Automatic Classification and Tracking of Solar Features

David C. Stenning,[1] Vinay Kashyap,[2], Thomas C. M. Lee,[3], David A. van Dyk,[4] C. Alex Young[5]

[1] Department of Statistics, University of California, Irvine; [2] Harvard-Smithsonian Center for Astrophysics; [3] Department of Statistics, University of California, Davis; [4] Statistics Section, Imperial College London; [5] ADNET Systems, Inc., NASA/GSFC

iCHASC

Background

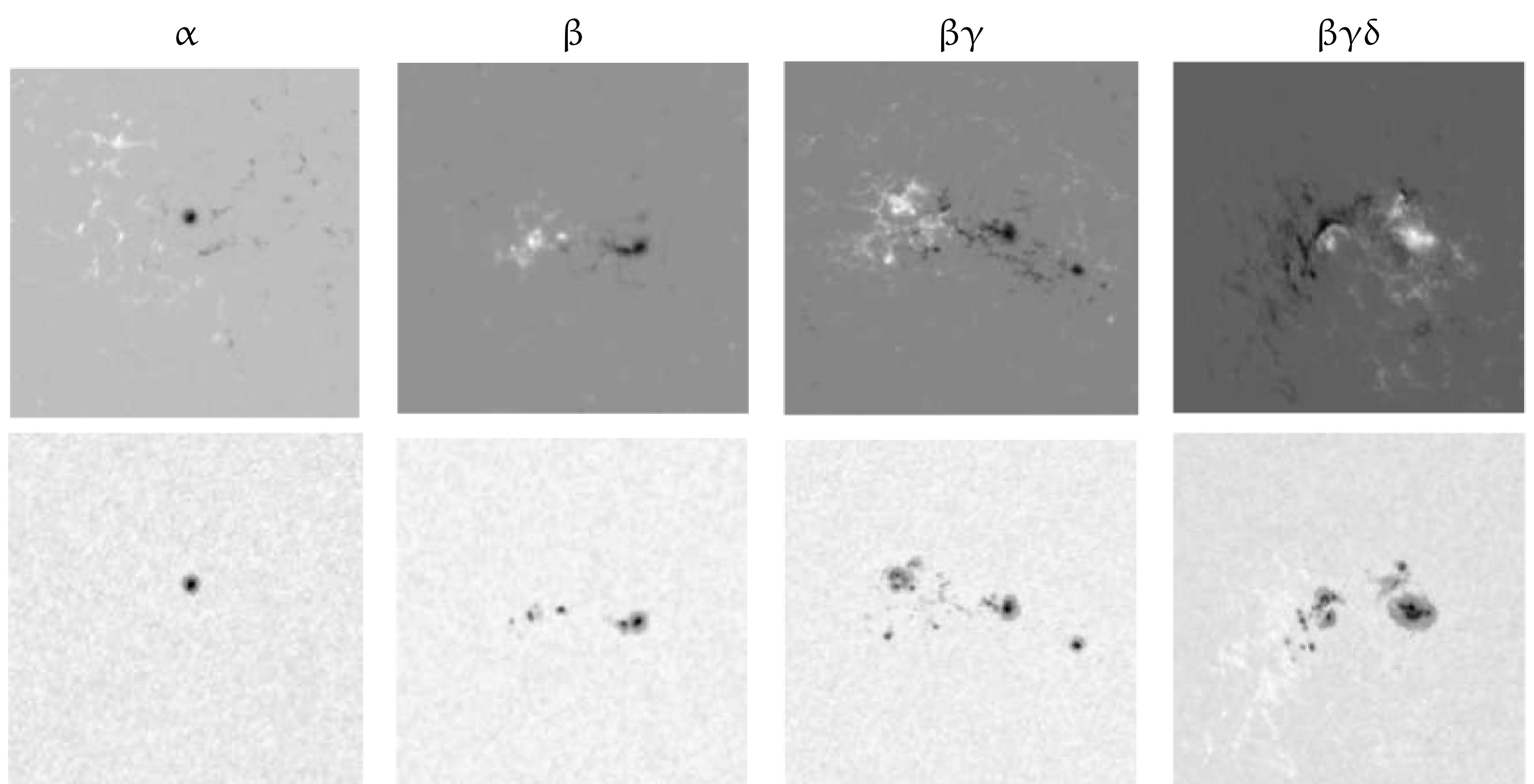
Sunspots appear as dark areas on the *photosphere*—the region of the Sun that emits the light that we see. They are formed when powerful magnetic fields inhibit convection, cooling the associated surface area which then appears as a dark spot in optical (white) light. In *magnetograms*—images of magnetic fields on the Sun—sunspots appear as areas of high magnetic flux. Energetic solar events such as *solar flares* and *coronal mass ejections* are related to sunspots through the complexity of associated magnetic active regions [1].

Currently, sunspot classification is performed manually via visual inspection by experts [2]. However, manual classification suffers from human observer bias stemming from the subjective and often ambiguous morphology of active regions [1]. That is, two experts looking at the same sunspot group may disagree as to the "correct" classification. Furthermore, with the advent of new scientific missions such as NASA's Solar Dynamics Observatory (SDO), manual classification is quickly becoming impractical. Sophisticated, robust, and automatic analysis procedures are required for handling the enormous volume of high cadence solar data that is soon to be available.

Mount Wilson Classification

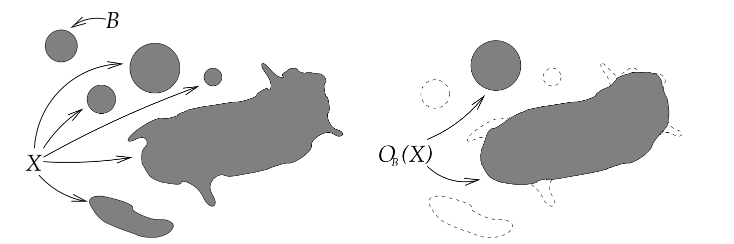
The Mount Wilson classification scheme groups sunspots into four classes based on the complexity of magnetic flux distribution in associated active regions [1]:

- α class: groups dominated by a single *unipolar* sunspot
- β class: a pair of sunspots of opposite magnetic polarity, but with a simple and distinct spatial division between the polarities
- βγ class: a bipolar group sufficiently complex that a single polarity inversion line cannot divide the two polarities
- βγδ class: a βγ group with *umbrae* of different polarity inside a single *penumbra*

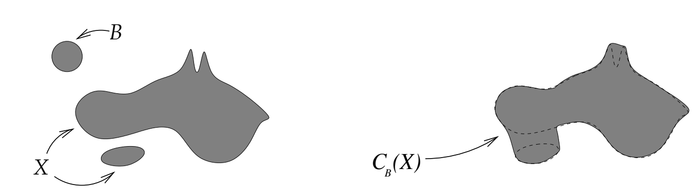


Magnetograms (top row) and white light images (bottom row) for the four sunspot classes.

Simple Morphological Operations

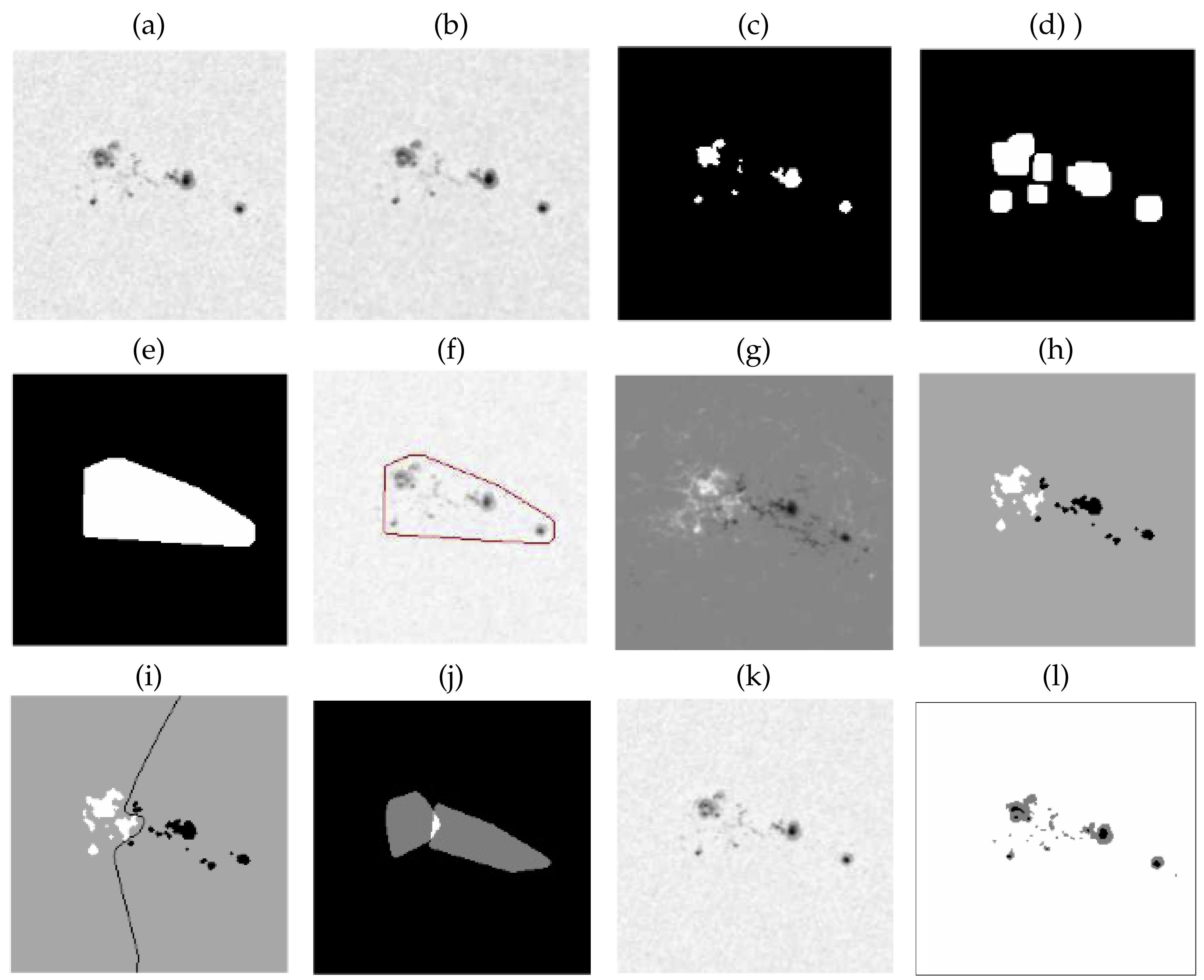


Opening of a set *X* by a disk *B*.



Closing of a set *X* by a disk *B*.

Science-Driven Feature Selection



The original βγ white light image (a) is cleaned (b) and thresholded to produce a binary representation of the sunspot group (c). This image is then dilated (d) and has a convex hull placed around the result (e) and the area inside the hull becomes the sunspot area (f) in the magnetogram. Then, in the βγ magnetogram (g), morphological opening followed by thresholding on both the image and inverse image yields the representation of the active region in (h). Region growing gives the separating boundary in (h). Convex hulls are utilized to measure polarity mixture in (j). We smooth the white light image in (k) and apply thresholding iteratively in (l) to produce a representation of the umbrae and penumbrae that can be used to detect delta spots.

Classification

Numerical summaries are calculated based on morphological representations of active regions:

- the *ratio* of pixels of opposite polarities
- the *amount of scattering* of the pixels for each polarity
- polarity inversion line *curvature*
- area of *opposite polarity mixture* for the convex hulls around each polarity region
- the *number of delta spots* detected

These summaries serve as features to a supervised learning algorithm based on a *random forest*—a state-of-the-art nonparametric classifier that utilizes an ensemble of individual decision trees [6]. To evaluate and illustrate our automatic classifier, we use a dataset consisting of 119 magnetogram and white light images capturing individual sunspot groups that have been manually classified according to the Mount Wilson scheme.

Results and Discussion

		Actual Class			
		α	β	βγ	βγδ
Predicted Class	α	8	2	0	0
	β	2	21	2	0
	βγ	0	0	2	1
	βγδ	0	0	2	4

Classification results on the test set with a random forest of 250 trees.

In our numerical results we split the data into 65% training and 35% testing data. The results of our random forest classifier on the test set are presented in the above table, and we notice that the the test sunspots are classified into either the correct class or an adjacent class. Since the classification "by eye" is prone to errors and inconsistencies, the true performance of our classifier is difficult to judge. A perfect classification rate is not necessarily the gold standard when automating a manual classification that is artificial and subjective. While manual classification schemes must necessarily rely on a discrete number of classes, the true morphology of active regions is continuous and sunspot groups can evolve smoothly from one class to another in short periods of time. As a result, there is often ambiguity as to the "true" classification of a particular sunspot group at a particular point in time. Thus, a misclassification by the random forest with respect to the manual classification may not actually be incorrect if the true class is indeterminate.

References

1. 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.
2. Tufan Colak and Rami S. R. Qahwaji, "Automated mcintosh-based classification of sunspot groups using mdi images," *Solar Physics*, vol. 248, no.2, pp. 277-296, 2009.
3. Jean Serra, *Image Analysis and Mathematical Morphology*, Academic Press, 1982.
4. Pierre Soille, *Morphological Image Analysis: Principles and Applications*, Springer, Berlin, second edition, 2003.
5. D. Stenning, V. Kashyap, T.C.M Lee, D.A. van Dyk, and C.A. Young, "Morphological image analysis and its application to sunspot classification," in *Statistical Challenges in Modern Astronomy V*, In Press.
6. Leo Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp.5-32, October 2001.

dstennin@uci.edu