SEGMENTATION OF EIT IMAGES USING FUZZY CLUSTERING: A PRELIMINARY STUDY

Vincent BARRA[1] , Véronique DELOUILLE[2] , Jean-François HOCHEDEZ[2] , and Pierre CHAINAIS[1]

[1]LIMOS, UMR CNRS 6158, Campus ces Cézeaux, 63177 AUBIERE CEDEX, FRANCE
[2]Royal Observatory of Belgium, Avenue Circulaire, 3, B-1180 BRUXELLES, BELGIUM

ABSTRACT

We present a new application of a signal processing technique called *fuzzy clustering* for automatic identification of various structures seen in EIT images. This technique gives for each pixel a probability of belonging to a particular class. By assigning each pixel to the class for which it has the greatest probability of belonging, we obtain image segmentations. In EIT 19.5 nm images we distinguish the Quiet Sun, Coronal Holes, and the Active Regions, whereas in EIT 30.4 nm we extract the plages and part of the network boundaries. We also show how a multiwavelength approach leads to an improved segmentation of the different coronal structures.

Key words: Irradiance study; EIT images; Fuzzy Clustering; Multiwavelength analysis.

1. INTRODUCTION

Identifying the respective contributions of different structures to the solar irradiance is now a key issue in solar physics, with implications to Sun-Earth relationships. In this paper, we propose an algorithm for the automatic segmentation of different structures seen in EIT images. The method relies on a fuzzy clustering on relevant feature descriptors representing image pixels. For coronal images, we are able to separate the Quiet Sun, Coronal Holes, and Active regions. In the transition region, we can distinguish the chromospheric network and the plages from the dark intra-cell regions.

Several methods of segmentation have been proposed in the solar physics literature, of which we name here but a few. Worden et al. (1999) propose an ad hoc method for segmenting 30.4 nm images, where the parameters of the procedure are derived by trial and errors. The aim there is to study the respective contributions of the different structures (plage, enhanced network, active network, quiet chromosphere) to the solar He II irradiance. Turmon et al. (2002) propose a Bayesian image-segmentation technique for treating SOHO/MDI data. Finally, Delouille et al. (2005) present a method for segmenting EIT images based on the local value of the wavelet spectrum.

This paper is organized as follows. Section 2 introduces the fuzzy segmentation method and describes its parameters. Section 3 presents the results, and Section 4 discusses future extension of the method.

2. THE FUZZY CLUSTERING ALGORITHM

The main idea behind fuzzy clustering is that an object can belong simultaneously to more than one class and does so to varying degrees called *memberships*. In this paper, we use the *possibilistic clustering*, see Section 2.2. The memberships generated by this algorithm offer several advantages. First, whenever the user overspecifies the number of clusters, the algorithm will replicate one class. The correct number of clusters will thus still appear in the segmentation, in contrast to classical clustering algorithms which typically divide artificially a cluster in two parts in such case. Second, the fuzzy nature of the clustering makes it possible to overcome the clear-cut nature of pattern descriptions: even if the chosen descriptors are not the best ones, the algorithm may still be robust to this choice. Third, experts have sometimes problems in determining precisely the borders of the different structures. It is thus convenient to define memberships, for which a large value indicates increased confidence that a pixel belongs to a particular structure. Finally, it is possible to include some human expertise knowledge in the method, in order, e.g. to alleviate some ambiguity near the edges of coronal structures.

The first fuzzy clustering method that was developed is the C-means fuzzy clustering algorithm (Bezdek (1981)), that we now recall.

2.1. The fuzzy C-means algorithm

In this preliminary study, we segment based on the intensity of the image expressed in DN. The pixel values x_j are used here as *features*, also called *descriptors* in pattern recognition terminology. In general, we denote by $X = \{x_j\}_{1 \le j \le N}$ the set of p-dimensional feature vectors $x_j \in \mathbb{R}^p$ of an image.

The fuzzy C-means algorithm (FCM) is an iterative method, which tries to separate the set X into C compact clusters. Every cluster is represented by its center, denoted b_i. FCM associates with every feature vector $x_j \in \mathbb{R}^p$ a fuzzy membership grade $u_{ij} \in [0,1], 1 \le i \le C$ in each of the C classes. This membership u_{ij} represents the degree to which x_j belongs to class i. The set $U = \{u_{ij}\}_{\{1 \le i \le C, 1 \le j \le N\}}$ is called the fuzzy partition of X. Let $B = \{b_i\}_{1 \le i \le C}$ be the set of cluster centers. The FCM uses an iterative optimization procedure to approximate the minima of a constrained objective function:

$$J(B,U,X) = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m d(x_j, b_i)$$

subject to

$$(\forall i \in \{1 \cdots C\}) \sum_{j=1}^{N} u_{ij} < N \qquad (1)$$

$$(\forall j \in \{1 \cdots N\}) \sum_{i=1}^{C} u_{ij} = 1 \qquad (2)$$

where m is a parameter that controls the degree of fuzzification ($m = 1$ means no fuzziness), and d is a metric in $\mathbb{R}^p$.

Several authors (Barra (2000), Krishnapuram et al. (1997)) showed that this algorithm creates *relative memberships*, interpreted as degrees of sharing the pixels between all the classes. These degrees are thus *not* representative of the true degree of belonging. Indeed, FCM leads to an analytic formulation of

$$u_{ij} = \left[\sum_{k=1}^{C} \left(\frac{d(x_j, b_i)}{d(x_j, b_k)} \right)^{\frac{2}{m-1}} \right]^{-1}$$

which depends on the distances of x_j to *all class* centers b_k and not only on $d(x_j, b_i)$. Following Zadeh (1978), we choose to remedy to this problem, by relaxing the constraint (2) using a Lagrangian relaxation. This leads to the *possibilistic clustering algorithm* (PCA).

2.2. The possibilistic clustering algorithm

Krishnapuram et al. (1997) propose to approximate the minima of the function

$$J(B,U,X) = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m d(x_j, b_i) + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} (1 - u_{ij})^m$$

where η_i is a weighted ponderation term that determines the distance at which the membership value of a feature in the cluster i becomes 0.5. Authors proved that this algorithm allows the memberships to be interpreted in terms of *absolute memberships* of features to clusters. In other words, u_{ij} now only depends on x_j and class i:

$$u_{ij} = \left(1 + \left(\frac{d(x_j, b_i)}{\eta_i} \right)^{\frac{1}{m-1}} \right)^{-1}$$

This is necessary in the case of strong ambiguity or uncertainty, which can happen in complex-shaped images, as shown by clear examples in Krishnapuram et al. (1997). Several parameters directly influence the PCA algorithm:

- Initialization: $J(B,U,X)$ is often not convex, and optimization may then only lead to a local minimum. Initialization of the method is thus a crucial step, and we choose to perform a few iterations of the FCM method to find first estimates of class centers B before running the PCA.
- Ponderation term η_i: η_i can be viewed as a penalization coefficient. We propose to compute η_i as the mean intra-class fuzzy distance

$$\eta_i = \frac{\sum_{j=1}^{N} u_{ij}^m d(x_j, b_i)}{\sum_{j=1}^{N} u_{ij}^m}$$

- Fuzzifier m: parameter $m > 1$ controls the degree of fuzzyness of U. If m is close to 1, U is almost 'clear-cut', that is, each x_j is assigned to one and only one class. On the contrary, U tends towards the uniform law on $[1 \cdots C]$ as m tends to infinity. Generally speaking, $m \ge N/(N-2)$ assures the algorithm convergence. In this paper, we choose $m = 2$.
- Feature vectors: a feature vector $x_j \in \mathbb{R}^p$ is a vector representing the image pixel $j, 1 \le j \le N$. The choice of these descriptors is a critical point in the algorithm, since they must capture all the relevant information allowing a good clustering. In this preliminary study, the component of x_j is either the gray-value of the pixel j, the log of its gray level or the square root of its gray level. In the next Section, we show some examples using either one image ($p = 1$), or several images ($p = 3$) coming from different coronal wavelengths.
- Number of classes: C is determined by the type of structures we await to see. For example, in coronal images, we expect to distinguish the Active Regions (AR), their surroundings, the Coronal Holes (CH), and the Quiet Sun (QS). Consequently, we choose $C = 4$ classes. If only three clusters are present in the feature space, then naturally the possibilistic clustering algorithm will duplicate one class, due to the relaxation of constraint (2).

The possibilistic clustering algorithm leads to the computation of fuzzy maps $(u_{i\cdot})$ (see e.g. Figure 3), where the gray level of a pixel j gives its membership value (u_{ij}) to class i. In this paper, the final segmented image is simply obtained by assigning each pixel to the class for which it has the greatest membership. More sophisticated physical schemes are under study.

3. SEGMENTATION OF EIT IMAGES

The images analyzed here have all been pre-processed using the standard *eit_prep* procedure of the *solar software* library. We first consider two 19.5 nm images, one taken on December 22, 1996, in period of solar minima (I_1), and the second on August 3, 2000 during maximal activity of the Sun (I_2). In order to avoid edge effect at the limb, we consider only the 'on-disc' part of the Sun, that is, the disk centered on the Sun and having a radius equal to $0.95R_{\odot}$.

We apply the algorithm with three types of descriptors: the gray level image, its log, and its square root. Note that taking the square root is close to doing an Anscombe transform, and normalizes somewhat the Poisson noise. Figure 1 shows the segmentations for I_1. We see that gray level image allows to segment active regions (AR) including their surrounding area, the quiet sun (QS) and coronal holes (CH), whereas the log of gray level allows to separate AR from their surrounding area and from QS, but does not recognize the coronal holes. Finally, the Anscombe transform of the image allows to separate AR including their surrounding area, the QS and the CH.

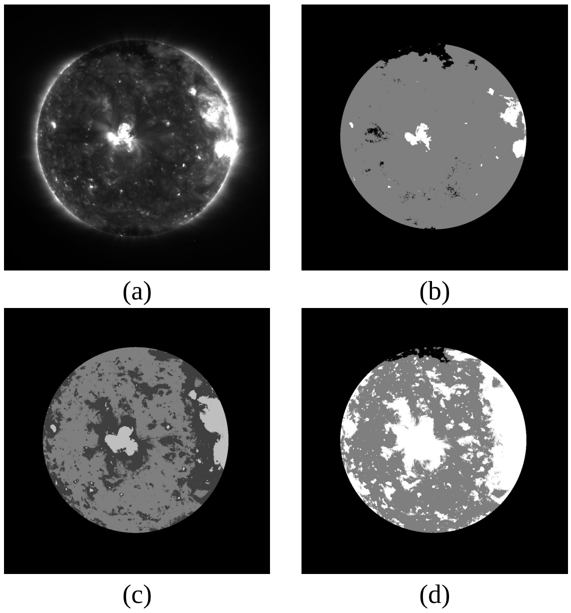


Figure 1. (a) EIT image I_1 recorded on December 22, 1996, together with the segmentation results on (b) gray levels, (c) logs and (d) square roots

The second image I_2 is represented in Figure 2 together with its segmentations. There again we see that the gray level feature is able to separate filaments and CH from AR and QS, whereas the logarithm of the gray level distinguish AR from their surroundings. As said above, we obtain these classes by associating a pixel to the class for which it has the largest membership value.

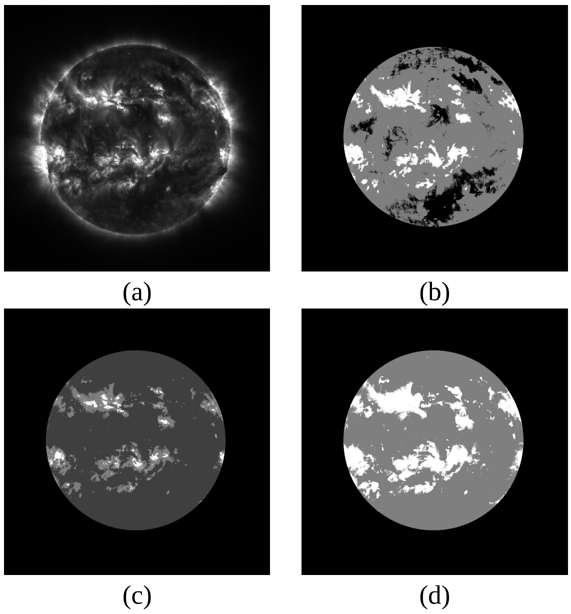


Figure 2. (a) EIT image I_2 recorded on August 03, 2000, together with the segmentation results on (b) gray levels, (c) logs, and (d) square roots

Figure 3 shows the four $(C = 4)$ fuzzy maps generated from the possibilistic clustering algorithm that leads to the segmentation of Figure 1(d). The brighter the gray level of a pixel in a class map, the higher its membership to that class.

We also segmented two 30.4 nm images (called I_3 and I_4), taken on the same day as the 19.5 nm images above. Figure 4 shows the results, where the plages and some of the network cell boundaries are extracted from the darker background. Note that the same images were segmented in Delouille et al. (2005) with a classical clustering technique, using as descriptors the fit of a local version of the wavelet spectrum. The separation obtained in Delouille et al. (2005) is less precise than the one of PCA.

Finally, in order to demonstrate the generic character of the method, we process a multiwavelength analysis, using the image I_1 together with the quasi-simultaneous 17.1

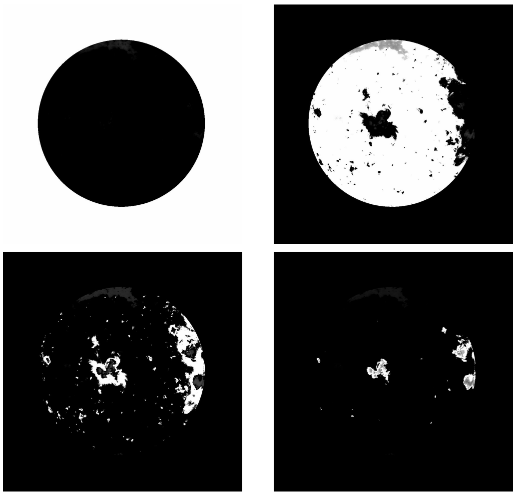
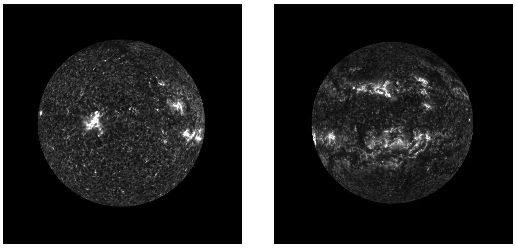
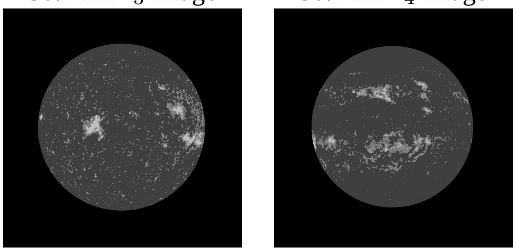


Figure 3. Fuzzy maps obtained when using as descriptor the square root of the gray level in the image I_1.



30.4 nm I_3 image 30.4 nm I_4 image



I_3 segmented image I_4 segmented image

Figure 4. Segmented images computed from the gray level pixel values of 30.4 nm images (I_3 and I_4).

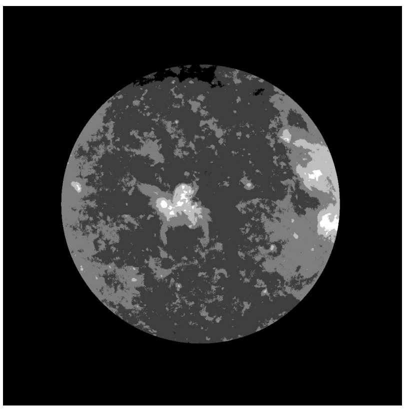


Figure 5. Multiwavelength approach: segmented image using 17.1, 19.5, and 28.4nm images from Dec. 22, 1996. AR and their surroundings, QS, and CH are well separated.

and 28.4 nm EIT images. Each feature vector x_j belongs thus to $\mathbb{R}^3$; its components are the gray level values in each of the three channels. We took $C = 5$ classes. Figure 5 shows the resulting segmentation. We are now able to distinguish the AR, the core of the AR, their surroundings, the CH and the QS.

4. FUTURE PROSPECTS

Numerous research and applicative extensions are now expected from this preliminary work:

- feature choice should be optimize: besides the intensity, we could use some preprocessed inputs which additionally inform on the local texture.
- fuzzy maps should be analyzed and memberships interpreted (in particular near the CH/QS and AR/QS interfaces)
- quantitative results from the segmentation should be processed: we aim in particular at studying the influence of AR, CH and QS on the variability of the total sun irradiance (Zhukov et al. (2002); Veselovsky et al. (2001)), by adding the pixel activities in each class and studying the time evolution of these quantities.

REFERENCES

Barra V., Boire JY.; 2000, Journal of Magnetic Resonance Imaging, 11:267–278.

Bezdek J., 1981; Pattern recognition with fuzzy objective function algorithms, Plenum Press

Delouille V., de Patoul J., Hochedez JF.; 2005, Solar Physics, 228:303–323.

Krishnapuram R., Keller J.M.; 1973, IEEE Transactions on Fuzzy Systems, 1:98–110.

Turmon M., Pap J.M., Mukhtar S.; 2002, ApJ, 568:396–407.

Veselovsky I. S., Zhukov A. N., Dmitriev A. V. , Tarsina M. V., Clette F., Cugnon P., Hochedez J. F.; 2001, Solar phys. 201:27–36.

Worden J., Woods T., Neupert W., Delaboudinière J.;1999, ApJ, 511:965–975.

Zadeh L.; 1978, International Journal of Fuzzy Sets and Systems,1:3–28.

Zhukov, A. N., Veselovsky I. S., Hochedez J.-F., Clette F., Panasenco O. A., Cugnon P. ; 2002, In: *ESA SP-508: From Solar Min to Max: Half a Solar Cycle with SOHO*. pp. 189–192.