Anisotropic Smoothing of Posterior Probabilities

Patrick C. Teo [1], Guillermo Sapiro [2], Brian A. Wandell [3]

Dept. of Computer Science, Stanford University, Stanford, CA 94305 [1]
Dept. of Electical and Computer Eng., University of Minnesota, Minneapolis, MN 55455 [2]
Dept. of Psychology, Stanford University, Stanford, CA 94305 [3]

Abstract

Recently, we proposed an efficient image segmentation technique that anisotropically smoothes the homogeneous posterior probabilities before independent pixelwise MAP classification is carried out [11]. In this paper, we develop the mathematical theory underlying the technique. We demonstrate that prior anisotropic smoothing of the posterior probabilities yields the MAP solution of a discrete MRF with a non-interacting, analog discontinuity field. In contrast, isotropic smoothing of the posterior probabilities is equivalent to computing the MAP solution of a single, discrete MRF using continuous relaxation labeling. Combining a discontinuity field with a discrete MRT is important as it allows the disabling of clique potentials across discontinuities. Furthermore, explicit representation of the discontinuity field suggests new algorithms that incorporate properties like hysteresis and non-maximal suppression.

1. Introduction

In [11], we proposed a new segmentation technique that was applied to segmenting MRI volumes of human cortex. The technique comprise three steps. First, the posterior probability of each pixel is computed from its likelihood and a homogeneous prior; i.e., a prior that reflects the relative frequency of each class but is the same across all pixels. Next, the posterior probabilities for each class are anisotropically smoothed (using a 3D-extension of the algorithm suggested by Perona and Malik [9]). Finally, each pixel is classified independently using the MAP rule. Fig. 1 compares the classification of cortical white matter with and without the anisotropic smoothing step. The anisotropic smoothing produces classifications that are qualitatively smoother within regions while preserving detail along region boundaries. The intuition behind the method is straightforward. Anisotropic smoothing of the posterior probabilities results in piecewise constant posterior probabilities which, in turn, yield piecewise "constant" MAP classifications.
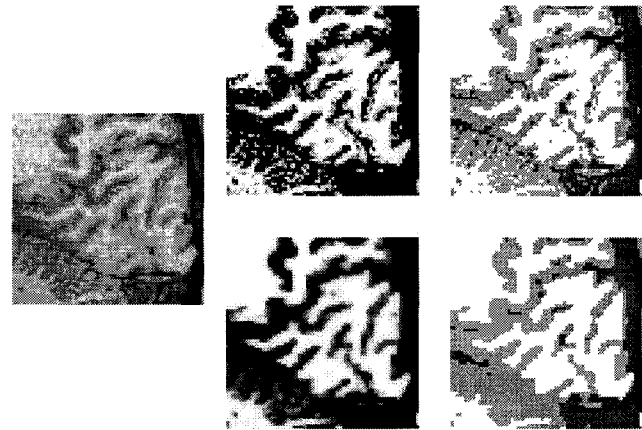
Figure 1: (Top row) Left: Intensity image of MRI data. Middle: Image of posterior probabilities corresponding to white matter class. Right: Image of corresponding MAP classification. Brighter regions in the posterior image correspond to areas with higher probability. White regions in the classification image correspond to areas classified as white matter; black regions correspond to areas classified as CSF. (Bottom row) Left: Image of white matter posterior probabilities after being anisotropically smoothed. Right: Image of MAP classification computed using smoothed posteriors.

In this paper, we explore the mathematical theory underlying the technique. We demonstrate that prior anisotropic smoothing of the posterior probabilities yields the MAP solution of a discrete MRF with a non-interacting, analog discontinuity field. In contrast, isotropic smoothing of the posterior probabilities is equivalent to computing the MAP solution of a single, discrete MRF using continuous relaxation labeling. Combining a discontinuity field with a discrete MRF is important as it allows the disabling of clique potentials across discontinuities. Furthermore, explicit representation of the discontinuity field suggests new algorithms that incorporate hysteresis and non-maximal suppression.

2. Isotropic Smoothing

In this section, we describe the relationship between maximum aposteriori probability (MAP) estimation of discrete Markov random fields (MRF) and continuous relaxation la-
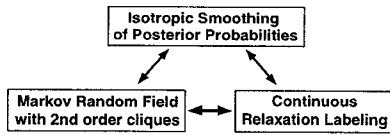
Figure 2: Equivalence between isotropic smoothing of posterior probabilities, Markov random fields with 2nd order cliques, and continuous relaxation labeling.

beling (CRL) [10]. This connection was originally made by Li *et. al.* [8]. We review this relationship to introduce the notation that will be used in the rest of the paper and to point out the similarities between this technique and isotropic smoothing of posterior probabilities. These relationships are depicted in Fig. 2.

We specialize our notation to MRF's defined on image grids. Let $\mathcal{S} = \{1, \ldots, n\}$ be a set of sites where each $s \in \mathcal{S}$ corresponds to a single pixel in the image. For simplicity, we assume that each site can take on labels from a common set $\mathcal{L} = \{1, \ldots, k\}$. Adjacency relationships between sites are encoded by $\mathcal{N} = \{\mathcal{N}_i | i \in \mathcal{S}\}$ where $\mathcal{N}_i$ is the set of sites neighboring site i. Cliques are then defined as subsets of sites so that any pair of sites in a clique are neighbors. In this paper, we will only consider 4-neighbor adjacency for images (and 8-neighbor adjacency for volumes) and cliques of sizes no greater than two. By considering each site as a discrete random variable f_i with a probability mass function over $\mathcal{L}$, a discrete MRF $\mathbf{f}$ can be defined over the sites with a Gibbs probability distribution.

If data $d_i \in \mathbf{d}$ is observed at each site i, and is dependent only on its label f_i, then the posterior probability is itself a Gibbs distribution and by the Hammersley-Clifford theorem, also a MRF, albeit a different one [5]: $P(\mathbf{f}|\mathbf{d}) = Z^{-1} \times \exp\{-E(\mathbf{f}|\mathbf{d})\}$ where

$$E(\mathbf{f}|\mathbf{d}) = \sum_{i \in \mathcal{C}_1} V_1(f_i|d_i) + \sum_{(i,j) \in \mathcal{C}_2} V_2(f_i, f_j) \quad (1)$$

and $V_1(f_i|d_i)$ is a combination of the single site clique potential and the independent likelihood. The notation (i,j) refers to a pair of sites; thus, the sum is actually a double sum. Maximizing the posterior probability $P(\mathbf{f}|\mathbf{d})$ is equivalent to minimizing the energy $E(\mathbf{f}|\mathbf{d})$.

Continuous Relaxation Labeling. The continuous relaxation labeling approach to solving this problem was introduced by Li *et. al.* [8]. In CRL, the class (label) of each site i is represented by a vector $p_i = [p_i(f_i)|f_i \in \mathcal{L}]$ subject to the constraints: (1) $p_i(f_i) \geq 0$ for all $f_i \in \mathcal{L}$, and (2) $\sum_{f_i \in \mathcal{L}} p_i(f_i) = 1$. Within this framework, the energy $E(\mathbf{f}|\mathbf{d})$ to be minimized is rewritten as

$$E(\mathbf{p}|\mathbf{d}) = \sum_{i \in \mathcal{C}_1} \sum_{f_i \in \mathcal{L}} V_1(f_i|d_i) p_i(f_i) + \quad (2)$$

$$\sum_{(i,j) \in \mathcal{C}_2} \sum_{(f_i, f_j) \in \mathcal{L}^2} V_2(f_i, f_j) p_i(f_i) p_j(f_j).$$

Note that when $p_i(f_i)$ is restricted to $\{0, 1\}$, $E(\mathbf{p}|\mathbf{d})$ reverts to its original counterpart $E(\mathbf{f}|\mathbf{d})$. Hence, CRL embeds the actual combinatorial problem into a larger, continuous, constrained minimization problem.

The constrained minimization problem is typically solved by iterating two steps: (1) gradient computation, and (2) normalization and update. The first step decides the direction that decreases the objective function while the second updates the current estimate while ensuring compliance with the constraints. A review of the normalization techniques that have been proposed are summarized in [8]. Ignoring the need for normalization, continuous relaxation labeling is similar to traditional gradient descent:
$p_i^{t+1}(f_i) \leftarrow p_i^t(f_i) - \frac{\partial E(\mathbf{p}|\mathbf{d})}{\partial p_i^t(f_i)}$ where

$$\frac{\partial E(\mathbf{p}|\mathbf{d})}{\partial p_i^t(f_i)} \doteq V_1(f_i|d_i) + 2 \sum_{j:(i,j) \in \mathcal{C}_2} \sum_{f_j \in \mathcal{L}} V_2(f_i, f_j) p_j^t(f_j). \quad (3)$$

and the superscripts $t, t+1$ denote iteration numbers. The notation $j : (i, j)$ refers to a single sum over j such that (i, j) are pairs of sites belonging to a clique. Barring the different normalization techniques could be employed, Eqn. 3 is found in the update equations of various CRL algorithms [10, 4, 7]. There are, however, two differences. First, in most CRL problems, the first term of Eqn. 3 is absent and thus proper initialization of $\mathbf{p}$ is important. We will also omit this term in the rest of the paper to emphasize the similarity with continuous relaxation labeling. Second, CRL problems typically involve maximization; thus, $V_2(f_i, f_j)$ would represent consistency as opposed to potential, and the update equation would add instead of subtract the gradient.

Isotropic Smoothing. A convenient way of visualizing the above operation is as isotropic smoothing. Since the sites represent pixels in an image, for each class f_i, $p_i(f_i)$ can be represented by an image (of posterior probabilities) such that k classes imply k such image planes. Together, these k planes form a volume of posterior probabilities. Each step of Eqn. 3 then essentially replaces the current estimate $p_i^t(f_i)$ with a weighted average of the neighboring assignment probabilities $p_j^t(f_j)$. In other words, the volume of posterior probabilities is linearly *filtered*. If the potential functions $V_2(f_i, f_j)$ favor similar labels, then the weighted average is essentially low-pass among sites with common labels and hi-pass among sites with differing labels.

3. Anisotropic Smoothing

Isotropic smoothing causes significant blurring especially across region boundaries. A solution to this problem is to
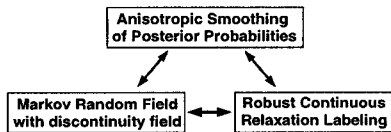
Figure 3: Equivalence between anisotropic smoothing of posterior probabilities, Markov random fields with discontinuity fields, and robust continuous relaxation labeling.

smooth adaptively such that smoothing is suspended across region boundaries and takes place only within region interiors. Anisotropic smoothing is often implemented by simulating nonlinear partial differential equations with the image as the initial condition [1, 9]. In this section, we show that while isotropic smoothing of posterior probabilities is the same as continuous relaxation labeling of a MRF, anisotropic smoothing of posterior probabilities is equivalent to continuous relaxation labeling of a MRF supplemented with a (hidden) analog discontinuity field. We also demonstrate that this method could also be understood as incorporating a robust consensus-taking scheme within the framework of continuous relaxation labeling. These relationships are depicted in Fig. 3.

We extend the original MRF problem to include a non-interacting, analog discontinuity field on a displaced lattice. Thus, the new energy to be minimized is:

$$E(\mathbf{f}, \mathbf{l}) = \sum_{(i,j)\in\mathcal{C}_2} \left[\frac{1}{2\sigma^2} V_2(f_i, f_j) \cdot l_{i,j} + (l_{i,j} - 1 - \log l_{i,j}) \right]$$

(4)

where $V_1(f_i)$ has been dropped for simplicity since the discontinuity field does not interact with it. The individual sites in the discontinuity field $\mathbf{l}$ are denoted by $l_{i,j}$ which represent either the horizontal or vertical separation between sites i and j in $\mathcal{S}$. When $l_{i,j}$ is small, indicating the presence of a discontinuity, the effect of the potential $V_2(f_i, f_j)$ is suspended; meanwhile, the energy is penalized by the second term in Eqn. 4. There are a variety of penalty functions that could be derived from the robust estimation framework (see Black [2]). The penalty function in Eqn. 4 was derived from the Lorentzian robust estimator.

The minimization of $E(\mathbf{f}, \mathbf{l})$ is now over both $\mathbf{f}$ and $\mathbf{l}$. Since the discontinuity field is non-interacting, $\mathbf{l}$ can be minimized analytically by computing the partial derivatives of $E(\mathbf{f}, \mathbf{l})$ with respect to $l_{i,j}$ and setting that to zero. Doing so and inserting the result back into $E(\mathbf{f}, \mathbf{l})$ gives us

$$E(\mathbf{f}) = \sum_{(i,j)\in\mathcal{C}_2} \log\left[1 + \frac{1}{2\sigma^2} V_2(f_i, f_j) \right].$$

(5)

Rewriting this equation in a form suitable for CRL, we get

$$E(\mathbf{p}) = \sum_{(i,j)\in\mathcal{C}_2} \log\left[1 + \frac{1}{2\sigma^2} \sum_{(f_i,f_j)\in\mathcal{L}^2} V_2(f_i, f_j) p_i(f_i) p_j(f_j) \right].$$

(6)

Note that when $p_i(f_i)$ is restricted to $\{0, 1\}$, Eqn. 6 reduces to Eqn. 5.

Anisotropic Smoothing. To compute the update equation for CRL, we take the derivative of $E(\mathbf{p})$ with respect to $p_i(f_i)$:

$$\frac{\partial E(\mathbf{p})}{\partial p_i(f_i)} \doteq \sum_{j:(i,j)\in\mathcal{C}_2} w_{i,j} \left[\sum_{f_j \in \mathcal{L}} V_2(f_i, f_j) p_j(f_j) \right]$$

(7)

where

$$w_{i,j} = 2\sigma^2 / \left[2\sigma^2 + \sum_{(f_i,f_j)\in\mathcal{L}^2} V_2(f_i, f_j) p_i(f_i) p_j(f_j) \right].$$

(8)

The term $w_{i,j}$ encodes the presence of a discontinuity. If $w_{i,j}$ is constant, then the above equation reverts to the isotropic case. Otherwise, $w_{i,j}$ either enables or disables the penalty function $V_2(f_i, f_j)$. This equation is similar to the anisotropic diffusion equation proposed by Perona and Malik [9].

Robust Continuous Relaxation Labeling. Each iteration of continuous relaxation labeling can be viewed as a consensus-taking process [12]. Neighboring pixels vote on the classification of a central pixel based on their current assignment probabilities $p_j(f_j)$, and their votes are tallied using a weighted sum. The weights used are the same throughout the image; thus, pixels on one side of a region boundary may erroneously vote for pixels on the other side. Anisotropic smoothing of the posterior probabilities can be regarded as implementing a robust voting scheme since votes are tempered by $w_{i,j}$ which estimates the presence of a discontinuity. The connection between anisotropic diffusion on continuous-valued images and robust estimation was recently demonstrated by Black et. al. [3].

4. Results and Discussion

The anisotropic smoothing scheme was used to segment white matter from MRI data of human cortex. Pixels a given distance from the boundaries of the white matter classification were then automatically classified as gray matter. Thus, gray matter segmentation relied heavily on the white matter segmentation being accurate. Fig. 4 shows comparisons between gray matter segmentations produced automatically
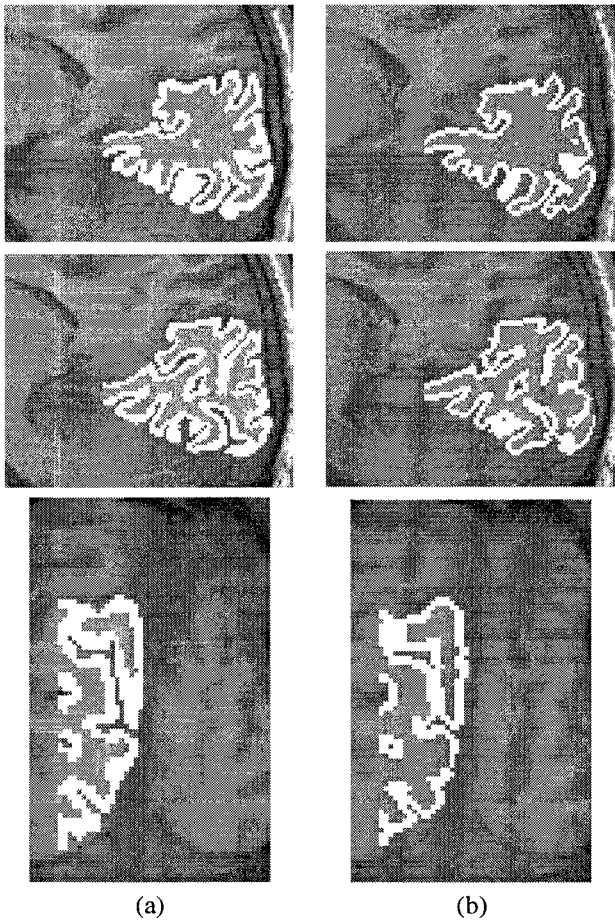
677

Figure 4: Left images show manual gray matter segmentation results; right images show the automatically computed gray matter segmentation.

by the proposed method and those obtained manually. More examples can be found in [11].

The technique being proposed bears some superficial resemblance to schemes that anisotropically smooth the raw image before classification [6]. Besides the connection between our technique and MAP estimation of Markov random fields, which is absent in schemes that smooth the image directly, there are two other important differences. First, anisotropic smoothing of the raw image does not take into consideration the discrete number of classes that are actually present. Second, anisotropic smoothing of the raw image is only applicable when the noise corrupting the image is additive and class independent. For example, if the class means were identical and the classes differed only in their variances, then anisotropic smoothing of the raw image is ineffective. On the other hand, applying anisotropic smoothing on the posterior probabilities is still feasible even when the class likelihoods are described by general probability mass functions.

The equivalence between anisotropic smoothing of posterior probabilities and MRF's with discontinuity fields also offers a solution to the problems of edge handling and missing data. These two issues can be treated in the same manner as in traditional regularization. Solving of the latter implies that MAP classification can be obtained even at locations where the pixel values are not provided.

5. References

[1] L. Alvarez, P. L. Lions, and J. M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numerical Analysis*, 29:845–866, 1992.

[2] M. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int'l J. Computer Vision*, 19(1):57–91, 1996.

[3] M. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *Submitted to IEEE Trans. Image Processing*, 1997.

[4] O. D. Faugeras and M. Berthod. Improving consistency and reducing ambiguity in stochastic labeling: an optimization approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3:412–423, 1981.

[5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(6):721–742, 1984.

[6] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Trans. Medical Imaging*, 11:221–232, 1992.

[7] R. A. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(2):267–286, 1983.

[8] S. Z. Li, H. Wang, and M. Petrou. Relaxation labeling of Markov random fields. In *Int'l Conf. Pattern Recognition*, pages 488–492, 1994.

[9] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[10] A. Rosenfeld, R. Hummel, and S. Zucker. Scene labeling by relaxation operations. *IEEE Trans. Systems, Man, and Cybernetics*, 6(6):420–433, 1976.

[11] P. C. Teo, G. Sapiro, and B. A. Wandell. A method of creating connected representations of cortical gray matter for functional MRI visualization. *To appear in IEEE Trans. on Medical Imaging*, 1997.

[12] Y. Weiss and E. Adelson. Perceptually organized EM: a framework for motion segmentation that combines information about form and motion. In *Int'l Conf. Computer Vision and Pattern Recognition*, pages 312–326, 1996.