Bayesian Inference for Identifying Solar Active Regions

Michael Turmon and **Saleem Mukhtar**
Machine Learning Group
Jet Propulsion Laboratory
Pasadena, CA 91109

Judit Pap
Dept. of Physics and Astronomy
University of California
Los Angeles, CA 90095

Abstract

The solar chromosphere consists of three classes — plage, network, background — which contribute differently to ultraviolet radiation reaching the earth. Solar physicists are interested in relating plage area and intensity to UV irradiance, as well as understanding the spatial and temporal evolution of plage shapes. We describe a data set of solar images, means of segmenting the images into constituent classes, and a novel high-level representation for compact objects based on a spatial 'membership function' defined via a triangulated planar graph. Segmentations are found using a discrete Markov random field setup, and the high-level representations are learned by a Markov chain Monte Carlo birth/death process on the triangulations.

Introduction

As observed in ultraviolet light (figure 1) the solar chromosphere roughly consists of three classes: plage (bright magnetic disturbances), network (hot boundaries of convection cells), and background (cooler cell interiors). Plages appear as irregular groups of clumps and experience a cycle of formation and dissipation, starting out as relatively compact regions and decaying over many days into a diffuse and broken-up cluster (Zirin, 1988, p. 317). The cell-structured network has little contrast with the background, is spatially homogeneous, and persists for tens of hours. The cells (difficult to see in this halftoned rendering) have a characteristic size, and it is thought that they arise due to convective processes in the plasma making up the solar atmosphere (Zirin, 1988, p. 126).

The three classes contribute differently to the ultraviolet (UV) radiation reaching Earth's upper atmosphere, with the plages and network giving the largest contribution. While this radiation cannot be sensed directly from the ground, the features giving rise to it can be. Such measurements are inputs to models of solar irradiance which are crucial to understanding phenomena such as global warming and atmopheric photochemistry (Withbroe and Kalkofen, 1994).

Also of interest is the evolution of plages. As mentioned above, a typical sequence has been described: from plage emergence as a shape of relatively smooth boundary, to expansion, and then eventual dissolution as an irregular, tentacled form. However, the understanding is of a qualitative and anecdotal sort (e.g. (Stix, 1991, p. 284) for related work), and a more quantitative description of anticipated plage shapes and the evolution of plage regions would be of value.

In both sorts of problem described above one reduces a series of images, comprising perhaps 500MB of data, to a time series of areas and intensities or plage descriptions. For example, per-class areas and mean intensities can be represented in about 48 bytes per 9MB image. The description of the features of interest for one plage might take a few KB per image, depending on the activity level. These time series distill the diffuse information in the large data set into a scientifically usable form, at least for the tasks at hand.

The primary source of data for this study is the set of CaII K full-disk spectroheliograms that has been taken on film, daily or as observing conditions permit, at Sacramento Peak National Solar Observatory in Sunspot, NM from the mid-sixties onward. An interval of these films, from the mid-eighties forward, has been digitized to 2K×2K pixels, at which point atmospheric blurring limits resolution.

Image Decomposition

First we discuss the problem of partitioning the image into plage, network, and background components. Generally, scientists either apply a threshold across the flattened image to determine plage areas, or manually surround the plages with polygons. The first method, while simple and objective, ignores all available spatial information. The second method uses substantial domain knowledge, but is also highly subjective, difficult to describe, and hard to repeat.

Due to the strong prior information available to us about the images, we adopt the Bayesian framework of inference of underlying pixel classes based on the

[2]This work was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

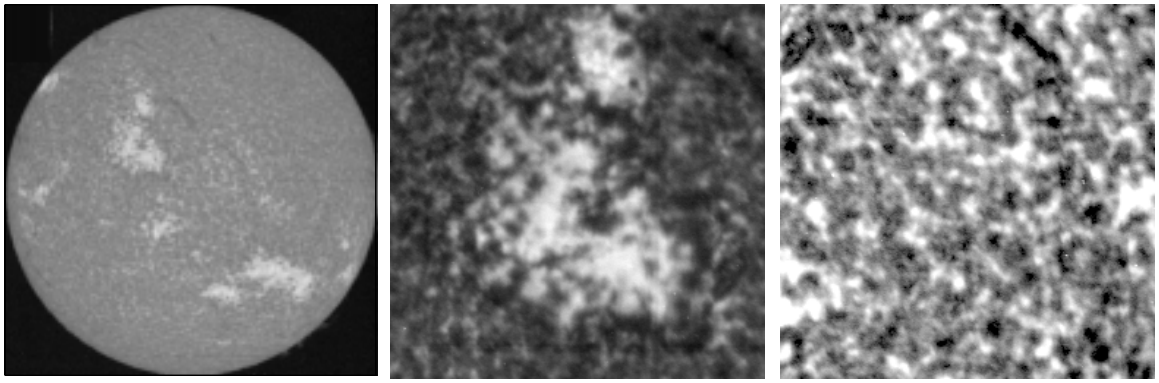

Figure 1: A chromospheric image from 15 July 1992, showing several plages. In the center is a detail image of the northwest plage pair; at right is a detail from disk center, with considerable contrast enhancement, showing the network and background.

observed intensity. Denoting pixel sites $s = [s_1\, s_2]$ in an image domain N, and defining matrices of class labels $\mathbf{x} = \{x_s\}_{s \in N}$ and observed intensities $\mathbf{y}$, the posterior probability of labels given data is

$$P(\mathbf{x} \,|\, \mathbf{y}) = P(\mathbf{y} \,|\, \mathbf{x}) P(\mathbf{x}) / P(\mathbf{y}) \propto P(\mathbf{y} \,|\, \mathbf{x}) P(\mathbf{x}) \ . \quad (1)$$

The maximum a posteriori (MAP) rule maximizes this probability:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \log P(\mathbf{y} \,|\, \mathbf{x}) + \log P(\mathbf{x}) \ . \quad (2)$$

The first term is the familiar likelihood function, telling how the data is gotten from the labels; the second is the prior probability of a given labeling. In practice, the first term forces fidelity to the data while the second penalizes unlikely rough labelings.

Prior models may be specified in many ways; we have used the smoothness priors

$$P(\mathbf{x}) = Z^{-1} \exp\big[-\beta \sum_{s' \sim s} 1(x_{s'} \neq x_s)\big] \quad (3)$$

introduced by Besag (Besag, 1974) and Geman and Geman (Geman and Geman, 1984). Above, Z is a constant normalizing the distribution, and the summation extends over all 'neighboring' pixels ($s \sim s'$) in N. Below we have taken the neighborhood relation to include all pixels strictly less than two units apart in Euclidean distance — each interior pixel has eight neighbors. For $\beta = 0$, this distribution is uniform on all $3^{\mathrm{card}(N)}$ labelings, and as β is increased, smoother and smoother labelings are favored.

The remaining ingredient is the likelihood

$$P(\mathbf{y} \,|\, \mathbf{x}) = \prod_{s \in N} P(y_s \,|\, x_s) \quad (4)$$

where we assume that intensities are independent conditional on the labels being known. The three densities $P(y \,|\, x = k)$ can be estimated from labeled data supplied by scientists. It is not surprising that the plage and network intensities have a heavy tail — making a normal distribution inappropriate. Nonparametric distributional tests confirm that the lognormal distribution is a good model for the per-class intensities.

The objective function of (2) becomes

$$-\sum_{s \in N} \left(\frac{(\log y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} + \log \sigma_{x_s} \right) - \beta \sum_{s \sim s'} 1(x_s \neq x_{s'})$$

The tradeoff between consistency of each observed intensity with the mean of its assigned class, and agreement of neighboring class labels, is apparent. If $\beta = 0$ and the class variances are identical we recover the threshold rule currently used in practice.

However, with $\beta > 0$, the optimization becomes coupled across sites, and is entirely intractable for our three-class problem. To tackle this problem we have followed the well-known numerical method known as the Gibbs sampler (Geman and Geman, 1984). In brief, this works by cycling through each site, computing $P(x_s \,|\, y_s,\, x_{N(s)})$ for each class, and choosing the next label from this distribution. For finite label spaces, the resulting (random) sequence of labelings converges in distribution to the posterior. To extremize the posterior, one sharpens the distribution by decreasing a scale parameter slowly to zero, and the resulting labeling is the MAP estimate.

Sample results are shown in figure 2. The first panel shows a detail of a spectroheliogram from January 1980; the plage is at lower-right. Beside this is the corresponding threshold segmentation. The abundant speckle is consistent with the implicit prior that is uniform over all labelings. In the final panel is the MAP segmentation with MRF prior at $\beta = 0.7$. The estimate is found by the standard Gibbs sampler approach with temperature lowered in steps over 800 image sweeps. The MAP/MRF segmentation eliminates many of the tiny gaps in the large plage and makes the network structure more apparent.

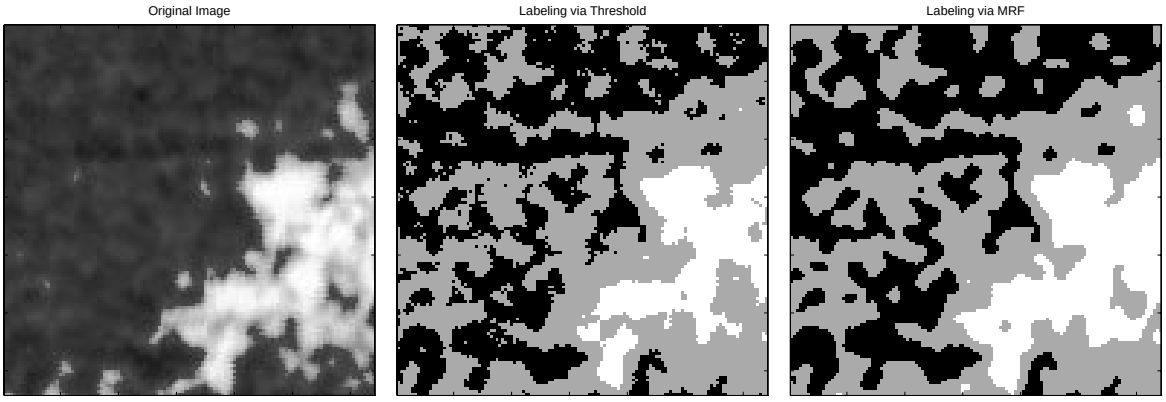

Figure 2: Original image detail, threshold segmentation, MAP/MRF segmentation with $\beta = 0.7$.

Spatial Descriptions

Now we address the second of the concerns raised in the introduction, that of representing and analyzing plage shape. In contrast to the essentially pixel-scale characteristics of the network/background interplay, plages are high-level phenomena which are not well-captured by pixel-level rules. Following the lead of Grenander (Grenander *et al.*, 1991), we pursue a hierarchical representation of plages. We will find it convenient to embed the pixel sites N in a bounded continuum $\bar{N} \subset R^2$. To represent a plage, or a cluster of related plages, we propose a tent-like structure defined by a triangulated planar graph

$$G = (V, E, h) \tag{5}$$

$V \subset \bar{N}$	a vertex set
$E \subset \bar{N}^2$	an edge relation
$h\colon V \to [0,1]$	a height function

The height function extends to all of $\bar{N}$ by linear interpolation across the faces of the pyramids (figure 3). This structure is intended to model the "degree of membership" of a given pixel in the plage class, and allows the binding of nearby plage regions into one coherent object. If the height function is thresholded, the resulting shape is a cluster of regions bounded by (not necessarily convex) polygons. This is similar to the way scientists currently delimit plage regions manually.

To define a probability distribution on these structures, we generate each as the Delaunay triangulation (Aurenhammer, 1991) of independently chosen points in $\bar{N}$. These points comprise V, and E is generated mechanically as the Delaunay triangulation of V. Heights are then assigned independently to the members of V to form tie-points. The probability density of such a height function h is

$$P(h) = Z^{-1} e^{-\gamma \operatorname{card}(V_h)} \tag{6}$$

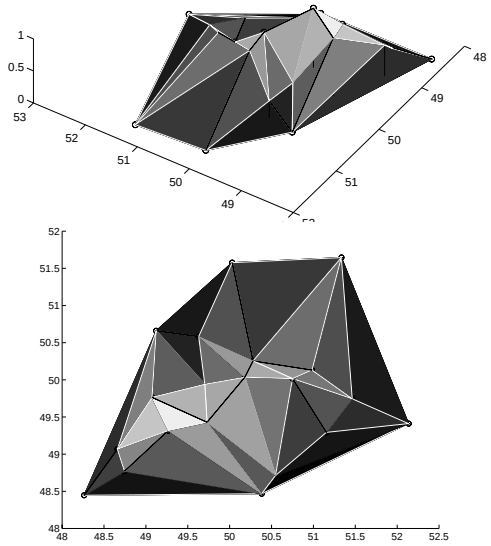

Figure 3: Top, a perspective view of a ridge structure; bottom, an air view of the same structure.

and zero if the height function is not generated as such a triangulation. We have assumed the members of V are chosen according to the uniform distribution on $\bar{N}$, and that the heights are uniform on $[0, 1]$. While refinements (self-avoiding vertices, correlated heights) are possible, their ultimate effect in the presence of data would be minimal and not worth the added model complexity. Another advantage of this distribution is that additions, deletions, and adjustments of one vertex have a simple effect on the cost, and a local effect on the triangulation and the resulting cost function.

Overlaying the new structure on the existing MRF model is simple; we wish to force agreement between

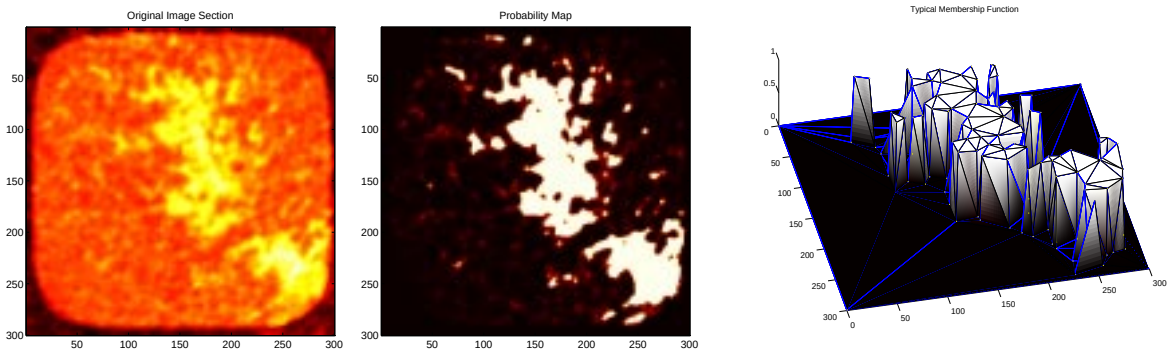

Figure 4: Left, original image, windowed to avoid edge effects; center, plage probability; right, membership function.

plage labels and the membership function

$$P(h \mid \mathbf{x}) = \frac{1}{Z} \exp\Big[-\sum_{s \in N} \big|h_s - 1(x_s = 3)\big| - \gamma \,\mathrm{card}(V_h)\Big]$$

The goal becomes to adapt h to achieve a parsimonious description while fitting the pixels which indeed belong to the object of interest.

A procedure related to the Gibbs sampler is followed to infer h for a given labeling $\mathbf{x}$. Earlier, movements in the parameter space were label-changes and were done via the Gibbs sampler; now such updates correspond to altering the plage graph, and are better done by the simpler Metropolis steps. Such a step proposes a new state h', computes $P(h' \mid \mathbf{x})/P(h \mid \mathbf{x})$, and probabilistically accepts or rejects h' on this basis. To propose a new state, one of three operators is chosen randomly at each iteration: vertex move, vertex raise (or lower), and vertex birth/death.

Move and raise are elementary as they are self-inverses as long as isotropic vertex-moves and symmetric vertex-raises are used. The birth/death pair is harder because such moves are not self-inverse, i.e. the inverse of a birth when $\mathrm{card}(V) = k$ is a death when $\mathrm{card}(V) = k+1$. To ensure equilibrium at the distribution above, the acceptance probability is chosen following the recent work of P. Green (Green, 1995). Finally, to speed the sampling process the indicator $1(x_s = 3)$ above is replaced with its expectation $P(x_s = 3 \mid y_s)$.

Some results are shown in figure 4. The first panel shows the original solar image, and the second panel is the 'probability map' or plage probability conditioned on the observed data. It is this map that the graph is intended to fit. The third panel shows a typical triangulation after a burn-in period of 15 000 successful Metropolis steps. It is clear that the triangulation has captured the essentials of the plage shape.

Conclusions

We described two scientific problems of relating solar active regions to solar irradiance, and understanding the evolution of active regions. Currently, scientists often label images manually, or by thresholding the observed intensities. The use of MRF image priors allows the controlled, objective incorporation of simple kinds of prior knowledge about the spatial coherence of labels. By using these priors in a Bayesian inference setup, images are segmented without the speckle artifacts associated with threshold labeling. Also, in an effort to understand the temporal evolution of plage shapes, we have proposed a representation of active regions in terms of a triangulated graph which gives rise to a membership function that is learned from image data.

References

F. Aurenhammer. Voronoi diagrams — A survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, 1991.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Jour. Royal Stat. Soc. Ser. B*, 36:192–236, 1974.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 6:721–741, 1984.

P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Technical report, Dept. of Mathematics, Univ. of Bristol, 1995.

U. Grenander, Y. Chow, and D. Keenan. *Hands: A Pattern-Theoretic Study of Biological Shapes.* Springer, 1991.

M. Stix. *The Sun: An Introduction.* Springer, 1991.

G. L. Withbroe and W. Kalkofen. Solar variability and its terrestrial effects. In *The Sun as a Variable Star: Solar and Stellar Irradiance Variations*, pages 11–19. Cambridge Univ., 1994.

H. Zirin. *Astrophysics of the Sun.* Cambridge Univ., 1988.