New approach of entropy estimation for robust image segmentation

O. Gutiérrez*, J. I. de la Rosa, *Member, IEEE*, J. Villa, E. González, and N. I. Escalante

***Abstract*—In this work we introduce a new approach for robust image segmentation. The idea is to combine two strategies within a Bayesian framework. The first one is to use a Markov Random Field (MRF), which allows to introduce prior information with the purpose of preserve the edges in the image. The second strategy comes from the fact that the probability density function (pdf) of the likelihood function is non Gaussian or unknown, so it should be approximated by an estimated version, and for this, it is used the classical non-parametric or kernel density estimation. This two strategies together lead us to the definition of a new maximum a posteriori (MAP) estimator based on the minimization of the entropy of the estimated pdf of the likelihood function and the MRF at the same time, named MAP entropy estimator (MAPEE). Some experiments were made for different kind of images degraded with impulsive noise and the segmentation results are very satisfactory and promising.**

***Index Terms*—Robust image segmentation, Markov random fields, Bayesian estimation, non-parametric density estimation, entropy minimization.**

I. Introduction

Segmentation is one of the most important tasks in image processing, it is considered the first step in object recognition, scene and image understanding. Some of its applications comprise industrial quality control, medicine, robot navigation, geophysical exploration, military applications, agriculture, among others. Nevertheless, digital images are usually affected by some degrading factors as blurring or noise coming from image acquisition systems, resulting in degraded or distorted images of the real world and producing, as a consequence, inadequate segmentation results.

A degradation process can be described as a degradation function H that, together with an additive noise term n, it operates on an input image x and produces a degraded image y, it means

$$y = Hx + n. \quad (1)$$

Fig. 1 illustrates this process.

An approach that have helped significantly to solve the problem of segmentation of degraded images is the use of Markov random fields (MRF) within a Bayesian framework [1-9]. This is because MRFs enables posing this problem, and many others in image processing, as statistical estimation problems [7] where the solution is going to be estimated from the degraded image. The basic premise is that neighborhood pixels are expected to have similar characteristics [8,10].

Usually, information data (input image) is not enough for an accurate estimation of the original image, so the regularization of the problem is necessary. This means that *a priori* information or assumptions about the structure of x need to be introduced in the estimation process [11]. The a priori knowledge is given in terms of a probability distribution. This distribution, together with a probabilistic description of the noise that corrupts the observations, allows the use of Bayes theory to compute the *posterior* distribution which represents the likelihood of a solution x given the observations y [10,12].

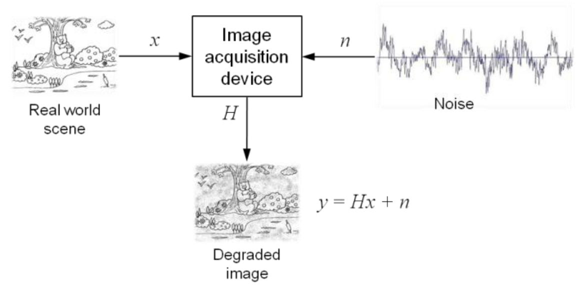

Fig. 1. Degradation process of an image.

The basic idea in Bayesian estimation is to construct a Maximum A Posteriori (MAP) by using MRFs. In the case of classical MAP filters, usually the additive Gaussian noise is considered, however in some applications this noise is non-Gaussian or unknown [13]. This becomes in a new source of information which imposes additional constraints in the image processing context (the spatial information) that represents the likelihood function or correlation between the intensity values of a well specified neighborhood of pixels.

The Bayes rule states that:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \quad (2)$$

where $p(x)$ corresponds to a probabilistic description of the real world or its properties, that we are trying to estimate, before collecting data; $p(y|x)$ is a description of the behavior of noise or stochastic characteristics that relate the original state x to the sampled input image or sensor values y; $p(x|y)$ is a probabilistic description of the current estimation of the

All the authors are with Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica, Laboratorio de Procesamiento Digital de Señales, Av. López Velarde 801, col. Centro, C. P. 98000, Zacatecas, Zacatecas, México.

*e-mail: osvaldo_gtz_mt@hotmail.com

 ISBN: 978-607-95476-6-0

original scene x, given the observed data y [14]; $p(y)$ is the density function of y and is constant if the observed image is provided [10,15].

The MAP estimator is defined by:

$$\begin{aligned}\hat{x}_{\mathrm{MAP}} &= \underset{x\in\mathbb{X}}{\operatorname{argmax}}\{p(x|y)\} \\ &= \underset{x\in\mathbb{X}}{\operatorname{argmax}}\{\log p(y|x) + \log g(x)\} \\ &= \underset{x\in\mathbb{X}}{\operatorname{argmin}}\{-\log p(y|x) - \log g(x)\},\end{aligned} \tag{3}$$

where $g(x)$ is a MRF function that models as a probability distribution the prior information of the phenomena to be estimated, $\mathbb{X}$ is the set of pixels capable to maximize $p(x|y)$ and $p(y|x)$ is the likelihood function from y given x [16].

In a previous work [10] it was introduced a new MRF model, named semi-Huber. There, it was demonstrated the advantages related to the model simplicity and the minor number of parameters to be tuned. In the present work we take as a starting point the use of the above mentioned MRF in the second term of (3). In that sense, the new approach of entropy estimation is going to be present in the variation of the first term of that expression. In [10], the first term in the MAP estimator was defined as a quadratic function of the differences between real and observed data, because of the additive noise regarded was Gaussian. For the experiments performed here, we considered a more degrading kind of noise, namely impulsive noise (*salt & pepper*) which does not follow a specific pattern. Actually, the main idea is to have a model that can be adapted to any kind of noise.

Thus, modeling in this new context lead us to assume a limited knowledge about the image noise pdf, so we propose to use the data itself to obtain a non-parametric Entropy Estimate (EE) of the log-likelihood pdf [17-19]. Then the log-likelihood will be optimized together with a log-MRF to obtain the MAP image segmentation. The rest of the paper is as follows: section II describes the definition of the approximation of the log-likelihood function by entropy estimation. Section III gives a brief background about the kernel structure for the density estimation function. The complete definition of the new MAP entropy estimator is presented in section IV. Some experiments and results are presented and discussed in section V, and in section VI some concluding comments are given.

II. LOG-LIKELIHOOD APPROXIMATED BY EE.

A. *The general problem of regression.*

A wide variety of applications in signal processing and instrumentation are based on statistical modeling analysis. The linear regression model is one of the most used

$$y_{i,j} = x_{i,j}^{\mathsf{T}}\theta_{i,j} + e_{i,j}, \qquad \text{con } e \sim p(e), \tag{4}$$

where y represents the response to x explicative variables for $i = 1, \dots, N$ and $j = 1, \dots, M$, and to a system parameterized by θ, a set of functional parameters associated to the data (y, x), which will be estimated by an identification procedure. The e variables are the errors, that model the system as a set of random processes which are independent and 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 $f(\cdot)$

$$y_{i,j} = f(x,\theta)_{i,j} + e_{i,j}, \qquad \text{con } e \sim p(e). \tag{5}$$

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 modeling when using nonlinear functions could be more realistic. The perturbations affecting the analyzed system are also modeled as stochastic processes [19].

There exist some classical techniques for the estimation of θ, for example Least Squares (LS), Maximum Likelihood (ML), among others. In this work it is proposed a MAP estimation based on the entropy minimization of an estimated version of the density of the errors ($\hat{p}_{n,h}(e)$).

B. *Likelihood pdf entropy estimators* (*EE*).

A classical procedure to estimate x when θ is known, is based in a cost function or criterion $\mathcal{J}(x)$ which varies in function $\psi(\cdot)$ of the residuals or noise $e(x)$, where

$$e_{i,j}(x) = y_{i,j} - f(x,\theta)_{i,j}. \tag{6}$$

Thus,

$$\mathcal{J}(x) = \sum_{i=1}^{N}\sum_{j=1}^{M}\psi\left(e_{i,j}(x)\right). \tag{7}$$

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

$$\hat{x}_{\mathrm{ML}} = \arg\min_{x\in\mathbb{X}}\left[-\sum_{\mathrm{i}=1}^{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{M}}\log p(e_{i,j}(x))\right]. \tag{8}$$

For an optimal performance, this estimator requires that all the information about the distribution $p(e)$ is accessible. When the knowledge about $p(e)$ is imprecise or wrong, the estimator $\hat{x}_{\mathrm{ML}}$ is possibly suboptimal [18, 19]. Moreover, under certain circumstances, in image processing, it results in an ill-posed problem or produces excessive noise and also causes smooth of edges. The regularization of the ML estimator provides a more effective approach, the Maximum A Posteriori (MAP) estimator, which reduces noise and smoothness at the same time.

Our proposition for a new MAP scheme is to use the semi-Huber MRF introduced in [10], together with a kernel estimator taken from [17-19] to obtain cost functions or criterions based on the entropy of the approximated likelihood function $\hat{p}_{n,h}(e)$ (first term of (3)). Thus, $-\log p(y|x)$ is built on the basis of the entropy of an estimated version $\hat{p}_{n,h}(e)$ of the distribution $p(e)$. A first proposition is due to Pronzato and Thierry [20-22], where the approximation is obtained using the classical kernel estimators which uses the empirical distribution of the random vector $e_{1,1}(x), \dots, e_{n,n}(x)$, as it is shown in the next expression:

$$\hat{p}_{n,h}(e) = \hat{p}_{n,h}\left(e|e_{1,1}(x), \dots, e_{n,n}(x)\right) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} K_h(e - e_{k,l}). \quad (9)$$

$K(\cdot)$ is a kernel weighted function which satisfies some imposed conditions treated in the work of Masry [23] and subsequently taken back by Devroye [24-27], Berlinet [28], and Loader [29] in some of their research work. The bandwidth $h = h_n$ is given in function of the sample size and can 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 $\hat{p}_{n,h}(e)$ and its convergence toward $p(e)$, depend on a convenient procedure of bandwidth selection [18]. A simple and faster procedure is the technique proposed and developed by Terrell [30, 31].

Assuming that $\hat{p}_{n,h}(e)$ converges and is consistent, such that $\hat{p}_{n,h}(e) \to p(e)$, then the entropy criterion over $\hat{p}_{n,h}(e)$ 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 an extended criterion based on the residuals or noise extended vector, given by:

$$e_E = \{e_{1,1}(x), \dots, e_{n,n}(x), -e_{1,1}(x), \dots, -e_{n,n}(x)\}, \quad (10)$$

and on a suitable choice of h:

$$\mathcal{J}_e(x) = \mathrm{H}_A\left(\hat{p}_{n,h}(e_E)\right) \approx -\log p(y|x), \quad (11)$$

where

$$\mathrm{H}_A(f) = -\int_{-A_n}^{A_n} f(x)\log f(x)\, dx. \quad (12)$$

then a fist version of the MAP Entropy Estimator (MAPEE) assuming unknown noise pdf, can be constructed from the fact that $-\log p(y|x)$ can be approximated by the entropy of an estimated version $\hat{p}_{n,h}(e)$ of the distribution $p(e)$, thus:

$$\hat{x}_{\text{MAPEE}} = \arg\min_{x\in\mathbb{X}}\left\{\mathrm{H}_A\left(\hat{p}_{n,h}(e_E)\right) - \log g(x)\right\}. \quad (13)$$

III. The Kernel Structure.

A function of the form $K(z)$ is assumed as a fixed kernel $K_h(z) = 1/(h^d)K(z/h)$, where $h > 0$, is a parameter 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. Taking as a reference the works [17-19], the Hilbert kernels [26] was selected here, this is because of the results presented in the referred papers and mainly because of their structure is such that they avoid the bandwidth selection and their performance depend on other parameters, which selection is very easy.

A. The Hilbert kernel.

The $K_h(z) = 1/(h^d)K(z/h)$, with $h > 0$, is considered equivalent to $K(u) = 1/\|u\|^d$, where the smoothing factor h is canceled, obtaining:

$$\hat{p}_n(z) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{\|z - z_{k,l}\|^d}. \quad (14)$$

The consistency of this class of estimators is proved in [26]. 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:

$$\hat{p}_n(z) = \sqrt{\frac{4}{V_d^2 \pi n(n-1)\log n}\sum_{1\le i<j\le n}\frac{1}{\mathrm{Den}_{i,j}}}, \quad (15)$$

where $\mathrm{Den}_{i,j} = \|z - z_i\|^{2d} + \|z - z_j\|^{2d}$, V_d is the volume of the unit ball in $\mathbb{R}^b$ and $\|\cdot\|$ denotes the L_2 metric on $\mathbb{R}^d$. Finally, it is assumed that $\hat{p}_n(z) \to p(z)$ at least in probability for almost all z. For a suitable choice of d and k, this estimator could be "blind asymptotically efficient".

IV. The MAP Entropy Estimator (MAPEE).

In this section it is obtained the complete cost function structure for the named $\hat{x}_{\text{MAPEE}}$ estimator derived from (3). The first term has been already described in sections II and III, corresponding to the new approach proposed here. The second term of (3), $-\log g(x)$, is based on a potential function, named semi-Huber MRF, introduced in [10, 32].

The Hammersley-Clifford theorem establishes that MRF are equivalent to Gibbs random fields [1,15, 33, 34]. A Gibbs distribution has the form:

$$g(x) = \frac{1}{Z}\exp\left(-\frac{1}{T}U(x)\right), \quad (16)$$

where Z is named the *partition function* and in practice is a normalization constant value. T is the *temperature* parameter, that controls the sharpness of the distribution [1] and in practice is assumed to take the value of 1 [33]. $U(x)$ is the energy function, given by:

$$U(x) = \sum_{c\in\mathbb{C}} V_c(x), \quad (17)$$

and determined as a sum of *clique potentials* $V_c(x)$ over all possible cliques $\mathbb{C}$ of the neighborhood [6, 15, 33]. These clique potentials are given in terms of the difference of the intensity values of neighboring pixels and have the general form $\rho(\lambda(x_i - x_j))$, which act on pairs of sites and λ is a constant that scales the difference between pixel values [10]. For the experiments presented here, the eight closest neighbors were considered.

Within this framework, the Huber-like norm or semi-Huber potential function, for the two dimensional case, is given by:

$$\log g(x) = -\lambda\left(\sum_{\{s,r\}\in\mathbb{C}} b_{sr}\rho_1(x)\right) + \mathrm{c}, \qquad (18)$$

where s is the site or pixel of interest, r corresponds to the local neighbors, b_{sr} is a constant that depends on the distance between pixels s and r, c is a constant term and

$$\rho_1(x) = \frac{\Delta_0^2}{2}\left(\sqrt{1+\frac{4(x_s - x_r)^2}{\Delta_0^2}} - 1\right), \Delta_0 > 0. \qquad (19)$$

A graphical representation of the behavior of the semi-Huber potential function is displayed in Fig. 2, where it can be seen that near zero the function is quadratic and with increasing of the differences, the function becomes practically linear. The linear region allows to preserve sharp edges, while convexity makes MAP estimate efficient to compute [10].

Now, substituting the particular expression (18) for $\log g(x)$ into (13), it can be obtained the complete form of the MAP entropy estimator for image segmentation degraded with non-Gaussian noise:

$$\hat{x}_{\mathrm{MAPEE}} = \arg\min_{\mathrm{x}\in\mathbb{X}}\left\{\mathrm{H}_A\left(\hat{p}_{n,h}(e_E)\right) + \lambda\sum_{\{s,r\}\in\mathbb{C}} b_{sr}\rho_1(x)\right\}. \quad (20)$$

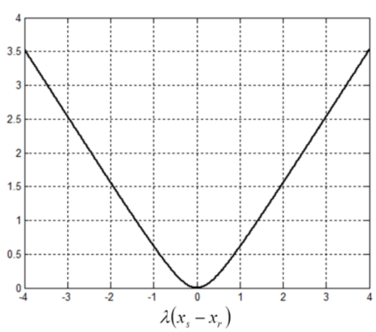

Fig. 2. Graphical representation of the semi-Huber cost function.

V. Experiments and Results.

In order to evaluate the performance of the new proposed approach of MAP entropy estimation applied to image segmentation, we present a set of experiments with some images, Fig. 3 shows the set of test images used. The first one is a synthetic image, with which we can make error measures and misclassified pixels count; The second one is an attempt to apply this new approach to medical imaging, and the third one is for the case of geographical imaging.

All the experiments was performed on a Mac Pro computer with a 2 × 2.8 GHz quad-core Intel Xeon processor and 2 GB at 800 MHz DDR2 RAM. The minimization process was made using the Levenberg-Marquardt algorithm provided in the optimization toolbox of MATLAB R2009a, where we needed to provide the initial value X_0 to start the search of the solution. The three images were degraded with impulsive noise: imnoise (X, 'salt & pepper', 0.15), and the aim is to obtain the segmentation of the image in spite of the noise present.

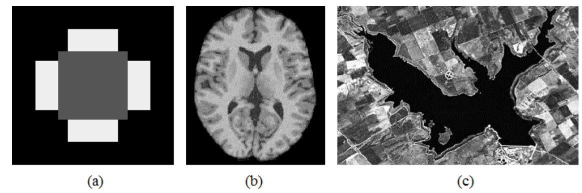

Fig. 3. Set of test images: (a) synthetic image, (b) MR image, (c) geographical image of a dam.

A first experiment was carried out with the synthetic image trying to separate as accurately as possible the three regions. Fig. 4 shows the segmentation results from the noisy image for parameter values $\lambda = 1$ and $\Delta_0 = 2$, with an initial value of $X_0 = 90$. Fig. 4(a) shows the image degraded with impulsive noise, as described in the previous paragraph, Fig. 4(b) shows the segmented image, Fig. 4(c) shows the difference: original image minus segmented image (X - Xs) and Fig. 4 (d) shows the difference: segmented image minus original image (Xs - X). This figures permit to count the number of misclassified pixels, obviously what we expect here is black images.

Table I includes information about times of computation, number of misclassified pixels (n) and error measures, namely the relative squared error (RSE) and the relative absolute error (RAE). These numerical results are compared with those obtained applying the segmentation process without considering the new approach of MAP entropy estimation for the log-likelihood function; this means, assuming Gaussian noise for the first term of the MAP estimator (3). Specifically, from [10]:

$$\hat{x}_{\mathrm{MAP}} = \arg\min_{x\in\mathbb{X}}\left\{\sum_{s\in\mathbb{S}}|y_s - x_s|^2 + \lambda\sum_{\{s,r\}\in\mathbb{C}} b_{sr}\rho_1(x)\right\}. \quad (21)$$

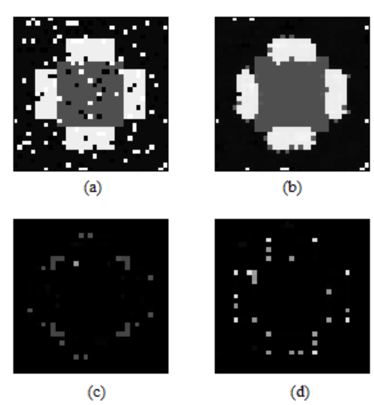

Fig. 4. Segmentation of the synthetic image applying MAP entropy estimation: (a) noisy image, (b) segmented image, (c)-(d) differences from original (X) and segmented images (Xs).

 ISBN: 978-607-95476-6-0

Fig. 5 shows the segmentation results of the synthetic image under the Gaussian assumption. It can be seen that visual results are not so good having impulsive noise present in the image. Numerical results also confirm the improvement produced with the new approach.

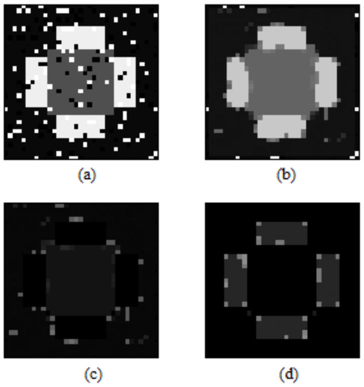

Fig. 5. Segmentation of the synthetic image with $X_0 = 90$, (a) noisy image, (b) segmented image, (c)-(d) differences from original (X) and segmented images (Xs).

TABLE I
NUMERICAL RESULTS FOR THE SEGMENTATION OF SYNTHETIC IMAGE

	X_0	Δ_0	Time (s)	n	RSE	RAE
MAPEE	90	2	21.4438	58	0.1179	0.1374
MAP	110	60	12.4741	84	0.1216	0.3388

By using the new approach of entropy estimation, time of computation increases, but in return the errors are reduced significantly. The values of X_0 and Δ_0 are different because they have to be adjusted in order to obtain the best result. With the same parameter values for MAP as in MAPEE the results obtained were disastrous.

For a second experiment we used a generic image of the brain, trying to separate three tissues: gray matter, white matter and cerebrospinal fluid (CSF). Fig. 6 shows the segmentation results obtained from the noisy image with both, MAP entropy estimation (20) and Gaussian assumption (21). Table II contains information about parameter values and times of computation.

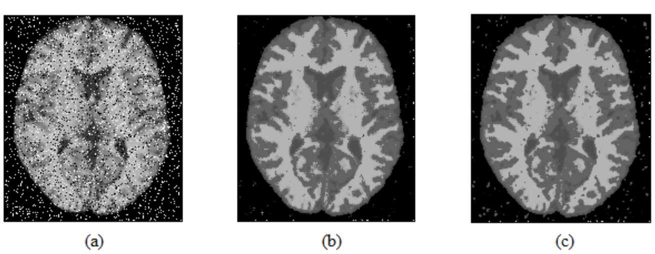

Fig. 6. Segmentation of an image of the brain: (a) image degraded by impulsive noise, (b) segmented image using MAPEE, (c) segmented image using MAP.

TABLE II
NUMERICAL RESULTS FOR THE SEGMENTATION OF BRAIN IMAGE

	X_0	Δ_0	Time (s)
MAPEE	90	110	457.9916
MAP	90	110	333.7739

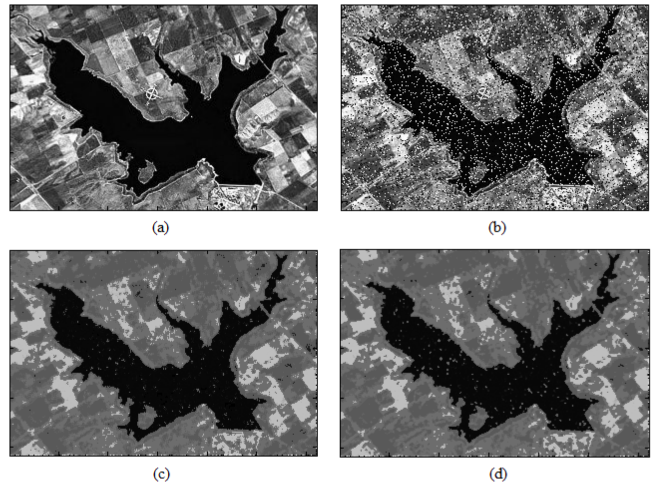

Fig. 7. Segmentation of a geographical image of a dam: (a) original image, (b) image degraded by impulsive noise, (c) segmented image using MAPEE, (d) segmented image using MAP.

It can be seen from Fig. 6 that the MAPEE produces a better result than MAP. In the black background the MAP result presents more and bigger gray spots, and in the part of white matter of the brain it can be seen also the same effect.

A third and last experiment was made with a geographical image of a dam named Paso de las Piedras, located in Argentina, taken from Google Earth. For this image the interest is on the segmentation of the water region, excluding all other elements. Figure 7 shows segmented images applying both approaches and Table III presents information about the realization of these two processes.

TABLE III
NUMERICAL RESULTS FOR THE SEGMENTATION OF GEOGRAPHICAL IMAGE

	X_0	Δ_0	Time (s)
MAPEE	100	130	957.1709
MAP	100	130	725.4942

As in the previous experiment, it is visually perceptible that MAPEE improves the segmentation result, for the case of impulsive noise, with respect to the previous approach of Gaussian noise assumption (MAP).

VI. CONCLUSIONS.

A new approach for image segmentation was proposed where one can, not only consider Gaussian noise, but also other kind of degradation factors. In this case it was used impulsive noise that is one of the most degrading and difficult to deal with. It was proved that this new approach produces very good results in the sense of robustness, adapting to the nature of the degradation present in the images. We are working in the improvement of this new proposal by adding additional filtering to enhance the final result.

VII. References

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration ofimages," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721-741, 1984.

[2] J. E. Besag, "On the statistical analysisof dirty pictures," *J. Roy.Stat. Soc. B*, vol. 48, pp. 259-302, 1986.

[3] K. Sauer and C. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 4, pp. 1144-1152, 1992.

[4] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no.3, pp. 296-310, 1993.

[5] K. Held, E. R. Kops, B. J. Krause, W. M. Wells III, R. Kikinis, and H. W. Müller-Gärtner, "Markov random field segmentation of brain MR images," *IEEE Trans. Med. Imaging*, vol. 16, no. 6, pp. 878-886, 1997.

[6] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imaging*, vol. 20, no. 1, pp. 45-57, 2001.

[7] S. Krishnamachari and R. Chellappa, "Multiresolution Gauss-Markov random field models for texture segmentation," *IEEE Trans. Image Process*, vol. 6, no. 2, pp. 251-267, 1997.

[8] D. A. Clausi and B. Yue, "Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 215-228, 2004.

[9] Y. Li and P. Gong, "An efficient texture image segmentation algorithm based on the GMRF model for classification of remotely sensed imagery," *Int. J. Remote Sens.*, vol. 26, no. 22, pp. 5149-5159, 2005.

[10] O. Gutiérrez, I. de la Rosa, J. Villa, E. González, and N. Escalante, "Semi-Huber potential function for image segmentation," *Optics Express*, vol. 20, no. 6, pp. 6542-6554, 2012.

[11] M. Rivera, O. Ocegueda, and J. L. Marroquin, "Entropy-controlled quadratic Markov measure field models for efficient image segmentation," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 3047-3057, 2007.

[12] J. Marroquin, S. Mitter, and t. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, no. 397, pp. 76-89, 1987.

[13] 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*, vol. 21, no. 12, pp. 2283–2291, 2004.

[14] R. Szeliski, "Bayesian modeling of uncertainty in low-level vision," *Int. J. Comput. Vision*, vol. 5, no. 3, pp. 271–301, 1990.

[15] X. Lei, Y. Li, N. Zhao, and Y. Zhang, "Fast segmentation approach for SAR image based on simple Markov random field," *J. Syst. Eng. Electron.*, vol. 21, no. 1, pp. 31–36, 2010.

[16] J. I. de la Rosa, J. J. Villa, and Ma. A. Araiza, "Markovian random fields and comparison between different convex criteria optimization in image restoration," in *Proc. XVII Int. Conf. on Electronics, Communications and Computers*, pp. 9 (CONIELECOMP, 2007).

[17] J. I. De la Rosa, and G. Fleury, "On the kernel selection for minimum-entropy estimation," *Proc. of the IEEE Instrumentation and Measurement Technology Conference* (IEEE, Anchorage, AK (USA)), vol. 2, pp. 1205–1210, 2002.

[18] J. I. De la Rosa, G. Fleury, and M.-E. Davoust, "Minimum-entropy, pdf approximation and kernel selection for measurement estimation," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1009–1020, 2003.

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

[20] L. Pronzato, and E. Thierry, "A minimum-entropy estimator for regression problems with unknown distribution of observation errors," *MaxEnt 2000* (Edited by A. Mohammad-Djafari, American Institute of Physics, 2000), pp. 169-180.

[21] L. Pronzato, and E. Thierry, "A minimum-entropy estimator for regression problems with unknown distribution of observation errors," Tech. Rep. 00-08, Laboratoire I3S, CNRS-Université de Nice-Shopia Antipolis, France, 2000.

[22] L. Pronzato, and E. Thierry, "Entropy minimization of parameter estimator with unknown distribution of observation errors," *Proc. of the IEEE International Conference in Acoustics, Speech and Signal Processing* (IEEE 2001), vol. 6, pp. 3993–3996.

[23] E. Masry, "Probability density estimation from sampled data," *IEEE Trans. on Inform. Theory*, vol. IT-29, no. 5, pp. 697-709, 1983.

[24] L. Devroye, "A note on the usefulness of superkernels in density estimation," *The Annals of Statistics*, vol. 20, pp. 2037-2056, 1992.

[25] L. Devroye, "The double kernel method in density estimation," Annales de l'Institut Henri Poincaré, vol. 25, pp. 533–580, 1989.

[26] L. Devroye, and A. Krzyżak, "On the Hilbert kernel density estimate," *Statistics and Probability Letters*, vol. 44, pp. 299-308, 1999.

[27] L. Devroye, "Universal smoothing factor selection in density estimation: Theory and practice," Test, vol. 6, pp. 223-320, 1997.

[28] A. Berlinet, and L. Devroye, "A comparison of kernel density estimates," Publications de l'Intitut de Statistique de l'Université de Paris, vol. 38, no. 3, pp. 3-59, 1994.

[29] C. M. Loader, "Bandwidth selection: classical or plug-in?," *The Annals of Statistics*, vol. 27, no. 3, pp. 415-438, 1999.

[30] G. P. Terrell, "The maximal smoothing principle in density estimation," *Journal of the American Statistical Association*, vol. 85, pp. 470-477, 1990.

[31] G. P. Terrell, and D. W. Scott, "Oversmoothed nonparametric density estimation," *Journal of the American Statistical Association*, vol. 80, pp. 209-214, 1985.

[32] J. I. de la Rosa and G. Fleury, "Bootstrap methods for a measurement estimation problem," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 3, pp. 820-827, 2006.

[33] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer-Verlag, 2009.

[34] S. Geman and C. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.

ISBN: 978-607-95476-6-0