Searchable Solar Feature Catalogues in EGSO

V V Zharkova[1], J. Aboudarham[2], **S Zharkov[1],** S S Ipson[1],

A K Benkhalil[1] and N. Fuller[2]

1 Department of Cybernetics, Bradford University, Bradford BD7 1DP, UK
2 LESIA, The Paris-Meudon Observatory, F92195 Meudon Principal Cedex, FRANCE
V.V.Zharkova@brad.ac.uk

Abstract

This paper describes a searchable Solar Feature Catalogue (SFC) created using automated pattern recognition techniques from digitized solar images. The techniques were developed for detection of sunspots, active regions, filaments and line-of-sight magnetic neutral lines using Ca II K1, Ca II K3 and Ha solar images from the Meudon Observatory and white light images and magnetograms from SOHO/MDI. A comparison of the results of automatic detection with manually generated synoptic maps shows good detection accuracy. Using the characteristics extracted from the recognized features a structured database of the Solar Feature Catalogues has been built on a mysql server and published with various pre-designed search pages on the Bradford University web site http://www.cyber.brad.ac.uk/egso/. The future SFC with 11 year coverage (1995-2005) is to be used for feature classification and short-term and long-term solar activity forecast. This research is a part of the European Grid of Solar Observations (EGSO) project.

1. Introduction

With a substantial increase in the size of solar image data sets, the automated detection and verification of various features of interest is becoming increasingly important for, among other applications, the reliable forecast of solar activity and space weather and data mining. However, this raises the accuracy and reliability requirements of the detection techniques applied for automated recognition that have to be significantly improved in comparison with the existing manual ones. One of the chief objectives for European Grid of Solar Observations (EGSO) Project Work Package 5 is production of Solar Feature Catalogues by means of automated feature recognition methods. Amongst such features of interest are sunspots, active regions, filaments and magnetic neutral lines.

The number of archives of digitized images of the Sun, taken from ground-based and space-based instruments in various wavelengths is growing steadily. These archives are available from different locations and are to be incorporated into a unified catalogue by the European Grid for Solar Observations (EGSO) project [1]. Digitized solar images from different sources have a variety of sizes, resolutions, dynamic ranges and instrumental and weather associated distortions. All are to be subjected to automated recognition processes in order to provide reliable data on the locations of features and their evolution at different times relative to solar rotation. This is 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 the development of reliable and fast techniques for feature recognition on solar disks and their presentation in Solar Feature Catalogues. These catalogues are intended to contain comprehensive statistics of active events (sunspots, active regions, filaments, flares, etc.) overlapping over time and to allow the extraction of physical characteristics, which are essential for the forecast of solar activity.

This requires the design of advanced image recognition techniques to identify individual features (sunspots, active regions, filaments, magnetic neutral lines, etc.) on images which may have imperfections. These can take the form of strongly varying background illumination caused by weather conditions in the terrestrial atmosphere, or by the presence of the solar atmosphere, irregularities in shape caused by instrumental errors or miscellaneous noise in images like strips or signatures etc. For added reliability, these algorithms have to use cross-referenced criteria at multiple wavelengths in order to correctly identify the features of interest while fully utilizing all the datasets linked into the Grid.

The following sections describe the techniques, which have been developed to date for the automated detection of sunspots, active regions, filaments and magnetic inversion lines.

2. Automated feature recognition techniques

2.1. Sunspot detection

The solar images considered here have a “quiet Sun” background on which are superimposed bright (faculae) and dark (sunspot) features. The sunspots are features in the solar photosphere and are the sites of strong magnetic fields at the surface of the Sun. Visually they consist of the two parts: a dark, roughly circular central disk called the umbra, and a lighter outer area called the penumbra. Sunspots are most clearly observed in “white light” images, but the stronger spots can also be detected in Ca II K1 images due to the absorption in that line.

Sunspot identification is required for a quantitative study of the solar cycle and this includes determining their locations, lifetimes, contrasts and other characteristics. Sunspot identification also plays an essential part in the modeling of the total solar irradiance (TSI) and the variations of sunspot properties with latitude and phase in the solar cycle. Sunspots are also part of solar Active Regions, and their local behavior is used in the study of Active Region evolution and for the forecast of solar activity

The sunspot detection process involves [2] first, if necessary, pre-processing the full disk high-resolution solar image by correcting the shape of the disk to a circular one and by removing limb-darkening [3]. Then a morphological gradient operator is applied to enhance the sunspot edges on the image, followed by thresholding in order to detect only strong edges. After removing the limb edge, an IDL 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. The regions' statistical properties are then used for the removal of false identifications such as, for example, the artefacts and lines, often present in the Meudon Observatory images. For the extraction of the area, shape, umbra/penumbra location of the detected sunspots and their basic classification, region growing, local contrast and contiguity techniques are used.

The automated sunspot detection technique was tested on Ca II K1 line images from the Meudon Observatory obtained over the month of April 2002 and achieved good correlation with the synoptic maps manually produced at the Observatory. The technique was also applied to observations over the same month of SOHO/MDI white light images, which confirmed the sunspots detected from the Meudon observations.

A quantitative comparison between the applications of the automated and manual techniques to the Meudon ground-based Ca II K1 observations was done by calculating the False Acceptance Rate (FAR) (where we detect a feature and they do not) and the False Rejection Rate (FRR) (where they detect a feature and we do not) for available observations. We introduced a Classifier Setting (CS) with value of 1 for all the sunspot candidates detected and with a value of 5 for those sunspot candidates of sizes greater than 8 pixel, mean intensity lower than the quiet sun intensity, ratio of principal axes less than 2.1 and mean absolute deviation greater than 21. As can be expected, the FRR is lowest for the sunspots with the classifier value of 1 which does not exceed 15.2 % difference from the total sunspot number detected on a day. On contrary, the FAR is lowest for the sunspots classified as value 5 and does not exceed 8.8% difference from the same number of sunspots.

Hence, one can conclude that the technique applied to sunspot detection on the Meudon Ca II K1 images performs well, particularly, in comparison with other earlier methods [4,5]. In most cases, using only the information present in the individual image, all sunspots are detected. We are also able to detect sunspots with excellent accuracy using SOHO/MDI white light images, combine them into groups using information extracted from SOHO/MDI magnetograms and extract sunspot locations and umbral/penumbral areas. Sunspot parameters have been extracted from the 4 months of SOHO/MDI data from April to July 2002 and incorporated into the SFC database.

2.2. Active Region Detection

A complex investigation of active regions in conjunction with flares, sunspots, coronal mass ejections (CMEs) and filaments is aimed at providing reliable forecasts of solar activity and space weather. Active regions are the basic reference features for solar activity. Their reliable automated detection will enable the building of a major database of solar active features and for the first time allow analysis of solar activity on a comprehensive database of active regions taken in various wavelengths. Techniques have been developed for the automated detection of Active Regions (ADAR) (plages) in solar images obtained from the Meudon Observatory, using the Hα and Ca II K3 spectral lines, aiming to replace the existing manual detection methods [6].

The automated technique starts 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 located segmented 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. The technique has been tested on full-disk solar images from the Meudon Observatory for the two months of April and July 2002 and compared with their manually generated synoptic maps.

A quantitative comparison was made between the results obtained using the ADAR technique, and those done manually at the Meudon Observatory [7] and those done manually by the National Oceanic and Atmospheric Administration Observatory (NOAA). In order to quantify the comparison, the FAR and the FRR were calculated for each day. Generally Meudon lists significantly more active regions than either ADAR or NOAA. For most days a higher number of active regions were detected by ADAR than by NOAA with an average FAR of 1.7 per day. The FRR of 0.2 was very low and there are only 5 days when ADAR failed to detect a region detected by NOAA. In some cases ADAR detects an active region while NOAA splits it into two regions. This affects the quantitative comparison. The reason for these different results is due to differences adopted for the definition of an active region. At Meudon all bright regions (plages) are detected, and these are defined as the regions in the chromosphere that are brighter than the normal "quiet" Sun background. At NOAA a detected active region is defined as a bright area on the Sun with a large concentration of magnetic field, often containing sunspots. However, not all plages contain a strong magnetic field as they might be decaying active regions with a weakening magnetic field [8].

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 Hα and Ca II K3 solar images from the Meudon Observatory, as described and also for full disk Fe XII 195Å solar images from SOHO. The structure of active regions at various levels of the solar atmosphere can provide a key to the understanding and reliable forecast of solar activity manifestations such as: solar flares, coronal mass ejections (CMEs), eruptive filaments etc. Their reliable automated detection will facilitate the building of a major database of solar active features and will enable the analysis of solar activity on a comprehensive database of active regions taken in various wavelengths.

The parameters of the active regions detected using the Meudon Ca II K1 data in the 4 months between April and July 2002 have been incorporated into the SFC database.

2.3. Filament Detection

Filaments are observed in solar images obtained in the Hα line and detected using a region growing technique applied to cleaned images. To find filaments seeds to initiate region growing, we enhance the contrast of the original image by applying a linear contrast stretch to the intensity range from the zero count 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 (see Fuller and Aboudarham, this edition).

The approach then is to first calculate the size of the area of the region found when region growing is stopped using the high threshold. If the size is bigger than a size limit (either a fixed size or function of the seed size), then the threshold is reduced until the size condition is satisfied. A second condition is that the final threshold can not be lower than half of the original one. With these two conditions large area filaments (greater than the size limit defined above) are retained and small ones, spread over the solar disk are avoided. As image intensity and sharpness can vary over the solar disk, threshold values are calculated from the neighborhood of the seed instead of from the whole image.

Filaments were also detected semi-automatically using an Artificial Neural Network technique [9]. The technique uses a feed-forward ANN with two hidden neurons, one responsible for filament pixels and the second for background pixels using a background approximated by a linear or parabolic function. The ANN was trained on solar image fragments labeled manually which is not feasible for automated detection. The technique has achieved a rather good accuracy of detection not less then 82 %. The technique can be used in combination with the region growing technique presented above, the results of which can be used for training the ANN. This will allow the fully automated detection of filaments with an ANN and reduce the total detection time.

The filament parameters have been extracted using the region growing technique applied to the 4 months of Hα data from Meudon between April and July 2002 and incorporated into the SFC database.

2.4. Searchable databases

The extracted parameters of detected features (sunspots, active regions and filaments) [11] were stored as ASCII files, which are used to populate the mysql searchable databases. The databases are published on the project website http://www.cyber.brad.ac.uk/egso/SFC/SFC_form.html. They can be searched by any of the extracted parameters and downloaded in ASCII or XML formats.

3. Conclusions

This paper surveys the efficient procedures used for the automated detection of solar features and extraction of their parameters into the Solar Feature Catalogues published on the Bradford University EGSO site. The techniques were applied to the detection of sunspots, active regions and filaments in Ca II K1, Hα images from the Meudon Observatory and MDI images from SOHO with a good accuracy and to magnetic neutral lines associated with these features. These techniques also allow us to automatically build synoptic maps (currently done manually) and to extract the data relevant to solar activity such as features numbers, area etc.

Work in progress includes extending the detection of sunspots, active regions and filaments to longer periods of observations, between 1995 and 2005 and the updating of the Solar Feature Catalogues to the full period of the 23rd solar cycle. By using the SFC to provide flags for solar activity and/or their combinations for every solar feature we hope to find the appropriate classifiers responsible for solar activity and to produce short-term and long-term forecasts of solar activity.

The authors would like to acknowledge the financial support of the European Commission which funds this research as part of the EGSO project in the IST Framework 5.

References

[1] Bentley, R.D., Proceedings of the Second Solar Cycle and Space Weather Euro-conference, 24 - 29 September 2001, Vico Equense, Italy, Edited by Huguette Sawaya-Lacoste, ESA Publication SP-477, 2002.

[2] Zharkov, S., Zharkova V.V., Ipson, S.S., Benkhalil A., Proc. AIISV, Aberystwyth, 7th-11th April 2003, pp. 74-84, ISBN 1-902956-33-1, 2003

[3] Zharkova, V.V., Ipson, S. S., Zharkov and 3 other authors Solar Physics, 2003; 214: 89-105

[4] Chapman, G. A. and Groisman, G., Solar Phys., 91: 45, 1984

[5] Chapman, G. A., Cookson, A. M., and Dobias, J. J., Astrophysical Journal. 432: 403-408, 1994

[6] Ali Benkhalil, Valentina Zharkova, Stanley Ipson and Sergei Zharkov, Proc. AIISV,, Aberystwyth, 7-11 April 2003, pp. 66-73, ISBN 1-902956-33-1, 2003.

[7] Mouradian Z., Synoptic Solar Physics ASP Conferences Series, vol. 140, pp. 181-204, 1998.

[8] Driel-Gesztelyi, L. V., Proc. SOLMAG the Magnetic Coupling of the Solar Atmosphere: Euroconference and IAU Colloquium 188, Santorini,Greece, 11-15 June 2002.

[9] Zharkova V. V. and Schetinin V. (2004) Special edition of *Neurocomputing*, Ed. D. Sanchez (in press)

[10] V.V.Zharkova, S.S. Ipson, R.S.Qahwaji, S.I. Zharkov and A. Benkhalil, Proceeding of Third International Workshop on Spectral Methods and Multirate Signal SMMSP 2003, pp. 115-121, ISBN 952-15-1062-5, 13-14/09/2003, Hotel Master, Barcelona, Spain

[11] Aboudarham J., Zharkov, S.., Zharkova V.., Benkhalil, A.K., Fuller, N.& Ipson, S.S. (2004), Feature parameters document,EGSO-WP5-IR3-1.8, http://www.cyber.brad.ac.uk/egso/SFC2/FeaturesParametershtml.htm