

Automatic Detection of Solar Filaments versus Manual Digitization

N. Fuller and J. Abouadarham

LESIA - Observatoire de Paris-Meudon, 5 place Jules Janssen, 92190 Meudon, France

[Fuller, Abouadarham}@obspm.fr](mailto:{Fuller, Abouadarham}@obspm.fr)

Abstract. The technique which is presented here is based on careful attention to image cleaning in order to achieve robust automatic filament detection and avoid confusion between filaments and image defects, due to particular shapes of filaments. The main part of the detection process is based on seed selection and region growing. The procedures developed have been tested on four months of full-disk images from the Meudon Observatory. The results are compared with those manually generated in Meudon, for several hundred filaments. A very good correspondence is found, showing the robustness of the method described.

1 Introduction

Full disk solar images (spectroheliograms) are obtained at the Paris-Meudon observatory every day, weather permitting, (approximately 300 days per year) for several wavelengths. The spectroheliograms in which the filaments are best seen are taken in the H α transition of Hydrogen (656.3 nm). Filaments appear as dark features on the solar disk, often very elongated and sometimes more compact (see figure 1). The monitoring of filaments is important in Solar physics because they are indicators of the Solar activity and also can be precursors of very energetic phenomena such as flares or coronal mass ejections which are very important in the frame of Solar-Terrestrial relationship.

Solar filaments have been manually detected for dozens of years from these daily observations. The increasing power of computers and the evolution of detection techniques now offer an opportunity to develop automatic detection codes. This has been done in the frame of the European Grid of Solar Observations (EGSO) [2], a project funded by the European Commission within the Information Society Technologies thematic priority of the Fifth Framework Program (IST-2001-32409).

Such automated detection methods have already been developed in order to track filaments in time. Gao, Wang and Zhou [4] used thresholding and region-based techniques, like region growing, to detect filament disappearances. Collin and Nesme-Ribes [3] used similar techniques to investigate the rotation of the Sun's surface. More recently, Shih and Kowalski [7] proposed a method based on directional filtering to identify thin and elongated objects and Wagstaff *et al.* [9] a region based

method to detect both filaments and sigmoids. All these methods give good results, but sometimes neglect the smallest, weakest or blob-like filaments. Our goal is to extract the most complete possible set of filaments from one observation, and we thus can't make any postulate about the features shapes. As region growing has proved to be a reliable means to investigate such features, we have based our method on it, improving the way seeds and thresholds are chosen.

Prior to any feature recognition, the solar images have to be pre-processed in order to correct them for geometrical or photometric distortions. These preliminary steps are described in Zharkova et al. [10]. Correction of ellipticity, centering the disk in the image, removing center-to-limb darkening, etc. all make the images more suitable for further processing. These standardized images are further enhanced by removing defects specific to the measurement conditions such as non-geometrical darkening due to clouds during the observation or dark lines across the disk due to dust particles. We present here techniques based on median filtering and on the Hough transform for removing them.

We finally compare our results to those manually obtained and show that the automated techniques developed give very good results on a large set of images, even if subjectivity is not included in our work.

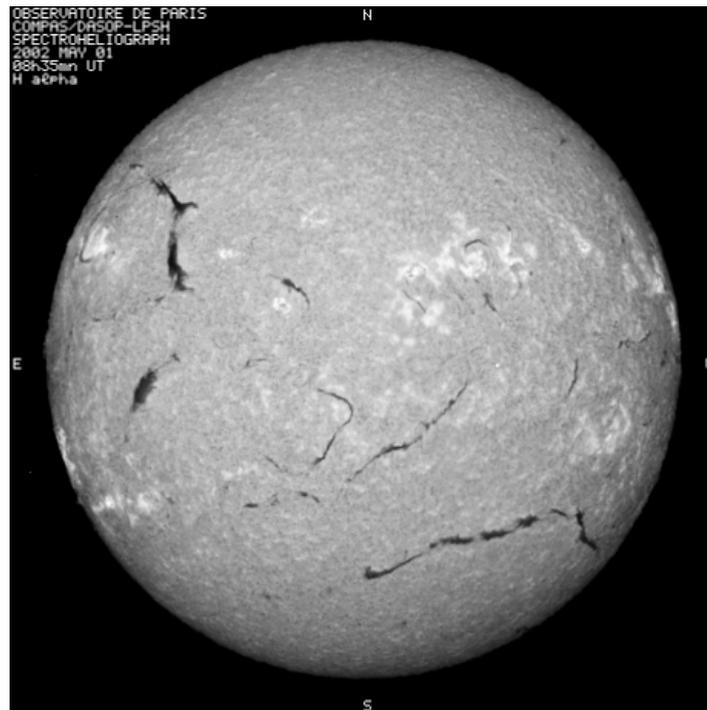


Fig. 1. Meudon spectroheliogram showing several filaments (dark features)

2 Image cleaning

2.1 Intensity Normalization

The first cleaning process we use consists of normalizing the intensity over the disk. As we use a region growing method (see section 3) to segment the filaments, and because some of them can be very long, the variations of the background intensity ('Quiet Sun') should be as small as possible.

To identify the large scale variations of the intensity we first use median filtering with a large window. The median filter will eliminate the highest and lowest values (corresponding to bright plages and dark filaments) to give a first approximation of the background fluctuations which can then be subtracted from the original image. From this first normalized image we can more efficiently locate pixels corresponding to bright and dark regions with two suitable thresholds. These thresholds are obtained from the new image histogram, and the pixels values are replaced by their corresponding values in the background map. By applying median filtering again (with a smaller window) we then get a more reliable estimate of the large scale variations. The following pseudo code contains the details of the algorithm and figure 2 provides an example of the resulting intermediate images:

```
begin
Rescale I to a smaller size:  $I_{small}$  // saves computer time
 $B_{small} = \text{median}(I_{small}, W_{size})$  //  $W_{size}$  is the filter size
 $I'_{small} = I_{small} - B_{small} + \text{mean}(B_{small})$  //Subtract and get back to original intensity level
Hist = histogram( $I'_{small}$ )
 $H_M = \max(\text{Hist})$ 
Let  $V_M$  be the intensity value corresponding to  $H_M$ 
Let  $V_1$  be the intensity value corresponding to  $H_M/a_1$  ( $V_1 < V_M$ )
Let  $V_2$  be the intensity value corresponding to  $H_M/a_2$  ( $V_2 > V_M$ ) //  $a_1, a_2$  are constants
Let be S the set of pixels in  $I'_{small}$  lower than  $V_1$  and greater than  $V_2$ 
 $I'_{small}[S] = B_{small}[S]$ 
 $B'_{small} = \text{median}(I'_{small}, W_{size}/2)$ 
Rescale  $B'_{small}$  to original size:  $B'$ 
 $I_n = I - B' + \text{mean}(B')$  //  $I_n$ : normalized image
end
```

2.2 Dust lines removal

As spectroheliograms are generated by scanning the disk horizontally, any dust grain on the entrance slit leads to a dark straight line on the image. These lines are

particularly difficult to differentiate from the filaments, especially if they overlap, and we thus need to remove them.

A binary image is first computed from the normalized image and its cumulative histogram used to control the ratio between the non-zero and zero pixels. A morphological thinning operator is then applied to thin the features with contiguous non-zero pixels to a thickness of one pixel. The Hough transform [1] of the thinned image is then thresholded at a value corresponding to the minimum number of points needed to identify a straight line. Finally the parameters of the identified straight lines give the locations of the lines of pixels whose values can now be corrected (see figure 3).

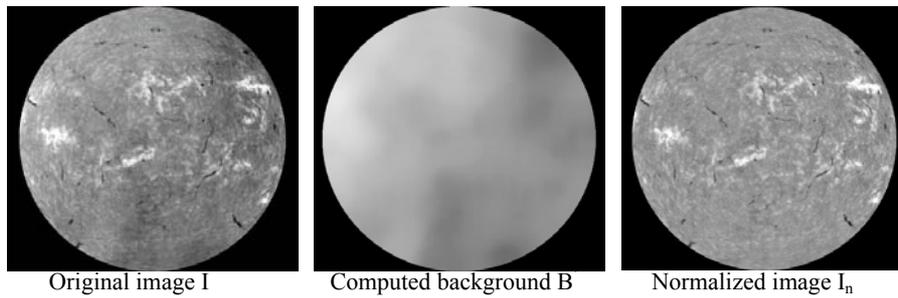


Fig. 2. Non-geometrical intensity normalization

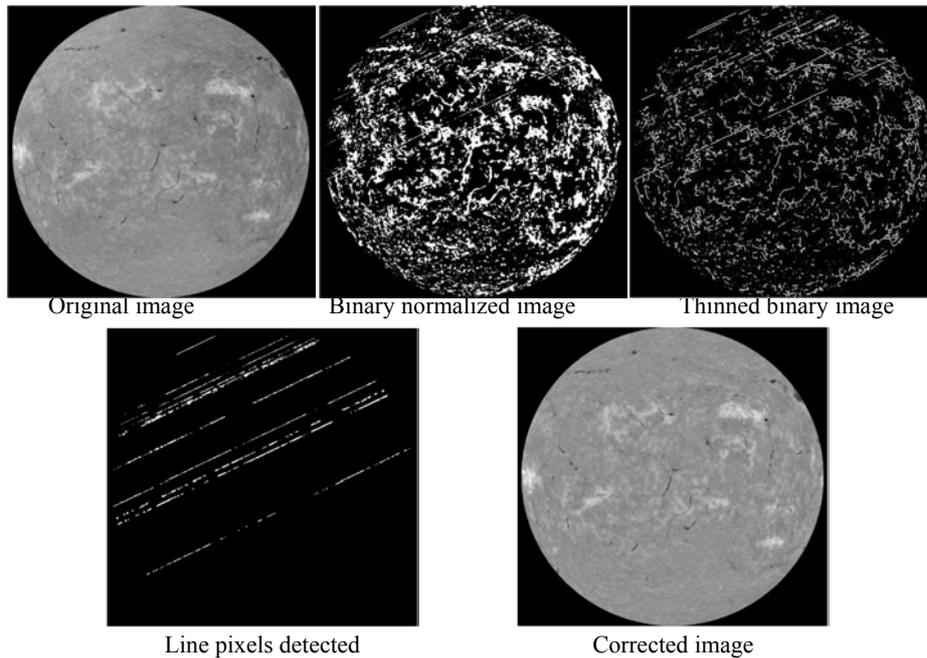


Fig. 3. Steps in the dust line removal procedure.

2.3 Sharpness enhancement

In order to better define filaments contours and to detect the thinnest parts of the filaments more efficiently (see figure 4) we enhance the image sharpness using the Laplacian method [6].

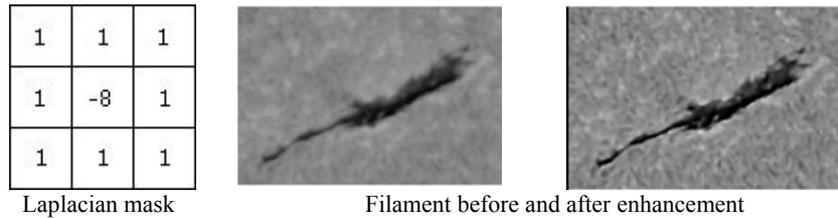


Fig. 4. Laplacian mask and the enhancement effect on filament's contrast

This sharpening is done by subtracting a Laplacian filtered version of the image from the original image. The Laplacian of an image is obtained by computing its second order derivative. The digital equivalent is to convolve the original image with a suitable mask (such a mask is shown in figure 4). Then, if A is the original image and B the result, the formulation is:

$$\begin{aligned} &\text{For each pixel } (i,j), \\ &B(i,j) = A(i,j) - A(i-1,j-1) - A(i,j-1) - A(i+1,j-1) - A(i-1,j) - A(i+1,j) - \\ &A(i-1,j+1) - A(i,j+1) - A(i+1,j+1) + 8*A(i,j) \end{aligned}$$

3 Filament detection

Once the image is fully cleaned, we can investigate the filaments using a region growing method based on the image grey level properties. The principle is to group pixels into larger regions if these pixels fall into a predefined intensity range. The procedure is started from a pixel or small region called a seed. This method is much more effective than applying a basic automatic threshold as it associates a grey level condition with a connectivity condition. The efficiency of the method will thus depend on the seed selection procedure and on the intensity range definition.

3.1 Seeds selection

As we said above, the seed selection is a major step in the procedure. Getting too many of them might lead to false detections. Alternatively, missing some would imply missing filaments. To select the seeds as precisely as possible we use a windowed threshold. In the first stage, pixels whose values are distant from the mean are dis-

carded from the calculation of a second mean M_w , which better reflects the ‘Quiet Sun’ intensity within each window (i.e. mean without brightest and darkest regions). Then the standard deviation σ_w is computed from the same set of pixels and the window threshold is given by:

$$T_w = M_w - \alpha_1 \times \sigma_w$$

[3] where α_1 is a constant obtained after testing a large set of images. All pixels with values lower than T_w are set to 1 and the rest are set to 0.

3.2 Region growing

The next major step consists in growing the seed sub-regions into larger ones. Once again we have to be very careful when computing the intensity criteria. As we did for seeds, we use windows again to be sure to get the most reliable local statistics. These windows are centred on the seeds and their sizes depend on the seed dimensions in both O_x and O_y directions. The lowest value of the intensity range is set to 0 and the highest T_{max} is obtained in the same way that we computed it for seeds:

$$T_{max} = M_w - \alpha_2 \times \sigma_w \quad \text{with} \quad \alpha_2 < \alpha_1$$

Each pixel connected to a seed is then investigated. If its value is within the range $[0, T_{max}]$ it is appended to the seed. Figure 5 gives an example of the final segmentation for one observation. Note that a minimum region size is also defined to remove regions that did not grow large enough.

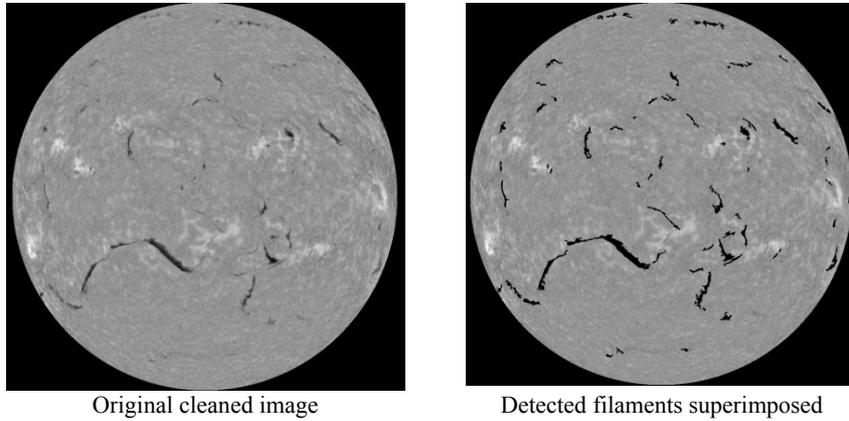


Fig. 5. Filament detection result

4 Comparison with manual detection

4.1 Shape description

In order to compare our results with what has been obtained manually we need to describe the filaments in a similar way. The manual digitization involves choosing a few points along the path of the filament and computing the full path by linking the points together. The closest way to represent our filaments is to compute their pruned skeletons. Such skeletons are the results of a thinning process based on the Hit-Or-Miss transform which is a combination of erosion and dilation operators that finds a match in foreground pixels according to a predefined structuring element [8]. 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_n , this operation is successively applied for each element S_n , and the whole process iteratively repeated until there are no more pixels to remove from the foreground regions (see figure 6).

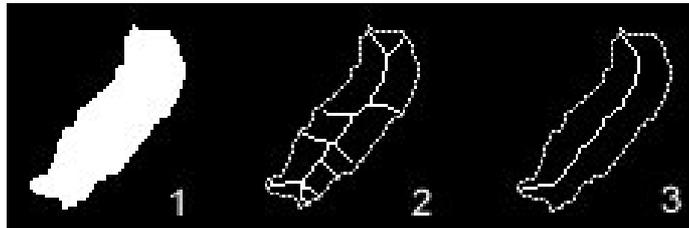


Fig. 6. Original shape (1), full skeleton (2) and pruned skeleton (3)

The algorithm used for pruning is based on the distance between node points in the skeleton tree and end of branches. From each node point, the closest end points are iteratively eliminated until there are only two end points left in the tree.

4.2 Comparison results

The procedure for automatic detection of filaments has been applied to the period between April and July 2002, on more than a hundred observations, because the filaments have already been manually detected, and this corresponds to a period where many (1265 manually detected) filaments were visible on the Sun. Examples of the manual detection can be seen on the Bass2000 web site (<http://bass2000.obspm.fr>).

Automatically and manually detected filaments (with size greater than 5°) have been compared one by one for the whole month of April, giving the following results:

- 89 % of automatically detected filaments match the manually detected ones
- 4 % of the 'automatic' filaments don't correspond to 'manual' ones. In fact, most of them have not been manually detected, but a careful inspection shows that they are real filaments but faint ones most of the time.
- 11 % of 'manual' filaments haven't been automatically detected. But the error in the total length of all filaments is only 7 % (i.e. non detected filaments were small ones). Again, when carefully looking at these filaments, it appears that these are also faint ones in general.

5 Conclusions

The complete comparison between manual and automatic detection of filaments shows that most of the differences between the two methods come from the subjective interpretation of the original image. The knowledge that a filament is often located at the border of bright regions helps the user to guess the presence of a filament. Moreover the knowledge that a filament was present on a previous day indicates that one may search for even a very thin one at the place it was previously seen.

The technique used here is robust for standard filament detection. It doesn't seem to be possible to improve it easily, taking into account the remarks above. Nevertheless the next step could be to teach the code how to guess whether a filament could be at some place or not. Neural networks could be well suited to this type of application.

References

- [1] Ballester P.: Hough transform for robust regression and automated detection, *Astron. Astrophys.* (1994) 286 1011.
- [2] Bentley R. D.: EGSO – the next step in data analysis, in Proceedings of the 2nd Solar Cycle and Space Weather Euro-Conference, Vico Equense, September, ESA SP-477, ESA Publications, (2001) 603.
- [3] Collin B. and Nesme-Ribes E.: Pattern recognition applied to H α spectroheliograms, Solar rotation and meridional circulation, in Proceedings of an ESA workshop on Solar Physics and Astrophysics at interferometric resolution, Paris, 17-19 Feb., (1992) 145.
- [4] Gao J., Wang H. and Zhou M.: 2002, *Solar Physics.* (2002) 205 93.
- [5] Gonzalez R. C. and Woods R. E.: Digital Image Processing, Second Edition, Prentice-Hall, Inc., Upper Saddle River, New Jersey, (2002) 613.
- [6] Russ, J. C.: The image Processing Handbook, Fourth Edition, CRC Press, (2002) 219.
- [7] Shih F. Y. and Kowalski A. J.: *Solar Physics.* (2003) 218 99.
- [8] Sonka M., Hlavac V. and Boyle R.: Image Processing Analysis and Machine Vision, Second Edition, PWS Publishing, (1999) 578.
- [9] Wagstaff K., Rust D.M., LaBonte B.J. and Bernasconi P.N.: Automated Detection and Characterization of Solar Filaments and Sigmoids, Solar Image Recognition Workshop Brussels, October 23-24, 2003.
- [10] Zharkova V.V, Ipson S.S., Zharkov S.I., Benkhalil A.K., Abouharham J. and Bentley R.D.: A full-disk image standardisation of the synoptic solar observations at the Meudon Observatory, *Solar Physics.* (2003) 214/1 89.