Recognizing Chromospheric Objects via Markov Chain Monte Carlo

Michael J. Turmon and Saleem Mukhtar
{turmon,mukhtar}@jpl.nasa.gov
Jet Propulsion Laboratory
Pasadena, CA 91109

Abstract

The solar chromosphere consists of three classes which contribute differentially to ultraviolet radiation reaching the earth. We describe a data set of solar images, means of segmenting the images into the constituent classes, and a novel high-level representation for compact objects based on a triangulated spatial 'membership function.' Such representations are fitted in a variable-dimension Markov chain Monte Carlo scheme.

1 Introduction

The solar chromosphere, observable (see figure 1) in ultraviolet light, roughly consists of three classes: plage (bright magnetic disturbances), network (hot boundaries of convection cells), and background (cooler interiors of these cells). Plages appear as irregular groups of clumps, seldom near the solar poles. Similar to sunspots, plages 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. The CaII-structured network has little contrast with the background, is spatially homogeneous, and persists for tens of hours. See [8] for more on chromospheric features.

The three classes contribute differently to the ultraviolet (UV) radiation reaching Earth's upper atmosphere, with the plages and magnetic network giving the largest contribution. This radiation cannot be sensed directly from the ground but the features giving rise to it can be; they are used as proxy inputs to models of solar irradiance. These models are crucial to understanding phenomena such as global warming and photochemical decomposition processes in the upper atmosphere [7].

Also of interest is the evolution of plages. Current understanding (outlined above) is of a qualitative and anecdotal sort and a more quantitative description of anticipated plage shapes and the evolution of plage regions would be of value.

The primary source of data for this study is the set of CaII K full-disk spectroheliograms that has been collected daily at Sacramento Peak National Solar Observatory from the mid-sixties onward. The images are recorded on photographic film, an interval of which (from the mid-eighties forward) has been digitized to 2K×2K pixels.

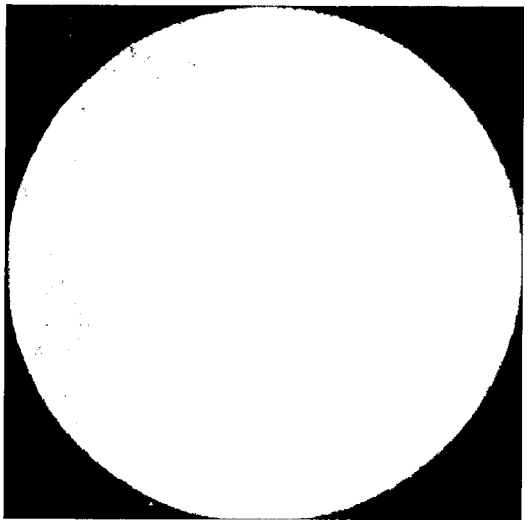

Figure 1: A full-disk chromospheric image from 15 July 1992, showing a decayed plage pair in the northwest quadrant of the sun, and a younger, more concentrated plage in the southeast.

2 Image Decomposition

First we discuss the problem of partitioning the image into plage, network, and background components. Scientists often 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 spatial information that is available. The second method clearly uses a large amount of side information possessed by the scientists, but is also highly subjective, difficult to even describe, and hard to repeat.

While the Bayesian framework is not universally appropriate for inference problems, in the situation at hand the prior information is so apparent that approximating it seems better than ignoring it. Accordingly, we establish the well-known Bayesian formalism. Denoting pixel sites $s = [s_1, s_2]$ in an image domain N, and defining matrices of class labels $\mathbf{X} = \{x_s\}_{s \in N}$, tak-

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

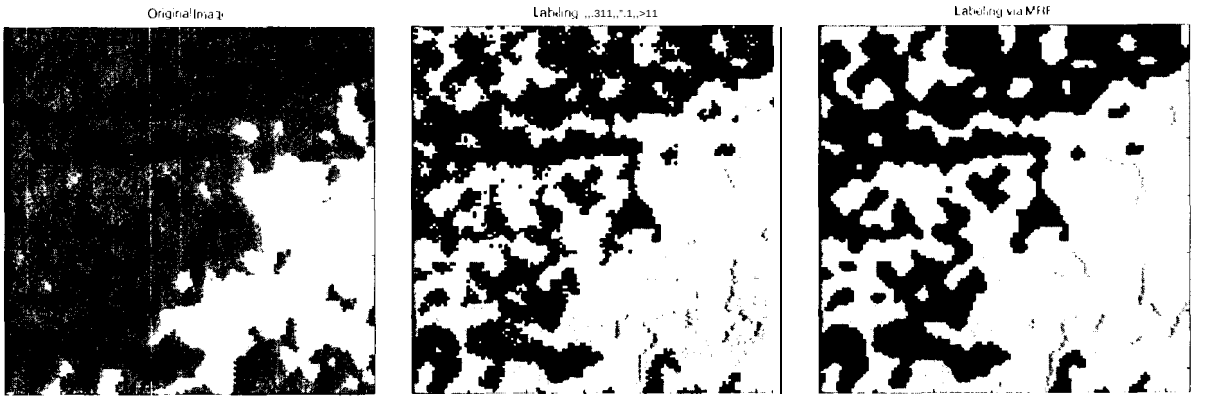

Figure 2: An image detail having network and plage elements is shown with threshold and MRF labelings.

ing values in the set {P, N, B}, and observed intensities y, the maximum a posteriori (MAP) decision rule is

$$x = \arg\max_{x} \log P(\mathbf{y} \mid \mathbf{x}) + \log P(\mathbf{x}) \qquad (1)$$

In practice, the first term forces fidelity to the data while the second penalizes unlikely rough labelings.

Prior models $P(\mathbf{x})$ may be specified in many ways. We have used the Markov field smoothness priors

$$P(\mathbf{x}) = Z^{-1} \exp\Big[-\beta \sum_{s' \sim s} 1(x_{s'} \neq x_s)\Big] \qquad (2)$$

for $\beta \geq 0$ [1]. The constant Z is chosen to normalize the probability mass function, and the sum extends over 'neighboring' sites in IV. 0110111 rectangular grid, sites are neighbors if they adjoin vertically, horizontally, or diagonally. As β drops, rougher labelings are allowed, and the uniform distribution is obtained at $\beta = 0$.

The remaining ingredient is the likelihood

$$P(\mathbf{y} \mid \mathbf{x}) = \prod P(y_s \mid x_s) \qquad (3)$$

The three densities $P(y \mid x)$ can be estimated from labeled images supplied by scientists. We have found that the lognormal distribution is a good model for the per-class intensities.

The objective function of (1) becomes

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

If the class variances are identical, and $\beta = 0$, we recover the threshold rule currently used in practice.

To tackle the optimization problem for general β we have followed the well-known Gibbs sampler scheme with Ripley's 'clock' modification [6, p. 99]. Sample results are shown in figure 2. The first panel shows a piece of a chromospheric image from January 1980 with a plage in the lower-right corner. 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 a temperature lowered in steps (discretized cooling with a geometric rate) over 800 image sweeps. We note that the MAP/MRF segmentation eliminates many of the tiny gaps in the large plage and makes the network St 1"(1('1111'(more apparent

3 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 Grenander (e.g., [4]), we pursue a hierarchical representation of plages. It is convenient to embed the pixel sites N in a continuum $\bar{N} = [0, 1]^2$. To represent a plage, or a cluster of related plages, we propose a tc;lt-like structure defined by a triangulated planar graph

$$G = (V, E, h) \qquad (4)$$

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

The height function extends to all of $\bar{N}$ by linear interpolation across the faces of the pyramids. This St P(let ure models the "degree of membership" of a given pixel in the plage class and allows the binding of nearby plage regions into one coherent object. We note that, if the height function is thresholded at a given level, the resulting shape is a cluster of regions 1)011 [1(1('(1 by polygons the same way scientists currently delimit plage regions manually. See figure 3.

To define a probability (distribution on membership functions, we generate each as the interpolated version

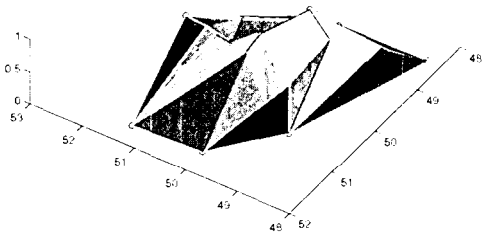

Figure 3: A perspective view of a ridge structure having about twenty vertices.

of the Delaunay triangulation of independently chosen points in $\bar{N}$. These points comprise V, and E is generated mechanically as the triangulation of V. Heights are then assigned independently to the members of V to form tie-points. The probability density of such a membership function is induced by the *one on V*:

$$P(h) = Z^{-1} e^{-\gamma \operatorname{card}(V_h)} \quad . \tag{5}$$

We have assumed the members of V are chosen according to the uniform distribution on $\bar{N}$, and that the heights are uniform 011 [0, 1]. A computational advantage of this scheme is that additions, deletions, and adjustments of one vertex have a local effect on the triangulation. Also, the penalty in log-probability paid by joining two separated graphs is the sum of component penalties, so that separated plages co-exist independently.

Incorporating the new structure into the existing MRF model is done, in brief, by letting the height function $h(s)$ favor the event $\{x_s = \mathrm{P}\}$. Specifically, let there be a Markov relationship between the three 1('1'('1S Of the Stochastic model so that

$$P(h, \mathbf{x}, \mathbf{y}) = P(h) P(\mathbf{x} \mid h) P(\mathbf{y} \mid \mathbf{x}) \tag{6}$$

In addition to $P(h)$ defined in (5), we let

$$\log P(\mathbf{x} \mid h) = K_h + \tag{7}$$
$$\beta \sum_{s \sim s'} 1(x_s \neq x_{s'}) + \alpha \sum_{s \in N} |1(x_s = \mathrm{P}) \; h(s)|$$

$$\log P(\mathbf{y} \mid \mathrm{x}) = \sum_{s \in N} \left(\frac{(\log y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} + \log \sigma_{x_s} \right) \tag{8}$$

where K_h is an appropriate normalizing constant (The corresponding constant $\log(2\pi)/2$ in the second equation has been dropped.) In this way the height function biases the corresponding label in favor of the plage Class. This conditional probability model allows us **to** generate random height functions, labelings, and images that are more physically reasonable than the unadorned MRF scheme, because larger-scale image characteristics are honored.

This describes the 'synthesis problem' (cf. [4]); the complementary 'analysis problem' focuses on the posterior

$$P(h, \mathbf{x} \mid \mathrm{y}) = P(h, \mathbf{x}, \mathbf{y}) / P(y) \propto P(h, \mathbf{x}, \mathbf{y})$$

As in section 2, we pursue the MAP estimate of the combined description. One technical difficulty is the normalizing constant K_h which figures in the posterior. In what follows, we have assumed that the variation of K_h with respect to h is negligible compared to the designed variation ill $P(h, \mathrm{X}, \mathrm{y})$, leading to an approximate posterior $\pi(h, \mathrm{X})$ with negative log-probability

$$\beta \sum_{s \sim s} 1(x_s \neq x_{s'}) + \alpha \sum_{s \in N} |1(x_s = \mathrm{P}) \; h(s)| +$$
$$\sum_{s \in N} \left(\frac{(\log y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} + \log \sigma_{x_s} \right) + \gamma \operatorname{card}(V_h) \tag{9}$$

having a *minimum* at $(\hat{h}, \mathrm{x})$. Inference proceeds in much the same way section 2, except that two sorts of variables must be varied. To simplify the discussion, we henceforth consider only updates to h with x held fixed; the other case proceeds as before.

Updates of h correspond to all ering the vertex list, and are done with simple Metropolis-Hastings steps [2] because the conditional (list ributions needed by the Gibbs sampler are intractable. Such a step proposes a new state h', computes $\rho(h, h') := \pi(h', \mathbf{x}) / \pi(h, \mathbf{x})$, and probabilistically accepts or rejects h' largely on this basis; this results in a Markov transition kernel $Q(v, dv')$ 011 the composite vertex-list set $\mathcal{V} = \cup_k \mathcal{V}_k$. 1 f Q is designed properly, it has the posterior π as its stationary distribution. Beyond the obvious restrictions t I lilt Q be aperiodic and irreducible, it is sufficient that Q maintains detailed balance: under π, the mass moving directly from A (V to B equals that moving in the reverse direction.

First we describe a set of ol)('rater's" complete enough to ensure irreducibility. A 'vertex move' operator M chooses a vertex at random and displaces it randomly. A 'vertex raise' operator R raises or lowers a vertex at random. To allow movement between the constituent spaces of $\mathcal{V}$, we have 'add' operators A_k, and corresponding 'kill' operators A'_k, which move back and forth between $\mathcal{V}_k$ and $\mathcal{V}_{k+1}$.

Next, we define a transition kernel Q on the basis of these operators; this kernel is a 'hybrid sampler' composed of each of the three move-types (M, R, A/A'). In each epoch in the simulation, one such move-type is chosen at random. Ensuring detailed balance within each move-type yields (detailed balance in the superposition. Obtaining detailed balance in types M and 11' is trivial provided the (list ribution of the ad- (I it ive displacement is symmetric. (Modular addition will eliminate edge conditions.) Operators M and 11' are accepted with probability $\min(1, \rho(h, h'))$.

Obtaining detailed balance **of** A_k, A'_k is more complex because the flow between two different Euclidean spaces must be equalized. Following recent work of 1'. Green [3], we find the chance of accepting a proposed deletion of v^* via A'_k should be the lesse(,1' of unity and

$$\mathrm{OO}/,///) \times \frac{P(\text{select } A_k)}{P(\text{select } A'_k)} \quad \frac{p_v(v^*)}{\text{'l / (k - l 1)}} \tag{lo}$$

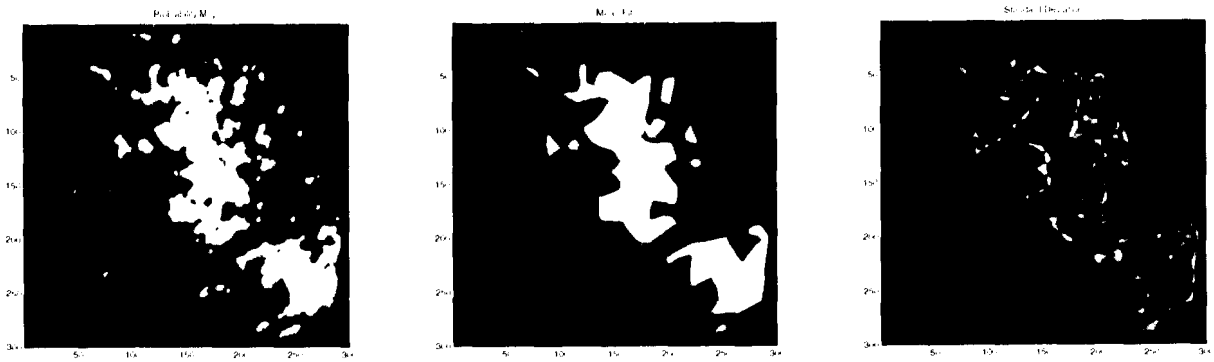

Figure 4: Plage probability; mean inferred membership function; standard deviation of membership functions

(Here p_v is a density used to choose a new point for an add operation; in practice it is used to focus attention on interesting parts of the image.) The intuition is simple: the more likely it is to attempt deletion, the less likely we must be to accept it. The more likely it is to add v^* back in, the more willing we are to delete it. The factor of $k+1$ comes from the random choice of which vertex to delete: when v^* is added via A_k, there is one chance in $k+1$ that a subsequent application of A'_k will consider v^* for deletion.

Initialization is important since a small feature may become hidden in a large triangle so that that π is not increased by any single vertex addition. The initialization procedure should therefore ensure locality of the effects of changes. A procedure that has proven effective is to initially replace the term of π enforcing agreement between h and the plage probability with one penalizing per-triangle inhomogeneity:

$$\sum_T |T|\, q_T(1 - q_T) \quad \text{, with} \quad q_T := |T|^{-1} \sum_{s \in T} 1(x_s = \mathrm{P})$$

and $|T|$ the number of pixels in triangle T. The modified criterion subdivides the image during an initial phase of 1000 epochs; then it is gradually replaced by the final criterion in a secondary stage twice this length. By the end of the second stage, a satisfactory basin of $\pi(h, \mathrm{x})$ has been found and the Metropolis iteration proceeds as described above.

Finally, to speed the sampling process the indicator $1(x_s = \mathrm{P})$ above is replaced with its expectation $P(x_s = \mathrm{P} \mid \mathrm{y})$. This is analogous to the use of conditional expectation in the ICE algorithm of A. Owen[5] and allows the sampler to directly access the uncertainty in the label, instead of reacting to its probabilistic fluctuations as Gibbs iterations proceed.

Sample results for fitting a rather complex plage pair are shown in figure 4. Fits with $\gamma = 2$, $\alpha = 0.4$ were obtained from a total of 30 000 Metropolis proposals taking 170 seconds of computation time on a Sun Ultrasparc. Roughly 175 proposals/sec are made by exploiting the significant cancellation in the quotient $\rho(h, h')$: only the changed triangles need be reconsidered. As desired, the membership function has suppressed the small-scale features and identified the two main objects and their principal outliers. The right-hand plot shows most of the variability in the fits is at the boundary, especially where a sharp projection occurs.

Acknowledgments

Thanks to Eric Mjolsness of JPL for helpful suggestions on the image representations used here, and to Judit Pap of the UCL Department of Astronomy and Astrophysics for help in understanding the scientific context of this problem. Delaunay triangulations were computed with the Triangle package authored by J. R. Shewchuck.

References

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

[2] J. Besag, P. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems. *Statistical Science*, 10:3–66, 1995.

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

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

[5] A. Owen. Discussion of Ripley, "Statistics, images, and pattern recognition". *Canad. Jour. Statist.*, 14:106–110, 1986. Article covers pp. 83–111.

[6] B. D. Ripley. *Statistical Inference for Spatial Processes*. Cambridge Univ., 1988.

[7] 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.

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