Automatic Identification of Active Regions (Plages) in the Full-Disk Solar Images Using Local Thresholding and Region Growing Techniques

Ali Benkhalil
Cybernetics Department
University of Bradford
BD7 1DP, UK
a.k.benkhalil@bradford.ac.uk

Valentina Zharkova,
Cybernetics Department
University of Bradford
BD7 1DP, UK
v.v.zharkova@bradford.ac.uk

Sergei Zharkov
Cybernetics Department
University of Bradford
BD7 1DP, UK
s.zharkov@bradford.ac.uk

Stan Ipson
Cybernetics Department
University of Bradford
BD7 1DP, UK
s.s.ipson@bradford.ac.uk

Abstract

The paper presents techniques developed for the automated detection of Active Regions comprising bright areas on a darker background using Hα and Ca KII 3 solar images from the Meudon observatory. Statistical information for each quarter of the solar full-disk is calculated and used to determine the intensity threshold values which are used for initial segmentation of the active regions within each quarter. Median filtering and morphological operations are applied to the resulting binary image to remove noise and merge the broken regions. The centroid of each labelled region is calculated for use as a seed for a region growing procedure. Statistically based local thresholding is used to calculate upper and lower threshold values defining the spatial extent of regions. This procedure has been tested on one month sequences (April 2002) of full-disk solar images from Meudon observatory and with active regions classified by the National Oceanic and Atmospheric Administration (NOAA) observatory in Hawaii. A very good correspondence of identified active regions was found.

1 Introduction

There are a growing number of the archives of digitised images of the Sun takes from ground-based and space instruments in various wavelengths. These archives are available from different locations and are to be included into a unified catalogue by the EGSO project (Bentley, 2001). The digitized solar images have different sizes, resolutions instrumental and weather distortions. All are to be subjected to the automated recognition process in order to provide a reliable location of features and their evolution on different times of solar rotation. With growing demand for solar activity forecasts by the space

weather project and by many industrial organisations there is a great need for the development of reliable and fast techniques for feature recognition on solar disks and their presentation in feature catalogues. These catalogues are intended to contain comprehensive statistics of active events (sunspots, active regions, flares, filaments), overlapping in a given period of time and to allow the extraction of physical characteristics which are essential for the solar activity forecast.

The complex investigation of active regions in conjunction with flares, sunspots, coronal mass ejections (CME's) and filaments is aimed at providing a reliable forecasts of solar activity and the space weather. Active regions are the basic reference features for solar activity. Their reliable automated detection will allow the building of a major database of solar active features and for the first time enable analysis of the solar activity on a comprehensive database of active regions taken in various wavelengths. The techniques are developed for automated detection of Active Regions (plages) using the Hα and Ca KII solar images from the Meudon observatory, aiming to replace the existing manual detection methods. This research is done as part of the European Grid of Solar Observations (EGSO) project.

Three different approaches have been identified in the literature for the automatic identification of bright active regions (plage). The first approach based on the selection of a threshold to separate an object from the background which straightforward if the intensity histogram is bimodal but otherwise can be difficult (Steinegger et al, 1998). The Second approach based on the region growing techniques of segmenting solar images in a number of wavelengths, including Hα, into bright and dark regions which have been used by a number of authors for a variety of purposes (Hill et al, 2001; Veronig et al,

2001). Finally the third approach using the Bayesian inference method for automatically identifying various surface structures on the Sun (Turmon et al, 1998). The Bayesian based methods were concluded to be very computationally expensive. Intensity threshold based methods are simple and fast but are very sensitive to noise which affects the reliability of the segmentation results obtained. As a result, in the current work region growing techniques are used in conjunction with intensity thresholds which are used in the initial processing to find seed location and in the subsequent active region segmentation. The method used and the results obtained for active region detection are described in section 2 and the conclusions are given in section 3.

2   The recognition techniques for active regions.

A flow chart of the segmentation procedure developed to detect active regions is shown in Figure 1. The presented technique first standardizes the full-disk high-resolution solar images using procedures that include limb fitting, shape correction and limb darkening removal which will convert them to a format that can be easily used by all subsequent algorithms (Zharkova V.V, 2002).

Because of this pre-processing the first assumptions made before started the active region recognition procedure are that the input images are free of radial limb darkening and of standardised size (1024 pixel×1024 pixel) and solar disk radius (420 pixel) with solar centre (511.5 pixel ×511.5 pixel). The second assumption made is related to the active region characteristics where the active regions are the brightest features on the solar disk which means that the intensity values inside the detected regions of interest are greater than the intensity values on the background.
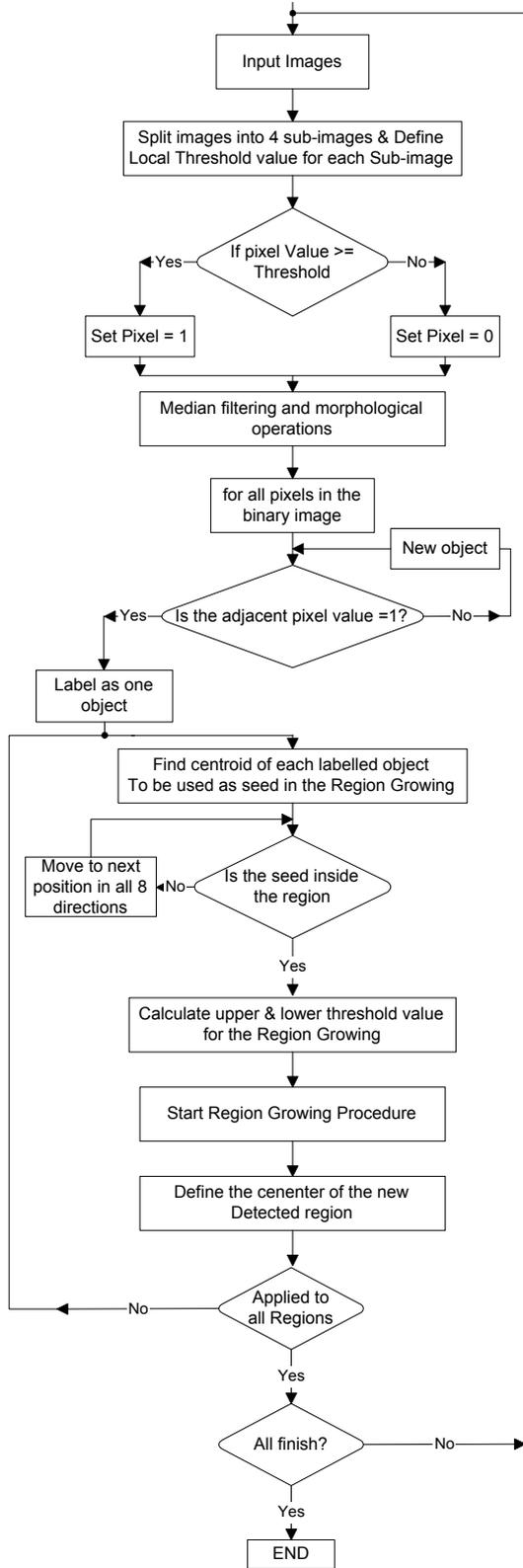
Figure 1: Flow chart of the used recognition technique for Active Regions detection

2.1 The Initial segmentation

The main stages of this technique using Ca KII 3 and Hα Meudon full-disk images are illustrated in Figure 2 and Figure 3 respectively. Figures 2(a) and 3(a) are the cleaned initial images while Figures 2(b) and 3(b) shows the results of remapping these cleaned images to polar coordinates. After the remapping, the statistically derived local intensity threshold values are used for an initial segmentation of the bright plages by setting each pixel to 1 whose intensity value is over this intensity threshold. Otherwise, the pixel value is set to zero. The choice of these local intensity threshold values is very importation because a value that is too high may lead to real features being missed, whereas a value that is too low may lead to noisier binary images and hence spurious features. The optimum global threshold value also varies with the images brightness level and the non-radial large scale intensity variation which are the main problems in some Meudon images. To overcome this problem and to obtain the optimal local intensity threshold values (*T*), these values are calculated in iterations, for the quarter-sized regions of an image as follows:

$$T_1 = \mu_1 + (1 + \Delta_1) \times \sigma_1 \qquad (1)$$

where μ_1 is the mean value for region one, σ_1 is the standard deviation for the same region and Δ_1 is a constant defining the range of values to look at, it is set to 0.4 after investigating more than 30 images. The results of this initial segmentation are shown in Figures 2(c) and 3(c).

2.2 The Median filtering, Morphological operations and Region labelling

The initial segmentation will include noise and unwanted small features caused by the

over-segmentation. The over-segmentation is preferable to under-segmentation as the former can be remedied using the Median filtering and Morphological operations while the latter could lose significant information. Firstly, in order to remove small features a 7×7 Median filter is used. The size was chosen through experimentation. Then the Morphological opening and closing operations are applied using a structure element of size 8×8. This smoothes the features and fills in holes. Figures 2(d) and 3(d) show the detected regions after applying Median and Morphological processing and transformation back to Cartesian coordinates. As can be seen, the noise and over-segmentation problems have been completely remedied.

The result of this initial segmentation is a set of segments, each of which corresponds to an active region on the solar disk. Each segment is labelled, and its centroid is calculated for use as a seed in the region growing procedure. Prior to this, the intensity at the seed location is checked in order to ensure that it is inside the region and, if not, its position is adjusted. In this case a new seed is selected as illustrated in Figure 4 by investigating pixel values in the eight nearest neighbour directions in order until a new seed is found inside the region from which the region growing procedure can start.

2.3   The Region growing

Region growing is a basic procedure that starts with a set of seed pixels. The aim is to grow a uniform and connected region from each seed. A pixel is added to a growing region if and only if:
- It has not been assigned to another region
- It is a neighbour of the growing region
- A new region created by the addition of a new pixel is still uniform

The algorithm used for the region growing takes as input a clean image and a set of points from which the region-growing process starts. The algorithm begins at each starting pixel and scans the neighbouring 8 pixels in a circular fashion, to determine membership of the region around the central pixel that fall within specified constraints.

Two forms of constraint have been considered. The first uses a fixed threshold range (with an upper and lower pixel value) and second uses a variable threshold range set to a multiple of the standard deviation of the current region pixel values. After experimentation the fixed threshold range was chosen as it gives more accurate control in defining the outer boundaries of regions while also reducing the occurrence of holes in regions. Figure 6 shows the two detected active region grown by using the two types of constraint. Using upper and lower threshold values, the regions grown include all 8-connected neighbouring pixels that fall within the given threshold range. This is not the case using variable threshold rang.

The upper and lower threshold values are determined by exploiting the statistical properties of the locally homogeneous background regions. As illustrated in the pseudo code below the lower threshold value is defined as $\mu - 0.3\sigma$ (where μ is the mean and σ is the standard deviation of that region) and the upper threshold is set to the maximum intensity value of the image. As pixels are added, the process repeats with the newly added pixels as the centre pixels. A list of tested pixels is maintained, in order to avoid unnecessary duplication of the tests. In this way, the region is constructed by using the membership criteria already discussed.
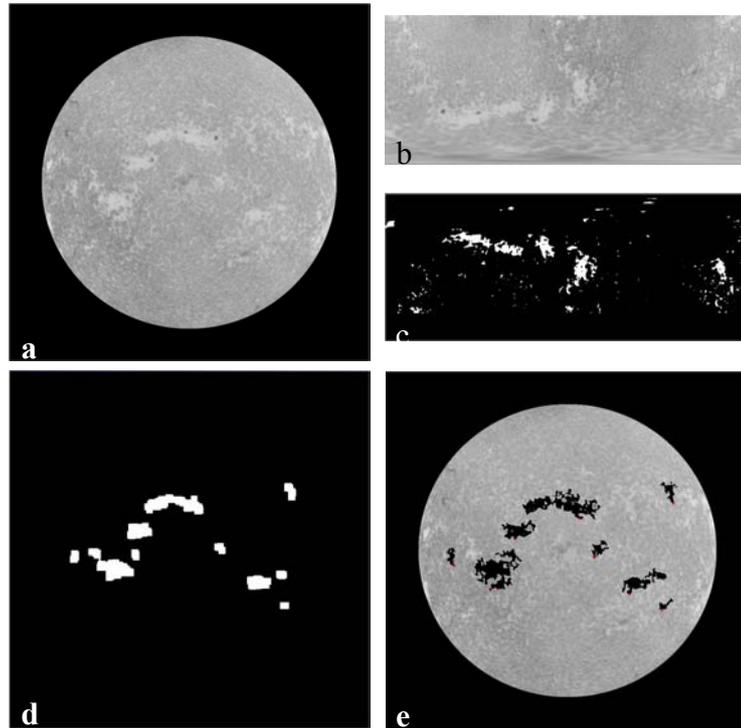
Figure 2: The segmentation procedure stages: a) an original Ca KII 3 image; b) after a transformation to Polar coordinates; c) after an initial thresholding; d) after a transformation back to the Cartesian coordinates, cleaning and morphological processing and e) a final result of the region growing.
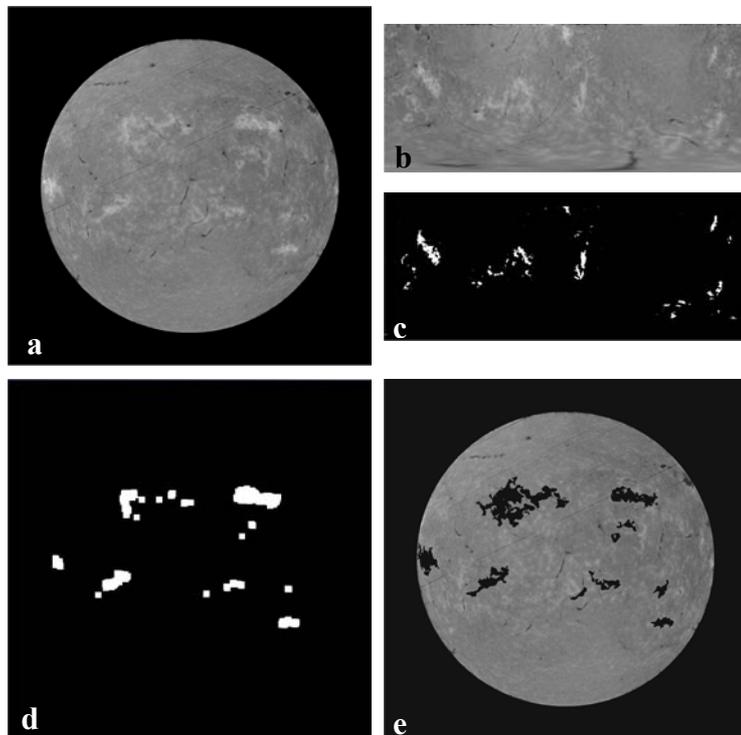

Figure 3: The segmentation procedure stages: a) an original Hα image; b) after a transformation to Polar coordinates; c) after an initial thresholding; d) after a transformation back to the Cartesian coordinates, cleaning and morphological processing and e) a final result of the region growing.

The following pseudo code defines the region growing algorithm:

Let f be an image, and $R_1, R_2,...R_n$ a set of regions each consisting of a single seed pixel.

Repeat
for i=1...n
 for each pixel p at the border of R_i
  for all neighbours of p
   Let x,y be the neighbour's coordinates
   Let lo_i be the lowest grey level of pixels in R_i
    $lo_i = \mu_i - 0.3 \times \sigma_i$
   Let hi_i be the highest grey level of pixels in R_i
   $hi_i = max(f)$
   if neighbour unassigned and
   $(hi_i \geq f(x,y) \geq (lo_i + \Delta))$
   Add neighbour to R_i, update hi_i and lo_i
   Until no more pixels are being assigned to regions

If in the binary image there are more than one seed pixels obtained from an active region, the region growing method merges the detected pixels and forms a single active region area. Figures 2(e) and 3(e) show results of applying the region growing procedure.
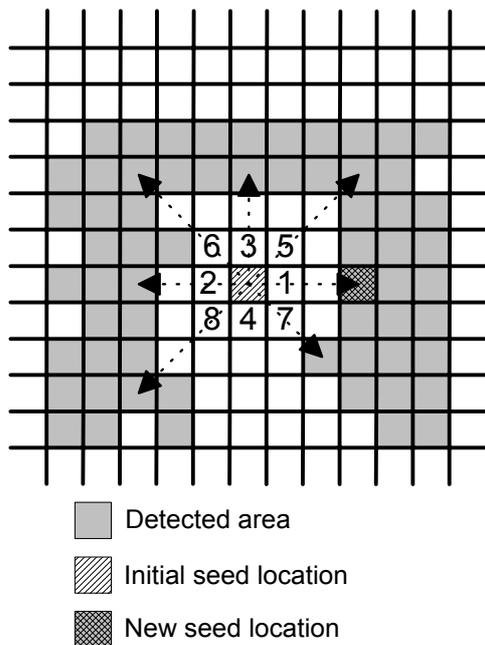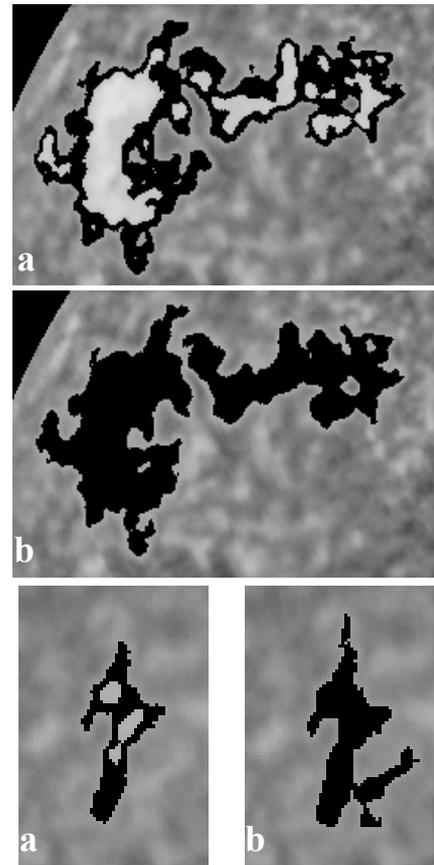

Figure 5: Results of region growing using the two forms of constraint: a) multiple of the standard deviation and b) threshold range

2.4   The accuracy of the technique

These procedures have been tested on the synoptic (28 days of April 2002) image sequences of full-disk solar images from the Meudon observatory. A comparison of the quantitative results obtained with our techniques with those done manually at the Meudon observatory (Mouradian Z., 1998) shows that in Meudon they detect 50% more active regions on most days. For example, on the 07/04/2002, there was 22 active regions in Meudon results while our procedure detected only 13 active regions and the NOAA observatory showed only 12 active regions. This was the case over the whole month (April 2002) except for 28/04/2002, when in Meudon there were detected 7, we detected 7 and

Figure 4: Seed pixel location

NOAA detected 8 and 29/04/2002, when Meudon detected 6, we detected 7 and NOAA detected 7. The reason for this is that in Meudon they detected the bright regions called plages, defined as the regions in the chromosphere that are brighter than the normal "quiet" Sun background, while NOAA detected the active region defined as a bright area on the Sun with a large concentration of magnetic field, often containing sunspots.

For further testing the results obtained using the Meudon images were compared and showed a reasonable agreement with those of the NOAA observatory (as illustrated in figure 6). The False Acceptance Rate (FAR) (where we detect an active region and they do not) and the False Rejection Rate (FRR) (where they detect an active region and we do not) were calculated on each day. In most cases there were more active regions detected by us than in NOAA with an average FAR of 1.8 per day. The FRR was very low, at 0.2 with only 5 days when we failed to detect region detected by NOAA. In some cases, and as shown in Figure 7, we detect an active region while NOAA splits it into two regions that does affect the quantitative comparison. In general, the agreement with NOAA is good, considering that NOAA bases its decisions on more information than we do in this stage

3   Conclusions

In this paper an efficient procedure of automated detection of solar active regions is presented. The statistically derived local intensity threshold values and region growing methods are used in the procedure. A local intensity threshold is applied to detect bright pixels that are considered to be the active region seed candidates. The subsequent region growing algorithm exploits statistical properties of the locally homogeneous

background regions in order to determine local upper and lower threshold values which control the spatial extent of the regions grown. The procedures used have demonstrated a satisfactory accuracy in the automated detection and segmentation of active regions from the Meudon Hα and Ca KII 3 full-disk solar images. Future work is aimed at processing coronal EUV solar images from SOHO/EIT and building pseudo 3D structure for the active regions from these three wavelengths (Hα, Ca KII 3 and SOHO/EIT).
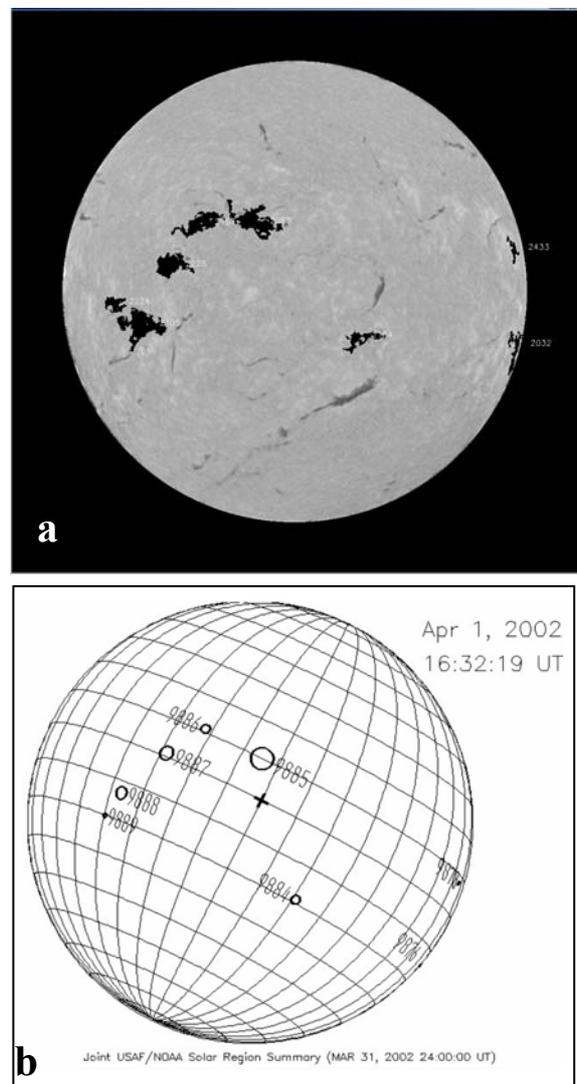

Figure 6: Active Region detection; a) Hα solar image from Meudon observatory c). NOAA observatory active region map on the same day 01/04/2002
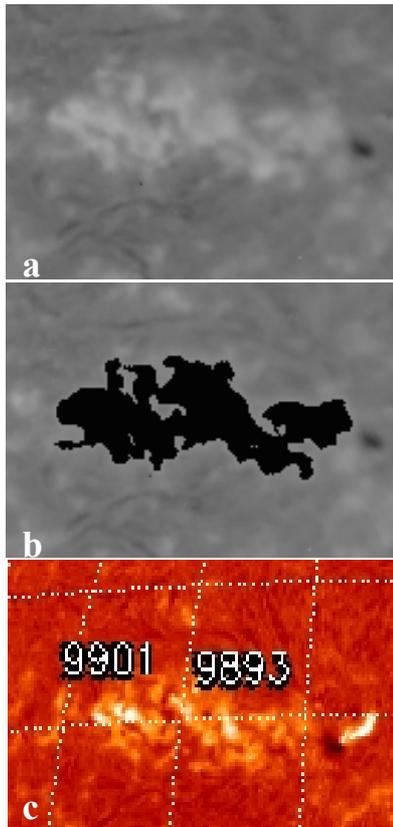
Figure 7: Two active regions merge: a) Meduon Hα input image, b) the same image after applying region growing and c) Big Bear observatory (BBSO) Hα image shown the same region as divided to two active regions

References

Bentley, R.D. EGSO - the next step in data analysis. *in Proceedings of the Second Solar Cycle and Space Weather Euroconference*, 24 - 29 September 2001, Vico Equense, Italy, Edited by Huguette Sawaya-Lacoste, ESA Publication SP-477, 2002.

Hill M, Castelli V, Chung-Sheng Li, Yuan-Chi Chang, Bergman L, Smith J.R. and Thompson B. Solarspire: querying temporal solar imagery by content. *International Conference on Image Processing*, Thessaloniki, Greece, 7-10 Oct. 2001, (1):834–837, 2001.

Mouradian Z., Synoptic Data Findings, *Synoptic Solar Physics ASP Conferences Series*, (140):181-204, 1998.

Steinegger, M and Brandt, P.N. On the determination of the quiet Sun centre-to-limb variation in Ca K spectroheliograms. *Solar Physic Journal,* (177):287-294, 1998.

Turmon**,** M., Pap J. M., and Mukhtar, S. Automatically Finding Solar Active Regions Using Soho/Mdi Photograms And Magnetograms, *Proc. SoHO 6/GONG 98 Workshop*, Structure and Dynamics of the Interior of the Sun and Sun-like Stars, Boston, 1998.

Veronig A., Steinegger M., Otruba W., Hanslmeier A., Messerotti M., and Temmer M. Automatic Image Processing In The Frame Of A Solar Flare Alerting System. *HOBUD7* 24(1):195-200, 2001.

Zharkova V.V., Ipson S. S, Zharkov S. I, Benkhalil A. K and Aboudarham J. A full disk image standardisation of the synoptic solar observations at the Meudon observatory. *Accepted for publication in the Solar Physics Journal.* 2002.