FILAMENT RECOGNITION AND IMAGE CLEANING ON MEUDON Hα SPECTROHELIOGRAMS

NICOLAS FULLER[1], JEAN ABOUDARHAM[1] and R. D. BENTLEY[2]

[1]*Observatoire de Paris, Section de Meudon, LESIA, F-92195 Meudon, France*
(e-mail: nicolas.fuller@obspm.fr, jean.aboudarham@obspm.fr)
[2]*MSSL, University College London, Hombury St. Mary, Dorking, Surrey RH5 6NT, U.K.*
(e-mail: rdb@mssl.ucl.ac.uk)

Abstract. This paper presents the techniques developed for the automated detection of filaments on Meudon Hα spectroheliograms, and, by extension, on any full-disk Hα Sun observations. Some cleaning processes are first applied to the images to correct them from defects characteristic of the instrument. Indeed, these defects may lead to spurious detections. From the cleaned images, filament areas are then segmented using a region growing method which efficiently returns the full extent of these dark areas. The filaments are finally described by means of their pruned skeleton. This representation allows one to compare the automatically segmented filaments with those manually recorded for Meudon Synoptic Maps. The very good agreement observed on a representative set of images confirms that this method can effectively be used in the frame of the EGSO (European Grid of Solar Observations) project in order to produce a reliable catalog dedicated to solar features.

1. Introduction

The increasing volumes of data in the field of solar astronomy makes it necessary to build powerful tools enabling access to a maximum amount of information from each site. The EGSO (European Grid of Solar Observations), as described in Bentley (2001) is based on the concept of a virtual solar observatory, providing integrated access to as many providers of solar observations as possible. EGSO has been designed to simplify the way solar data can be located and retrieved. One of its major components, the Solar Feature Catalog, will facilitate the use of features as a search criterion, rather than just date, time and location. The catalogue has been built from feature detection tools applied to a selected set of synoptic data, like the Hα images for filaments. Within the framework of the project, the automatic detection of sunspots and active region plages were already subject of publications by Zharkov *et al.* (2003) and Benkhalil *et al.* (2003).

This study is based on the images obtained by the spectroheliograph of Meudon in the Hα line, in particular because of their availability over a long period. Tests have also been carried out on Hα images from Big Bear Solar Observatory[1] (Denker *et al.*, 1999) and Yunnan Astronomical Observatory[2] (Steinegger *et al.*, 2001). The

[1]http://www.bbso.njit.edu
[2]http://www.ynao.ac.cn

ability to adapt the techniques to other data sets ensures the completeness of the catalogue in terms of continuity of the observations. The temporal coverage could also be increased by processing data from observatories at different longitudes.

The general definition of a prominence is an object in the chromosphere or corona that is denser and cooler than its surroundings (Tandberg-Hanssen, 1995). When seen against the intense Hα light of the disk, the prominences appear as dark features in absorption, i.e., filaments. The extraction process used in this work is therefore based on the segmentation of the darkest regions in the image. Even if filaments, especially quiescent ones, are commonly described as elongated features, their shape is not taken into account in this work. Indeed, the variety in filament forms due to the wide range of possible magnetic field structures and sources of condensed material (Gibson, 1973) prohibits any definitive postulate about their shapes, as long as completeness is a priority. The many blob-like filaments, or filament parts, observed in Hα could be missed if such a condition was applied. The counterpart is that a small number of sunspots are also falsely detected and have to be removed using criteria other than morphology.

Such automated detection methods have already been developed in order to follow the evolution of a filament in time. Gao, Wang, and Zhou (2002) used it to detect filament disappearances, and Collin and Nesme-Ribes (1992) to investigate solar surface rotation. To extract the filaments they use either thresholding and region-based techniques and particularly region growing. In these studies the simple matching of a filament from one observation to another can be sufficient and the exact shape of the filament is not required. More recently, Wagstaff *et al.* (2003, private communication) presented the results of a region-based method to detect both filaments and sigmoids, Shih and Kowalski (2003) described a method based on directional filtering to identify thin and elongated objects (the features obtained are then considered as seeds for a region growing operation) and Zharkova and Schetinin (2003) explained another method using an artificial neural network.

The goal is to build a catalog describing the data as precisely as possible, in terms of accuracy and completeness. While the catalogue is valuable in itself, it must also demonstrate a favorable comparison with features determined by visual detection. In Section 2, we describe the corrections made to the images in order to avoid, as far as possible, false detections. The region growing process is the subject of Section 3, and Section 4 introduces the morphological operators used to describe the segmented shapes. The skeleton representation allows us to retrieve descriptive parameters, and also to compare our results with those manually obtained in Meudon in the form of synoptic maps – these are the subjects of Section 5.

2. Need for a Cleaning Process

Prior to any processing, the images are standardized using dedicated software also written within the framework of the EGSO project. These preliminary steps are
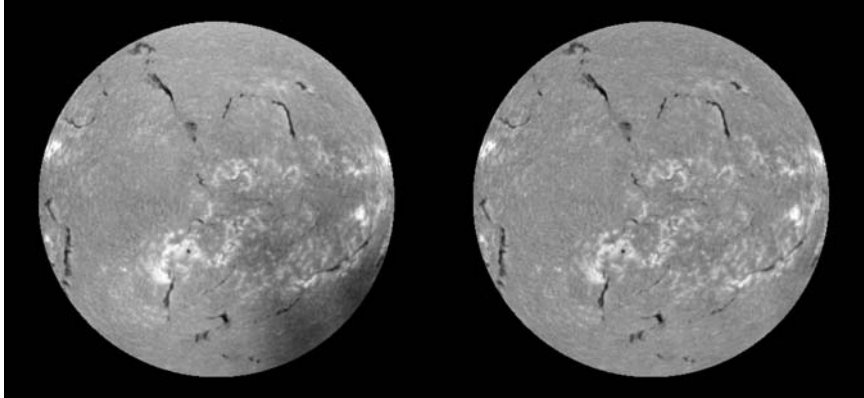
Figure 1. Meudon Hα spectroheliogram before and after intensity equalization (4th of January 2001 14:28:00 UT).

detailed in Zharkova *et al.* (2003). Several geometric and photometric issues are considered: ellipticity, centering, center-to-limb darkening, etc.

As input, preprocessed images available in FITS format are used – these typically have a size of 1024 by 1024 pixels. Working with these images, it became evident that additional corrections would reduce the spurious detections and enhance the segmentation process. Specific defects can appear on a restricted, but significant, set of images. For example, the cloud cover can locally darken the disk (see Figure 1). Likewise, the variation of the seeing (atmospheric turbulence) will produce a decrease in image sharpness (see Section 3.1). Factors independent of the weather conditions can also alter the image. Dust particles on the optical components of the instrument result in dark lines across the disk that are particularly difficult to differentiate from the filaments, especially if they are superimposed on them. Another defect comes from the use of photographic plates where dust can settle speckling the image with dark points – this last issue can easily be dealt with median filtering.

Below, we summarize how the intensity is normalized to remove local darkenings, and the process used to suppress dark lines.

2.1. INTENSITY NORMALIZATION

To optimize conditions for the region growing method (see 3.2), and because some filaments can be very long, the background intensity should be as flat as possible. To correct the image, an approximation of its large scale fluctuations is subtracted from it. A first estimate of the background is obtained with a large median filter. It is then refined with a smaller median filter applied to the intermediate flattened image, from which dark and bright solar features were roughly removed. An example is given in Figure 1. One should note that this method only gives a rough estimate of the effective background, only suitable for segmentation purposes.

2.2. REMOVAL OF DUST LINES

Images corrupted by dust on the optics show lines in one predominant direction. This makes it possible to develop a simple method to detect them, the main advantage of which, when compared to the Hough transform (Ballester, 1994), is that it is much faster. First a binary image is computed, setting as many as possible of the dark line pixels to 1 and the others to 0, thereby thresholding the image. The pixel discrimination is refined by applying a thinning operator to the binary image (Sonka, Hlavac, and Boyle, 1999). The thinning operator erodes large areas and only keep their skeletons, whereas the erosion of lines will be negligible. By summing pixel values of the thinned binary image in any direction from the disk center, the specific angle corresponding to the main orientation can be found as well as the line locations. The line pixel values are finally replaced by the mean value of a box centered on them.

3. Automated Segmentation

3.1. IMAGE ENHANCEMENT

Because an image can sometimes be more or less blurred, a sharpening spatial filter will improve the filament segmentation. Such a filter will better define the filament contours, and as a consequence increase the efficiency of the region growing algorithm. The best results were obtained using a second derivative filter like the Laplacian (Russ, 2002), whose mathematical formulation is given by

$$g(x, y) = f(x, y) - \nabla^2 f(x, y), \quad \text{where} \quad \nabla^2 f = \frac{\partial^2 f}{\partial x2} + \frac{\partial^2 f}{\partial y2}.$$

Its digital implementation consists of subtracting the image convolved with a suitable mask from the original image. The chosen mask is shown in Figure 2, together
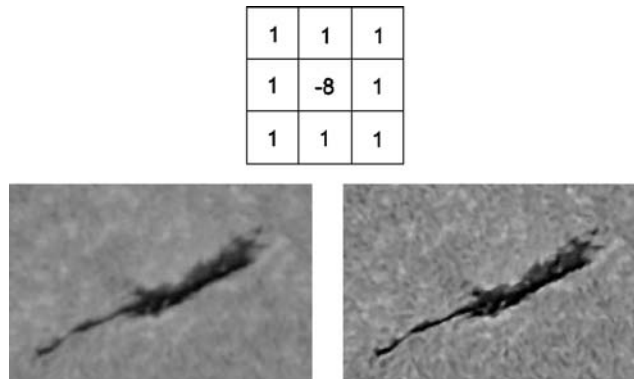
Figure 2. Digital Laplacian mask and result of the sharpening spatial filter on a filament seen on 1st of April 2001.

with an example for a single filament. As it highlights fine details, this filter facilitates the detection of the filament's thinnest parts.

3.2. REGION GROWING

The region growing procedure (Gonzalez and Woods, 2002) consists of grouping pixels together by appending similar pixels to a seed region. The seed can be either a single pixel or a group of contiguous pixels within the region of interest and may be chosen by the user or by means of an automated process. In this work, the seed pixels are selected by applying an appropriate threshold to the image in order to extract the darkest ones. Since the objective is to try to segment dark areas in the image, a similarity criterion based on a pixel's gray level value is necessary to append neighboring pixels to the seed pixels. Once the whole image has been processed, every pixel belonging to a filament is set to 1, and otherwise to 0.

In the following section, the way the seeds are chosen and the intensity criteria defined are described more precisely.

3.2.1. *Detection of Seeds*

To find the seed regions which are to be grown, a global thresholding of the image would be inefficient in many cases, giving many spurious seeds, even using the intensity normalization described earlier. This could be avoided with a higher threshold, but at the risk of missing important features. To overcome this problem, the image is divided into local square windows. Initially, window pixels whose values deviate from the mean are discarded from the calculation and a local quiet Sun intensity, QS_{win}, is computed. Then, from the same set of pixels, the standard deviation σ_{win} is calculated and the threshold for each sub-image is given by

$$T_{\text{win}} = QS_{\text{win}} - \alpha \times \sigma_{\text{win}},$$

where α is a constant. The chosen window size must be small enough to properly match standard deviation variations over the disk, but large enough to give significant statistics. After investigating a large number of Meudon images, the window size was set to 1/6 the image size, and α set to 3.7. Some spurious seeds are still found with these threshold values, but they are reduced in size (a few pixels) and can be removed, for example, by only keeping seeds whose size is greater than 20 pixels (see Figure 3b).

3.2.2. *Growing Process*

The next step is to grow the seeds according to the statistics of the local pixels. The constraint on the threshold is less stringent here because there is also a spatial requirement: the pixel values must be within a range defined by the mean and standard deviation of their neighborhood, *and* also have to be connected to the seed region. A bounding rectangle is defined from the location and size of the seed. Quiet Sun statistical parameters are calculated within the rectangle QS_{br}, σ_{br}, where
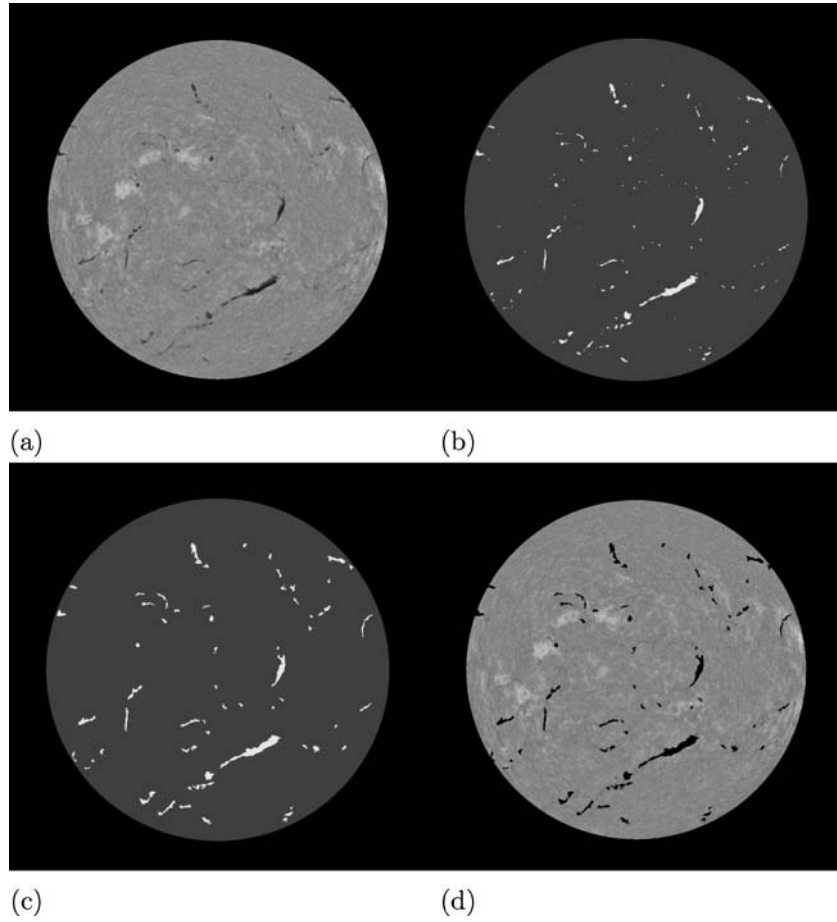
Figure 3. Segmentation results for the 1st of April 2001, 08:45:00 UT: (a) original Hα image from Meudon, (b) seeds detected, (c) region growing results and (d) original and results superimposed.

br stands for the bounding rectangle. The upper and lower intensity limits for the growing process of each seed are then given by

$$T_{\text{upp}} = Q S_{\text{br}} - \beta \times \sigma_{\text{br}} \, (\beta < \alpha),$$
$$T_{\text{low}} = 0.$$

For a set of images, β is set to 1.5. An example of the segmentation results with an image from Meudon is given in Figure 3 and another one with an image from BBSO is shown in Figure 4, where all filament pixels are set to 2, other Sun pixels to 1 and background pixels to 0.
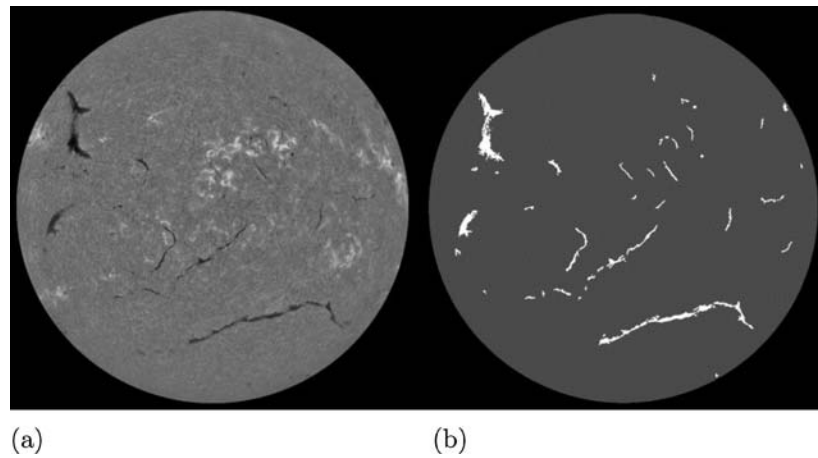
Figure 4. Segmentation results for the 30th of April 2001, 16:55:55 UT: (a) original Hα image from BBSO and (b) region growing results.
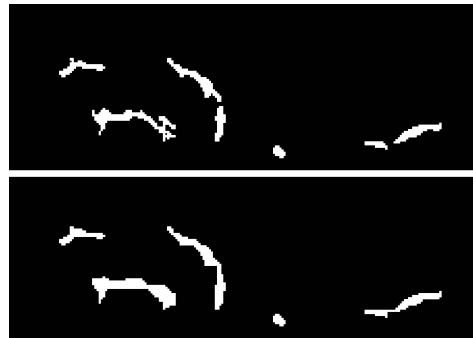

Figure 5. The effect of the morphological closing operator on a set of segmented filaments (1st of April 2001).

3.3. MORPHOLOGICAL CLOSING

The last step of the segmentation process consists in applying a morphological operator to the binary image. A closing operator makes the segmented regions more blob-like, by filling small lakes and long, thin gulfs. It will also merge close regions which could be considered as a single filament by blocking up the narrow channels between them (Serra, 1988). Figure 5 gives an example of the closing operator effect, which makes the filaments easier to describe.

4. Shape Representation: the Skeleton

The final goal of this work is to extract information from the detected filaments in order to provide entries for the Solar Feature Catalog. To this end, descriptors

such as length, location, mean intensity, perimeter, etc. need to be retrieved from the segmented regions. Some can easily be found from the segmented set of pixels, like the perimeter or the area. For others, we first need to shrink the shape to a simpler representation. Given the pruned skeleton of the region, we reduce it to a line. It is then easy to retrieve the length and the center of the filament, and compute other descriptors, e.g., the curvature. To produce the best pruned skeleton attention must be paid to two particular constraints. Firstly, all skeleton parts should be connected. Second, the full extent of the region must be contained in the skeletonized representation, to reproduce the filament's length as accurately as possible.

In the sections below, the algorithms used to fulfill these requirements are described.

4.1. THINNING METHOD

The thinning process is based on the HitOrMiss transform which is a combination of erosion and dilation operators that allows one to match a foreground pixel according to a predefined structuring element (see Sonka, Hlavac, and Boyle, 1999). Such structuring elements are shown in Figure 6a where 1 and 0 values correspond respectively to foreground and background pixels (blanks correspond to "don't care" values). The thinning of array A by the structuring element S can be written: $\text{Thin}(A, S) = A - \text{HitOrMiss}(A, S)$. In the thinning process, every matched pixel is set to 0 when this logical difference is computed ($X - Y = X \cap Y^c$). Given a set of structuring elements S, this operation is successively applied for each element S_n, and the whole process is iteratively repeated until there are no more pixels to remove from the foreground regions.

4.2. NUMERICAL IMPLEMENTATION

A succession of convolutions of the binary image with the kernels corresponding to the chosen structuring elements are then applied to the segmented pixels. Figure 6b
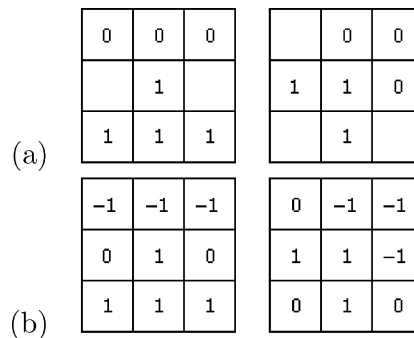
(a)

0	0	0
	1	
1	1	1

	0	0
1	1	0
	1	

(b)

−1	−1	−1
0	1	0
1	1	1

0	−1	−1
1	1	−1
0	1	0

Figure 6. (a) First two HitOrMiss structuring elements and (b) corresponding kernels.
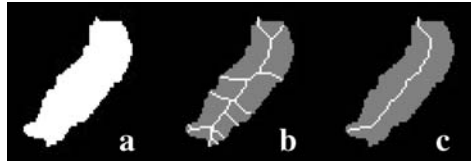
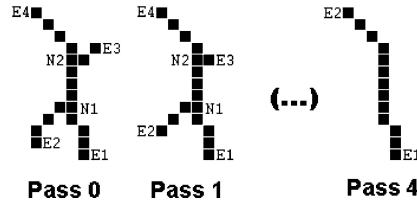*Figure 7.* Thinning and pruning results: (a) original shape, (b) full skeleton, and (c) pruned skeleton.

Figure 8. Pruning method iterations.

gives the first two kernels; six others are given by all the 90° rotations of them. This means that each pass consists of eight convolutions. A pixel is set to 0 if the convolution result is equal to the sum of elements set to one in each array (in this case 4 and 3, respectively). When the number of foreground pixels is the same before and after a pass (convergence), the process is stopped. This procedure produces the skeleton of the region (Figure 7b).

4.3. PRUNING

Removing branches of the skeleton tree, i.e. pruning, can easily be computed by iteratively detecting and removing the end points of the tree, until there are only two of them left. This method successfully removes short spurs from the skeleton but it can also discard end points of the skeleton main path, especially when branches are long. The main path is required to maintain its full length, i.e. to connect two-pixel positions on the boundary of the original shape. To compute it, an algorithm based on the Euclidean distance between the end points of the skeleton and its nodes is used. For each node N, the distance between the node and each end point E is calculated (see Figure 8). The end point with the minimum distance is then removed. By this technique, only the longest of all the branches starting from a given node are kept. This process is done iteratively until there are no more nodes, i.e. only two end points left. Figure 7 shows the result of the whole process, and Figure 9 the superimposition of a skeleton on the original image.

5. Comparison with Manual Digitization

Comparing the original images with the resulting segmentations, it can be seen that there is good agreement with the filaments. Nevertheless, a more precise
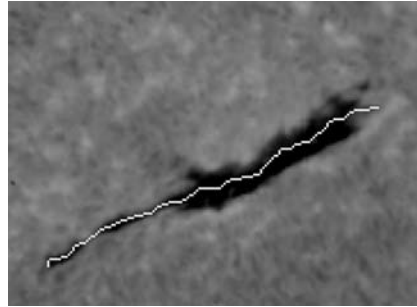
Figure 9. Filament skeleton superimposed on the original image (1st of April 2001).

criterion is needed to get a reasonable idea of the efficiency of the automated detection.

Manual digitized images from Meudon were readily available to the authors – these are used to draw the synoptic maps synthesizing the location of the filaments over one Carrington rotation (Mouradian, 1998). The maps and daily digitizations are freely available via the Bass2000[3] database and it is interesting to compare the results of this work with what can be done by the human eye.

5.1. METHODOLOGY

For each day in April 2002, the filaments automatically and visually detected are compared. To make it clearer, the following quantities were defined: the FPC (false positive count) corresponds to the filaments which were detected with the automatic process but do not appear on the synoptic maps, and the FNC (false negative count) corresponds to the filaments which were not detected by the automatic process but appear on the synoptic maps, i.e. missed ones. The matched filaments are those detected with both methods. The following rules are applied:

1. The FPC is given only for filaments whose length is greater than or equal to $5°$.
2. If one filament appears to be split in two or more parts in one detection, and not in the other, the lengths of the different parts are added, and the fragmented filament is considered as a single one.

The first rule is set because the smallest filaments are usually neglected in the manual process and also because the automated process may also detect a small number of sunspots, which would then be misidentified as small filaments. We thus start from the same set of detectable features. The second rule is set to compare the lengths of the filaments with a reliable statistic.

[3]http://bass2000.obspm.fr/home.php

TABLE I

Cumulated results for April 2002 (see text for description).

	Synoptic	Automatic	FPC	FNC
Count	1232	1149	51	134
Length	10719	11084	438	823

5.2. RESULTS OF THE COMPARISON

Table I shows the results of the comparison for the month of April 2002 (only three missing observations). In the first row, the first column gives the total number of filaments which were manually recorded (synoptic maps). The second one gives the number of filaments automatically detected and columns three and four respectively give the FPC and the FNC. The second row gives the corresponding cumulated length in degrees.

Checking the filaments corresponding to the FPC, it was found that these were quite often real filaments, but very faint or short ones that were not recorded on the synoptic maps. This can be verified from Figure 10 where these filaments are indicated with black squares for the day with the maximum FPC (6th of April).

An important value here is the FNC, which depicts the efficiency of the automated method compared to the manual one, taken as a reference. For the whole month we get an average of 4.9 false negatives a day. This means that 10.9% of the filament *population* were missed. The same calculation with length instead of count only gives 7.6%, equivalent to 7.6% of the filament *material*. However, looking at
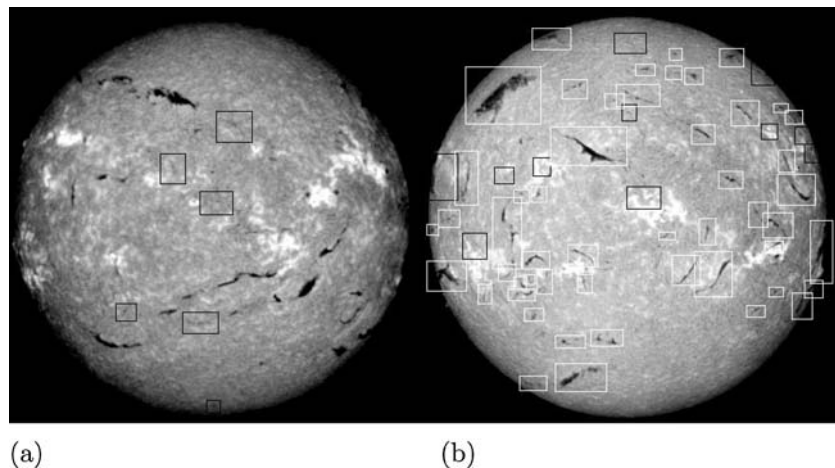

Figure 10. (a) False positive filaments for the 6th of April 2002 and (b) false negative filaments (*black squares*) for the 18th of April 2002 (*white squares* are matched filaments).

Figure 10, showing the observation for the 18th of April (maximum FNC), it can be seen no important filaments are missed, except one on the east (which is too close to the limb to be detected).

The main limit of the automatic detection described here, compared to the visual detection, is that it does not include some additional objective and subjective criteria. Tracking a filament from day to day is one of these objective criteria, which could make the operator to *see* a filament (or a part of a filament) whereas it is too faint to be automatically detected. However, Figures 3, 4 and 10 illustrate quite well the global efficiency of the method.

6. Conclusion

Starting from standardized Hα images, the goal of this work was to detect and describe filaments as precisely as possible. Specific tools to achieve this task were developed. First, by removing defects from the images using median filtering, advanced thresholding or morphological operators, it was ensured that the final set of detected features is as free as possible from spurious detections. Then, working with well established segmentation methods, the filaments were exhaustively detected (as compared to the manual detection). Finally, using morphological operators, skeletons were obtained that accurately describe the shape of the filament through a small set of pixels, which can be reduced to a chain code for storage purpose.

As the segmentation approach is not based on any assumption of the shape, all dark feature are detected, not just long and thin ones. This ensures that blob-like shapes that can also be part of the filament material are not missed, but a counterpart to this is that a few sunspots which are visible in Hα are also (falsely) detected as filaments. These could be removed in a further step, by combining the results with those obtained for sunspots at other wavelengths (Zharkov *et al.*, 2003). Time tracking of the features might also be a way to enhance the detection. It could be helpful to find very faint filaments which can be more easily detected in previous or later data. Nevertheless, one should keep in mind that the image quality remains the main factor leading to an efficient segmentation. These points, together with the extension of this work to limb prominences, will be the subject of future developments in the frame of the EGSO project.

Acknowledgements

References

Ballester, P.: 1994, *Astron. Astrophys.* **286**, 1011.

Benkhalil, A. K., Zharkova, V. V., Ipson, S. S., and Zharkov, S. I.: 2003, in *Proceedings of the AISB'03 Symposium on Biologically-Inspired Machine Vision, Theory and Application*, University of Wales, Aberystwyth 7–11th April 2003, Proceeding ISBN 1-902956-33-1, p. 66.

Bentley, R. D.: 2001, in *Proceedings of the 2nd Solar Cycle and Space Weather Euro-Conference, Vico Equense*, September 2001, ESA Publications, ESA **SP-477**, 603.

Collin, B. and Nesme-Ribes, E.: 1992, in *Proceedings of an ESA workshop on Solar Physics and Astrophysics at Interferometric Resolution*, Paris, 17–19 February 1992, p. 145.

Denker, C., Johannesson, A., Marquette, W. H., Goode, P. R., Wang, H., and Zirin, H.: 1999, *Solar Phys.* **184**, 87.

Gao, J., Wang, H., and Zhou, M.: 2002, *Solar Phys.* **205**, 93.

Gibson, E. G.: 1973, *The Quiet Sun*, NASA **SP-303**, 49.

Gonzalez, R. C. and Woods, R. E.: 2002, *Digital Image Processing*, 2nd edn., Prentice-Hall, Inc., Upper Saddle River, New Jersey, p. 613.

Mouradian, Z.: 1998, in *Synoptic Solar Physics*, ASP Conference Series **140**, 197.

Russ, J. C.: 2002, *The Image Processing Handbook*, 4th edn., CRC Press, p. 219.

Serra, J.: 1988, *Image Analysis and Mathematical Morphology*, Vol. 1, Academic Press Limited, London, p. 52.

Shih, F. Y. and Kowalski, A. J.: 2003, *Solar Phys.* **218**, 99.

Sonka, M., Hlavac, V., and Boyle, R.: 1999, *Image Processing Analysis and Machine Vision*, 2nd edn., PWS Publishing, p. 578.

Steinegger, M., Denker, C., Goode, P. R., Marquette, W. H., Varsik, J., Wang, H., Otruba, W., Freislich, H., Hanslmeier, A., Luo, G., Chen, D., and Zhang, Q.: 2001, in *Proceedings of the First SOLSP Euro-Conference 'The Solar Cycle and Terrestrial Climate'*, ESA **SP-463**, 617.

Tandberg-Hanssen, E.: 1995, *The Nature of Solar Prominences*, Kluwer Academic Publishers, Dordrecht, p. 3.

Wagstaff, K., Rust, D. M., LaBonte, B. J., and Bernasconi, P. N.: 2003, *Solar Image Recognition Workshop*, Brussels, 23–24 October 2003.

Zharkov, S. I., Zharkova, V. V., Ipson, S. S., and Benkhalil, A. K.: 2003, in *Proceedings of the AISB'03 Symposium on Biologically-Inspired Machine Vision, Theory and Application*, University of Wales, Aberystwyth 7–11 April 2003, Proceeding ISBN 1-902956-33-1, p. 74.

Zharkova, V. V., Ipson, S. S., Zharkov, S. I., Benkhalil, A. K., Aboudarham, J., and Bentley, R. D.: 2003, *Solar Phys.* **214**, 89.

Zharkova, V. V. and Schetinin, V.: 2003, in *Proceedings of the Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems KES'03, Part I*, Proceeding ISBN 3-540-40804-5, p. 148.