrt{\pi})}$. Under mild conditions, the kernel density estimates based on the over-smoothing principle are consistent and for sufficiently large sample sizes, they will display all information present in the underlying errors $p(e)$ density.

3.2 Cosine based weights functions

A sequence of special weight cosine functions $c_q(z) = 1/(A_q)\cos^{2q}(z/2)$ requires only $O(q^2 n)$ arithmetic operations, where q is a slowly growing function that increases without bound together with n (i.e. $q(n) < n$ or $q(n) > n$). This estimator is similar to the kernel classical estimators and evaluates a series expansion in a more efficient way. The role of h is replaced by q ($c_q(z) \equiv K_h(z)$) and the efficiency is attained (according to the time of calculus), since the selection of q is generally simpler than h. Thus (9) is equivalent to

$$\widehat{p}_{n,q}(z) = \frac{1}{n^2 A_q} \sum_{k=1}^{n} \sum_{l=1}^{n} \left(\frac{1 + \cos(z - z_{k,l})}{2}\right)^q, \quad (14)$$

where the value of A_q could be approximated by

$$A_q \approx \frac{2\sqrt{\pi}}{\sqrt{q}}. \quad (15)$$

Sufficient consistency and convergence conditions are given in [13] for the one dimensional case.

J. Europ. Opt. Soc. Rap. Public. **8**, 13047 (2013)

J. I. de la Rosa, *et al.*

3.3 The Hilbert kernel

Finally, the Hilbert kernel estimate is used. The function $K_h(z) = 1/(h^d)K(z/h)$ is considered equivalent to $K(u) = 1/\|u\|^d$, where the smoothing factor h is canceled obtaining:

$$\widehat{p}_n(z) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{1}{\|z - z_{k,l}\|^d}. \quad (16)$$

The Hilbert estimates are viewed as a universally consistent density estimate whose expected performance (L_1, L_∞, pointwise) is monotone in n (at least in theory) for all densities. The consistency of this class of estimators is proved in [21] (see theorem 2). The Hilbert density estimate of order k ($k > 0$) is a redefined subclass that avoids the infinite peaks produced during estimation, in one dimensional case and using the value of $k = 2$ the kernel estimate is given by:

$$\widehat{p}_n(z) = \sqrt{\frac{4}{V_d^2 \pi n(n-1) \log n} \sum_{1 \le i < j \le n} \frac{1}{\text{Den}_{i,j}}}, \quad (17)$$

where $\text{Den}_{i,j} = \|z - z_i\|^{2d} + \|z - z_j\|^{2d}$ and V_d is the volume of the unit ball in $\mathbb{R}^b$. This last expression is also called Cauchy density estimate, due to its similarity to the multivariate Cauchy density, $\|\cdot\|$ denotes the L_2 metric on $\mathbb{R}^d$. Finally, it is assumed that $\widehat{p}_n(z) \to p(z)$ at least in probability for almost all z. For a suitable choice of A_n and alternatively of h_n, q, or d and k, these estimators could be "blind asymptotically efficient". The asymptotic properties and the strong consistency of the truncated entropy estimators were analyzed in [16]. More over, in recent works the powerfulness of these nonparametric tools have been largely used for different signal processing problems [11, 12, 25].

4 GENERALIZED GAUSSIAN MRF (GGMRF)

The MAPEE approach proposed here takes into account the proved robustness in presence of outliers of the minimum entropy estimators proposed in [8]–[10]. Obtaining now the complete cost functional structure for the $\widehat{x}_{\textbf{MAPEE}}$ estimator from the point of view of the MRF, the $\log g(x)$ used is based on a generalized Gaussian MRF introduced in [4, 7].

If one considers to generalize the Gaussian MRF (when $p = q = 2$ one has a Gaussian MRF), as proposed by Bouman [4], where the generalized potential functions can be limited such as

$$\rho(\Delta) = |\Delta|^p, \quad \text{for } 1 \le p \le 2 \quad (18)$$

obtaining the GGMRF

$$\log g(x) = -\lambda^p \left(\sum_{s \in \textbf{S}} a_s x_s^p + \sum_{\{s,r\} \in \textbf{C}} b_{sr} |x_s - x_r|^p \right) + \text{ct}, \quad (19)$$

where theoretically $a_s > 0$ and $b_{sr} > 0$, s is the site or pixel of interest and $\textbf{S}$ is the set of sites into the whole MRF, and r corresponds to the local neighbors. In practice it is recommended to take $a_s = 1$, since the likelihood term is not given in terms of quadratic $q = 2$ functional. In order to relax the convexity

problem, the following equation has been used

$$\log g(x) = -\lambda^p \left(\sum_{s \in \textbf{S}} a_s x_s^2 + \sum_{\{s,r\} \in \textbf{C}} b_{sr} |x_s - x_r|^p \right) + \text{ct}, \quad (20)$$

and from Eq. (2), $\log p(y|x)$ is strictly convex and so $\widehat{x}_{\textbf{MAPEE}}$ is continuous in y, and in p. The choice of the power p is capital, since it constrains the convergence speed of the local or global estimator, and the quality of the restored image, small values for p allows abrupt discontinuities modeling while large values smooth them.

Now, substituting this particular $\log g(x)$ into the Eq. (11) one could obtain at least three MAPEE estimators given by

$$\widehat{x}_{\textbf{MAPEE}_m} = \arg \min_{x \in \textbf{X}} \left\{ \text{H}_A \left(\widehat{p}_{n,h}(e_E) \right) \right.$$

$$\left. -\lambda^p \left(\sum_{s \in \textbf{S}} a_s x_s^2 + \sum_{\{s,r\} \in \textbf{C}} b_{sr} |x_s - x_r|^p \right) \right\}, \quad (21)$$

for $m = 1, 2, 3$ according to the three previous kernels: Normal, cosine and Hilbert.

5 RESULTS FOR NOISE FILTERING

Continuing with the problem of filtering noise, some estimation results were obtained when images are contaminated by Gamma, Beta, Uniform and impulsive noise, and there are no other type of distortions (all $\theta_{i,j} = 1$). The observation equation then could be

$$y = x + e, \text{ where } e \sim \mathcal{G}(\alpha, \beta), e \sim \mathcal{B}(\alpha, \beta), \dots$$

The first experiment was made considering Gamma noise where $\alpha = 0.5, 1.5, 2.5$ and $\beta = 1, 2, 3$, and also two factors of amplification of noise were used $\sigma_a = 5, 10$ ($\sigma_a \mathcal{G}(\alpha, \beta)$). The values of α and β are given such that the obtained degradation is perceptible and difficult to eliminate. Some results presented in this section were concerned experimenting extensively with four images: synthetic, Lena, Cameraman, and Boat (see Figure 1), to probe the performance of the presented
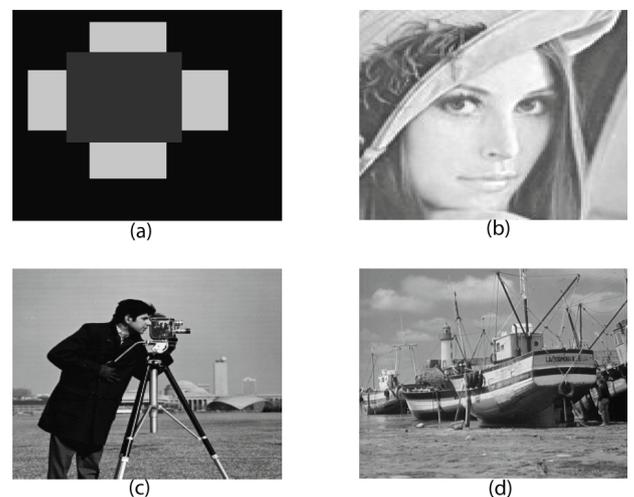

(a)

(b)

(c)

(d)

FIG. 1 The figure shows: (a) synthetical probe image; (b) the classical Cameraman image; (c) the classical Lena image; (d) the classical Boat image.

	$\alpha = 1.5, \beta = 2$	MAPEE1	MAPEE2	MAPEE3
Im. synthetic	PSNR noise	22.2	22.4	22.6
35×35	PSNR filtered	24.1	24.1	24.4
	Time (sec)	10.6	12.1	9.3
Im. Lena	PSNR noise	16.3	16.2	16.3
120×120	PSNR filtered	17.9	17.9	18
	Time (sec)	123.6	141.6	109.8
Im. Cameraman	PSNR noise	16.4	16.4	16.4
256×256	PSNR filtered	18.5	18.5	18.5
	Time (sec)	563.2	644.8	499.9
Im. Boat	PSNR noise	16.3	16.4	16.4
512×512	PSNR filtered	18.7	18.7	18.8
	Time (sec)	2,340.1	2,569.8	2,063.5

TABLE 1 Results obtained in evaluating the filtering capacity of the different MAPEE estimators for Gamma $\mathcal{G}(\alpha, \beta)$ noise, with $\sigma_a = 10$.

	$\alpha = 2.5, \beta = 1$	MAPEE1	MAPEE2	MAPEE3
Im. synthetic	PSNR noise	24.6	24.6	24.7
35×35	PSNR filtered	25.6	25.5	25.6
	Time (sec)	9.8	11.6	8.8
Im. Lena	PSNR noise	24.5	24.7	24.6
120×120	PSNR filtered	24.9	24.8	25.0
	Time (sec)	115.3	135.9	102.4
Im. Cameraman	PSNR noise	24.6	24.6	24.6
256×256	PSNR filtered	25.3	25.1	24.9
	Time (sec)	524.8	618.6	467.6
Im. Boat	PSNR noise	24.6	24.6	24.6
512×512	PSNR filtered	25.1	25.2	25.0
	Time (sec)	2,099.2	2,465.3	1,829.6

TABLE 2 Results obtained in evaluating the filtering capacity of the different MAPEE estimators for Beta $\mathcal{B}(\alpha, \beta)$ noise, with $\sigma_a = 20$.
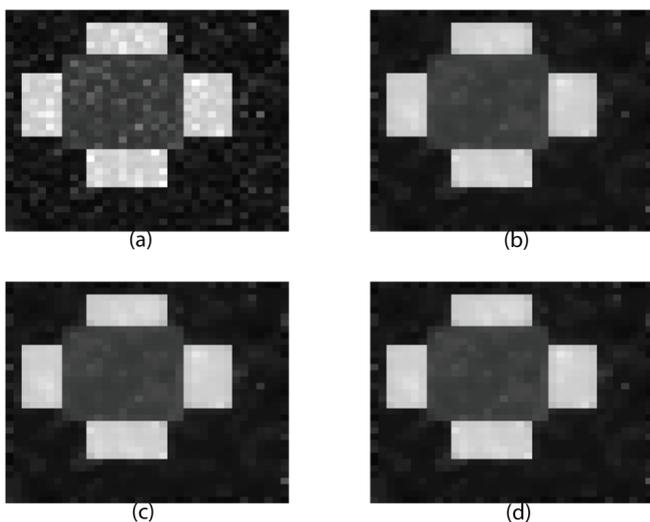

FIG. 2 Results for synthetical probe image: (a) describes the noisy image, for Gamma noise with $\alpha = 1.5$ and $\beta = 2$, and $\sigma_a = 5$; (b) filtered image using MAPEE1 (Normal kernel); (c) filtered image using MAPEE2 (cosine kernel); and, (d) filtered image using MAPEE3 (Hilbert kernel).
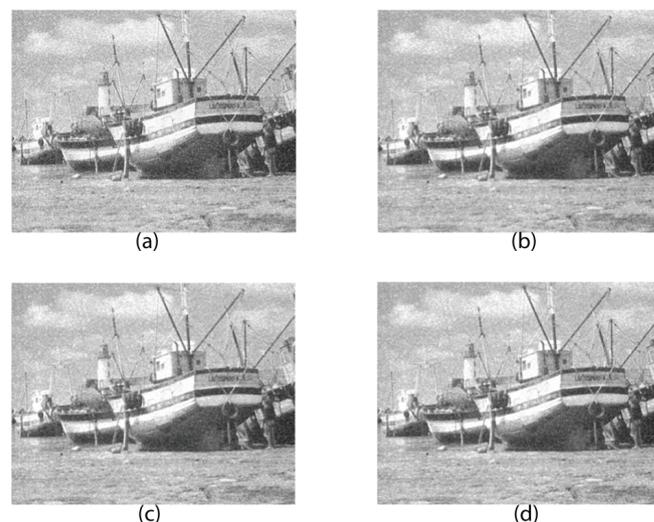

FIG. 3 Results for Boat image: (a) describes the noisy image, for Gamma noise; (b) filtered image using MAPEE1 (Normal kernel); (c) filtered image using MAPEE2 (cosine kernel); and (d) filtered image using MAPEE3 (Hilbert kernel).

MAPEE estimators. Figure 2 shows the filtering of the synthetical image using the three MAPEE estimators for parameters $\lambda = 10$, $p = 1.2$ of the GGMRF. Generally, with the three estimators the filtering task gives good visual results (see also Figure 3), but the time of computation is different between them, the fastest estimator was MAPEE3 (Subfigure 2(d), see also Table 1). On the other hand, some objective measurements were obtained, such as the PSNR depicted in Table 1 which agree to the visual results. Also times of computation are shown in Table 1.
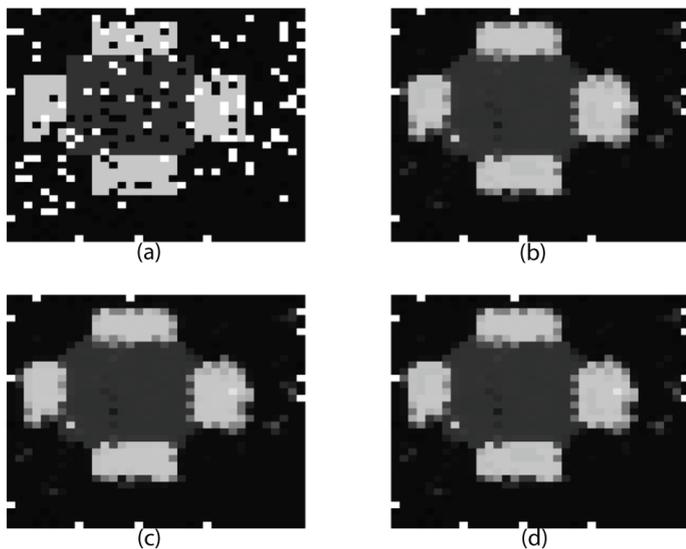
J. Europ. Opt. Soc. Rap. Public. **8**, 13047 (2013)

J. I. de la Rosa, *et al.*

FIG. 4 Results for synthetical probe image: (a) describes the noisy image, for impulsive noise: imnoise(X,'salt & pepper',0.18); (b) filtered image using MAPEE1 (Normal kernel); (c) filtered image using MAPEE2 (cosine kernel); and (d) filtered image using MAPEE3 (Hilbert kernel).
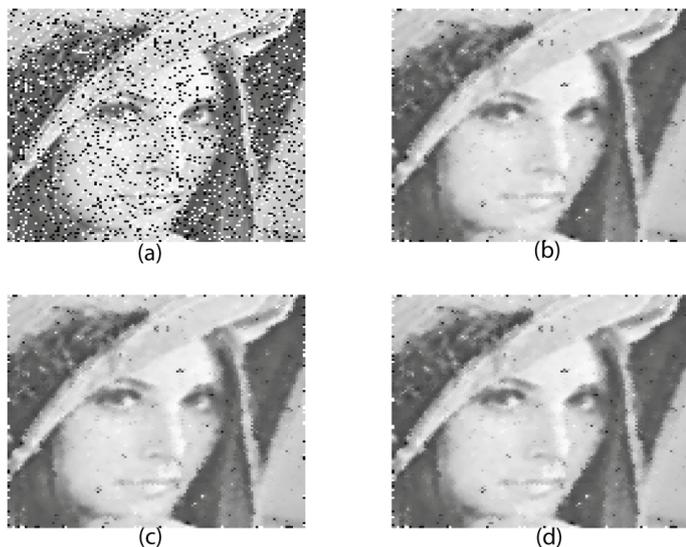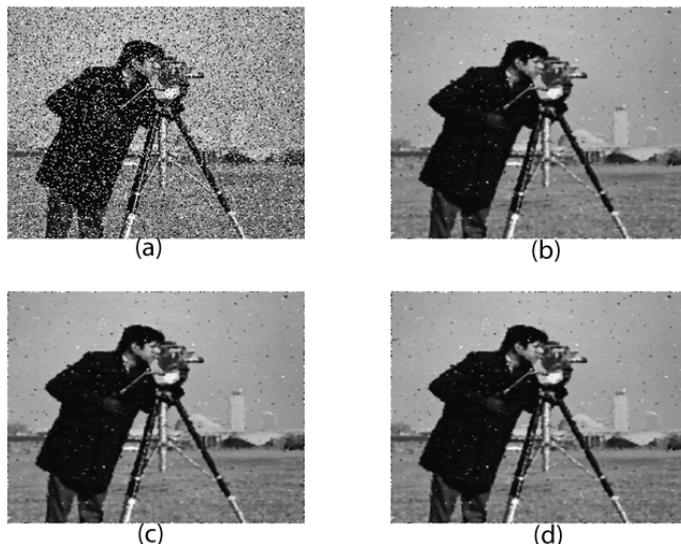

FIG. 6 Results for Cameraman image: (a) describes the noisy image, for impulsive noise; (b) filtered image using MAPEE1 (Normal kernel); (c) filtered image using MAPEE2 (cosine kernel); and (d) filtered image using MAPEE3 (Hilbert kernel).

FIG. 5 Results for Lena image: (a) describes the noisy image, for impulsive noise; (b) filtered image using MAPEE1 (Normal kernel); (c) filtered image using MAPEE2 (cosine kernel); and (d) filtered image using MAPEE3 (Hilbert kernel).

Some other experiments were conducted for Beta and impulsive noise. For the case of Beta noise generation we choose $\alpha = 1.5, 2.5, 5, 7.5$ and $\beta = 1$, and also two factors of amplification of noise were used $\sigma_a = 10, 20$ ($\sigma_a \mathcal{B}(\alpha, \beta)$). The filtering results are promising for non-Gaussian noise, the obtained results for the Beta noise filtering also confirm the robustness of the proposed approach, where the PSNR depicted in Table 2 shows the obtained improvement. Moreover, as one can see for impulsive noise in Figures 4, 5 and 6 the robustness is assured when using the three MAPEE estimators, where for instance, in the synthetic image case the PSNR was improved from 11.45 dB to 17.1 dB. The iterated method used in this work to minimize the obtained criterions for the MAPEE was the Levenberg–Marquardt method of MATLAB 2008 running in a computer with CORE i7 processor, and 4 Gbytes of RAM memory.

6 CONCLUSIONS

The selection among the three different kernel options, permits the performance improvement of the MAPEE estimators which could be classified in terms of simplicity and in terms of filtering quality. The obtained results for the three proposed estimators are favorable in general in the sense of robustness, and the fastest convergence is obtained for the MAPEE1 and MAPEE3 estimators (in the case of the MAPEE2 estimator $q = 40$, and for MAPEE3, $k = 2$, $d = 1$, and $V_d = 0.7071$. The value of q was changed alternatively in the range of $30 \leq q \leq 90$, and some performance improvement has been noticed). Moreover, the times of computation depends also on the type of noise to be filtered. A general scheme for MAPEE estimators has been introduced and it works in real frameworks, where the nonlinearity conditions of some systems could be present. For future works it exists the interest to implement procedures of MAPEE estimation into high level programming that will be characterized into algorithms to be used in DSP cards, and tasks such as image reconstruction and segmentation, also one can change the MRF and the optimization procedures to decrease the times of computation. The final aim is to use MAPEE to process real signals issued from real instrumentation or signal processing problems.

7 ACKNOWLEDGEMENTS

References

[1] N. Bertaux, Y. Frauel, P. Réfrégier, and B. Javidi, "Speckle removal using a maximum-likelihood technique with isoline gray-level regularization," J. Opt. Soc. Am. A **21(12)**, 2283–2291 (2004).

[2] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," J. Royal Stat. Soc. Ser. B **B-36**, 192–236 (1974).

[3] J. E. Besag, "On the statistical analysis of dirty pictures", J. Royal Stat. Soc. Ser. B **B-48**, 259–302 (1986).

[4] C. Bouman, and K. Sauer, "A generalizaed Gaussian image model for edge-preserving MAP estimation," IEEE T. Nucl. Sci. **99(4)**, 1144–1152 (1992).

[5] K. Champagnat, and J. Idier, "A conection between half-quadratic criteria and EM algorithms," IEEE Signal Proc. Let. **11(9)**, 709–712 (2004).

[6] S. Geman, and C. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," IEEE T. Pattern Anal. **6**, 721–741 (1984).

[7] K. Sauer, and C. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," IEEE T. Image Process. **2(3)**, 296–310 (1993).

[8] J. I. De la Rosa, and G. Fleury, "On the kernel selection for minimum-entropy estimation," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, 1205–12010 (IEEE, Anchorage, 2002).

[9] J. I. De la Rosa, G. Fleury, and M.-E. Davoust, "Minimum-entropy, pdf approximation and kernel selection for measurement estimation," IEEE T. Instrum. Meas. **52(4)**, 1009–1020 (2003).

[10] J. I. De la Rosa, "Convergence of minimum-entropy robust estimators: Applications in DSP and instrumentation," in *Proceedings of the XIV International Conference on Electronics, Communications, and Computers - CONIELECOMP'04*, 98–103 (IEEE, Veracruz, 2004).

[11] P. Milanfar "A Tour of Modern Image Filtering," IEEE Signal Proc. Mag. **30(1)**, 106–128, Jan. (2013).

[12] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," IEEE T. Image Process. **16(2)**, 349–366, Feb. (2007).

[13] Ö. Eğecioğlu, and A. Srinivasan, "Efficent non-parametric density estimation on the sphere with applications in fluid mechanics," SIAM J. Sci. Comput. **22(1)**, 152–176 (2000).

[14] G. P. Terrell, "The maximal smoothing principle in density estimation," J. Am. Stat. Assoc. **85**, 470–477 (1990).

[15] G. P. Terrell, and D. W. Scott, "Oversmoothed nonparametric density estimation," J. Am. Stat. Assoc. **80**, 209–214 (1985).

[16] L. Pronzato, and E. Thierry, "A minimum-entropy estimator for regression problems with unknown distribution of observation errors," in textslMaxEnt 2000, Eds. A. Mohammad-Djafari, 169–180 (American Institute of Physics, College Park, 2000).

[17] L. Pronzato, and E. Thierry, "Entropy minimization of parameter estimator with unknown distribution of observation errors," in *Proceedings of the IEEE International Conference in Acoustics, Speech and Signal Processing*, 3993–3996 (IEEE, Salt Lake City, 2001).

[18] E. Masry, "Probability density estimation from sampled data," IEEE T. Inform. Theory **IT-29(5)**, 697–709 (1983).

[19] L. Devroye, "A note on the usefulness of superkernels in density estimation," Ann. Stat. **20**, 2037–2056 (1992).

[20] L. Devroye, "The double kernel method in density estimation," Ann. I. H. Poincare **25**, 533–580 (1989).

[21] L. Devroye, and A. Krzyzák, "On the Hilbert kernel density estimate," Stat. Probabil. Lett. **44**, 299–308 (1999).

[22] L. Devroye, "Universal smoothing factor selection in density estimation: Theory and practice," Test **6**, 223–320 (1997).

[23] A. Berlinet, and L. Devroye, "A comparison of kernel density estimates," Publications de l'Intitut de Statistique de l'Université de Paris **38(3)**, 3–59 (1994).

[24] C. M. Loader, "Bandwidth selection: classical or plug-in?," Ann. Stat. **27(3)**, 415–438 (1999).

[25] E. Wolsztynski, E. Thierry, and L. Pronzato, "Minimum entropy estimation in semi parametric models," Signal Process. **85**, 937–949 (2005).