**Journal of Signal Processing Systems manuscript No.**
(will be inserted by the editor)

# Image Segmentation using Excess Entropy

**A. Bardera, I. Boada, M. Feixas and M. Sbert**

Graphics and Imaging Laboratory, University of Girona

Received: date / Revised version: date

**Abstract** We present a novel information-theoretic approach for thresholding-based segmentation that uses the excess entropy to measure the structural information of a 2D or 3D image and to locate the optimal thresholds. This approach is based on the conjecture that the optimal thresholding corresponds to the segmentation with maximum structure, i.e., maximum excess entropy. The contributions of this paper are severalfold. First, we introduce the excess entropy as a measure of the spatial structure of an image. Second, we present an adaptive thresholding method based on the maximization of excess entropy. Third, we propose the use of uniformly distributed random lines to overcome the main drawbacks of the excess entropy computation. To show the good performance of the proposed segmentation approach different experiments on synthetic and real brain models are carried out.

## 1 Introduction

The main objective of image segmentation is to divide an image into regions that can be considered homogeneous with respect to a given criterion such as color or texture. Segmentation is an essential part of any image analysis system and especially in medical environments, where segmented images provide valuable information for diagnosis. Image thresholding, which segments an image by dividing its intensity histogram into a set of thresholds, is one of the most used segmentation techniques, because of its simplicity and efficiency [13, 19]. In the case of bimodal images, two main approaches of this technique can be considered: (i) the global thresholding that compares all the image pixels to a same value, which may be constant or be chosen from the image histogram [13], and (ii) the local thresholding that computes an independent threshold for each pixel over a local window whose center is the pixel being binarized. In this paper we will focus on a global thresholding not restricted to bimodal images.

The excess entropy, first introduced by Crutchfield and Packard [9], is a measure of global correlation or structure for spatial systems in any dimension. It is commonly used and well understood in one dimension, but some difficulties are found in its extension to higher dimensions. In this paper, we propose a thresholding technique that uses the excess entropy to find the optimal thresholds of a 2D or 3D image automatically. Our approach is based on the assumption that a medical image is structured in regions, such as tissues or organs, and the conjecture that the optimal thresholding corresponds to the segmentation with maximum structure, i.e., maximum excess entropy. Hence, we present a thresholding method

that uses the maximization of excess entropy to determine the optimal thresholds. The computation of the excess entropy for a 3D dataset requires, for each voxel, an adequate selection of a sequence of neighbor voxels. Depending on how these neighbor voxels are selected, different problems such as high-dimensionality, sparsity, and non-invariance to rotation appear. To overcome these problems, excess entropy is computed using uniformly distributed random lines [3]. Experimental results analyze the behavior of our approach for different image modalities.

This paper is organized as follows. In Section 2 some background and related work are reviewed. In Section 3 the excess entropy is introduced as a measure of the spatial structure of an image. In Section 4, a new method is presented for thresholding segmentation using the excess entropy. In Section 5, different experiments on synthetic and real medical images are carried out. Finally, conclusions are summarized in Section 6.

## 2 Background and Related Work

We review some basic information-theoretic measures [7, 11] and the image segmentation method based on thresholding.

### *2.1 Information-Theoretic Measures*

Let $\mathcal{X}$ be a finite set, let $X$ be a random variable taking values $x$ in $\mathcal{X}$ with distribution $p(x) = Pr[X = x]$. Likewise, let $Y$ be a random variable taking values $y$

in $\mathcal{Y}$. The *Shannon entropy* $H(X)$ of a random variable $X$ is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

and measures the average uncertainty of random variable $X$. If the logarithms are taken in base 2, entropy is expressed in bits.

We review now the definitions of entropy rate and excess entropy. The notation used here is inspired by the work of Feldman and Crutchfield [11].

Given a chain $\ldots X_{-2}X_{-1}X_0X_1X_2\ldots$ of random variables $X_i$ taking values in $\mathcal{X}$, a block of $L$ consecutive random variables is denoted by $X^L = X_1 \ldots X_L$. The probability that the particular $L$-block $x^L$ occurs is denoted by the joint probability $p(x^L) = p(x_1, x_2, \ldots, x_L)$. The Shannon entropy of length-$L$ sequences or *L-block entropy* is defined by

$$H(X^L) = -\sum_{x^L \in \mathcal{X}^L} p(x^L) \log p(x^L), \tag{1}$$

where the sum runs over all possible $L$-blocks. The *entropy rate* is defined by

$$h = \lim_{L \to \infty} \frac{H(X^L)}{L} = \lim_{L \to \infty} h(L), \tag{2}$$

where

$$h(L) = H(X^L) - H(X^{L-1}) \tag{3}$$

is the entropy of a symbol conditioned on a block of $L-1$ adjacent symbols. The two terms of Equation (2) are shown to be equivalent in [8, 11]. The entropy rate of a sequence quantifies the average amount of information per symbol $x$ and the optimal achievement for any possible compression algorithm [7]. The entropy rate

is always equal or lower than the Shannon entropy and is only equal when there is no correlation between consecutive symbols.

A complementary measure to the entropy rate is the *excess entropy*, which is a measure of the *structure* of a system. The *excess entropy* is defined by

$$E \equiv \sum_{L=1}^{\infty} (h(L) - h) \tag{4}$$

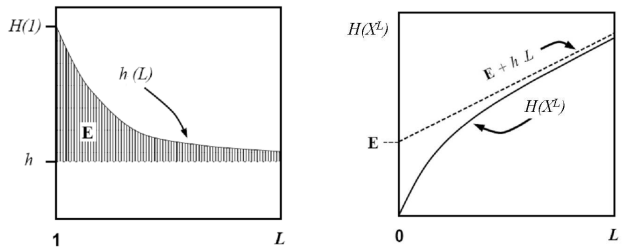
and captures how $h(L)$ converges to its asymptotic value $h$. Figure 1(a) is a graphical representations of the excess entropy measure, which is represented by the shaded area, corresponding to the sum of differences between $h(L)$ and the limit $h$.

If one inserts Equation (3) into Equation (4), the sum telescopes and one arrives at an alternate expression for the excess entropy [11]:

$$E = \lim_{L \to \infty} [H(X^L) - h \cdot L]. \tag{5}$$

Hence, excess entropy is the y-intercept of the straight line to which $H(X^L)$ asymptotes as indicated in Figure 1(b).

It is important to note that, when we take into account only a few number of symbols in the entropy computation, the system appears more random that it actually is. This excess randomness tells us how much additional information must be gained about the configurations in order to reveal the actual uncertainty $h$ [12].



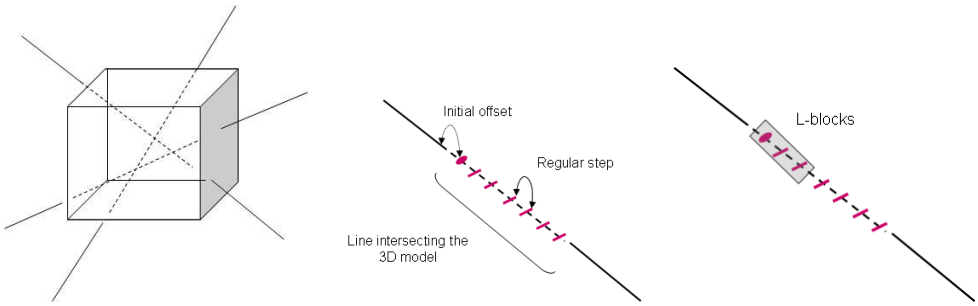

(a) (b)

**Fig. 1** Two different graphical representations of the excess entropy measure, corresponding to the Equations (4) and (5), respectively. Images obtained from [12].

### *2.2 Image Segmentation based on Thresholding*

Because of its simplicity and efficiency, thresholding is a widely used technique for image segmentation, especially in applications where speed is an important factor. The main objective in thresholding is the selection of the threshold values that properly isolate the different parts of the image. To locate the thresholds, parametric and nonparametric approaches can be used. In the first case, a gray level distribution of a region is assumed, while, in the second case, no assumptions are made. In this case, the thresholds are obtained in an optimal manner according to some criteria. The nonparametric methods are more robust, and usually faster than the parametric ones. Thresholding is best suited for bimodal distribution, such as solid objects resting upon a contrast background [19] and it can be also effectively used as the initial step in more sophisticated image analysis tasks.

Information theory has been applied to define some of the thresholding criteria. A first method was proposed by Pun [15] and enhanced by Kapur et al. [14]. The histogram is separated into independent classes so that its entropy is maximized.




(a) (b) (c)

**Fig. 2** (a) Global lines are cast from the walls of the bounding box, (b) intensity values are captured at evenly spaced positions over the global lines from an initial random offset and (c) neighbor intensity values are taken in $L$-blocks.

Brink [4] extended this method to two dimensions by introducing spatial information. More recently, Rigau et al. [16] introduced an algorithm for medical image segmentation based on the mutual information maximization of the information channel between the histogram bins and the regions of the partitioned image.

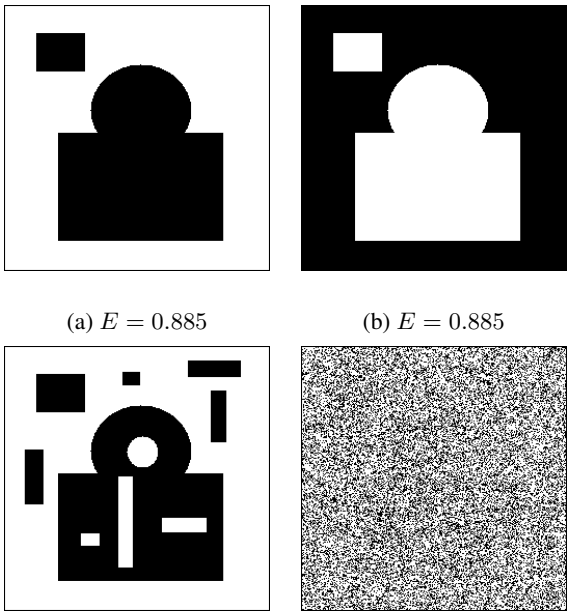
## 3 Spatial structure of an image

In this section, the excess entropy is introduced as a measure of the spatial structure of a 2D or 3D image. Structure here is taken to be a statement which expresses the degree of correlation between the components of a system. Excess entropy, which provides us with a measure of the regularities present in an image, can be also interpreted as the the degree of predictability of a pixel (or voxel) given its neighbors. From the concepts introduced in Section 2.1, we analyze how the excess entropy can be computed from Equation (5).

In the context of an image, $\mathcal{X}$ represents the set of clusters or bins of the image histogram and $x^L$ is given by a set of $L$ *neighbor* intensity values. In order to compute the excess entropy, two main considerations have to be taken into account:

- The first is the definition of the neighborhood concept pixel or voxel. While neighborhood is unique and unambiguous in 1D, its extension to 2D or 3D introduces ambiguity, since a sequence of L-block neighbor pixels or voxels can be selected in different manners [11].
- The second is the computation of $L$-block entropies when $L \rightarrow \infty$. In practice, $L$-block entropies for high $L$ are not computable, since the number of elements of the joint histogram (required to compute joint probabilities $p(x^L)$) is given by $N^L$, where $N$ is the cardinality of the system. Note that in our case, $N$ is the number of clusters or bins of the segmented image histogram, i.e., the number of colors of the image. Thus, a tradeoff between the accuracy of the measure, given by $L$, and the number of clusters $|\mathcal{X}|$ is required.

To overcome the neighborhood problem, uniformly distributed random lines, also called *global lines* [18] are used. Global lines sample the 3D-volume stochastically in the sense of integral geometry, i.e., invariant to translations and rotations [17]. These lines are generated from the walls of a convex bounding box containing the volume [5]. This can be done taking a random point on the surface of the convex bounding box and a cosinus distributed random direction as it is illustrated in Figure 2(a). The sequence of intensity values (L-block $X^L$) needed to estimate the joint probabilities is captured at evenly spaced positions over the global lines from an initial random offset, that ranges from 0 to the step size (see



(a) $E = 0.885$ (b) $E = 0.885$

(c) $E = 0.782$ (d) $E = 0.001$

**Fig. 3** Synthetic images and their excess entropy values.

Figure 2(b)). Points chosen on each line provide us with the intensities to calculate the $L$-block entropies, required to compute the excess entropy (see Figure 2(c)). In this manner, the 3D-neighborhood problem is reduced to 1D, where the concept of neighborhood is well defined. In our implementation, $N$ is taken as an input parameter of the algorithm, while $L$ is determined from $N$ such that the computation of the joint histogram is attainable.

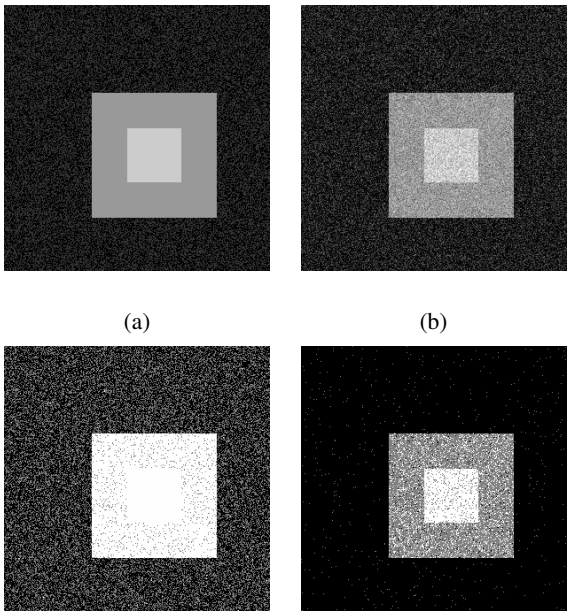
To illustrate the behavior of the excess entropy as a measure of the image structure, we use the 2D images of Figure 3. The two first images (a) and (b) represent the same scene with the colors interchanged. In this case, the excess entropy values are the same since the structure of the image is not dependent on the colors. In the third image (c), some additional shapes are added to the original image (a),

keeping the same probability for each color. Because of the higher variability of the obtained image, the excess entropy measure decreases, reflecting a lower spatial structure. The last image (d) has been generated by swapping 200000 points of image (a). Each swapping has been done by choosing two random points of the image and interchanging their intensity values. Observe that now the image has not spatial structure (no shape can be detected) and, therefore, the excess entropy is close to 0. It is important to remark that the values of the Shannon entropy of all the images of Figure 3 are the same, since the probabilities of each color have remained unaltered.

## 4 Thresholding based on Excess Entropy

In this section, a thresholding segmentation method that uses the excess entropy to obtain the optimal thresholds is presented.

We propose a nonparametric approach to segment the image without a priori assumptions about the underlying data. As it has been described in the previous section, excess entropy can be used as a measure of the structure of a system. From the assumption that an image is structured in regions, we conjecture that the optimal thresholding should provide us with the maximum structure. Consequently, the selection of thresholds will be formulated as a histogram quantization problem using the maximization of excess entropy. That is, the optimal histogram quantization should correspond to the maximum excess entropy of the resulting image. Figure 4 shows an example to illustrate the behavior of our method compared to the k-means algorithm [10]. Figure 4(a) is a four-colored synthetic image



(a) (b)

(c) (d)

**Fig. 4** Preliminary experiment: (a) 4-colored image, (b) test image, (c) segmentation result of the k-means algorithm and (d) segmentation result of the proposed method.

with two nested squares on a background with two different colors randomly distributed. Figure 4(b) has been obtained by adding $1\%$ of Gaussian noise to the original image. Figures 4(c) and 4(d) show the segmentation of the image (b) in three different clusters obtained by the k-means algorithm, one of the most popular clustering algorithms, and the excess entropy method applied to the noisy image, respectively. Observe that the k-means method separates the background in two colors and merges the two nested square shapes. This behavior is due to the fact that the number of background pixels is greater than the number of pixels of the squares and this causes that the intensity error classification is lower if the background is split into two different clusters. Using our method, the spatial

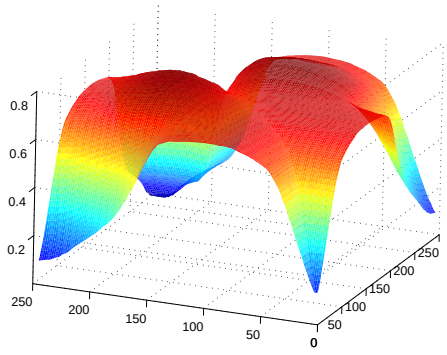



**Fig. 5** Excess entropy values of the image 4(b) with 3 colors for different threshold levels.

information is taken into account and, hence, the square shapes have been mainly preserved. Observe that the misclassified pixels are caused by the Gaussian noise of the input image (Figure 4(b)).

The kernel of our method is the computation of excess entropy given by the Equation (4), where $N$, the number of colors of the segmented image, and $L$, the block length, are input parameters of the segmentation algorithm. If $C$ is the number of colors of the original image and $N-1$ the number of thresholds, the quantity of possible combinations of different threshold levels is $\frac{C!}{(C-(N-1))!}$. Note that for standard medical images (with a minimum of 256 colors) this number becomes intractable when $N$ is greater than 4 or 5. Therefore, except for the case of thresholding in two or three groups, it will be impossible to explore all the possible solutions. For instance, in the segmentation process of Figure 4, all possible threshold levels have been checked. In the cases that the number of threshold levels is high, certain optimization criterion must be used.

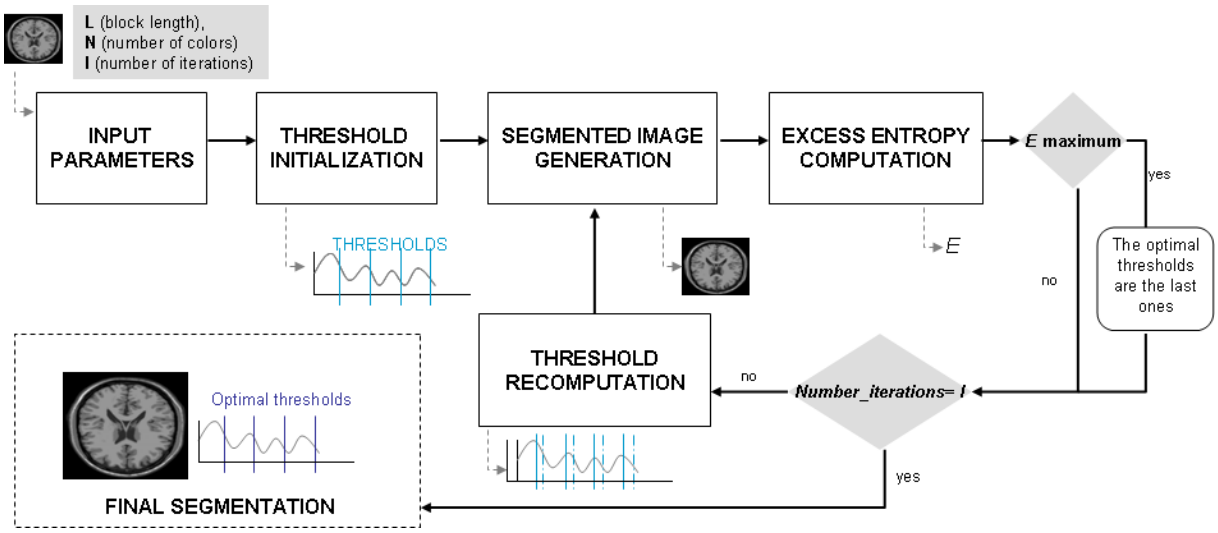



**Fig. 6** Block diagram showing the excess entropy-based thresholding method.

In Figure 5, we plot the values of excess entropy values of the image 4(b) segmented with two different threshold values, ranging from 0 to 255. As it can be seen, the function has a smooth behavior and therefore an optimization process can easily converge to the optimal solution. In our implementation, in order to properly locate the thresholds avoiding local minima, that can appear for a higher number of optimization parameters, a global genetic-based strategy is applied, since this is robust to these local minima. If high accuracy results are required, a second optimization process based on a gradient descent approach is applied.

Taking into account all these considerations we propose the iterative algorithm represented in Figure 6 and described below:

1. *Input parameters:* the image, the number of colors ($N$), the block length ($L$) and the number of iterations ($I$).

2. *Threshold initialization.* This module computes the image histogram and distributes uniformly $N-1$ thresholds.
3. *Segmented image generation*. The input image is segmented using the last computed thresholds.
4. *Excess entropy computation.* This module computes the excess entropy $E$ of the previous segmented image.
5. If $E$ is maximum, the thresholds of the image are stored as the optimal thresholds.
6. If the number of iterations is not reached, we *recompute thresholds* by adding a noise term to the optimal thresholds. This noise is generated following a gaussian distribution. Then we go to step 3.
7. *Final segmentation*. The input image is segmented using the optimal thresholds.

At the end of the process, if high accurate threshold locations are needed, the gradient descent algorithm is applied.

## 5 Results

The proposed segmentation approach has been developed using ITK [1] and VTK [2] libraries, and integrated in the medical image visualization and processing framework developed in our laboratory. It has been evaluated on different synthetic and real sequences of brain images.

An important limitation to be considered when segmenting brain images is the overlapping intensity values between different tissues. To illustrate such a limi-

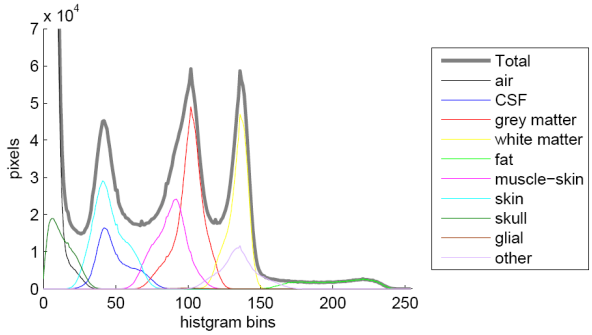



**Fig. 7** Histogram of the T1 image with 3% of noise.

tation (see Figure 7), the histogram of a T1 image from Brainweb database [6] is plotted. In this plot, each one of the ten tissues represented in the T1 image is shown with a different color. Note how different intensity values overlap and consequently thresholding methods are not capable to isolate one from the other. For instance, both cerebro spinal fluid and skin intensity values range from 20 to 80 and, hence, they are not distinguishable. Due to this limitation, the main purpose of segmentation techniques is not the separation of all real brain tissues but only the most important ones which in general correspond to background, cerebro spinal fluid, grey matter, white matter and skull. Therefore, in our experiments the number of clusters to be considered has been 4, 5 or 6 (i.e. $4 \leq N \leq 6$). The $L$ parameter has been set to 6 and the number of iterations to 200.

The first experiment has been designed to evaluate the performance of the proposed approach for images with different levels of noise. With this purpose a set of synthetic magnetic resonance (MR) images from the Brainweb database [6] is used. The method with $N = 5$ is applied to T1 and T2 image modalities with three different levels of noise $0\%$, $3\%$ and $7\%$, respectively.

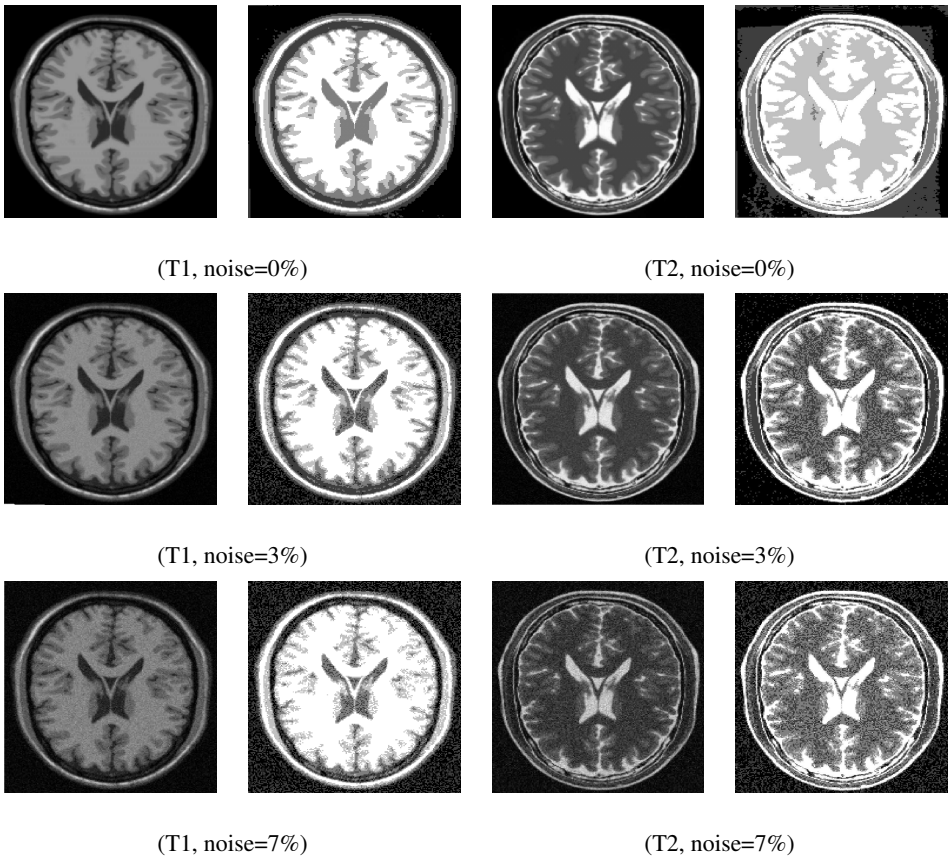



**Fig. 8** T1 (first column) and T2 (third column) images of Brainweb database with different degrees of noise. In the second and fourth columns, the corresponding segmented images with $N = 5$.

The obtained results are illustrated in Figure 8. In the first row, we show, from left to right, the T1 image without noise and the obtained segmentation, and the T2 image without noise and its corresponding segmentation. In a similar way, the second and third rows represent the results obtained from the T1 and T2 images with $3\%$ and $7\%$ of noise, respectively. Observe that, in the case of T1 without noise, background, skull and cerebro spinal fluid (CSF) are correctly separated,

while the classification gets worse when noise increases. Such a behavior is typical of the thresholding approach. Similar results are obtained with noisy T2 images. In the case of T2 without noise, CSF and grey matter are grouped in the same cluster, while background is split into two clusters. Observe also the rectangular shape that surrounds the brain in the segmented image. This shape affects the excess entropy measure leading to an uncorrect segmentation. This undesirable effect disappears in the case of noisy images, achieving a correct segmentations. We want to remark that these patterns do not appear in real images and hence such a bad behavior is not expected for real data.

In our second experiment, the segmented images of the first experiment are compared with the true standard that corresponds to a brain image segmented in ten tissues [6]. The goal of this comparison is to evaluate which tissues and how the tissues of the true images have been grouped into the five clusters used for the image segmentation. To illustrate the obtained results for each one of the test images we have generated a bar plot (see Figure 9). Each bar represents one of the ten tissues of the true standard, from left to right, air, CSF, grey matter (GM), white matter (WM), fatty tissue, muscle and skin, mostly skin, skull, glial tissue and other tissue. To represent the five clusters used for segmentation five different colors have been considered. The bars have been filled with one or more colors according to the $\%$ of pixels of the corresponding tissue in the different clusters. For instance, in plot 9(a) corresponding to T1 image without noise, we can see from the first bar that air pixels have been grouped into two different clusters, $75\%$ into cluster 1 and $25\%$ into cluster 2.

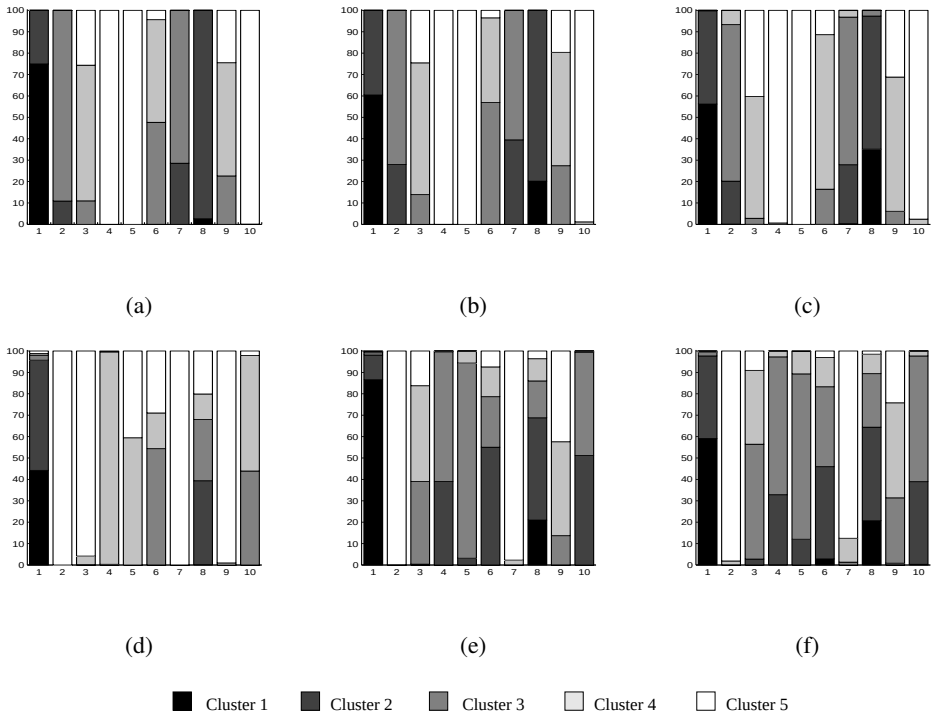



**Fig. 9** Comparison of the segmented images of Figure 8 with the true image. T1 (first row) and T2 (second row) results with $0\%$, $3\%$ and $7\%$ of noise are plotted. Columns of each plot represent: (1) air, (2) cerebro spinal fluid, (3) grey matter,(4) white matter, (5) fatty tissue, (6) muscle and skin, (7) mostly skin, (8) skull, (9) glial tissue and (10) other tissue.

In the first row of Figure 9, the T1 images with $0\%$, $3\%$ and $7\%$ of noise are plotted. Observe that in the case of T1 without noise (plot 9(a)) the four main tissues (CSF, GM, WM and skull) and air are well classified, in the sense that there is a good correspondence between the clusters of the segmented image and the main tissues of the true image. In particular, $90\%$ of CSF pixels are of cluster 3 and $10\%$ of cluster 2; $65\%$ of GM pixels are of cluster 4, $25\%$ of cluster 5 and $10\%$ of cluster 3; $100\%$ of WM pixels are of cluster 5; $95\%$ of skull pixels are of cluster

2 and the rest of cluster 1; and $75\%$ of air is of cluster 1 and $25\%$ of cluster 2. As it was expected the quality of the T1 segmentation decreases when the percentage of noisy increases. In spite of this, it can be seen that the method still groups the main tissues in the same clusters (see plots 9(b) and 9(c)).

In the second row of Figure 9, the T2 images with $0\%$, $3\%$ and $7\%$ of noise are plotted. In the case of T2 without noise (plot 9(d)), air is separated in two main clusters (corresponding to the two background regions represented in Figure 8 (T2, noise=$0\%$), while CSF and grey matter are grouped in the same cluster. In the case of noisy T2 images (see plots 9(e) and 9(f)), it can be seen that the main tissues are well classified.

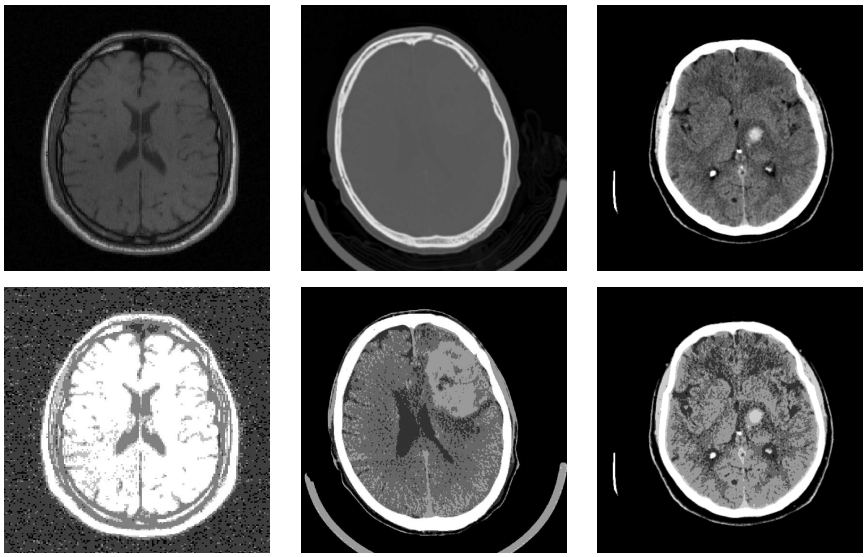
In the next experiment we evaluate the performance of the proposed approach for different values of $N$. For the test we use the T1 image with $3\%$ of noise and $N = 4..6$. The results are listed in Table 1 where rows represent the ten tissues of the true standard and the columns the clusters generated for each one of the $N$ values. The listed values represent the distribution of the tissue in each cluster. If we analyze the distribution of tissues for the different cases, it can be seen that white matter is perfectly segmented. Observe that the cluster containing white matter also contains other brain tissues. This is an expected result due to the overlapping intensity values (see Figure 7).

Finally, a set of real computed tomography and magnetic resonance brain images obtained from the Hospital Josep Trueta of Girona has been segmented. The method has been tested with $N = 6$ and $L = 6$. The obtained results are shown in Figure 10. In the case of MR images, represented in the first column, the main tis-

| Brain | N=4 | | | | N=5 | | | | | N=6 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tissue | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (5) | (1) | (2) | (3) | (4) | (5) | (6) |
| air | 60.4 | 39.6 | | | 60.4 | 39.6 | | | | 60.4 | 38.0 | 1.6 | | | |
| CSF | | **94.2** | 5.8 | | | 28.0 | 72.0 | | | | | 48.9 | 51.1 | | |
| GM | | 0.6 | 56.9 | 42.5 | | | 13.9 | 61.6 | 24.5 | | | | 15.4 | 60.1 | 24.5 |
| WM | | | | **100** | | | | | **100** | | | | | | **100** |
| FT | | | | **100** | | | | | **100** | | | | | | **100** |
| MS | | 8.4 | 84.4 | 7.2 | | | 56.9 | 39.6 | 3.5 | | | | 60.8 | 35.7 | 3.5 |
| SKN | | **98.6** | 1.4 | | | 39.5 | 60.5 | | | | 0.1 | 57.6 | 42.3 | | |
| SKL | 20.2 | 79.8 | | | 20.2 | 79.8 | | | | 20.2 | 54.2 | 25.6 | | | |
| GT | | 1.4 | 66.8 | 31.8 | | | 27.4 | 53.0 | 19.0 | | | | 29.8 | 50.6 | 19.6 |
| other | | | 0.3 | **99.7** | | | | 1.2 | **98.8** | | | | | 1.2 | **98.8** |

**Table 1** Distribution of brain tissues in the different clusters after applying our method with $N = 4$, $N = 5$ and $N = 6$. Rows represent, from top to bottom: air, cerebro spinal fluid (CFS), grey matter (GM), white matter (WM), fatty tissue (FT), muscle and skin (MS), mostly skin (SKN), skull (SKL), glial tissue (GT) and other tissues.

sues have been correctly separated. When segmenting the CT image of the second column, the method not only has isolated the main tissues but it also has enhanced a hidden pathological region. Finally, in the third column, a CT image of a patient with an intracranial hemorrhage, where the image intensities have been appropriately rescaled to a better visualization, is presented. In the segmented image, the different image tissues, the lesion and the background are correctly separated, giving a good representation of the anatomical structures.



(a) (b) (c)

**Fig. 10** Original MR and CT brain images (first row) with their corresponding segmentations (second row).

## 6 Conclusions

An information-theoretic measure, called excess entropy, has been introduced to quantify the image structure and to obtain a new adaptive thresholding method for image segmentation. The main novelties of this paper are the use of excess entropy as a measure of structural information of an image, the search for optimal thresholds by maximizing the excess entropy, and the use of uniformly distributed random lines to compute this measure. Experimental results have shown the good behavior of the presented approach. Further research will be done in the application of excess entropy to other image processing areas. We will also investigate the use of other sampling strategies to calculate the excess entropy.

## 7 Acknowledgements

Our project is funded in part by Spanish Government grants number TIN2007-68066-C04-01 and TIN2007-67982-C02.

## References

1. ITK Insight Toolkit, http://www.itk.org.
2. VTK Visualization Toolkit, http://www.vtk.org.
3. A. Bardera, M. Feixas, I. Boada, and M. Sbert. Medical image registration based on random line sampling. In *IEEE International Conference on Image Processing (ICIP'05)*, Genova, Italy, September 2005.
4. A. Brink. Minimum spatial entropy threshold selection. *IEE Proceeding-Vision, Image, and Signal Processing*, 142(3):128–132, June 1995.
5. F. Castro and M. Sbert. Application of quasi-monte carlo sampling to the multipath method for radiosity. In *Proceedings of 3rd International Conference on Monte Carlo and Quasi-Monte Carlo methods in Scientific Computing*, Claremont (CA), USA, June 1998.
6. C. Cocosco, V. Kollokian, R.-S. Kwan, and A. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4), 1997.
7. T. M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., 1991.
8. J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: The entropy convergence hierarchy. *Chaos*, 15:25–54, 2003.
9. J. P. Crutchfield and N. Packard. Symbolic dynamics of noisy chaos. *Physica D*, 7:201–223, 1983.

10. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons Inc., 2nd edition, 2001.
11. D. Feldman and J. Crutchfield. Structural information in two-dimensional patterns: Entropy convergence and excess entropy. *Physical Review E*, 67, 2003.
12. D. P. Feldman. A brief introduction to: Information theory, excess entropy and computational mechanics. Technical report, Department of Physics, University of California, Berkeley (CA), USA, October 2002.
13. R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River (NJ), USA, 2002.
14. J. Kapur, P. Sahoo, and A. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics and Image Processing*, 29(3):273–285, March 1985.
15. T. Pun. A new method for gray-level picture thresholding using the entropy of the histogram. *Signal Processing*, 2:223–237, 1980.
16. J. Rigau, M. Feixas, M. Sbert, A. Bardera, and I. Boada. Medical image segmentation based on mutual information maximization. In *Lecture Notes in Computer Science (MICCAI 2004)*, pages 135–142, Rennes-Saint Malo, France, September 2004.
17. L. A. Santaló. *Integral Geometry and Geometric Probability*. Cambridge University Press, 1976.
18. M. Sbert. An integral geometry based method for fast form-factor computation. *Computer Graphics Forum (Proceedings of Eurographics'93)*, 12(3):409–420, 1993.
19. M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, January 2004.