A need for Automated Tools for Extraction and Visualization of the data from Solar Orbiter

V. V. Zharkova

Department of Computing, Bradford University, Bradford BD7 1DP, UK

Abstract

The Solar Orbiter instruments are expected to provide a large amount of full disk solar images, which processing by individual users can delay the scientific advances of the mission. Recently developed automated techniques applied for detection of sunspots, active regions, filaments in full disk solar images in Ca II K1, Ca II K3 and Ha images (Meudon Observatory), EUV images (SOHO/EIT) and white light images (SOHO/MDI) revealed a good accuracy with the manual synoptic maps and NOAA data and can be used for the automated image standardization and feature extraction from the Solar Orbiter images. The extracted parameters of active features can be also automatically populated into the extended relational database of the Solar Feature Catalogues (SFCs) http://solar.inf.brad.ac.uk. In addition to the original images, this will allow delivering to the users the major activity features extracted with a sufficient accuracy and daily updating the fully digitized database of solar features, which continuation and consistency is valuable for the solar activity models.

Keywords: Sun: image standardization-- Sun: active regions -- Sun: Sunspots -- Sun: filaments-- Sun: automated recognition

1. Introduction

With a substantial increase in size of solar image data sets, the automated detection and verification of various features of interest is becoming increasingly important for higher value of the data representation to the users who can use it for a reliable forecast of the solar activity and space weather.

The archives from the Solar Orbiter mission will be contributing to a growing number of archives of digitized images of the Sun, taken from ground-based and space instruments in various wavelengths included into the unified catalogues by various grids (VSO -USA, EGSO -Europe, Astrogrid -UK). In addition, there is a new type of the archive, the Solar Feature Catalogues, that comprises of the solar activity features (sunspots, active regions and filaments) automatically extracted from the full disk solar images taken from the observations and automatically standardized in shape and intensity (Zharkova et al., 2003).

These catalogues are intended to contain comprehensive statistics of active events overlapping in a given period of time and allowing the extraction of physical characteristics, which are essential for the solar activity forecast. This was aimed partly at the growing demand for solar activity forecasts by the space weather project and by many industrial organizations, which have a great need for usage of reliable and fast techniques for feature recognition on solar disks and their presentation in Solar Feature Catalogues (Zharkova et al., 2005).

The SFCs allowed to identify individual features (sunspots (Zharkov et al., 2005); active regions (Benkhalil et al., 2006); filaments (Fuller et al., 2005 and magnetic neutral lines (Ipson et al., 2005)) on the images with strongly varying background caused by different terrestrial atmosphere observing conditions of solar atmosphere activity period, irregularities in shape caused by instrumental errors or any other noise in images like strips or signatures etc. For added reliability, these algorithms were cross-referenced with the manual detections in Meudon and NOAA in order to correctly verify the detection of the features of interest. These automated techniques are valuable additions to the tools available for observers for processing the data with a cadence higher than 1 days required for many other studies.

The statistical analysis of the daily features from the SFCs has revolutionized the view on the solar activity features in the cycle 23 and their contribution into unveiling the nature of the solar activity and the solar dynamo mechanisms. These include the persistent North-South asymmetry and short term (2-2.5 years) periodicity (in addition to the long-term one of 9-11 years defined by the basic solar dynamo mechanism) in the total and cumulitative areas of sunspots and active regions (Zharkov et al., 2005, Zharkov et al., 2006), their magnetic tilts and polarity separation (Zharkov and Zharkova, 2007), detection of a persistent longitude (200° and persistent latitudes (5, 15, 25) in the sunspot and flare occurrences (Zharkova et al., 2007) and their strong correlation with the solar background magnetic field (Zharkov et al., 2007,). The new trends in the solar activity features revealed in the cycle 23 need to be tested in current cycle in order to complete the full 22 year investigation.

Therefore, before they are delivered to users, the data from the future solar missions including the Solar Orbiter, need to be processed with the adjusted automated recognition tools that can provide the validated detection parameters to all users and an opportunity to include these features into the studies of a particular features evolution during the passage over the solar disk. Also long-term variations of the activity features can contribute substantially to understanding the nature of the solar dynamo mechanisms at the verious levels in the solar interior.

2. *Automated image standardization technique*

There are number of difficulties that can occur with a solar image as it is demonstrated in Fig. 1: errors in FITS header information ; image shape (ellipse), error in the centre and the pole coordinates ; weather transparency (clouds) and different thickness of atmosphere ; centre-to-limb darkening ; defects in data (strips, lines, intensity). These original observation data with the header are stored in an ***observation table*** of the SFC database (see Section 4).

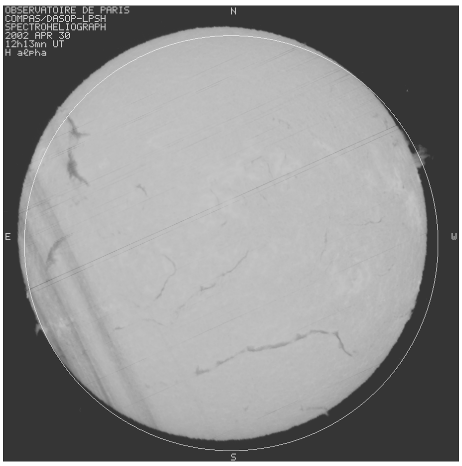

Fig. 1. A sample solar image demonstrates the distortions in intensity (strips, limb darkening) and elliptical shape. The white line circle shows the solar disk position taken from the image header.

When the geometrical information provided in image headers is not correct and the external photometric effects are to be removed this requires the automated procedures to be developed. These geometric and intensity effects are checked with the robust techniques developed for full disk solar images in order to convert them into a standardized form of a 'virtual solar image', or a flat image, free of geometrical distortions and limb darkening (Zharkova et al, 2003).

The technique was applied for H_α and Ca K lines full disk images taken at the Meudon Observatory, 173 A and 195 A lines of SOHO/EIT and for the SOHO/MDI white light images. This provided a fully standardized approach to the features detected in any images from these archives. Although, the technique adjustment based on the nature of distortion is required for any other archives produced by different instruments.

The parameters used for an image preprocessing, or standardization, including a code version and procedures applied are stored in a ***pre-processing table*** of the SFC database (see Section 4).

3. Automated Feature Recognition Techniques

3.1. Sunspot detection

After images are standardized, a morphological gradient operator is applied for edge enhancements, followed by thresholding in order to detect only strong edges. After removing the limb edge, a watershed operator is applied to the binary image in order to fill the sunspot area enclosed by the edges. Further median filtering is used to eliminate noise and smaller features. For the extraction of the area, shape, umbra/penumbra location of the detected sunspots and their basic classification the Canny edge detection technique is used.

The detected sunspot areas were compared with the average sunspot numbers from Sunspot Index Data Centre (SIDC), which confirmed a very high (0.86) correlation coefficient between these two sets despite their principal differences (area-numbers) (Zharkov et al, 2005). A comparison of the temporal variations of daily sunspot areas extracted from the Solar Feature Catalogue with those available from the digitized sunspot drawings at NOAA revealed the highest correlation coefficient of 96 % (Zharkov et al., 2005). The sunspot parameters are extracted from the SOHO/MDI data are populated into a ***sunspot feature table*** of the database discussed in Section 4.

3.2. Active Region Detection

The automated technique start with an initial segmentation of active regions, which is achieved using intensity thresholds determined using statistical information obtained for each quarter of a full disk solar image. Median filtering and morphological operations are applied to the resulting binary image to remove noise and to merge broken regions. Seed pixels selected in each of the initially segmented located regions are used to initiate more accurate region growing procedures. Statistically based local thresholding is applied to calculate upper and lower threshold values which control the spatial extents of the final detected regions.

A quantitative comparison was also made between the results with those done manually at the Meudon Observatory and the National Oceanic and Atmospheric Administration Observatory (NOAA). The procedures developed for the automated detection of active regions have achieved a satisfactory accuracy in the detection and segmentation of active regions using full disk Ha and Ca II K3 solar images from the Meudon Observatory and full disk Fe XII 195Å solar images from SOHO (Benkhalil et al., 2005, 2006). The parameters extracted from active regions are populated into the ***active region feature table*** of the SFC discussed in Section 4.

The developed automated technique for detection of bright active regions can be used for detection of any types of bright features in the images from Solar Orbiter payload (plages, flares, bright points etc.).

3.3. Filament Detection

Filaments are automatically detected in Ha line full disk solar images obtained at the Meudon Observatory. At first, the seeds of filaments are to be found in order to initiate the region growing. This is achieved by enhancing a contrast of the original image by applying a linear contrast stretch to the intensity range from zero to the value which excludes the top 1% of the area of the image histogram near maximum intensity. This standardization has the effect of putting the intensities of the darker areas near 0, so that a low intensity threshold can be applied to get seeds. A high threshold value, used by the region growing process, is also calculated from the histogram. As image intensity and sharpness can vary over the solar disk, threshold values can be calculated from the neighborhood of the seed instead of from the whole image.

This region growing technique combined with the edge detection developed for sunspots can be used for automated detection of coronal holes and other dark features in the images from Solar Orbiter.

4. Searchable database of the SFCs

Extracted parameters of the detected features (sunspots, active regions and filaments) were stored as ASCII files in the relevant format according to the Feature Parameter (FP) document (Aboudarham and Zharkov, 2004), which are used to populate the mysql searchable database. The database was designed to include the parameters describing the pre-processing and feature-detection code that was used for the extraction of the feature parameters as well as observational and individual feature parameters themselves. The database is published on the project website with the search page http://solar.inf.brad.ac.uk. This allows to search by time, location and size of a feature and can be downloaded in the ASCII or XML formats.

The detection process for each can be summarized as follows:

First the initial observation (a full disk image from the archive) is pre-processed using the automated cleaning technique (Zharkova et al., 2003) described in Section 2. The cleaning code setup, generally, depends on the source of the observation, or chosen archive. Then the features of interest are detected using the feature recognition methods described in section 3; these include sunspot, active regions (plages) and filaments. For each type of feature a number of parameters described in FP (Aboudarham and Zharkov, 2004) are extracted. Each feature boundaries are stored as a set of pixels in the pre-processed image being either a Bounding Rectangle Raster Scan for sunspots or a Chain Code for filaments and active regions (plages).

Hence, the database contains the following tables related to original observation (Fig. 3):

<u>Observation table:</u>

- Observations, which includes the observational parameters as related to the original observation;
- Observatory, which contains parameters related to the Observatory/Instrument (linked to Observations)

<u>Pre-processing table:</u>

- Pre-Processing Info - contains information about pre-processing code version, where it was run etc
- Pre-Processing Setup - which describes pre-processing code settings and input parameters.
- Pre-Processing Output - which contains the parameters which have been extracted or amended in the pre-processing stage, such as (where applicable) quiet sun intensity, image size, resolution, Solar Disk Radius,
- Ellipse fitting parameters.

Feature tables:

This table contains the individual feature parameters describing each detected feature: Sunspots, Faculae, Filaments.

- Feature Recognition Code Info - provides the information describing the code used for the extraction of feature parameters.
- Feature parameters themselves: location, size, area, intensity ratios to the quiet sun intensities and many others, variable for different features (for details see Aboudarham and Zharkov, 2004)

The database structure is organized in such a way that each feature has an identification number related to a processed observation (one record in Observation Output), to pre-processing setting (one record in Pre-Processing Output) and to the Feature Information (one record in the Feature output). Each original observation (image) is associated with one entry from the Observation table including an observatory, observation time, wavelength, cadence etc. Each pre-processed image is related to its original observation with the observation ID that also provides an additional entry from the Pre-Processing Setup where the Pre-Processing information is recorded.

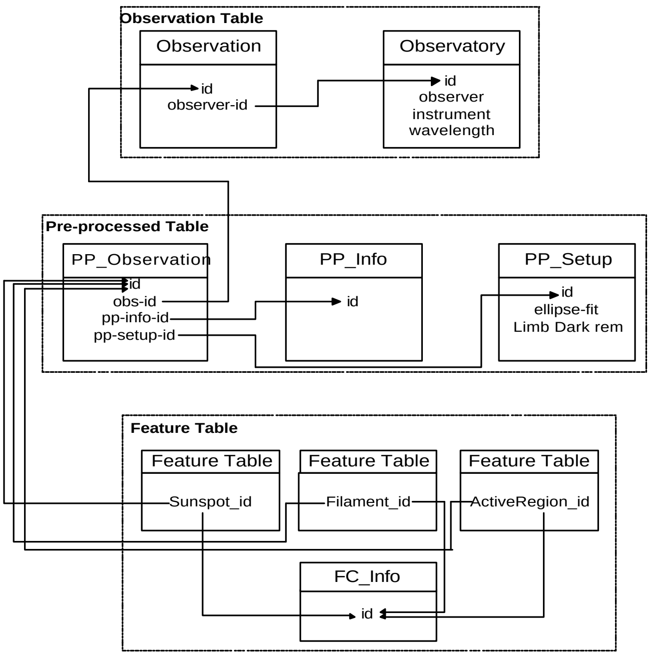

Figure 2. The SFC database structure: the observation, pre-processing and feature tables.

Then for each observation the features are detected and their identification numbers for a given date are assigned with different identification numbers assigned to sunspots, active regions (plages) and filaments. Each of these three IDs has entries to the pre-processing and observation tables as shown in the diagram in Fig 2.

Hence, this allows to uniquely identify every feature appearance in the original image and to relate it to the original observation time. The database can be extended to more features and extra duration from those already selected and the new instruments from the Solar Orbiter payload. The process can be fully automated allowing to extract the features and populate their parameters into the updated database.

5. Conclusions

Recently developed automated techniques applied for detection of sunspots, active regions, filaments in full disk solar images in Ca II K1, Ca II K3 and Ha images (Meudon Observatory), EUV (SOHO/EIT) and white light images (SOHO/MDI) revealed a good accuracy comparing to the manual synoptic maps and NOAA data. The developed procedures for the automated image standardization and feature extraction can be adjusted for the recognition in the Solar Orbiter images.

The extracted parameters of active features can be also automatically populated into the extended relational database of the Solar Feature Catalogues (SFCs) http://solar.inf.brad.ac.uk. This will allow delivering the major activity features extracted with a sufficient accuracy, in addition to the original images. This will also enable updating of the fully digitized database of solar features and ensure its continuation and consistency that is valuable for the solar activity modeling and forecast.

References

Aboudarham J and Zharkov, S. EGSO-WP5-IR3-2.3,http://www.cyber.brad.ac.uk/egso/Doc/Doc.html, 2004

Benkhalil, A.K., Zharkova, V., Ipson, S. and Zharkov S., ISCA, **12**/1, 21, 2005.

Fuller, N. Aboudarham, J. Automatic detection of Solar Filaments:, Solar Phys., 227, 234, 2005

Zharkov, S.I., Zharkova, V.V. and Ipson, S.S., Solar Phys., 228/1, 401, 2005
Zharkov, S.I., Zharkova, V.V.: Adv. Space Res., 38/5, 568, 2006
Zharkov, S.I., Zharkova, V.V.: Adv. Space Res., subm., 2007
Zharkov, S.I., Gavryuseva, E.V. and. Zharkova, V.V.., Solar Phys., subm, 2007
Zharkova, V.V., Ipson, S.S., Zharkov, S.I., Benkhalil, A.K., Aboudarham, J. Bentley, R.D., Solar Phys., 214/1, 89, 2003.
Zharkova, V.V., Zharkov, S.I., Aboudarham, J. Ipson, S.S., and Fuller, N., Solar Phys., **228**/1, 139, 2005
Zharkova, V.V. and Zharkov, S.I., Adv Space Res.., subm, 2007