A Non-Extensive Entropy Algorithm for Multi-Region Segmentation: Generalization and Comparison

Paulo S. Rodrigues* and Gilson A. Giraldi
National Laboratory for Scientific Computing,
Petropolis, Rio de Janeiro, Brazil
{pssr,gilson}@lncc.br

Abstract

Since eighties, the concept of Shannon entropy has been applied in the field of image processing and analysis for image binarization. The general idea underlining this type of image segmentation is the maximization of the information (entropy) associated to background and foreground pixels. Under the context of mechanical statistic, physical systems that may be well described by Shannon entropy are called extensive systems. It is well known, however, that several physical systems have non-extensive behavior, letting the Shannon entropy inadequate to describe such systems. A generalization of the Shannon entropy is the so called q-entropy, proposed for non-extensive systems. This paper presents an algorithm for image segmentation base on this new kind of entropy. We show that, applying recursively the proposed algorithm, it is possible to highlight the main image regions without oversegmentation. In order to validate our proposal, we compare it with four well known approaches for image clustering, namely: bootstrap, fuzzy-c-means, k-means and self-organized maps. Besides, we have used synthetic images to quantitatively compare these methods. Our results show that the proposed algorithm overcomes the traditional Shannon entropy for image binarization and gets multi-regions with similar performance in terms of precision and time of execution.
I. INTRODUCTION

Image segmentation plays an important role on the basis of computational vision tasks, such as image analysis, recognition and tracking, to name a few. Image segmentation is a basic but important and complex problem which has been intriguing the researches for decades. The issue behind image segmentation is to decompose the image into regions of coherent properties in an attempt to identify objects and their parts. Gray level image segmentation techniques can be classified into the following categories: thresholding, methods based on feature space, edge detection based methods, region based methods, fuzzy logic techniques and neural networks. Besides, most of these methods can be extended to color images by representing color information in an appropriate color spaces. In addition, it is possible to combine more than one approach to achieve better performance.

Among these methods, the thresholding ones are fundamental for this work. Thresholding is a large class of segmentation techniques that are based on the assumption that the objects can be distinguished and extracted from the background by their gray levels. The output of traditional thresholding operations is a binary image whose intensity pattern distinguish the foreground (gray level 0, for example) from the background (gray level 255). A survey can be found in references [12] and [27]. In general, threshold selection can be categorized into two classes, local methods and global methods. In the global thresholding methods a single threshold is selected through the image histogram and applied for the segmentation task, while in the local methods the image is partitioned into a number of subimages and a specific threshold selected for each one. The global thresholding techniques are computationally cheaper than local methods as well as easy for implementation. However, its efficiency is highly dependent of the histogram distribution, specifically peaks separated by a valley, which limits their applications. In this sense, global methods tend to be more efficient, although
more computational expensive also, as a general threshold estimation is not necessarily an easy task. However, a threshold can be based on different criterions, such as Otsus method [23], minimum error thresholding [15], and entropic methods [14], [25], to name a few.

Applying the concept of entropy in order to segment a digital image is a common practice since PUN’s work [25] shows how to find out a threshold that maximizes the information measure, the Shannon entropy, of the resulting binary image. Other works following the same philosophy were proposed, e.g., Kapur et al [14] maximized an upper bound of the total a posteriori entropy in order to obtain the threshold level. Abutaleb [1] extended the method using two-dimensional entropies. Li and Lee [18] and Pal [24] used the directed divergence of Kullback for the selection of the threshold, and Sahoo et al. [27] used the Rényi entropy model for image thresholding.

The concept of Shannon entropy was proposed for Theory of Information based on Boltzaman-Gibbs entropy for the context of classic thermodynamic. However, for several decades it is well known that this concept fails to explain some phenomena having complex behaviors such as long range interactions and long-time memories [35], [37]. Such systems are called “non-extensive systems”, and those following the BGS formalism are called “extensive systems”.

In 1988 Tsallis proposed a new formalism for the generalization of BGS entropy, which is called $q$-entropy or Tsallis entropy. This new entropy has reached relative successful in explanation complex phenomena for several applications. A complete list of nonextensive entropy is vast and can be fully find in [37].

In 2004, Albuquerque et al [2] applied the concept of non-extensive entropy for mamographic gray scale images. They assume a probability distribution of gray scale luminance, one for background and other for foreground classes of pixels. Then, they had taken the threshold that maximizes the separation between these two classes.

However, in some classes of images, where the region around the boundaries has important information for the application, a simple thresholding or clustering method does not guarantee a
good separation between the classes. Examples of such images are mamographic ones from ultrasound devices having breast tumor, where, on the boundary of such lesions (e.g. malignant ones), dead cells are among healthful ones, generating a third region in the image. As these regions have generally a small narrowband, simple threshold methods may fail in capturing adequately their information content. On the other hand, most sophisticated methods, which combines several basic approaches may have a high computational time.

In this paper, we proposed a extension of the work of Albuquerque et al. [2]. Applying in a recursive way the $q$-entropy concept at each partition (background and foreground). Our proposal gets more information (in terms of number of retrieved regions) at a low computational time. Also, we have total control over the number of recursions and, as a consequence, over the number of retrieved regions.

The contributions of our paper are: 1) proposing a recursive algorithm from that proposed in [2]; 2) compare the non-extensive approach against other prominent methodologies: k-mean [5], [19], [29], Bootstrap clustering technique [6], Fuzzy c-mean (FCM) [4] and self organizing maps (SOM) (proposed by Kohonen [16]).

The paper is organized as follows. In Section II we indicate some related works; in Section III we introduce the $q$-entropy under the context of non-extensive systems; the proposed method is seen in Section IV; in Section V we compare our methodology against other well known methods from a point of view of natural and synthetic images. Finally, in Section VI we offer some discussions and conclusions about the results.

II. RELATED WORKS

The mean-shift algorithm is a general nonparametric technique proposed by Comaniciu and Meer [8] for clustering of complex multimodal feature space. It randomly tessellates the space with search windows, and moves until convergence is achieved at the nearest mode of the underlying probability distribution of density gradients. Several application of this algorithm for color clustering can also be found in [20], [26], [38].
Bootstrap clustering technique is similar to other resampling schemes, such as cross-validation and jackknifing. A bootstrap is obtained by sampling with replacement from an empirical distribution function from training set. Chen et al [6] applied a bootstrap implementation to computer-aided diagnosis in breast ultrasound images. Dutendas et al [11] presented an Bayesian approach combined with a bootstrap algorithm in order to segment images from the retina. Another two applications applying bootstrap techniques on image segmentation can be found in [21], [40].

The Watershed transform is a reliable tool for initial image segmentation. A significant advantage of Watershed segmentation and a reason behind its extended utilization is that boundaries on the image plane are always guaranteed to be connected and closed, and each gradient minimum corresponds to one region [22]. A nice explanation about Watershed algorithm can be found also in [13].

As an unsupervised clustering algorithm we can cite the Fuzzy c-mean (FCM) which has been applied successfully to a number of problems involving feature analysis, clustering and classifier design. It has been applied to wide variety of applications such as agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis and target recognition [4]. Unlabeled data are classified by definition of a norm, cluster prototype and by minimizing an objective function. Although the description of the original algorithm dates back from 1973 [3], [10], further variations have been described with modified definitions for the norm and prototypes for the cluster center [17] and [9].

One of the most used methods in the last years to image clustering is the so called self organizing maps (SOM), proposed by Kohonen [16]. The SOM neural network consists of two layers, and for every neuron in the input layer, there is a link to every neuron in the output layer. During the training process of SOM network, for each input vector we get one best matching neuron in the output layer. Here a competitive learning algorithm is used to adjust weight vectors in the neighborhood of best matching neuron. The adjustment decreases as the time and the range of neighborhood increasing.
III. TSALLIS ENTROPY

The entropy is an idea born in the heart of the classic thermodynamic, not as some fundamentally intuitive, but as some fundamentally quantitative, defined though an equation, which has been known as Boltzman-Gibbs (BG) entropy. Later, Shannon redefined the concept of BG entropy (now called BGS entropy) as an uncertainty measure associated to the content of system information. This traditional form of entropy is well known by the following equation:

\[ S = - \sum_i p_i \ln(p_i) \]  

(1)

Generically speaking, systems having statistics such as BGS type are called extensive systems and have an additive property, defined as follows. Let \( P \) and \( Q \) be two random variables, with probability densities functions \( P = (p_1, \ldots, p_n) \) and \( Q = (q_1, \ldots, q_m) \), respectively, and \( S \) be the entropy associated with \( P \) or \( Q \). If \( P \) and \( Q \) are independent, under the context of the Probability Theory, the entropy of the composed distribution\(^1\) verifies the so called additivity rule:

\[ S(P \ast Q) = S(P) + S(Q) \]  

(2)

This traditional form of entropy is well known and for years has achieved relative success to explain several phenomenon if the effective microscopic interactions are short-ranged (i.e., close spatial connections) and the effective spatial microscopic memory is short-ranged (i.e., close time connections) and the boundary conditions are non(multi)fractal. Roughly speaking, the standard formalism are applicable whenever (and probably only whenever) the relevant space-time is non(multi)fractal. If this is not the case, some kind of extension appears to became necessary. We can make a complete analogy with Newtonian mechanics, when it becomes only an approximation (an increasingly bad one) when the involved velocities approach that of light or the masses are as small as say the electron mass;

\(^1\)we define the composed distribution, also called direct product of \( P = (p_1, \ldots, p_n) \) and \( Q = (q_1, \ldots, q_m) \), as \( P \ast Q = \{ p_i q_j \}_{i,j} \), with \( 1 \leq i \leq n \) and \( 1 \leq j \leq m \)
the standard statistical mechanics do not apply when the above requirements (short-range microscopic interactions, short-ranged microscopic memory and (multi)fractal boundary conditions) are not the case.

Then, recent developments, based on the concept of non-extensive entropy, also called Tsallis entropy, have generated a new interest in the study of Shannon entropy for Information Theory [28], [36]. Tsallis entropy (or $q$-entropy) is a new proposal for the generalization of Boltzmann/Gibbs traditional entropy applied to non-extensive physical systems.

The non-extensive characteristics of Tsallis entropy has been applied through the inclusion of a parameter $q$, which generates several mathematical properties and the general equation is the following:

$$S_q(p_1, \ldots, p_k) = \frac{1 - \sum_{i=1}^{k} p_i^q}{q - 1}$$  (3)

where $k$ is the total number of possibilities of the whole system and the real number $q$ is the entropic index that characterizes the degree of non-extensiveness. In the limit $q \to 1$, Equation (3) meets the traditional BGS entropy defined by Equation (1). These characteristics give to $q$-entropy flexibility in explanation of several physical systems. On the other hand, this new kind of entropy does not fail to explain the traditional physical systems since it is a generalization.

Furthermore, a generalization of some theory may suppose the violation of one of its postulates. In the case of the generalized entropy proposed by Tsallis, the additive property described by Equation (2) is violated in the form of Equation (4), which apply if the system has a non-extensive characteristic. In this case, the Tsallis statistics is useful and the $q$-additivity describes better the composed system which is defined as:

$$S_q(P \ast Q) = S_q(P) + S_q(Q) + (1 - q)S_q(P)S_q(Q)$$  (4)

In this equation, the additive property does not hold unless in the limit $q = 1$. 
Considering $S_q \geq 0$ in the pseudo-additive formalism of the Equation (4), the following classification for entropic systems is defined: sub-extensive system (when $q > 1$ and $S_q(P \ast Q) > S_q(P) + S_q(Q)$); extensive system (when $q = 1$ and $S_q(P \ast Q) = S_q(P) + S_q(Q)$) and superextensive (when $q < 1$ and $S_q(P \ast Q) < S_q(P) + S_q(Q)$).

Taking into account the similarities between the formalisms of Shannon and Boltzmann/Gibbs entropy, it is interesting to investigate the possibility of the generalization of Shannon entropy to the case of the information theory, as has been recently shown by Yamano [39]. This generalization may be extended to image segmentation tasks, by applying Tsallis entropy, which has nonadditive information contents.

Albuquerque et al. [2] proposes an algorithm using the concept of $q$-entropy to segment US images. Since this concept may be naturally applied over any statistical distribution, in this paper we propose a natural extension of the algorithm proposed by Albuquerque et al. [2] which yields to a recursive procedure by applying to each distribution $P$ and $Q$ the concept of $q$-entropy. As said, we have named our extended algorithm as NESRA.

The motivations to use the $q$-entropy are: 1) managing only a simple parameter $q$ yields to a more controllable system; 2) as suggested in [2] for mammographic images, it is interesting to study the behavior of non-extensive segmentation under several others class of images; 3) it is simple and makes the implementation easy having a low computational time.

In the following section, we fully describe the NESRA proposal.

IV. THE NON-EXTENSIVE SEGMENTATION APPROACHES

Applying the concept of entropy in order to segment a digital image is a common practice since PUN [25] showed how to maximize the difference between the foreground and background using the Shannon entropy over a gray level distribution. Then, other works following the same line were proposed, e.g, Kapur et al [14] maximized un upper bound of the total a posteriori entropy in order to obtain the threshold level. Abutaleb [1] extended the method using two-dimensional entropies. Li
and Lee [18] and Pal [24] used the directed divergence of Kullback for the selection of the threshold, and Sahoo et al. [27] used the Reiny entropy model for image thresholding.

In 2004, Albuquerque et al [2] presented the concept of non-extensive entropy applied to mammographic gray scale images. They assume a probability distribution, one for background and other for foreground and take the threshold that maximizes the non-additivity characteristic given by Equation (4). However, as several methods designed to produce binary images, this approach does not work for multiregion segmentation as well. Then, we proposed an extension of the method presented in [2] applying recursively the maximization of Equation (4) over the background and the foreground in order to achieve multi-regions of homogenous gray level distribution.

In this section, we formalize the NESRA algorithm. Firstly, we will review the non-extensive procedure for image segmentation proposed in [2].

A. Non Extensive Segmentation Algorithm for Image Binarization

Suppose an image with \( k \) gray-levels, let the probability distribution of these levels be \( P = \{ p_i = p_1; p_2; \ldots; p_k \} \). Now, we want to consider two probability distribution from \( P \), one for the foreground \( (P_A) \) and another for the background \( (P_B) \). We can make a partition at luminance level \( t \) between the pixels from \( P \) into \( A \) and \( B \). In order to maintain the constraints \( 0 \leq P_A \leq 1 \) and \( 0 \leq P_B \leq 1 \) we must to re-normalize both distribution as: \( P_A : \frac{p_1}{p_A}, \frac{p_2}{p_A}, \ldots, \frac{p_t}{p_A} \) and \( P_B : \frac{p_{t+1}}{p_B}, \frac{p_{t+2}}{p_B}, \ldots, \frac{p_k}{p_B} \), where \( p_A = \sum_{i=1}^{t} p_i \) and \( p_B = \sum_{i=t+1}^{k} p_i \).

Now, we calculate the a priori Tsallis entropy for each distribution such as

\[
S_A = \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_A})^q}{q - 1} \quad (5)
\]

\[
S_B = \frac{1 - \sum_{i=t+1}^{k} (\frac{p_i}{p_B})^q}{q - 1} \quad (6)
\]

We can observe that the Tsallis entropy represented by Equations (3), (5) and (6), depend on directly the parameter \( t \) for the foreground and background, and it is formulated as the sum of each entropy,
allowing the pseudo-additive property for statistically independent systems, defined in Equation (7).

\[ S_{A+B}(t) = \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{A1}})^q}{q-1} + \frac{1 - \sum_{i=t+1}^{k} (\frac{p_i}{p_{A2}})^q}{q-1} + (1-q) \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{B1}})^q}{q-1} + (1-q) \frac{1 - \sum_{i=t+1}^{k} (\frac{p_i}{p_{B2}})^q}{q-1} \]  

(7)

To accomplish the segmentation task, in [2] the information measure between the two classes (foreground and background) is maximized. In this case, the luminance level \( t \) is considered to be the optimum threshold value (\( t_{opt} \)), which can be achieved with a cheap computational effort of

\[ t_{opt} = \text{argmax}[S_A(t) + S_B(t) + (1-q)S_A(t)S_B(t)] \]  

(8)

Note that the value \( t \) which maximizes Equation (8) depends on mainly the parameter \( q \). This is an advantage due to its simplicity. In the next section, we present the recursive formulation for this algorithm.

B. The Non-Extensive Segmentation Recursive Algorithm (NESRA)

Following the definitions and formulations from last section, we can take each distribution \( P_A \) and \( P_B \) and subdivide them into two news distribution, \( P_{A1}, P_{A2}, P_{B1} \) and \( P_{B2} \), as following: \( P_{A1} : \)

\[ \frac{p_1}{p_{A1}}, \frac{p_2}{p_{A1}}, \ldots, \frac{p_t}{p_{A1}}; \quad P_{A2} : \frac{p_{t+1}}{p_{A2}}, \frac{p_{t+2}}{p_{A2}}, \ldots, \frac{p_k}{p_{A2}}; \quad P_{B1} : \frac{p_{t+1}}{p_{B1}}, \frac{p_{t+2}}{p_{B1}}, \ldots, \frac{p_k}{p_{B1}} \quad \text{and} \quad P_{B2} : \frac{p_{t+1}}{p_{B2}}, \frac{p_{t+2}}{p_{B2}}, \ldots, \frac{p_k}{p_{B2}}. \]

having the constraints \( p_{A1} = \sum_{i=1}^{t} p_i, p_{A2} = \sum_{i=t+1}^{k} p_i, p_{B1} = \sum_{i=t+1}^{q} p_i, p_{B2} = \sum_{i=q+1}^{k} p_i. \) For each one of these four distributions we can compute its respective non-extensive entropy as follows:

\[ S_{A1} = \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{A1}})^q}{q-1} \]  

(9)

\[ S_{A2} = \frac{1 - \sum_{i=t+1}^{k} (\frac{p_i}{p_{A2}})^q}{q-1} \]  

(10)

\[ S_{B1} = \frac{1 - \sum_{i=q+1}^{k} (\frac{p_i}{p_{B1}})^q}{q-1} \]  

(11)

\[ S_{B2} = \frac{1 - \sum_{i=v+1}^{k} (\frac{p_i}{p_{B2}})^q}{q-1} \]  

(12)
Using Equations (4), (9) and (10) to compute $S(A) = S(A1 + A2)$ and, similarly (11) and (12) to compute $S(B) = S(B1 + B2)$, we have:

$$
S_q(P + Q) = \left( 1 - \sum_{i=1}^{t} \left( \frac{p_{A_{i1}}}{p_{A_{i2}}} \right)^q \right) + \left( 1 - \sum_{i=t+1}^{v} \left( \frac{p_{B_{i1}}}{q_{B_{i2}}} \right)^q \right) + \left( 1 - \sum_{i=v+1}^{k} \left( \frac{q_{B_{i1}}}{q_{B_{i2}}} \right)^q \right) + \left( 1 - \sum_{i=k+1}^{w} \left( \frac{q_{B_{i1}}}{q_{B_{i2}}} \right)^q \right).
$$

In this case, we have a set of optimal values $t_{opt} = \{t, q, v, k\}$ which maximizes the Equation (13) which is equivalent of calculation of the following expression:

$$
t_{opt} = \text{argmax}[S_{A1} + S_{A2} + (1 - q) \cdot S_{A1} \cdot S_{A2} \cdot (S_{B1} + S_{B2}) + (1 - q) \cdot S_{B1} \cdot S_{B2} \cdot (1 - q) \cdot (S_{A1} + S_{A2} + (1 - q) \cdot S_{A1} \cdot S_{A2} \cdot (S_{B1} + S_{B2} + (1 - q) \cdot S_{B1} \cdot S_{B2})]
$$

The Equation (13) is simple, although with several terms. By developing a third recursion would yielding the number to sixteen, which is not necessary for our discussion. Now, two observations may be done. First, the experimental results show that it is not necessary more than two or three recursions in order to obtain results which are equals or better than traditional methods. Second, the growing of the number of recursion does not enlarge the algorithm complexity or computation, since this growing accompanies a dropping of states to be computed at each recursion. Since at each recursion the data are partitioned into two new groups, the computational complexity of the NESRA algorithm is $O(cN)$, where $c$ is the number of recursions and $N = \text{lines} \times \text{columns}$ is the size of image. Also, we can preview the maximum number of achieved regions as $2^c + 1$. When $c = 0$ the NESRA outputs two regions (background and foreground) only, and behaves as the proposal of
Albuquerque et al. [2], letting NESRA be a generalization of [2].

The recursive algorithm for the previous formulation is simple and we give it in the following.

\textbf{Algorithmus 1} Non-Extensive Segmentation Recursive Algorithm, NESRA procedure

\begin{itemize}
  \item \textbf{Input:} \( H = \) image histogram, \( i = \) first histogram bin, \( k = \) last histogram bin
  \item Call NESRA\((H, i, k)\) procedure
  \item \textbf{if} Histogram \( H \) is homogeneous \textbf{then}
    \item Goto FIM
  \item \textbf{end if}
  \item \textbf{for all} \( t = i \) until \( k \) \textbf{do}
    \item compute normalization for background
    \item compute normalization for foreground
    \item compute q-entropy for background according to Equation (5)
    \item compute q-entropy for foreground according to Equation (6)
    \item compute composed q-entropy according to Equations (7) and (8)
  \item \textbf{end for}
  \item \( topt = \) argmax of the composed q-entropy
  \item Call NESRA\((H, i, topt)\) procedure
  \item Call NESRA\((H, topt + 1, k)\) procedure
  \item FIM: there is nothing to do, return to calling procedure
\end{itemize}

It can be argued that any parametric or non parametric method with the objective of finding an ideal threshold, producing a binary image, can be recursively used in the foreground and in the background in order to achieve multi-regions. However, our approach, based on non-extensive entropy, out performs several well known approaches under the same conditions as we show in section V.

\textbf{V. Experimental Results}

To show the robustness of our proposed multi-region segmentation algorithm as well as the range of its application, we have experimented it over four classes of gray scale images: a natural image (sunset), mammographic ultrasound image, object with homogeneous background and lenna. The sunset image (Fig. 1.a) is suitable to test the method under low contrast intensity patterns. On the other hand, medical image applications have been a source of scientific investigation with several open challenges too, which demand for well segmentation algorithms. Therefore, we have chosen
a well known class of medical images to test the robustness, namely: mammographic images (Fig. 1.b). The ultrasound images have as main characteristics high speckle noises, low resolution, many spurious regions and bed region of interest definition, demanding for specific algorithms. Then, it also is suitable to test any algorithm for image segmentation.

Fig. 1.c we have a photo of the Columbia database, which was taken under controlled light and having homogeneous black background, which facilitates the segmentation task. However, the target object shape and gray scale patterns are complex. Also, we tested our algorithm over the well known lenna image since it is a broadly used image in the literature (Fig. 1.d).

A. Non-Extensive × Extensive Results

The range of \( q \) values is a key issue in our proposal. In this paper, we have tested the \( q - entropy \) segmentation (recursive and non-recursive) for a range \( 0 < q \leq 10.0 \). All experiments were carried-out under three different values of \( q \) in the Equation (13). The main proposal of this section is to show the results comparing our proposed algorithm with the non recursive version.

![Fig. 1. The four classes of images used in our experiments.](image)

The first row of Fig. 2 (images a-c) presents the results of the segmentation of Fig. 1.a with Tsallis entropy (non-recursive) for values of \( q = 1.0, 6.0 \) and 10.0, respectively. Clearly, this image has five main regions: the solar circumference, the solar crown, the sky, mountains and buildings. However, only two regions were retrieved by the algorithm: with threshold \( t = 101, q = 1.0 \) (Fig. 2.a); \( t = 113, q = 6.0 \) (Fig. 2.b), and \( t = 55, q = 10.0 \) (Fig. 2.c).
The second row of Fig. 2 (images d-f) presents the results of the segmentation of Fig. 1.a with our NESRA algorithm (the recursive version) also for values of $q = 1.0, 6.0$ and $10.0$, respectively. In this case, three more regions were outlined and the results, under the same $q$ values, are clearly better than those on the first row. The Fig. 2.d shows four regions which were retrieved with threshold $t = \{101, 135, 55\}$, all for $q = 1.0$. Similarly, Fig. 2.e shows four regions retrieved with threshold $t = \{113, 161, 63\}$, all with $q = 6.0$; and Fig. 2.f shows four regions retrieved with threshold $t = \{55, 79, 40\}$, for $q = 10.0$.

Below each image are the sequence of retrieved thresholding. For example, Fig. 2.d has three values (101, 135, 55) in this sequence. The first value (101) is the threshold achieved by NESRA in the first image partition (regarding the pseudo additivity of Equation (8)). Then, NESRA is applied again to first distribution (those pixels with luminance level lower than 101) achieving $t_{opt} = 55$ and to the second distribution (those pixels with luminance level higher than 101) achieving $t_{opt} = 135$, generating four regions in the image limited by the thresholds 55, 101 and 135. Then, we apply different luminance levels for each range: $0 \leq t_1 \leq 55, 55 < t_2 \leq 101, 101 < t_3 \leq 135$ and $135 < t_4 \leq 256$. The same is analog to Fig. 2.e and Fig. 2.f.

We can find more regions if we apply the NESRA one second recursion, finding up to $2^{2+1} = 8$ regions. The third row of Fig. 2 (images g-i) shows the results when we use the same $q$ values as before. In this case, the sequence of generated thresholds is as follows. The first application of NESRA achieves the value of partition $t_{opt} = 101$. Then, the process is repeated (first recursion) to the range $t < 101$, achieving $t_{opt} = 55$ and $t \geq 101$, achieving $t_{opt} = 135$. Then, a second recursion is applied to the ranges $t \leq 55$, achieving $t_{opt} = 39, 55 \leq t < 101$, achieving $t_{opt} = 81, 101 \leq t < 135$, achieving $t_{opt} = 117$, and $t \geq 135$, achieving $t_{opt} = 164$. Therefore, the image of Fig. 2.g has the following eight ranges of luminance distributions: $0 \leq t_1 < 39 \leq t_2 < 55 \leq t_3 < 81 \leq t_4 < 101 \leq t_5 < 117 \leq t_6 < 135 \leq t_7 < 164 \leq t_8 < 256$. We have applied arbitrary luminance values for each one of these ranges. The same reasoning can be carried out to the images of Fig. 2.h and Fig. 2.i,
but for values of $q = \{6.0, 10.0\}$, respectively. A third recursion, which could yield up to $2^{3+1} = 16$ regions is not necessary for this image, as no more salient region could be retrieved.

Fig. 2. result segmentation of Fig. 1.a: with Tsallis entropy non-recursive algorithm (a-c) and with our extended proposed algorithm (NESRA) (d-f) for three different $q$ values and two recursions. $t$ stands for the retrieved threshold.

A first important observation to this experiment is that the Fig. 2.h, generated with two recursions and $q = 6.0$ (subextensive system), seems to be the most interesting result as it delineates better the target regions. A second important observation is that the regions given by a recursion $i$ always preserves the separation achieved by recursion $i - 1$, which may be a nice characteristic of NESRA algorithm as the user can get more control over local details. This control is achieved handling the number of recursions. This is specifically useful in applications where we can have coarse transitions in the boundary of main regions, such as ultrasound images which have large central regions (representing lesions) with mixing of dead and healthful cell on the lesion’s boundaries. In this case, we may have a good separation between the lesion’s nucleus, the background and the narrow region boundary.
(transition). This will be well observed in the next experiment.

The second class of images we have tested with the NESRA algorithm is from mammographic of ultrasound devices (Fig. 1.b is an example). This image shows a benign region at the center of the image with a low SNR background and a coarse transition on the lesion boundary. Adequate extraction of the breast lesion from ultrasound images is an important and open problem as breast cancer ranks second in the list of women’s cancer and Computer Aided Diagnosis (CAD) has become even more powerful, reducing the physician’s work load (see [34] and references there in).

The first row of Fig. 3 shows three segmentation results (without recursion) of Fig. 1.b. The given

---

Fig. 3. Result segmentation of Fig. 1.b: with Tsallis entropy non-recursive algorithm (a-c) and with our extended proposed algorithm (NESRA) (d-f) for three different q values and two recursions. t stands for the achieved thresholds.

---

thresholds are 91, 52 and 10 for \( q = 1.0, 6.0 \) and 10.0, respectively. As the last experiment, we can preserve the partition given by the first NESRA application and can find more regions applying NESRA recursively. The results are shown in the second row of Fig. 3. The values of \( t \) below the images of Fig. 3.d-f are the given thresholds for this first recursion. The best result seems to be that of Fig. 3.e as we clearly have a good segmentation of lesion nucleus separated from the background by a narrow boundary with distinct threshold as a third region. This is particularly useful if the CAD
systems want to analyze separately the boundary region in order to decide if it belongs to the lesion region or to the background. Also, it may help to extract the lesion’s nucleus from the image. We can also note that the thickness of boundary region may be controlled by setting the $q$ value. As $q \to 1.0$ (extensive systems) the region of interest increases and tends to merge with the background. On the contrary, as $q > 1.0$ increases (subextensive systems), the region of interest is easier to define but diminishing missing (important) information.

The third data set we have used in our experiments is composed of synthetic images from the Columbia database [7]. Although these images have homogeneous background, which helps the foreground extraction, several applications need the recognition of parts of objects. Then, the binarization of this kind of images may not work. This fact may be observed in the segmentation of Fig. 1.c outlined in Fig. 4.a-c. In this figure, the unique acceptable segmentation seems to be that of Fig. 4.a, when $q = 0$. When $q = 0.5$ or $1.0$ the results are clearly inferior. When applying our proposed method, however, (Figs.4d-f.) we can reach better results for all values of $q$. The best result seems to be that of Fig. 4.d ($q=0$).

Finally, we have experimented our proposed algorithm over the lenna image, with the same previous $q = \{0, 0.5, 1.0\}$ values. Then, in Fig. 5.a-c, it is shown the application of the extensive algorithm; and, by applying the proposed non-extensive algorithm over the same original image, we have better results in terms of segmentation as may be seen in the corresponding images of Fig. 5.d-f. In this case, an unique recursion was applied, which was sufficient to get four main regions.

\textbf{B. Comparisons with Other Methods}

In this section we compare our proposed approach with four well known algorithms which were briefly described in Section II, namely: bootstrap, fuzzy c-means, k-means and SOM. In Fig. 6 there are four columns and five rows. Each column corresponds to an image from Fig. 1, and each row corresponds to a different method (named from up to bottom): our proposed NESRA algorithm, bootstrap, fuzzy c-means, k-means and SOM.
Fig. 4. Segmentation of a Columbia database image (Fig. 1.c). The first row shows three binarization with \( q = \{0, 0.5, 1.0\} \). The second row shows the corresponding three segmentations when we applied our proposed method for the same \( q \) values as before.

Fig. 5. The segmentation of lenna image for \( q = \{0, 0.5, 1.0\} \). The first row presents the simple non-extensive algorithm and second row presents the corresponding recursive version. \( t \) stands for the retrieved threshold value.
Fig. 6. Results of application of the five approaches in four types of images: row 1: the proposed NESRA method; row 2: bootstrap; row 3: fuzzy c-means; row 4: k-means; and row 5: som. Columns 1, 2, 3 and 4 correspond to images from Fig. 1.
The NESRA was applied with two recursions, which means that there are at most four main regions. Bootstrap, fuzzy c-means and k-means was set in order to find up to five clusters, and the SOM network was set to spread their input data among a $3 \times 2$ rectangular grid of neurons. However, similar results were obtained with a $4 \times 4$ and $6 \times 6$ rectangular grids. Regarding these set up, we can make the following discussion about the results presented in Fig. 6.

In the first column of Fig. 6 (which correspond to Fig. 1.a) we can see that the NESRA and KM give similar results in comparison to Bootstrap, FCM and SOM, since NESRA and KM give four main salient regions, which correspond to the light around the sun, the sky, the horizon and buildings. Additionally, NESRA seems to delineate better these four regions.

The second column of Fig. 6 shows the application of all algorithms over a mammographic images. All algorithms traced well the tumor (central circular region). KM obtained more homogeneous regions for the background. All results are similar for background for the remaind methods.

In the third column (a cup with some drawings with homogeneous background corresponding to Fig. 1.c), we may discard only SOM results (row 5) and analyze only the other four, which seems to generate similar results. Among them, FCM generates small noises, NESRA, Bootstrap and KM seems to produce similar (good) segmentation.

In the last column shows the results for lenna image. All approaches except NESRA and bootstrap generated an oversegmentation, but NESRA has clearly low noise and homogeneous main regions.
C. Synthetic Images

Finally, in order to quantify a comparison of the methods, we have ran each algorithm with a synthetic image consisting of two concentric circles with radius 100 and 50. The intensities of the background and the outer and inner circle are 150, 100 and 50, respectively. These intensities do not have large distances in their histograms. This image is seen in Fig. 7 with three examples of increasing application of gaussian noise with \( \mu = 0 \); \( \sigma^2 = 0.01, 0.05 \) and 0.1.

The experiment consists of applying in this synthetic image a gaussian noise with zero mean and variance \( \sigma^2 > 0 \). After, a 2D \( n \times n \) adaptive noise removal filter was applied, which causes noise dropping but blurring the frontiers between the inner and outer circles and the outer circle and the background as well. Although the value of \( n = 9 \) has been chosen empirically, it is the same for all algorithm in order to put them under the same conditions. Also, in the case of NESRA method, the value of \( q = 0.001 \) was applied for the synthetic image. This value was chosen empirically and the good results indicate the system non-extensiveness.

After, a procedure based on the Door-In-Door-Out algorithm was implemented so that to extract the circles’s boundaries and further their coordinates. Our goal is to measure the robustness of the methods in achieving the original boundaries under increasing values of noise variance. Then, we use the original boundary as a ground truth and compute how far from it is the estimated one. The coordinates of the inner and outer original curves (ground truth) were obtained in a straightforward manner by analytical calculation, given their radius and center.

The comparison between an original and an estimated curve was carried out through the PolyLine Distance Measure (PDM) method, proposed by Suri et al [30]–[33]. The PDM is defined as the closest distance from the each estimated boundary point to the ideal/ground-truth region of interest. The closest distance of each estimated boundary point can be the perpendicular distance (shortest Euclidian distance) to one skin-line, or can be one of the end boundary points joining the points of the closest interval. It is a measure of the average polyline distance of all boundary points of the
Letting $B_1$ be the first boundary and letting $B_2$ be the second boundary, we can derive the PDM measure, $d_{poly}^{Error}$, as follows:

$$d_b(A, B_2) = \min_{S \in \text{sides}B_2} d(A, S)$$  \hspace{1cm} (15)

$$d_{vb}(B_1, B_2) = \sum_{A \in \text{vertices}B_1} d_b(A, B_2)$$  \hspace{1cm} (16)

$$d_{Error}^{poly} = \frac{d_{vb}(B_1, B_2) + d_{vb}(B_2, B_1)}{\#\text{vertices} \in B_1 + \#\text{vertices} \in B_2}$$  \hspace{1cm} (17)

Also, the PDM Equation (17) does not need the two boundaries $B_1$ and $B_2$ having the same length.

Note the two main characteristics of the PDM equation: first, as $B_1 \rightarrow B_2$, $d_{Error}^{poly} \rightarrow 0$; second, it is a distance in pixel units. If $N \times M$ is the image dimension, it is straightforward to show the maximum value for $d_{Error}^{poly}$ as:

$$d_{max} = |R - r|$$  \hspace{1cm} (18)

where $R$ is the larger image’s diagonal and $r$ is the average radius of the estimated curve. For a fixed $N$ and $M$, and dividing (17) by (18), we can define the robustness of an algorithm in achieving the ideal boundary as:

$$R = 1 - \frac{d_{Error}^{poly}}{d_{max}}$$  \hspace{1cm} (19)

which is a percentage of the maximum error. Note that $R \rightarrow 0$ when the error is maximum and $R \rightarrow 1$ when the error is minimum. Therefore, in our experiments, we compute $R$ as a function of increasing values of $\sigma^2$ in order to see the general performance of the methods while achieving the ideal boundary. We also comparing them under the same conditions and scenarios.

In order to visually compare the methodology’s robustness, after applying the 2D adaptive filtering, we superpose the estimated curves on the original images. To give a visual idea of the result, we choose to illustrate this paper with the higher used variance ($\sigma^2 = 0.1$) in the segmented synthetic images.
Fig. 8. The result segmentation of the five algorithms used in this paper. We applied in the original image a gaussian noise with zero mean and $\sigma^2 = 0.1$ which is the highest noise used, and after, a $9 \times 9$ 2D adaptive filter was used for smoothing the noise. In the specific case of NESRA algorithm we use the parameter $q = 0.001$ since it generated the best visual result with more homogeneous and noiseless regions.

Fig. 9. The estimated (black) and original (white) curves superimposed over the original image corresponding to the segmentations of Figure 8.

In the Fig. 8, we show the segmentation applied in the synthetic image for the five methodologies. In this figure, all the segmented image’s regions, only for viewing, were assigned to a different gray scale intensity: black for background, gray for outer circle and white for inner circle.

In the Fig. 9 we show the original (in white) and estimated (in black) curves superimposed over the original image with gaussian noise and the application of the 2D adaptive filter.

Fig. 10. Comparative performance of the five used methods as a function of increasing gaussian noise. Left image is for the outer circle and right one is for inner circle. The $x$-line is the $\sigma^2$ and $y$-line is $R$ according to Equation (19). std is the standard deviation.
Fig. 11. Comparative performance for the five used methods according to the estimated area inside inner, outer and background regions. The performance percentage is an average of the estimated area of the three regions. The $x$-line is the $\sigma^2$ and $y$-line is the average of estimated area (for the three regions) divided by real area. std is the standard deviation.

A complete numerical result can be see in the Fig. 10-left and 10-right, where we plotted $R$ as a function of the complete range for $\sigma^2 = \{0.01, 0.02, \ldots, 0.1\}$. Also, Fig. 10 shows the comparative performance for the used algorithms. Fig. 10-left shows the results for the outer curve and Fig. 10-right is the same for the inner curve.

A final graphic is showed in Fig. 11, where we plotted an average percentage of the original area for the inner, outer and background areas as a function of increasing gaussian noise.

The running time of the NESRA algorithm is proportional of the binary procedures, $O(\log N)$, where $N$ is the image dimension. This is due to as we split the image at each iteration, the number of pixels to be evaluated drops proportionally according to $N$.

VI. Discussion and Conclusion

We have presented a new method for image segmentation based on non-extensive entropy, which is a recursive version, called NESRA, of that proposed in [2].

The NESRA algorithm is a segmentation procedure for basic segmentation, which should be used in an initial segmentation. As such, it was compared with other algorithms under the same conditions, namely: Bootstrap, Fuzzy C-Means, K-Means and SOM network. All of them were ran for several classes of real and synthetic images.
The results show that the NESRA procedure is able to retrieve all the main clusters for all tested images. The quality of clusters retrieved is also appreciable. The proposed technique gives results at least similar to the best performance for three images (sunset, object and ultrasound) and clearly overcomes the others approaches for lenna image. This is not only an indication that the NESRA is a good strategy for image segmentation, but also, by considering an image as a non-extensive system, q-Entropy is a promisor line of investigation for image clustering. The proposed technique gives best results with the tested images and it has less time complexity, since it reduces the number of pixels examined at each recursion.

The NESRA algorithm retrieved more homogeneous and noiseless regions avoiding oversegmentation. It is due to the NESRA always splits the data set into two new regions only, yielding to less sensitiveness to noise and spurious regions.

It may be argued that any algorithm for binary segmentation (e.g., such as the well known iterative threshold) may be extended to a recursive version and applied like NESRA. However, it always splits the luminance space into two blocs, generating a background and a foreground regions maximizing the information (associated entropy). In comparison with the traditional method proposed by PUN [25], which uses the Shannon entropy, the NESRA overcomes for all images presented in Fig. 2-5. This is an indication that the NESRA’s advantage is not only its recursion, but also a better separation between the foreground and background which may be given due to the q parameter flexibility.

Regarding the choice of the q parameter, an automatic value can not be computed yet. Then, as future work, we proposed the investigation of the optimal q through an iterative method, since the entropic equations are functions of q. Up to this moment, depending on the class of image, the value of q may be taken from a small discrete range between 0 ≤ q ≤ 1. This is the case for all experiments of this paper. As we can see for all results, there is little changing in the results when q varies. Generically speaking, the q value may be used as a tuning setting when we consider the system as non-extensive.
According to Fig. 8, under the maximum gaussian noise ($\sigma^2 = 0.1$), only NESRA, Fuzzy C-Means and K-Means gave good segmentation. In the cases of Fuzzy and K-Means, they gave a quite identical results, but both with a hight background noise compared with NESRA. The NESRA and Bootstrap were less sensitive to background noise; however, the NESRA segmented bed the outer region, while the Bootstrap did not retrieve any inner circle. On the other hand, the SOM generated oversegmentation results. Similar results also can be seen in the Fig. 9, where the NESRA, Fuzzy C-Means and K-Means are compatible in achieving the boundaries. The contrary occurs with the SOM and Bootstrap, which, in turn, did not retrieve the central circle.

When these results are quantitatively analyzed, as observed in the Fig. 10 and 11, Fuzzy C-Means, K-Means and NESRA are stable with slight advantage to Fuzzy C-Means and K-Means, in achieving the inner circle. On the other hand, also the Bootstrap is stable in achieving the outer circle (Fig. 10-left) but has low performance in achieving the inner circle (Fig. 10-right).

When we compare the estimated areas with the original ones (Fig. 11) the performance of all algorithms are quite similar. The Fuzzy C-Means and K-Means have similar good performance and NESRA are compatible with them, while SOM and Bootstrap have low performance.

Finally, we can conclude that the NESRA algorithm has compatible performance with the best tested algorithms for the synthetic image and the best general performance of all methods in real images when we analyze qualitatively. This is due to the NESRA algorithm gives better homogeneity of the target regions. It is less sensitive to noise than all the others tested methods with the additional advantage of low time complexity. Then, NESRA algorithm is well indicated to segment general images even at real time.

In the future, we will test NESRA algorithm inside a complete system for image segmentation. An alternative is to use NESRA with active contour models in order to smooth the given boundaries. Also, a possible future direction is to study the best class of images to apply the NESRA methodology.
ACKNOWLEDGMENT

The authors are grateful to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazilian agencies for Scientific Financing, for the support of this work.

REFERENCES

[1] A. S. Abutaleb. A new method for gray-level picture thresholding using the entropy of the histogram. *Comput. Graphics Image Process.*, 47:22–32, 1989.

[2] M. P. Albuquerque, M. P. Albuquerque, I.A. Esquef, and A.R.G. Mello. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, 25:1059–1065, 2004.

[3] J. C. Bezdek. *Fuzzy Mathematics in Pattern Classification*. PhD thesis, Cornell University, Ithaca, NY, USA, 1973.

[4] J. C. Bezdek. *Partition Structures: A tutorial*. In: *The Analysis of Fuzzy Information*. 1987.

[5] C. W. Chen, J. Luo, and K. J. Parker. Image segmentation via adaptive k-mean clustering and knowledge-based morphological operations with biomedical applications. *IEEE Transactions on Image Processing*, 7(Issue 12):1673–1683, December 1998.

[6] D.R. Chen, W. J. Kuo, R.F. Chang, W. K. Moon, and C. C. Lee. Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound. *Ultrasound Med. Biol.*, 28(7):897–902, July 2002.

[7] Columbia object image library (coil-100). Department of Computer Science. Columbia University, http://www.cs.columbia.edu/CAVE/research/softlib/coil-100.html.

[8] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(Issue 5):603–619, May 2002.

[9] R. N. Dave and K. Bhaswan. Adaptive fuzzy c-shells clustering and detection of ellipses. *IEEE Transaction on Neural Networks*, 3(5):643–642, 1992.

[10] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters. *Journal of Cybernetics*, 3:32–57, 1974.

[11] D. Dutendas, L. Moreau, F. Ghorbel, and P. M. Allioux. Unsupervised bayesian segmentation with bootstrap sampling application to eye fundus image coding. In *Nuclear Science Symposium and Medical Imaging Conference, IEEE Conference Record*, volume 4, pages 1794 – 1796, October-November 1994.

[12] S. K. Fu and J. K. Mu. A survey on image segmentation. *Pattern Recognition*, 41:3–16, 1981.

[13] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison Wesley, 1992.
[14] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Comput. Graphics Image Process.*, 29:273–285, 1985.

[15] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19:41–47, 1986.

[16] T. Kohonen. *Self-Organizing and Associated Memory, 3th Edition*. Springer-Verlag, 1989.

[17] R. Krishnapuram, O. Nasraoui, and J. Keller. The fuzzy c spherical shells algorithm: A new approach. *IEEE Trans. Neural Networks*, 3(5):663–671, 1992.

[18] C. H. Li and C. K. Lee. Minimum cross entropy thresholding. *Pattern Recognition*, 26:617–625, 1993.

[19] M. Luo, M. Yu-Fei, and Z. Hong-Jiang. A spatial constrained k-means approach to image segmentation. In *Proceedings of the 2003 Joint Conference on Information, Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia*, volume 2, pages 738–742, December 2003.

[20] Q. Luo and T. M. Khoshgoftaar. Efficient image segmentation by mean shift clustering and mdl-guided region merging. In *16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI-04*, pages 337–343, November 2004.

[21] I. Kochand G. Marshall. Bootstrap coverage plots for image segmentation. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 2, pages 447–451, August 1996.

[22] F. Meyer. Color image segmentation. In *4th IEEE Conference on Image Processing Applications*, volume 354, pages 303–306, 1992.

[23] N. Otsu. A threshold selection method from gray level histogram. *IEEE Transaction on Systens, Man and Cybernetics, SMC-8*:62–66, 1978.

[24] N. R. Pal. On minimum cross entropy thresholding. *Pattern Recognition*, 26:575–580, 1996.

[25] T. Pun. Entropic thresholding: A new approach. *Comput. Graphics Image Process.*, 16:210–239, 1981.

[26] Feng Wei; Z. Rong-Chun. Non-rigid objects detection and segmentation in video sequence using 3d mean shift analysis. In *International Conference on Machine Learning and Cybernetics*, volume 5, pages 3134–3139, November 2003.

[27] P. K. Sahoo, S. Soltani, and A. K. C. Wong. A survey of thresholding techniques. *Comput. Vis. Graphics Image Process.*, 41:233–260, 1988.

[28] C. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1948.

[29] M. Singh, P. Patel, D. Khosla, and T. Kim. Segmentation of functional mri by k-means clustering. *IEEE Transactions on Nuclear Science*, 43(Issue 3, Part 2):2030–2036, June 1996.

[30] J. Suri. Error and shape measurement tools for cardiac projection images: a closer look. *International Conference in Applications of Patterns Recognition*, pages 125–134, 1998.

[31] J. Suri, R. M. Haralick, and F. H. Sheehan. Greedy algorithm for error reduction in automatically produced boundaries from low contrast ventriculogram. *Proceedings of International Conference in Pattern Recognition (ICPR)*, 4(1):361–
[32] J. Suri, R. M. Haralick, and F. H. Sheehan. Greedy algorithm for error reduction in automatically produced boundaries from low contrast ventriculogram. *International Journal of Pattern Analysis and Applications*, 3(1):39–60, 2000.

[33] J. Suri, S. K. Setarhedran, and S. Singh. *Advanced algorithm approaches to medical image segmentation: state of art applications in cardiology, neurology, mammography and pathology*. Springer-Verlag, 2002.

[34] J. S. Suri and R. M. Rangayyan. *Recent Advances in Breast Imaging, Mammography and Computer Aided Diagnosis of Breast Cancer*. SPIE Press, 2006.

[35] C. Tsallis. Nonextensive statistics: Theoretical, experimental and computational evidences and connections. *Brazilian Journal of Physics*, 29(1), March 1999.

[36] C. Tsallis. *Nonextensive entropy Statistical Mechanics and its Applications*. Series of Lecture Notes in Physics. Springer, Berlin, 2001. The bibliography of the subject is regularly updated at http://tsallis.cat.cbpf.br/biblio.htm.

[37] C. Tsallis. Nonextensive statistical mechanics and its applications. *Series Lecture Notes in Physics*, Springer, Berlin, 2001. The bibliography of the subjects is regularly updated at http://tsallis.cat.cbpf.br/biblio.htm.

[38] G. Xian-Jiu and W. Wei. Multiscale segmentation algorithm based on subdivision and mean shift. In *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, volume 7, pages 4427–4431, August 2005.

[39] T. Yamano. Information theory based in nonadditive information content. *Entropy 3*, pages 280–292, 2001.

[40] T. Zheng, W. Xian-Bin, and L. Wei. Multiscale segmentation of sar imagery with bootstrap sampling. In *Proceedings of 7th International Conference on Signal Processing (ICSP-04)*, volume 1, pages 713–716, August 2004.