

# Automatic Sunspot Classification for Real-Time Forecasting of Solar Activities

T. Colak<sup>1</sup>, R.Qahwaji<sup>2</sup>

<sup>1</sup>Department of Electronic Imaging and Media Communications, t.colak@bradford.ac.uk  
Chesham B1.14, Richmond Road, Bradford BD7 1DP, England, U.K.

<sup>2</sup>Department of Electronic Imaging and Media Communications, r.s.r.qahwaji@bradford.ac.uk  
Horton A2.4 B, Richmond Road, Bradford BD7 1DP, England, U.K.

**Abstract-** Solar imaging is currently an active area of research. A fast hybrid system for the automated detection and classification of sunspot groups on MDI Continuum images using active regions data extracted from MDI Magnetogram images is presented in this paper. The system has three major stages: sunspots detection from MDI Continuum images, sunspots grouping and McIntosh classification of sunspot groups. Image processing and machine learning are integrated in all these stages.

## I. INTRODUCTION

The importance of space weather is increasing day after day because of the way solar activities affect life on Earth as we rely more and more on different communication and power systems. Induced electric fields and currents can affect the normal operation of ground-systems such as, high voltage power transmission grids, pipelines, telecommunications cables and railway signaling. Wireless communications systems and satellites also suffer from these activities.

The solar activity is the driver of space weather. Thus it is important to be able to predict the violent solar eruptions such as coronal mass ejections (CME) and solar flares[1]. According to reference [2], “*The most important social and economic aspects of space weather are related to being aware of and possibly avoiding the consequences of space weather events either by system design or by efficient warning and prediction systems allowing for preventive measures to be taken*”. Eruptions on the Sun travel to Earth in about 8 minutes in forms of light, radio waves, or X-rays. However the electrically charged particles from the Sun’s eruptions are the actual cause of the magnetic storms and they take many hours to a few days to reach Earth. Thus, proper warning of magnetic storms on Earth can be initiated if proper instruments to observe the Sun, the intervening space, and the Earth’s magnetic field, are combined with efficient data processing techniques.

Previous research on solar flares showed that they are mostly related to sunspots and active regions [3], [4], [5], [6], [7]. Sunspots are part of active regions, and their local behavior is used for the forecast of solar activity [8].

In this work we present a system that uses SOHO/MDI continuum and magnetogram images to detect, group, cluster

and then classify sunspots based on the McIntosh classification system [7]. The development and behavior of these groups is important to determine whether they may cause significant flares or not. This process is mainly manual and subjective, which means that it is a labor intensive and time consuming process and its accuracy depends on the experience of the human analyst. There is an apparent need for automated systems that can process solar data to extract useful knowledge and provide reliable predictions for the possible occurrence of solar activities that may affect us on Earth. The need is more obvious if we consider that the rate of solar data acquisition will shortly increase by more than 1000 times because of the very recent space missions (Hinode and STEREO) and the future space missions (SDO).

Previous attempts for the detection of sunspots are reported in [9] [10], and [11]. In [10] an automated system for the detection of sunspots on the Ca K1 and SOHO/MDI white light images was presented and a detection rate of 98% was achieved for MDI images when compared with the detection results of Locarno Solar Observatory. In [11], image processing, and clustering methods were applied to SOHO/MDI white light images for the recognition and classification of sunspots according to the modified Zurich class of McIntosh system. Testing involved 128 sunspot groups. Although 100% correct classification rate was achieved for the modified Zurich classes C and H (25% of test data), only 60%, 19% and 21% correct classification rates for D, E and F were obtained respectively. Full disk white light images were used in [9] to automatically detect and cluster sunspots into groups using morphological imaging. Neural networks were used to classify them. However, no good results were reported for grouping.

Previous research shows that accurate detection of sunspots has been achieved on white light solar images. However, no good results for the grouping and clustering of sunspots were reported. This is the biggest challenge facing the creation of an accurate and automated sunspot classification system [11], [9].

This paper is organized as follows: The automated detection and grouping of sunspots is introduced in Sec. II. The classification of sunspot groups is described in Sec. III, while the results and the concluding remarks are given in Sec. IV.

## II. SUNSPOT DETECTION & GROUPING

Several stages are involved in the detection and grouping of sunspots, such as: preprocessing, initial detection of features (sunspots and active regions) and clustering (Fig.1). All these stages are described below.

### A. Pre-processing

The pre-processing is divided into two stages. The first stage is applied to continuum and magnetogram images and is called Stage-1 processing. This stage is described below:

- Apply the filtering process reported in [12] to detect the solar disk, determine its radius, center and to create a mask. Using this mask, remove any irrelevant information.
- Calculate the Julian date and solar coordinates (the position angle, heliographic latitude, heliographic longitude) for the image using the equations of [13].

Stage-2 processing is applied only to magnetogram images and it is important because it enables us to correlate both MDI images. Usually there is a time difference (mostly less than 30 minutes) between magnetogram and continuum images, and the size of the solar disk on both images could differ. These problems has to be tackled to align and then correlate these images. To achieve this, magnetogram images need to be resized so that both images share the same center and radius. The time difference is equalized as follows:

- Map the magnetogram image from Heliocentric-Cartesian coordinates to Carrington Heliographic coordinates [14].
- Re-map the image to Heliocentric-Cartesian coordinates [14]. Use center, radius and solar coordinates of the continuum image as the new center, radius and solar coordinates of the magnetogram image.

### B. Initial Feature Detection

Initial detection of sunspots from continuum images and active regions from magnetogram images is carried out using intensity filtering and region growing methods, in a manner similar to reference [12].

The intensity filtering threshold value ( $T_f$ ) for each image is found automatically using (1), where,  $\mu$  is the mean,  $\sigma$  represents the standard deviation, and  $\alpha$  is a constant that is determined empirically based on the type of the features to be detected and images.

$$T_f = \mu \pm (\sigma \times \alpha) \quad (1)$$

In order to detect sunspot candidates from continuum images (Fig. 2.D):

- Apply histogram stretching to continuum image.
- Apply Gaussian filtering to resulting image.
- Apply intensity filtering using a threshold value calculated (1) with the minus (-) sign and 2.5 as the value of  $\alpha$ . If the value of pixel on image is less than the threshold value, mark it as sunspot candidate.

In order to detect active region candidates from magnetogram images (Fig. 2.C):

- The first threshold value is determined using (1) with a plus sign (+) and  $\alpha$  equals two. All pixels that have intensity values larger than this threshold are marked as active region seeds with “north” polarity.

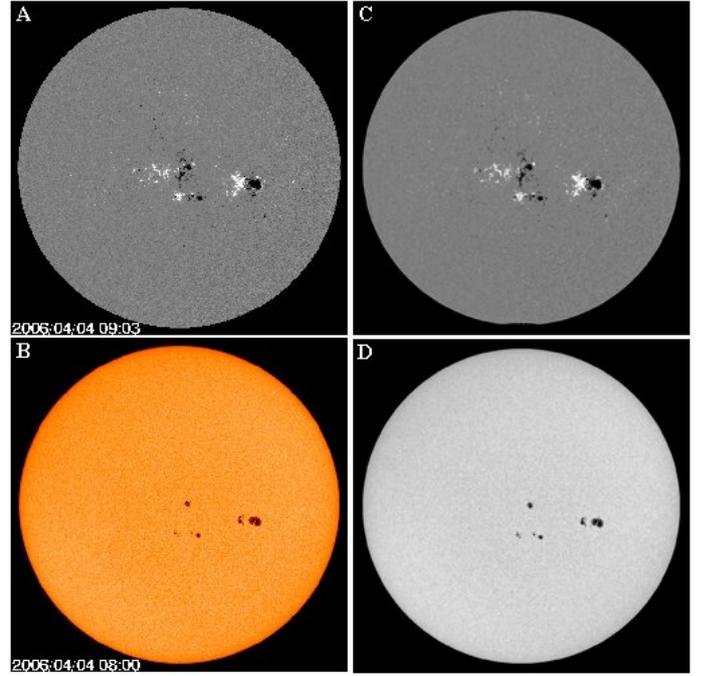


Fig. 1. Continuum and Magnetogram Images after preprocessing stage. A is the magnetogram and B is the Continuum image. C is the resulting magnetogram image after stage-1 and stage-2 processing. D is the resulting continuum image after stage-1 processing.

- In the same manner, the second threshold is determined using (1) with the minus (-) sign and  $\alpha$  equals two. Any pixel with an intensity value less than this threshold is marked as an active region seed with “south” polarity.
- Apply a simple region growing algorithm. Place a  $9 \times 9$  window on every seed and mark every pixel inside this window that has a similar intensity to the seed’s intensity ( $\pm 20\%$ ) as an active region candidate.

### C. Sunspot Grouping (Clustering)

After detecting the initial candidates for sunspots and active regions, the resulting images are combined to cluster sunspots into groups. Using this method the exact locations of the active regions and sunspots are determined and grouped. This method can be summarized as follows:

- Get a pixel marked as a candidate ( $P_{\text{spotcan}}$ ) on the sunspot candidate image (Fig. 2.D):
- If the active region candidate image (Fig. 2.C) has an active region candidate ( $P_{\text{actcan}}$ ) at the same location, create a new image for active regions and mark it as an active region ( $P_{\text{act}}$ ) with the same pixel value (dark or bright) of  $P_{\text{actcan}}$  and continue processing, otherwise return to the first step for processing another  $P_{\text{spotcan}}$ .
- Place a circle on  $P_{\text{act}}$  with “ $\alpha$ ” degree radius on the active region candidate image and mark all the  $P_{\text{actcan}}$  within this circular region as  $P_{\text{act}}$  on the newly created active region image. In this work, the value of  $\alpha$  is determined empirically and it equals 10.

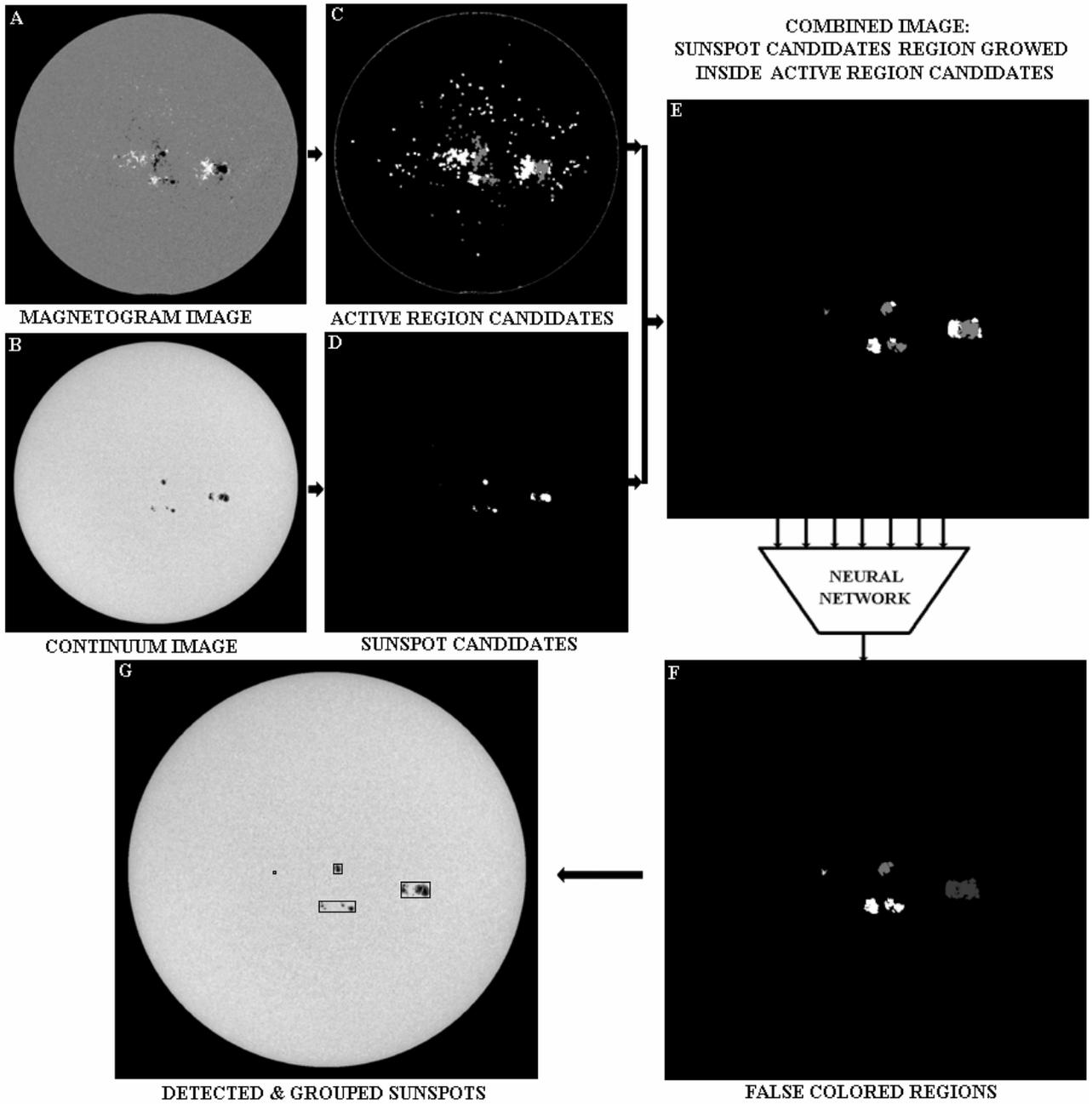


Fig. 2. Initial feature detection and grouping stages.

- Marked regions with the same color will be considered to belong to the same active region and these regions will be combined by filling the gaps between them.
- Finally this image will be ANDed with the original sunspot candidate image to group the detected sunspots. In the final image every sunspot belonging to the same group will have the same intensity values.

We used Neural Networks (NN) to combine regions of opposite magnetic polarities in order to determine the exact boundaries of sunspot groups. NN is applied to two opposite polarity regions to decide if they are part of the same active region or not.

NN training vector consists of seven inputs and one output

showing the relation between each opposite polarity magnetic field pairs. The calculations required for inputs and outputs for the NN are given in Table I, where,  $A_a$ ,  $A_b$  are the areas in pixels for each region,  $d$  is the distance between the two regions in heliographic degrees,  $d_{lon}$ ,  $d_{lat}$  are the longitude and latitude difference between the two regions and  $I_{ab}$  is the intersecting area between the two regions in pixels. The training vector is constructed using nearly one hundred examples. Several experiments are carried to optimize the NN in a manner similar to [15]. It was found that the best learning performance is obtained with back propagation training algorithm and using the following NN topology: 7 input nodes, one hidden layer with 8 nodes and one output node.

TABLE I  
INPUTS AND OUTPUT FOR NN TRAINING VECTOR FOR ACTIVE REGION DECISION

INPUTS	DESCRIPTION
$\text{Min}(A_a, A_b) / \text{Max}(A_a, A_b)$	Ratio of the smallest area to biggest area of regions
$I_{ab} / A_a$	Ratio of intersecting area to area of first region
$I_{ab} / A_b$	Ratio of intersecting area to area of second region
$d_{lon} / d$	Ratio of the longitude difference between regions to distance between regions
$d_{lat} / d$	Ratio of the latitude difference between regions to distance between regions
$d / 180$	Ratio of distance between regions to 180 degrees.
0.1 or 0.9	If two regions are intersected by same sunspot candidate it is 0.9 otherwise 0.1.
OUTPUT	DESCRIPTION
0.1 or 0.9	If two regions are part of the same active regions it is 0.9 otherwise 0.1

After deciding the active regions and sunspots, the spots belonging to same groups are marked as detected groups (Fig. 2 image G). All the stages after pre-processing are shown in Fig.2. The detected groups are processed further for determining their McIntosh classes.

### III. MCINTOSH CLASSIFICATION OF SUNSPOT GROUPS

After grouping the detected sunspots, each sunspot group is classified based on the McIntosh classification system which is the standard for the international exchange of solar geophysical data. The classification depends on the size, shape and spot density of sunspots. It is a modified version of the Zurich classification system, which has improved definitions and added indicators of size, stability and complexity [7]. The general form of the McIntosh classification is  $Zpc$  where,  $Z$  is the modified Zurich class, “ $p$ ” is the type of penumbra on largest spot, and “ $c$ ” is the degree of compactness in the interior of the group.

#### A. Computing the Modified Zurich Class - $Z$

The modified Zurich classes are defined depending on the polarity of sunspot group, presence of penumbra, distribution of the penumbra, and the length of the sunspot group [7]. To compute the Modified Zurich Class the following factors are considered:

- Determine if the group is Unipolar or Bipolar.

The polarity of sunspot groups is decided according to the separation between sunspots within the group. If there is a single spot or compact cluster of spots in a group and the greatest separation is smaller than 3 degrees, the group is considered to be Unipolar; if the separation is higher the group is considered to be bipolar.

- Determine if a penumbra exist in any of the spots.

In white light images, large sunspots have a dark central umbra surrounded by the brighter penumbra, and small spots have a dark central umbra surrounded by a dark penumbra. In order to decide if a sunspot has penumbra or not we apply equation (2) to find its threshold, where,  $\mu$  is the mean,  $\sigma$  represents the standard deviation of all the spots on the image,  $P_N$  is the total number of pixels for the spot under consideration, and  $P_T$  is

the total number of pixels for all spots on the image.

$$T_N = \mu - (\sigma \times P_N / P_T) \quad (2)$$

If the sunspot pixel is smaller than this threshold value  $T_N$ , it is considered as part of the umbra, otherwise it is considered to belong to the penumbra (Fig. 3).

- Determine if the spots with the penumbra are located on one end or both ends.

For deciding this all the sunspots are sorted from east to west depending on their central longitudes. The first and last sunspots within the sunspot group are searched to determine whether they have a penumbra or not.

- Calculate the length of the group in absolute heliographic degrees.

After finding all the necessary data, the modified Zurich class for the sunspot group is found using a decision tree.

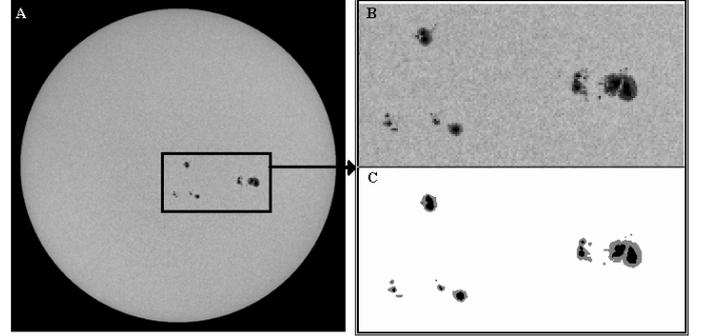


Fig. 3. Results of calculating the penumbra and the umbra of the spot. A is the original image, B is the zoomed area from A, and C is the output showing umbra of the spots with black and penumbra of the spots with grey.

#### B. Computing the Type of Largest Spot Class - $p$

The largest spot in a sunspot group can be classified depending on its type, size and symmetry of the penumbra [7]. The penumbra can either be rudimentary (partially surrounds the umbra) or mature (completely surrounds the umbra). Its size is the value of the north to south diameter. Symmetry of the penumbra depends on the irregularity of the outline of the penumbra (Fig. 4). The size of the spot can be easily calculated by finding the difference between its north and south latitudes. However, finding the symmetry and type of penumbra is a real challenge because it depends mostly on the subjective judgment. In order to determine the type of largest spot:

- Decide if the penumbra of the largest spot is rudimentary or not.

The Type of the penumbra is decided by calculating the ratio of penumbra and umbra pixels in a group. If the number of penumbra pixels is more than umbra pixels, the sunspot is assumed to have a mature penumbra; otherwise penumbra is assumed to be rudimentary.

- Decide if the penumbra of the largest spot is symmetric or not.

Apply horizontal and vertical integral projections to the sunspots. The vertical integral is computed by finding the average energy contained in every row, while the horizontal integral is computed by finding the average energy of each column. After projections are applied, energy of each column and each row are compared with the energy of adjacent

columns and rows in order to find the minima and maxima in the energy spectra. This provides accurate detection for every umbra within the penumbra. Compare the numbers of horizontal and vertical umbra peaks. The sunspot is assumed to be symmetric, if the number of peaks on both projections is equal to one, otherwise it is assumed to be asymmetric.

- Calculate the value of the north to south diameter in heliographic degrees.

After finding all the necessary data, they are applied to a decision tree to find the type of largest spot class in the sunspot group.

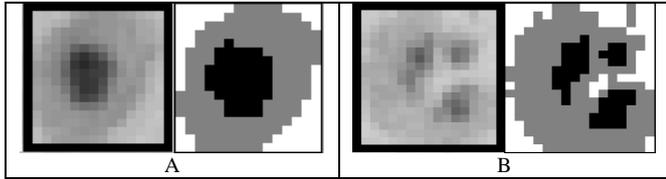


Fig. 4. A Symmetric sunspot, B Asymmetric sunspot.

### C. Computing Sunspot Distribution – c

The sunspot distribution depends on the compactness of the sunspot group [7]. In order to determine the distribution of sunspots:

- Determine the compactness of sunspots within the group. Calculate the ratio of the total sunspots area to the sunspot group area. If this ratio is less than 40% then the sunspot group is assumed to be “open” [7].
- Determine if there exist a spot with a mature penumbra in the group besides the leader and the follower.

If there is a spot with mature penumbra besides the leading and ending spots of the sunspot group, then the group is assumed to be “compact” otherwise “intermediate”.

## IV. RESULTS AND CONCLUSIONS

A computer platform using C++ .Net was created for the automated detection and classification of sunspots using SOHO/MDI continuum and magnetogram images in GIF format. A publicly available library: “corona.dll”<sup>1</sup> is used for reading all the GIF images. The program for training and applying the NN is also created and implemented in C++. The whole system works with 1024 × 1024 images and the detection of sunspots, detection of active regions and classification of sunspot groups takes approximately four seconds per image depending on the complexity of features, using P4-2.8 GHz PC with 1 GB RAM.

The detection, grouping and classification algorithms are tested on a total of 31 intensitygram and 31 magnetogram images mostly available from 01/01/2006 to 01/02/2006 and some individual images with complex sunspot groups for better verification of the algorithms.

The output from our system is compared to a total of 96 sunspot groups which are manually detected and classified with the help of the publicly available sunspot catalogues from NGDC/USAF that are created by solar physicists from

different observatories.

Our algorithms managed to detect 105 groups and 89 of them matched the manually extracted groups. This means that there is nearly a 92% correct match for sunspot groups and 15% of the groups that are detected by our algorithms are not present on the manually detected groups. Almost all the sunspot groups that are not present are the sunspot groups with only one or two sunspots.

Although almost 99% of the detected sunspots are correct, we found that there are some miss detections of very small sunspots. All the initial sunspot candidates are compared with their corresponding magnetic activity on magnetograms images. This reduces the probability for wrong detection of sunspots. This also shows that most of the error is caused by wrong grouping of our algorithms and/or miss detections of sunspots by solar experts.

Our algorithms have clustered some sunspots into separate groups despite the fact that they belong to the same group. This applies in particular to sunspots that are separated by large distances compared to their areas. This causes their magnetic traces to be separate from each other and as a result the NN clusters them as separate groups. Sometimes two or three small sunspots that are part of the same group can be clustered as two different sunspot groups which explains some of the wrong grouping.

Also, the lack of visibility by ground observatories at the time of sunspots detection and human error (Some small sunspots are very hard to determine by human eye) can cause the miss detections of sunspots. Furthermore, sunspots forming or deforming can be hard to notice. We came across some examples, where some sunspot groups are detected by our system in their early development stage and are not reported in the sunspots catalogues till they have matured a little.

An example for grouping and miss detection on NGDC/USAF catalogue is shown on Fig. 5. In this figure, the detected groups, and classification results from our algorithms under the automatic classification headline and classification results from NGDC/USAF catalogue is given. Our algorithms detected four sunspot groups on the image 04/04/2006 taken at 08:00, the closest classification on the available USAF sunspot catalogue is at 08:35 by San Vito observatory. This observatory detected three sunspot groups and the newly forming group (2) that is detected by our algorithms is not one of them. This group is only available on the sunspot groups detected nearly 9 hours later by Holloman AFB observatory classified as BXO. This group will be counted as wrong grouping on our test results although it is proven that it is correct by the next observation results.

Out of the 89 matched sunspot groups, the correct classification rates for modified Zurich class (Z), type of largest spot (P), and group distribution (C) are 79.7% (71 correct), 47.1% (42 correct), and 84.2% (75 correct), respectively. Although the modified Zurich class ratio and group distribution ratio results are satisfactory, we can not say the same thing for the type of the largest spot ratio. Deciding the type of the largest spot is a very hard task, even for an experienced solar physicist because it involves subjective

<sup>1</sup> <http://corona.sourceforge.net/>

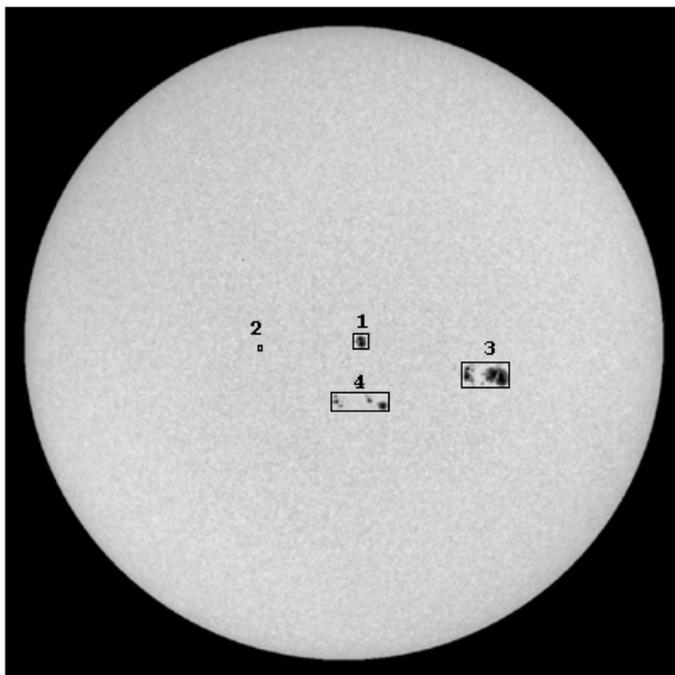
judgment on the degree of symmetry and maturity.

For future work, we would like to improve the classification rates, especially for determining the type of the largest spot. This can be achieved by using machine learning techniques for deciding the symmetry and type (mature or rudimentary) of penumbra.

We would also like to classify the sunspot groups according to Mt. Wilson classification which can be done with higher accuracy when we take into account that the polarity of each sunspot can be easily be determined from magnetogram images. Our major aim is to combine the output data from this system with a machine learning system, as described in [15] to provide an automated platform for the short-term prediction of major solar flares using neural networks and/or support vector machines .

## REFERENCES

- [1] H. Koskinen, L. Eliasson, B. Holback, L. Andersson, A. Eriksson, A. Mälkki, O. Norberg, T. Pulkkinen, A. Viljanen, J.-E. Wahlund, and J.-G. Wu, "Space Weather and Interactions with Spacecraft," ESA Technology Research Programme 1999.
- [2] H. Koskinen, E. Tanskanen, R. Pirjola, A. Pulkkinen, C. Dyer, D. Rodgers, and P. Cannon, "Space Weather Effects Catalogue," FMI, QinetiQ, RAL Consortium 2001.
- [3] H. Künzel, "Die Flare-Häufigkeit in Fleckengruppen unterschiedlicher Klasse und magnetischer Struktur," *Astronomische Nachrichten*, vol. 285, pp. 271, 1960.
- [4] A. B. Severny, "On the changes of magnetic fields connected with flares," in *Stellar and Solar Magnetic Fields, International Astronomical Union. Symposium no. 22*, R. Lust, Ed. Amsterdam, Holland: North-Holland Pub. Co., 1965, pp. 358.
- [5] C. S. Warwick, "Sunspot Configurations and Proton Flares," *Astrophysical Journal*, vol. 145, pp. 215, 1966.
- [6] K. Sakurai, "On the magnetic configuration of sunspot groups which produce solar proton flares," *Planetary and Space Science*, vol. 18, pp. 33, 1970.
- [7] P. S. McIntosh, "The Classification Of Sunspot Groups," *Solar Physics*, vol. 125, pp. 251-267, 1990.
- [8] D. Hathaway, R. M. Wilson, and E. J. Reichmann, "The Shape of the Sunspot Cycle," *Solar Physics*, vol. 151, pp. 177, 1994.
- [9] J. J. Curto, M. Blanca, and J. G. Solé, "Automatic detection of sunspots and group classification from white full disc images," presented at Solar Image Recognition Workshop, Brussels, 2003.
- [10] S. Zharkov, V. Zharkova, S. Ipson, and A. Benkhalil, "Automated recognition of sunspots on the SOHO/MDI white light solar images," in *Knowledge-Based Intelligent Information And Engineering Systems, Pt 3, Proceedings*, vol. 3215, *Lecture Notes In Artificial Intelligence*, 2004, pp. 446-452.
- [11] S. H. Nguyen, T. T. Nguyen, and H. S. Nguyen, "Rough Set Approach to Sunspot Classification Problem," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, vol. 3642, D. Slezak, J. Yao, J. F. Peters, W. Ziarko, and X. Hu, Eds.: Springer Berlin / Heidelberg, 2005, pp. 263-272.
- [12] R. Qahwaji and T. Colak, "Automatic Detection And Verification of Solar Features," *International Journal of Imaging Systems & Technology*, vol. 15, pp. 199 - 210, 2006.
- [13] J. Meeus, *Astronomical Algorithms - Second Edition*. Virginia: Willmann-Bell, Inc., 1998.
- [14] W. T. Thompson, "Coordinate systems for solar image data," *Astronomy & Astrophysics*, vol. 449, pp. 791-803, 2006.
- [15] R. Qahwaji and T. Colak, "Automatic Short-Term Solar Flare Prediction Using Machine Learning and Sunspot Associations," *Solar Physics*, 2007.



AUTOMATIC CLASSIFICATION							
#	DATE	TIME	COOR	CLASS	VIS	OBS	
1	20060404	0800	S06W03	HAX	S	AUTO	
2	20060404	0800	S06E15	AXX	S	AUTO	
3	20060404	0800	S11W26	DKO	S	AUTO	
4	20060404	0800	S16W03	EA0	S	AUTO	

USAF CLASSIFICATION							
#	DATE	TIME	COOR	CLASS	VIS	OBS	
1	060404	0832	S05W03	HSX	3	SVTO	
2	060404	1723	S07E10	BX0	3	HOLL	
3	060404	0832	S11W27	DKC	3	SVTO	
4	060404	0832	S16W03	DS0	3	SVTO	

Fig. 5. Results of sunspot grouping and comparison of automated classification with classification from observatories.

## ACKNOWLEDGMENT

This work is supported by an EPSRC Grant (GR/T17588/01), which is entitled "Image Processing and Machine Learning Techniques for Short-Term Prediction of Solar Activity".