Morphological Image Analysis and Its Application to Sunspot Classification

D. Stenning, V. Kashyap, T. C. M. Lee, D. A. van Dyk, and C. A. Young

Abstract The morphology of sunspot groups is predictive both of their future evolution and of explosive associated events higher in the solar atmosphere, such as solar flares and coronal mass ejections. To aid in this prediction, sunspot groups are manually classified according to one of a number of schemes. This process is both laborious and prone to inconsistencies stemming from the subjective nature of the classification. In this paper we describe how mathematical morphology can be used to extract numerical summaries of sunspot images that are relevant to their classification and can be used as features in an automated classification scheme. We include a general overview of basic morphological operations and describe our ongoing work on detecting and classifying sunspot groups using these techniques.

David Stenning
Department of Statistics, Donald Bren Hall, 2nd Floor University of California, Irvine, CA 92697-1250 e-mail: dstennin@ics.uci.edu

Vinay Kashyap
Smithsonian Astrophysical Observatory, 60 Garden St., Cambridge, MA 02138 e-mail: vkashyap@cfa.harvard.edu

Thomas C. M. Lee
Department of Statistics, 4118 Mathematical Sciences Building, University of California, One Shields Avenue, Davis, CA 95616 e-mail: tcmlee@ucdavis.edu

David A. van Dyk
Department of Statistics, Donald Bren Hall, 2nd Floor University of California, Irvine, CA 92697-1250 e-mail: dvd@ics.uci.edu

C. Alex Young
ADNET Systems, Inc., NASA/GSFC, Mail Code 671, Greenbelt, MD, 20771 e-mail: c.alex.young@nasa.gov

1 Scientific Background and Motivation

The Sun's *photosphere* is the region that emits the light that we see. The deeper regions are opaque and the higher and much less dense *corona* is only one-millionth as bright as the photosphere in visible light. Sunspots are dark areas on the photosphere that result from intense magnetic fields. The magnetic fields inhibit convection, cooling the corresponding surface regions. Areas on the photosphere where the surface temperature has been reduced then appear as dark spots when viewed in optical light. Sunspots can also be seen in *magnetograms* which are images that represent variations in the strength of magnetic fields in the Sun's photosphere [3]. In magnetograms, sunspots correspond to high flux regions that appear as areas of opposite magnetic polarity.

The classification and tracking of sunspots is an active undertaking of solar-physicists hoping to untangle connections between sunspot activity and various solar phenomena. Recent studies, for example, suggest that solar flares are related to the magnetically active regions around sunspot groups [5]. As a result, various sunspot classification schemes aim to characterize magnetic flux content in the active-regions on the solar disk [4]. One particular scheme—the Mount Wilson classification—puts solar active-regions into four classes based on the complexity of magnetic flux distribution. When combined with space weather data, this scheme can be used to predict activity in the solar corona such as highly energetic solar flares and massive bursts of solar wind known as coronal mass ejections [4]. While precise precise predictions remain elusive, the complexity of the magnetic flux distribution of sunspot groups can be used to infer trends and tendencies in the patters of solar flares and coronal mass ejections

Recently launched NASA missions such as the Solar Dynamics Observatory—with its continuous science data downlink rate of 130 Megabits per second—are producing an unprecedented volume of solar data. Nonetheless the majority of sunspot classification is still performed through visual inspection by experts [2]. This is a laborious process and, as with all manual procedures, is susceptible to bias from the human observer [4]. Since the morphology of sunspot groups form a continuous spectrum rather than a set of discrete and obvious classes, there is a level of subjectivity in manual classification. One of the attractions of the Mount Wilson scheme is its reliance on a relatively simply set of classification rules. While this may aid manual classification it introduces artificial dichotomies that may hinder scientific understanding. Even with the relatively straightforward Mount Wilson scheme, trained experts do not always agree on classifications. As a result, there is a need for an automated, objective and reliable procedure for detecting and classifying sunspot groups.

The Mount Wilson classification scheme divides sunspot groups into four classes. The simplest morphologically is the α class which consists of groups that are dominated by a single *unipolar* sunspot, i.e., a sunspot with a magnetic field that is dominated either by a positive or a negative polarity. The second class, the β class, is made up of groups with both polarities, but with a simple and distinct spatial division between the polarities. In particular a straight line can be drawn through

the group that nearly divides the negative from the positive polarities. Groups in the third class, $\beta\gamma$, are also *bipolar*, but are sufficiently complex that a straight line cannot divide the positive and the negative polarities. Finally, in the fourth class, $\beta\gamma\delta$, the positive and negative polarities are scattered throughout the region and cannot be easily separated. Example of the sunspot groups from the four classes appear in Figure 1.

Because this classification scheme is defined in terms of the morphology of the sunspots, we propose to use methods from mathematical morphology to extract features from the magnetograms that can be used in an automated classification technique, such as a classification tree, support vector machine or some other common method, to reconstruct the Mount Wilson classification. We use a data set that consists of magnetogram images collected by the Solar and Heliospheric Observatory/Michelson Doppler Imager (SOHO/MDI). Each magnetogram includes the date and time the image was taken, the location on the solar disk, and the identification number of the sunspot group given jointly by the U. S. Air Force and the National Oceanic and Atmospheric Administration (USAF/NOAA). The manual classification of the sunspot group by USAF/NOAA according to the Mount Wilson scheme is also provided.

The primary goal of this article is to make progress toward an automatic sunspot classification method that relies on features extracted using techniques from mathematical morphology. We begin in Section 2 with an overview of the mathematical morphology methods that we employ. In Section 3 we describe how we compute relevant numerical summaries of the magnetogram images using mathematical morphology and methods for using these summaries for classification. Finally in Section 4 we discuss the road forward toward automated sunspot classification.

2 Mathematical Morphology

Mathematical morphology is a powerful tool for image analysis, which was developed about forty years ago. Unlike other tools (e.g., Fourier methods), morphological operators relate directly to shape. When used appropriately, morphological operations can simplify images by preserving their essential shapes and eliminating noise. For detailed descriptions of the subject, see [6, 7].

2.1 Binary and Greyscale Images

Objects in digitized images are only approximations to their counterparts in the real world. One reason is simply because their domains are defined in different spaces: images are pixelated and thus “discrete” while the object itself is “continuous” in nature. We will use $\mathbb{Z}^2$ to denote the space of objects in binary images. That is $\mathbb{Z}^2$ can be thought of as a two dimensional grid of pixels that is infinitely tall and

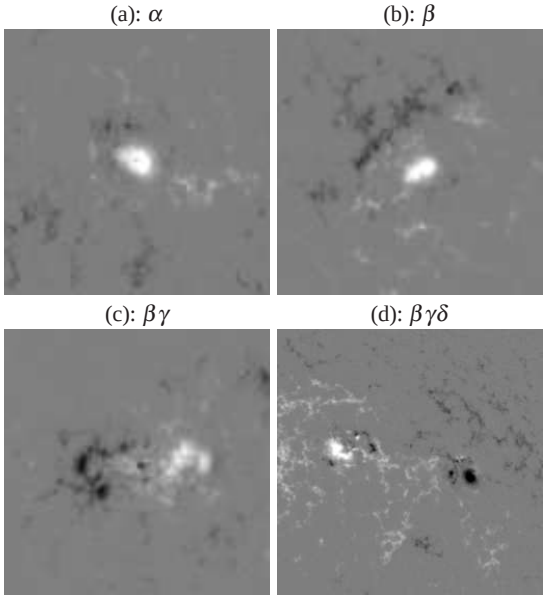


Fig. 1 Examples of the four classes of sunspot groups used in the Mount Wilson scheme. The α class (a) is dominated by a single pole that appears black or white in the magnetogram, depending on the polarity (positive or negative). The β class (b) has regions of both positive and negative polarity that can be separated by a straight line. The $\beta\gamma$ class (c) also exhibits both polarities but they cannot be easily separated into two regions. In the $\beta\gamma\delta$ class (d) the two polarities are scattered throughout the region.

infinitely wide. We can treat $\mathbb{Z}^2$ as the discrete version of the Euclidean plane $\mathbf{R}^2$, and represent it as a two dimensional Cartesian square grid.[1]

A *binary image*, f, is a image where each pixel is either black or white. For example we can assign the value 1 (i.e., black) to a pixel if it belongs to an object, otherwise the value 0 (i.e., white). Notice that we can always consider objects (i.e., the "black" parts) in a binary image as sets and the image itself as the union of all such sets. See Figure 2(a) for a binary image. Mathematically, we can write a binary image as a mapping, which maps each pixel of a subset $\mathscr{D}_f$ of $\mathbb{Z}^2$ into the couple $\{0,1\}$:

$$f : \mathscr{D}_f \subset \mathbb{Z}^2 \longrightarrow \{0,1\},$$

where $\mathscr{D}_f$ is some subset of $\mathbb{Z}^2$ and is called the definition domain of f.

More generally, a *greyscale image*, f, is a mapping which maps each element in a subset $\mathscr{D}_f$ of $\mathbb{Z}^2$ into the set of non-negative integers $\mathbf{N}_0$:

[1] Originally mathematical morphology was defined in the d-dimensional Euclidean space $\mathbf{R}^d$, but there is no great difficulty in translating the this theory from $\mathbf{R}^d$ to its discrete version $\mathbb{Z}^d$. In our discussion about mathematical morphology, we use $\mathbb{Z}^2$, but understand that that the development would work equally well for either $\mathbb{Z}^d$ or $\mathbf{R}^d$.

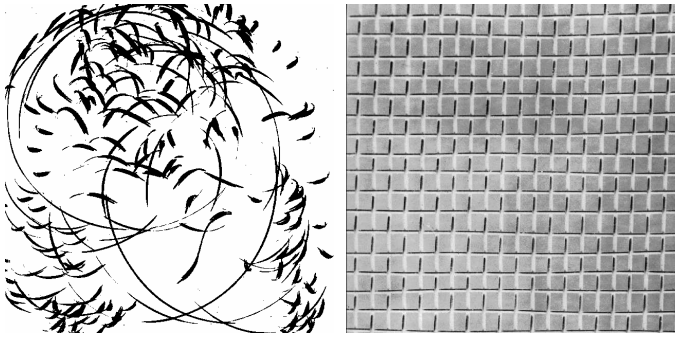


Fig. 2 (a): a binary image and (b) a greyscale image.

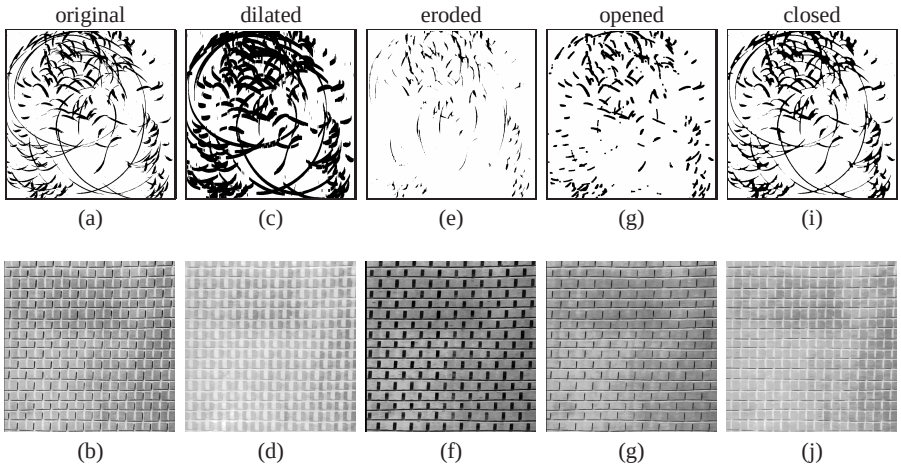


Fig. 3 Top row: (a) a binary image that has been (c) dilated, (e) eroded, (g) opened, and (i) closed. Bottom row: (b) a greyscale image that has been (d) dilated, (f) eroded, (h) opened, and (j) closed. For the binary image a vertical line was used as the SE in the morphological operations. For the greyscale image, a rectangle was used.

$$f : \mathscr{D}_f \subset \mathbb{Z}^2 \longrightarrow \mathbf{N}_0.$$

Very often the set of non-negative integers under consideration is $\{0,\ldots,255\}$, where the larger the value, the brighter the pixel is. In mathematical morphology, it is useful to treat the pixel values of a greyscale image as the heights of a surface above the image plane. See Figure 2(b) for a greyscale image.

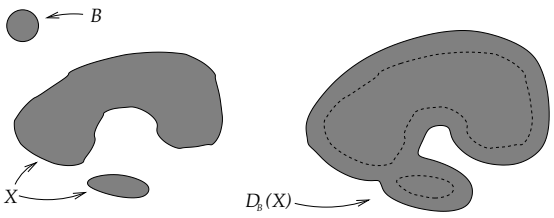


Fig. 4 Dilation of a set X by a disk-shaped structuring element B.

2.2 Dilation and Erosion

In mathematical morphology there are two basic operations: *dilation* and *erosion*. These are the basic building blocks and many other morphological operations can be expressed in terms of dilation and erosion. We first define dilation.

Suppose we have a set $X \subset \mathbb{Z}^2$ and a curser B that scrolls across $\mathbb{Z}^2$. If we record the location of B whenever it intersects or "runs into" X the result is called the dilation of X by B, denoted by $D_B(X)$. This is illustrated in Figure 4. Notices that the dilation of X is a bloated version of X, where the degree and character of the bloating is determined by the shape and size of B. The dilation of X by B is the answer to the question: "What is the location of B when B *hits* X?" (We define A *hits* B as $A \cap B \neq \phi$.) In other words, $D_B(X)$ is the set of all points $\mathbf{x}$ such that B *hits* X when the location or *origin* of B is at $\mathbf{x}$.

We call B a *structuring element* (SE). Generally speaking a SE is a subset of $\mathbb{Z}^2$ with a known shape and origin. SE elements are used to examine or transform the image f under study. As with dilation, all morphological operators treat the image as a set (i.e., a binary image) and use one or more SEs to examine it. We could also say these operators use the shape(s) of the SE(s) to transform f. Notice that the SE B is arbitrary, hence one can always choose a suitable SE to perform the desired task. This gives the user a great flexibility in applying morphological methods. Usually SEs are regular and small in size when compared to the image. For example, in the case of a binary image in Figure 4, B is a disk with a small radius and with its center as the origin.

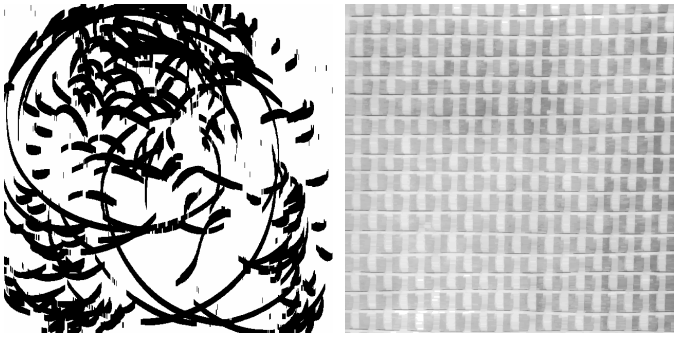
The formal definition of dilation is:

$$D_B(X) \equiv \{\mathbf{x} \in \mathbb{Z}^d \mid B_\mathbf{x} \cap X \neq \phi\},$$

where $B_\mathbf{x}$ is the SE B placed with its origin at $\mathbf{x}$. Figure 5 shows the dilation of the images displayed in Figure 2.

The erosion of X by B, denoted by $E_B(X)$, is the answer to the question: "Where is the origin of B when B *fits* wholly inside X?" That is, $E_B(X)$ is the set of points $\mathbf{x}$ such that B *fits* wholly inside X when the origin of B is at $\mathbf{x}$. The formal definition of erosion is:

$$E_B(X) \equiv \{\mathbf{x} \in \mathbb{Z}^d \mid B_\mathbf{x} \subset X\}.$$



(a) (b)

Fig. 5 (a): dilation of the image in Figure 2(a) with a vertical line as SE; (b) dilation of the image in Figure 2(b) with a rectangle as SE.

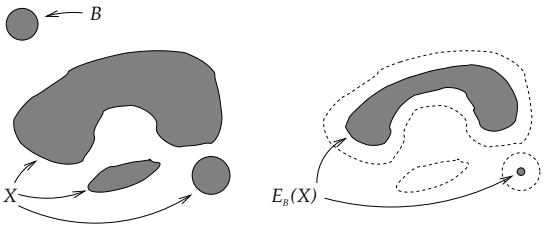


Fig. 6 Erosion of a set X by a disk-shaped structuring element B.

See Figure 6 for an example, and Figure 7 for examples of eroded images.[2]

2.3 Opening and Closing

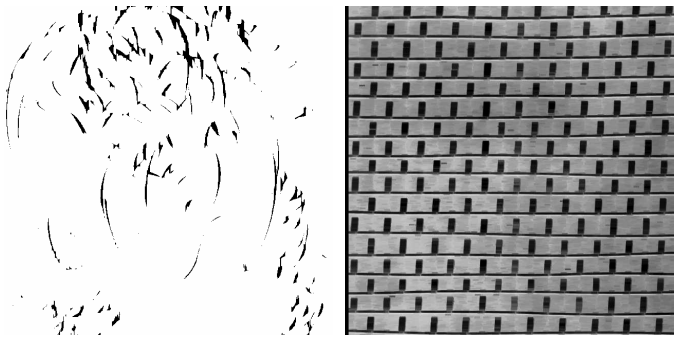
Dilation and erosion remove information and in general the lost information cannot be retrieved. The search for an operation that attempts to revert the effects of dilation and erosion leads to the definition of, respectively, morphological *closing* and *opening*. We first give the definition of opening, and for that, we define the reflection $\check{A}$ of a set A: $\check{A} \equiv \{-\mathbf{a} \mid \mathbf{a} \in A\}$. That is, $\check{A}$ is the mirror image of A about the origin.

The opening of X by B, denoted by $O_B(X)$, is defined as the erosion of X by B followed by the dilation by $\check{B}$. That is:

[2] It is easy to verify that dilation and erosion form a pair of dual transformations:

$$D_B(X) \equiv \{E_B(X^c)\}^c.$$

This duality property means that, when using the same SE, the dilation of a set X is equivalent to the complement of the erosion of the complement (i.e., the "background") of the set X.



(a) (b)

Fig. 7 (a): erosion of the image in Figure 2(a) with a vertical line as SE; (b) erosion of the image in Figure 2(b) with a rectangle as SE.

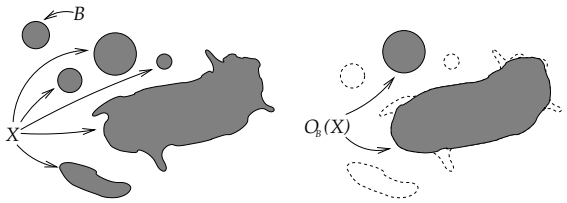


Fig. 8 Opening of a set X by a disk-shaped structuring element B.

$$O_B(X) \equiv D_{\check{B}}\{E_B(X)\}.$$

Figure 8 is an example of opening. Notice that X has been rounded by B from the inside, and that those disks which are smaller in size than the SE B *vanish* after opening.

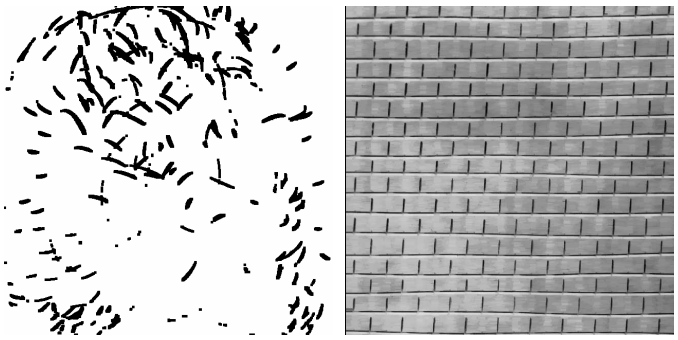
Also notice the filtering effect of opening: those image structures that cannot contain the SE B are removed from the image. Therefore the size and shape of B should be carefully chosen for the information to be extracted from the image. For example, if one wants to remove linear features but not disk shaped structures, B should be chosen as a disk of a suitable size. Examples of opened images can be found in Figure 9.

The closing of X by B, denoted by $C_B(X)$, is defined as the dilation of X by B followed by the erosion by $\check{B}$. That is:

$$C_B(X) \equiv E_{\check{B}}\{D_B(X)\}.$$

See Figure 10 for an example of closing. As opposite to opening, closing rounded the objects "from outside". See also Figure 11 for examples of closed images.[3]

[3] Opening and closing also share a dual property: $O_B(X) = \{C_B(X^c)\}^c$.



(a) (b)

Fig. 9 (a): opening of the image in Figure 2(a) with a disk as SE; (b) opening of the image in Figure 2(b) with a rectangle as SE.

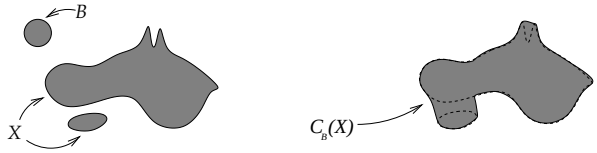
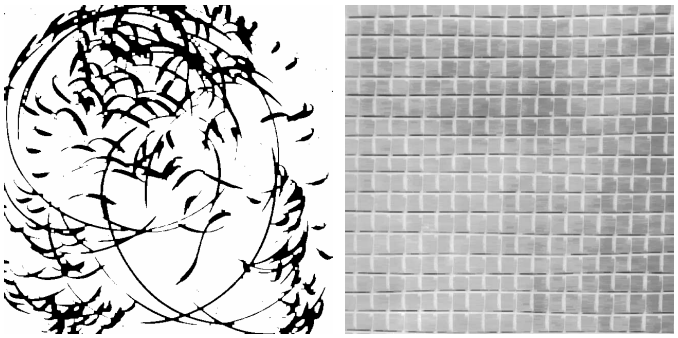


Fig. 10 Closing of a set X by a disk B.



(a) (b)

Fig. 11 (a): opening of the image in Figure 2(a) with a disk as SE; (b) opening of the image in Figure 2(b) with a rectangle as SE.

In practice the choice between opening or closing depends on the types of objects or noise to be extracted/removed. For example, the removal of "salt noise" —white dots in the image—requires opening, while "pepper noise"—black dots in the image—requires closing.

2.4 Other Morphological Operations

There are other useful morphological operators, but due to space limitation, we omit their detailed descriptions here. One such operation is *skeletonization*: the skeleton of an binary object is defined as the union of the centers of all the maximal balls inside the object. It is useful for extracting summary features to represent the object. Another useful operator for detecting object boundaries is *morphological gradient*, typically defined as the arithmetic difference $D_B(X) - E_B(X)$.

3 Detection and Classification of Sunspot Groups

We aim to develop an automatic procedure for detecting and classifying sunspot groups according to the Mount Wilson scheme. Given the complexity of the magnetogram images, we adopt an imaging-oriented modular approach. That is, the ultimate problem of detection and classification is broken into a sequence of sub-problems, and simple and effective imaging techniques are applied to sequentially solve these sub-problems.

Since the Mount Wilson scheme relies on characterizing the shape and distribution of magnetic flux in sunspot groups, mathematical morphology is utilized to extract scientifically meaningful features from the available magnetograms. That is, the morphological operations described in Section 2 are used to examine the distribution of positive and negative magnetic polarities visible in the magnetogram. In particular, we characterize the complexity of the sunspot group based on the scatter of magnetic flux and the separation of the two polarities. In this way, our procedure tailors a classifier to utilize expert knowledge in constructing an interpretable and effective classifier. Another approach to classification, at the other extreme, is to generate a large set of numerical summaries to use as features in a "blackbox" classifier. While this approach can also yield an effective classifier, the results tend to be much more difficult to interpret in terms of the underlying science.

3.1 Science-Driven Feature Extraction

In this section we describe the procedure that we employ to extract numerical summaries of the magnetogram images that will serve as features in the ultimate classification. Our strategy is to derive features that are tailored to distinguish between the four classes in the Mount Wilson scheme. Since all four classes are defined in terms of the distribution of the positively and negatively oriented magnetic fields, we begin by using morphological operators to identify the regions of positive and negative polarity in a magnetogram.

To do this we first clean the image using a morphological opening operation with a spherical structuring element of radius 2. This smooths the white sunspots—the re-

gions of positively oriented magnetic field that appear white in the magnetograms—so that smooth boundaries can be obtained after thresholding. After cleaning we extract the white sunspot by selecting pixels with magnetogram intensity greater than a given threshold, namely greater than $\bar{x} + 2.5s$, where $\bar{x}$ and s are, respectively, the mean and sample standard deviation of all the pixel values in the image. Next we aim to extract the black sunspots—the regions of negatively oriented magnetic polarities that appear black in the magnetograms. To do this,we invert the original image by multiplying by negative one so that it looks like a film negative, and then clean and threshold the inverted image in exactly the same we did with the original image when extracting the white sunspots.

Figure 12 illustrates our feature extraction routine for α, β, $\beta\gamma$, and $\beta\gamma\delta$ sunspot groups. In this figure, the first row is the original magnetogram that appears in Figure 1, the second row is the cleaned magnetogram, the third row is the extracted white sunspot, and the fourth row is the extracted black sunspot. The columns represent α, β, $\beta\gamma$, and $\beta\gamma\delta$ types, respectively. We will describe the final two rows below.

Given the extracted white and black sunspots, we are in a position to define a feature that aims to identify sunspot groups in the α class. Since this class is defined by "A unipolar sunspot group", an extreme ratio of the number of extracted pixels the white and black sunspots (N_W and N_B, respectively) should be indicative of an α group. This ratio is denoted $|N_W/N_B|$ and is given, for each representative magnetogram, beneath its respective column in Figure 12.

The difference between the β, $\beta\gamma$, and $\beta\gamma\delta$ classes is the degree of separation between the white and black sunspots. In the β class they can be largely separated by a straight line, in the $\beta\gamma$ class they can be largely separated, but not by a straight line, and in the $\beta\gamma\delta$ class they are mixed. Thus to distinguish between these groups we aim to identify the best boundary between the the white and black sunspots and to access the quality of this boundary. We do this by combining the extracted white and black sunspots into the same image and using a standard region growing operation to produce the separating boundary. In Figure 12, the fifth row shows the combined image, with the white and black sunspots plotted in blue and yellow, and the sixth row illustrates the resulting separating boundary. Notice that the boundary becomes more complex for the $\beta\gamma$ group than β group and even more so for the $\beta\gamma\delta$ group.

A natural way to distinguish β groups and $\beta\gamma$ groups is to measure the "roughness" of the separating line. A good example of roughness measure is the averaged second derivative, which we compute using second differencing. In some cases the region growing operation results in more than one separating line, indicating poor separation between the white and black sunspots. In this case the group should be classified as a $\beta\gamma\delta$ group.

To help identify sunspot groups in the $\beta\gamma\delta$ class we must quantify the degree of scatter or mixture of the region's positive and negative polarities. In order to do this we introduce a spatial complexity measure. In particular, let $\mathscr{W}$ be the set of pixels in an extracted white sunspot. We then compute the center of mass, c, of $\mathscr{W}$. For each pixel $w \in \mathscr{W}$, the number of pixels that a line segment from w to c passes

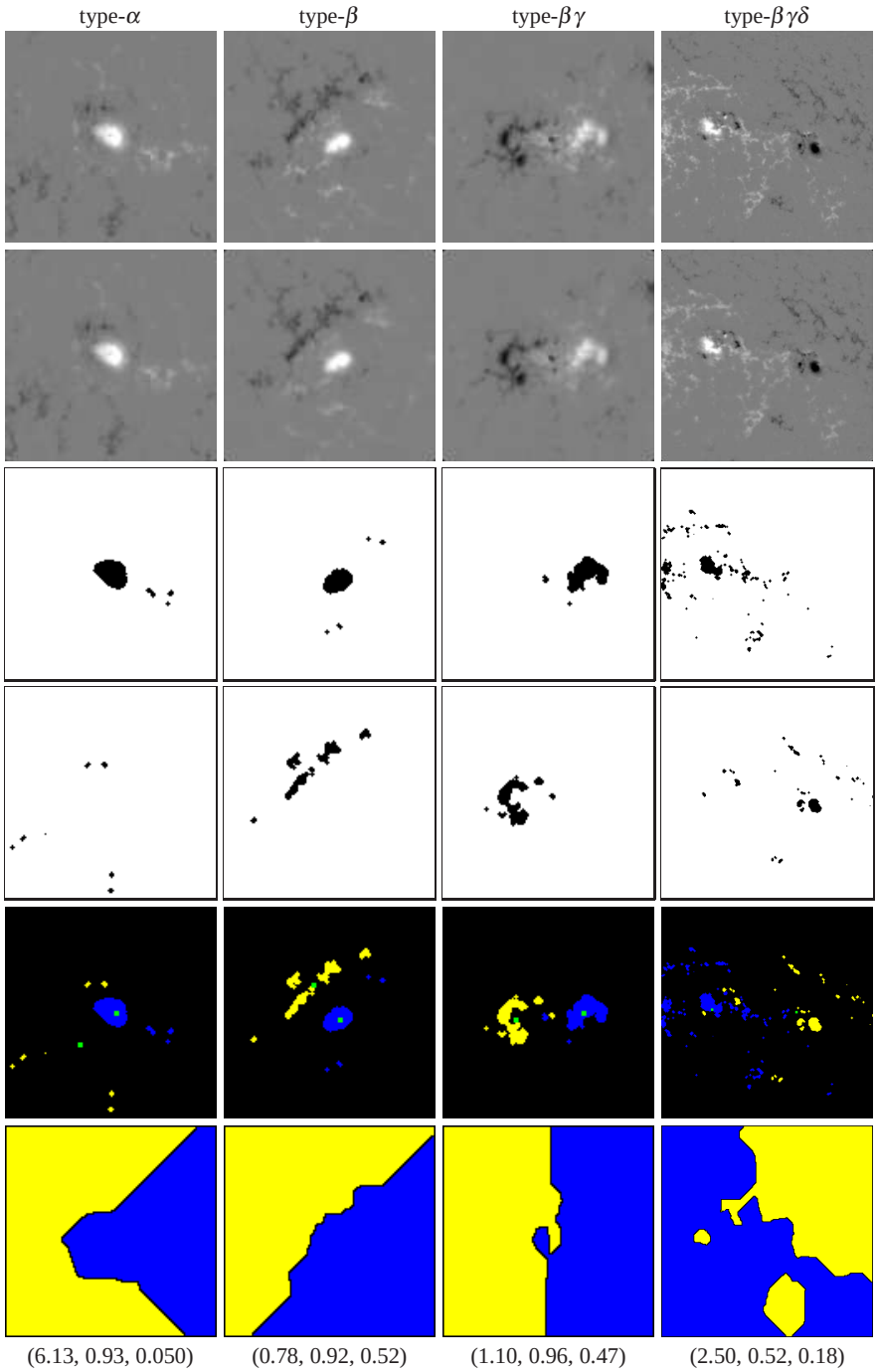


Fig. 12 Top row: original magnetograms for four types of sunspots. Second row: morphologically cleaned magnetograms. Third row: extracted white sunspot. Fourth row: extracted black sunspot. Fifth row: detected white (in blue) and black (in yellow) sunspots. The green dots are their centers of mass. Bottom row: separating line(s) between the white and black sunspots. The parenthetical summaries at the bottom are the area ratio of white to black sunspots and the spatial complexity measure $A(\cdot)$ values for the white and for the black sunspots. We expect the area ratio to be more extreme for α groups and the complexity measurements to be smaller for the $\beta\gamma\delta$ groups than for β or $\beta\gamma$ groups.

through is denoted $L(w)$ and of these, the number of blue pixels is denoted $l(w)$. (Recall that blue pixels correspond to the white sunspots.) The spatial complexity measure, $A(\mathscr{W})$, is computed as

$$A(\mathscr{W}) = \frac{1}{|\mathscr{W}|} \sum_{w \in \mathscr{W}} \frac{l(w)}{L(w)},$$

where $|\mathscr{W}|$ is the number of pixels in $\mathscr{W}$. Notice that $L(w) \geq l(w)$ and $0 \leq A(\mathscr{W}) \leq 1$. To see why $A(\mathscr{W})$ can be used as a spatial complexity measure, observe that if the white sunspot pixels are scattered (and disconnected) around in the image, then for most $w \in \mathscr{W}$, $l(w)$ is small relative to $L(w)$, and thus a small value of $A(\mathscr{W})$ indicates high spatial complexity of $\mathscr{W}$.

A similar quantity $A(\mathscr{B})$ can be computed for the set of pixels in an extracted black sunspot $\mathscr{B}$. The $A(\mathscr{W})$ and $A(\mathscr{B})$ values for each of the representative magnetograms are given beneath the columns in Figure 12. The green dots in the fifth row of Figure12 are the centers of mass of $\mathscr{W}$ and $\mathscr{B}$.

The full procedure for computing the features is outlined in Table 1.

Table 1 Feature Extraction Procedure

1. Clean the original magnetogram image using morphological operations.
2. Extract the "white sunspots" by thresholding the cleaned image.
3. Apply the above steps to the negative of the image to extract the "black sunspots".
4. Compare the relative areas of the white and black sunspots (for discriminating α from the other three types).
5. Compute the separating line for the white and black sunspots (for discriminating β and $\beta\gamma$).
6. Compute the complexity measures $A(\mathscr{W})$ and $A(\mathscr{B})$ (for discriminating $\beta\gamma\delta$ from the rest).

3.2 Classification

Given the set of four features described in Section 3.1 along with their quadratic and interaction terms, we can use a standard classification (supervised learning) technique to derive a classification rule. There are numerous possible method, but we focus mainly on the technique known as *random forests* [1] because it is relatively immune to over-fitting, meaning we have to worry less about the classifier being over-sensitive to spurious relationships in the data, even when including a large number of features. (Four features grows to 14 features if we include quadratic and interaction terms.)

A random forest is a state-of-the-art nonparametric classifier that is an ensemble of a set of *decision trees*. The individual trees are grown by finding the best split of the training cases into the classes based on a set of features. The classification in each of the resulting subgroups is improved using new separate classification rules.

In a case with N training cases and p features, the number of features used to make a decision at each node of a tree is set at r, where r is much less than p (one common technique is to set $r = \sqrt{p}$). The ensemble of trees is created by randomly selecting N cases with replacement from the original N training cases. Each tree is grown by randomly choosing r features at each node and making a split based on the selected features. Each tree is grown to completion without pruning , and the random forest combines the individual decision trees based on the majority vote of the trees.

As an illustration we randomly divided a data set consisting of 128 magnetograms into a training set of 90 (70%) magnetograms and test set of 38 (30%) magnetograms. We fit a random forest of 250 trees using the `randomForest` routine in `R` to the training set and used the resulting classification rule to separately classify both the training and test sets. While the training set had a 100% correct classification rate, 58% of the test set was correctly classified, based on the USAF/NOAA classification. All of the misclassified sunspot groups were classified into a class neighboring the USAF/NOAA classification (i.e., all α sunspot groups where classified as either α or β, all β groups as β or $\beta\gamma$, all $\beta\gamma$ as β or $\beta\gamma$, and all $\beta\gamma\delta$ as $\beta\gamma$ or $\beta\gamma\delta$.)

A difficulty that arrises when we try to evaluate the quality of our proposed features for sunspot classification is that the USAF/NOAA classification is not particularly reliable. An examination of the magnetograms that appear to be missclassified by our method more often than not reveals that the USAF/NOAA classification is incorrect or that the sunspot groups is marginal and does not clearly belong to any one of the four classes. This is of course problematic not only for evaluating the classifier but also for training the classifier because the USAF/NOAA classifications in the training set are no more reliable than those in the test set. The problem stems from the lack of true discrete classes. There is a continuum between the α class that is "dominated by a single unipolar spot" and the bipolar β class, as the second polarity grows from negligible to equal in importance. Likewise there is a continuum from the β to the $\beta\gamma$ and to the $\beta\gamma\delta$ class as the bipolar group ranges from simple distinct regions of positive and negative polarity to a group with positive and negative polarities scattered throughout. The lack of a distinct underlying classification lead to subjective assessments as to the proper classification of a group and an inherent inconsistency in the human classification. It is both difficult and ultimately fruitless to automatically reproduce such a human classification.

4 Discussion

Our ultimate goal is to provide numerical descriptions and summaries of sunspot images that capture physical characteristics in sunspot development and evolution and can be use to predict turbulent events such as solar flares and coronal mass ejections. Research suggests that the morphology of the sunspot groups is relevant to the evolution of the group and predictive of such events. Thus our work has focused on developing morphological summaries that in the first place capture scientific theo-

ries about formation and evolution and secondly may be able to be used to reproduce existing classification schemes. An immediate goal is to develop new classification schemes and/or continuous numerical summaries that better represent the observed variability in sunspot images and are more correlated with solar activity. Current classification schemes are based on static sunspot groups. A more interesting classification would characterize not just the static morphology, but also the development, evolution, and track of the group. The goal is to automatically track the formation and evolution of sunspot groups using the massive solar data sets that are now coming online—and for this tracking to be in terms of sunspot features that are most pertinent to the ultimate scientific objectives.

Acknowledgements If you want to include acknowledgments of assistance and the like at the end of an individual chapter please use the `acknowledgement` environment – it will automatically render Springer's preferred layout.

References

1. L. Breiman. Random forests. *Mach. Learn.*, 45:5–32, October 2001.
2. T. Colak and R. S. R. Qahwaji. Automated mcintosh-based classification of sunspot groups using mdi images. *Solar Physics*, 248(2):277–296, 2009.
3. H. J. Hagenaar. Ephemeral regions on a sequence of full?disk michelson doppler imager magnetograms. *The Astrophysical Journal*, 555(1):448–461, 2001.
4. J. Ireland, C. A. Young, R. T. J. McAteer, C. Whelan, R. J. Hewett, and P. T. Gallagher. Multiresolution analysis of active region magnetic structure and its correlation with the mt. wilson classification and flaring activity. *Solar Physics*, 2008.
5. R. S. R. Qahwaji and T. Colak. Automatic short-term solar flare prediction using machine learning and sunspot associations. *Solar Physics*, 2009.
6. J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
7. P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer, Berlin, second edition, 2003.