Available online at www.sciencedirect.com
ScienceDirect
Advances in Space Research 42 (2008) 917–925
ADVANCES IN SPACE RESEARCH (a COSPAR publication)
www.elsevier.com/locate/asr

Segmentation of extreme ultraviolet solar images via multichannel fuzzy clustering

Vincent Barra [a], Véronique Delouille [b], Jean-François Hochedez [b,*]

[a] LIMOS, UMR CNRS 6158, Campus ces Cézeaux, 63177 Aubiere Cedex, France
[b] Royal Observatory of Belgium, Circular Avenue 3, B-1180 Brussels, Belgium

Received 10 November 2006; received in revised form 23 October 2007; accepted 23 October 2007

Abstract

The study of the variability of the solar corona and the monitoring of its traditional regions (Coronal Holes, Quiet Sun and Active Regions) are of great importance in astrophysics as well as in view of the Space Weather and Space Climate applications. Here we propose a multichannel unsupervised spatially constrained fuzzy clustering algorithm that automatically segments EUV solar images into Coronal Holes, Quiet Sun and Active Regions. Fuzzy logic allows to manage the various noises present in the images and the imprecision in the definition of the above regions. The process is fast and automatic. It is applied to SoHO–EIT images taken from February 1997 till May 2005, i.e. along almost a full solar cycle. Results in terms of areas and intensity estimations are consistent with previous knowledge. The method reveal the rotational and other mid-term periodicities in the extracted time series across solar cycle 23. Further, such an approach paves the way to bridging observations between spatially resolved data from imaging telescopes and time series from radiometers. Time series resulting form the segmentation of EUV coronal images can indeed provide an essential component in the process of reconstructing the solar spectrum.

Keywords: Solar corona; Solar cycle; Solar rotation; Ultraviolet; Fuzzy clustering; Multispectral; Segmentation

1. Introduction

The EIT telescope (Delaboudinière et al., 1995) onboard the SoHO ESA-NASA solar mission takes daily several datasets composed of four images (17.1, 19.5, 28.4 and 30.4 nm), all acquired within 30 min. They are thus well spatially registered and it provides for each pixel a collection of four intensities that potentially permit to recognize the standard solar atmosphere region to which it belongs. The commonly identified regions are Coronal Holes (CH), Quiet Sun (QS) and Active Regions (AR) although their boundaries are not well defined and other regions may be present (e.g. prominences).

The problem of coronal image segmentation in general and the detection and tracking of regions of interest in solar images in particular has been addressed in many ways in the last decade. Published procedures include: classical image processing methods classified as edge-based algorithms (Steinegger et al., 1997; Veronig et al., 2000, 2001; Fuller et al., 2005; Young et al., 2003; Revathy et al., 2005; Robbrecht et al., 2006), region-based methods (Bornmann et al., 1996; Pettauer and Brandt, 1997; Preminger et al., 1997; Steinegger et al., 1997, 1998; Worden et al., 1999; Hochedez et al., 2000; Veselovsky et al., 2001; Hill et al., 2001; Gao et al., 2002; Nieniewski et al., 2002; Turmon et al., 2002; Benkhalil et al., 2003; Wagstaff et al., 2003; Zharkova et al., 2003, 2004; Berrili et al., 2005; Ortiz, 2005) or hybrid and cooperative approaches (Bratsolis and Sigelle, 1998; Portier-Fozzani et al., 2001; Qahwaji, 2003).

Pattern recognition methods – when applied to multidimensional data – have proved to increase the level of

* Corresponding author. Tel.: +32 2373 0302; fax: +32 2374 9822.
E-mail addresses: vincent.barra@isima.fr (V. Barra), verodelo@sidc.be (V. Delouille), hochedez@sidc.be (J.-F. Hochedez).
URL: http://www.isima.fr/vbarra (V. Barra).

0273-1177/34.00
doi:10.1016/j.asr.2007.10.021

confidence in image segmentation as compared to gray-scale approaches using single images, e.g. in MRI processing (Vannier et al., 1991) or remote sensing images (Rangsanseri et al., 1998). The superiority of a multichannel approach is expected when some information relevant to the segmentation process cannot be found clearly in any individual channel. To our knowledge, only (Dudok de Wit, 2006) proposed a multispectral segmentation of EUV images; his method is based on supervised classification, thus requiring human intervention to train a dataset. In contrast, we present in this paper a multispectral unsupervised classification approach (also called 'clustering') for EUV images. It is based on the combination of three fundamental aspects: (1) the use of fuzzy logic, allowing to overcome the uncertainty in the images and the imprecision of the regions definition, (2) the genericity of the pixel representation permitting adequate feature descriptions in the algorithm and (3) the integration of a spatial regularization term in the clustering algorithm.

This paper is organized as follows. Section 2 introduces the theoretical background and justifies the construction of the implemented fuzzy clustering algorithm. In Section 3, the results of its application to a large fraction of the EIT archive are given and discussed; in particular we focus on the temporal evolution of the area and intensity of the coronal regions. We conclude in Section 4.

2. Multispectral fuzzy clustering algorithm

2.1. Theoretical background

The information provided by an EUV solar image is uncertain. It contains Poisson and readout noise, as well as cosmic ray hits. In addition, it is likely to have observational biases (line-of-sight integration of a transparent volume) and is subject to interpretation (the apparent boundary between regions is a matter of convention). Size, shape, and precise location of areas of interest are not easy to determine with only a single piece of information. To tackle this problem, we choose to model these data with a theory that is able to manage uncertainty and imprecision. More precisely, we resort to *Possibilistic logic*. Possibilistic logic was introduced by Zadeh (1978), following its former works in fuzzy logic and fuzzy set theory (Zadeh, 1965). Zadeh meant to simultaneously represent and manage imprecise and uncertain knowledge. In fuzzy set theory, a fuzzy measure is a representation of the uncertainty, giving for each event to be analyzed a coefficient in $[0,1]$ assessing the degree of certitude for this event. In possibilistic logic, when these events are singletons (e.g. in our application: "Pixel x belongs to an Active Region"), this measure is called a *distribution of possibility*, commonly denoted π, and satisfying $x \mapsto \pi(x) \in [0,1]$, such that $\sup_x \pi(x) = 1$. Consider an event $x :=$ "*pixel x belongs to S*", where S is either a Coronal Hole (CH), Quiet Sun (QS) or an Active Region (AR). We aim at assessing the degree to which this statement is correct by giving a quantitative evaluation through the attribution of a value $\pi(x) \in [0,1]$. The spatial distribution of CH, QS and AR are modeled in this context as distributions of possibility. These can be represented by fuzzy maps, *cfr.* Section 3.4.

2.2. Fuzzy clustering algorithm

Since its introduction by Bezdek (1981), the Fuzzy C-Means (FCM) algorithm was widely used in pattern recognition and image segmentation, in various fields including medical imaging (Philipps et al., 1995; Bezdek et al., 1997), remote sensing images (Rangsanseri et al., 1998; Melgani et al., 2000), color image segmentation in vision (Baker et al., 2003) or speaker authentification using lip image analysis (Leung et al., 2004).

FCM and its variants are iterative methods that search for C compact clusters in a multidimensional dataset of size N. They associate with every p-dimensional feature vector dataset $X = \{\vec{x_j}, 1 \leqslant j \leqslant N, \vec{x_j} \in \mathbb{R}^p\}$ a fuzzy partition matrix $U = (u_{ij})$, $1 \leqslant i \leqslant C$, $1 \leqslant j \leqslant N$, whose coefficients are in $[0,1]$, $u_{i,j} = u_i(\vec{x_j})$ being the membership degree of $\vec{x_j}$ to class i.

In the application to EIT images described below, we will aim at defining $C = 3$ clusters: CH, QS and AR. p could be as high as 4 by using all 4 EIT passbands but we restrict ourselves to 2 in the current study. With the upcoming SDO-AIA suite the multispectrality prospects of the method will be enhanced. j spans the N indexes of the on-disc pixels.

The idea behind FCM is the minimization of the total intracluster variance:

$$J_{\mathrm{FCM}}(B,U,X) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^m d\left(\vec{x_j},\vec{b_i}\right)$$

subject to

$$(\forall i \in \{1\dots C\}) \sum_{j=1}^{N} u_{ij} < N \text{ and} \tag{1}$$

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

where $B = \{\vec{b_1},\dots,\vec{b_C}\}$ is the set of cluster centers, m is a parameter that controls the degree of fuzzification ($m = 1$ means no fuzziness), and d is a metric in $\mathbb{R}^p$.

Krishnapuram and Keller (1993, 1996) showed that this algorithm creates relative memberships, interpreted as degrees of sharing pixels amongst all classes, that are thus unrepresentative of the true degree of belonging. In other words, FCM produces an analytic formulation of u_{ij} that depends on the distances of $\vec{x_j}$ to all class centers $\vec{b_k}$ and not only on $d(\vec{x_j},\vec{b_i})$. To solve this issue, Krishnapuram and Keller (1993, 1996) proposed a new version, called the Possibilistic Clustering Algorithm (PCA) that allows u_{ij} to depend only on $d(\vec{x_j},\vec{b_i})$. The objective function is now