COMPARISON OF ENERGY MINIMIZATION METHODS FOR 3-D BRAIN TISSUE CLASSIFICATION

Subrahmanyam Gorthi, Jean-Philippe Thiran, Meritxell Bach Cuadra

Signal Processing Laboratory (LTS5),
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
{subrahmanyam.gorthi, jp.thiran, meritxell.bach}@epfl.ch

ABSTRACT

This paper presents 3-D brain tissue classification schemes using three recent promising energy minimization methods for Markov random fields: graph cuts, loopy belief propagation and tree-reweighted message passing. The classification is performed using the well known finite Gaussian mixture Markov Random Field model. Results from the above methods are compared with widely used iterative conditional modes algorithm. The evaluation is performed on a dataset containing simulated T1-weighted MR brain volumes with varying noise and intensity non-uniformities. The comparisons are performed in terms of energies as well as based on ground truth segmentations, using various quantitative metrics.

Index Terms— Energy minimization, Markov random fields, medical image segmentation, brain tissue classification.

1. INTRODUCTION

Recent developments in energy minimization methods for Markov random fields (MRF) have resulted in faster and efficient global (or strong local) minimization algorithms [1]. Graph cuts (GC) [2, 3, 4], loopy belief propagation (LBP) [5, 6] and tree reweighted message passing (TRW) [7, 8] are among such most notable algorithms. The applications of these algorithms are spread over a wide variety of early vision problems. Szeliski et al. [1] compared these algorithms in the domains of stereo matching, image stitching, 2-D binary image segmentation, denoising and inpainting, for different types of smoothness-based priors. They demonstrated the potential of these global minimization algorithms over the older yet widely used iterated conditional modes (ICM) [9] algorithm. While many of these global minimization algorithms have received the much deserved attention in the domains like stereo matching, they are less explored in the context of medical imaging. For instance, to the best of our knowledge, TRW algorithm, which is found to give consistently strong results in various early vision problems [1], has never been evaluated in the context of medical image segmentation.

The main contribution of this paper is the convergence study of these optimization methods for tissue classification of Magnetic Resonance (MR) brain imaging. Brain tissue classification plays an important role in many applications. For instance, this is essential for the quantitative study and analysis of several brain disorders like alzheimer's disease, as well as in understanding the development process of the brain. Further, brain tissue classification is also used as a preprocessing step for many applications like voxel-based morphometry. Among numerous approaches, MRF models are widely used for performing automated 3D brain tissue classification [10]. This paper evaluates the recent MRF energy minimization methods on the widely used finite Gaussian mixture MRF (FGMMRF) model [11].

We note that except in very few works [12], none of the global (or strong local) optimization methods are evaluated for brain tissue classification. [12] compared brain tissue classification results obtained from ICM with graph cuts. However, in their comparison, while graph cuts-based method was using tissue-priors information, this information is not used with ICM-based method; further, the 3-D segmentation in [12] is done slice by slice on a 2-D grid, but not on the original 3-D grid. Thus, no comparisons of MRF energy minimization methods for brain tissue classification are available in the literature that are performed under identical parameters.

The following algorithms are evaluated in this paper: (i) two most popular versions of GC, known as *expansion-move* and *swap-move* algorithms [2, 3, 4], (ii) an LBP implementation derived by Kolmogorov from TRW-S [8], called *BP-S* algorithm, (iii) an improved version of the original TRW algorithm [7], called the *sequential TRW (TRW-S)* [8], and (iv) *ICM* algorithm [9].

Regarding the implementation of these algorithms, thanks to the MRF library [1] [1], it has served as a basis for our current implementation. The above mentioned MRF library can, however, handle neighborhood priors only on a 2-D grid which is not suffice for 3-D medical imaging applications. We have now enhanced it to handle 3-D grid and also integrated it with ITK [2], which is a widely used open-source tool in medical imaging.

The rest of the paper is organized as follows. In Section 2, we briefly present the FGMMRF model and the optimization methods. Evaluation results are presented in Section 3. Discussion and conclusions are presented in Section 4.

2. METHODOLOGY

2.1. Energy Model

Let ν represent the set of all voxels in a given image, and X_p be the label (tissue-class) assigned to the p^{th} voxel. Let N be the number of voxels in the set ν. Let X be the set containing the labels assigned to ν, i.e., $X = \{X_1, \cdots, X_N\}$. Then, the brain tissue classification is often formulated as an energy minimization problem of the form:

$$E(X) = \sum_{\forall p \in \nu} \psi_p(X_p) + \beta \sum_{\forall p \in \nu} \sum_{\forall q \in \aleph_p} \psi_{pq}(X_p, X_q),$$

This work is supported in part by the Swiss National Science Foundation under Grant 205321-124797 and by IBM of the Geneva–Lausanne Universities and EPFL, as well as the foundations Leenaards and Louis-Jeantet.

[1] http://vision.middlebury.edu/MRF/
[2] http://itk.org/

where the first term is a data term (unary function) representing the intensity modeling of the tissue classes, and the second term is a smoothness term (pairwise function) representing the neighborhood-prior modeling. β is a weighting parameter between the data term and smoothness term.

a) Data term: Let I_p be the observed intensity of the p^{th} voxel, and $I = \{I_1, \cdots, I_N\}$. Let $L = \{l_1, \cdots l_D\}$ be the set of labels to be assigned. Let $\theta_i = \{\mu_i, \sigma_i\}$ are mean and variance associated with label l_i, and $\theta = \{\theta_1, \cdots, \theta_D\}$. For the brain tissue classification, intensities of the tissue classes are generally modeled with Gaussian distribution [11]. Hence, the likelihood that a given labeling X and θ have produced I is given by:

$$P\left(I|X,\,\theta\right) = \prod_{\forall p\,\in\,\nu} \left(\frac{1}{\sqrt{2\pi}\,\sigma_{X_p}} \exp\left(\frac{-\left(I_p - \mu_{X_p}\right)^{\,2}}{2\,\sigma_{X_p}^2} \right) \right).$$

By taking a negative logarithm of the above equation, the problem of maximizing the likelihood can be transformed into an equivalent energy minimization problem with the following unary function:

$$\psi_p(X_p) = \frac{\left(I_p - \mu_{X_p}\right)^2}{2\,\sigma_{X_p}^2} + \log(\sqrt{2\pi}\,\sigma_{X_p}).$$

b) Smoothness term: The smoothness term for FGMMRF model is given by the discontinuity preserving Pott's model:

$$\psi_{pq}\left(X_p, X_q\right) = \begin{cases} 0, & \text{if } X_p = X_q; \\ 1, & \text{otherwise.} \end{cases}$$

Note that the above smoothness term does not penalize when neighborhood voxels are assigned the same label. The finite Gaussian mixture MRF (FGMMRF) model studied here [11] aims to classify the image voxels into one of the three brain tissue types: cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM). The background (BKG) is represented by an additional label. Thus, $D = 4$ for this model, and the set of labels is given by $L = \{\text{BKG, CSF, GM, WM}\}$.

2.2. Optimization Methods

We briefly mention here the general characteristics of the optimization methods that we evaluate on the energy model described in the preceding subsection. For a more detailed discussion, the readers are referred to [1]. The classical ICM algorithm uses a greedy approach, and converges faster than the rest of the algorithms evaluated in this paper. However, ICM provides local minima, and thus it is very sensitive to initialization of the labeling. Graph cuts, if applicable, generally gives very accurate results. They are however applicable to limited pairwise potentials. For example, for obtaining a guaranteed global optimum with graph cuts, the pairwise potential term should be a metric in case of GC-Expansion and semimetric for GC-Swap. On the other hand, LBP is applicable to any energy function, but it does not always converge. Regarding TRW-S, it is noted in [1] that it can be a good competitor to graph cuts in certain applications. TRW-S is applicable to any energy function. However, TRW-S is computationally slower than graph cuts, and convergence of energy is not guaranteed.

3. RESULTS

The evaluation is performed on the simulated brain volumes from the *Brain Web* database of the McConnell Brain Imaging Centre [13]. We use a dataset of 12 brain volumes that contain images with varying noise levels (0%, 1%, 3%, 5%, 7% and 9%), and two levels of intensity non-uniformities (INU 20% and 40%).

Among different parameters to be set, the most sensitive parameter in obtaining accurate tissue classification is the *mean intensity* of each tissue class (μ_l). Hence, we have iteratively updated the mean intensity values of each label using the expectation maximization approach in [10], by computing the a posteriori probabilities of the labels, and is given by:

$$\mu_l^{(k)} = \frac{\sum\limits_{\forall p\,\in\,\nu} (I_p)\, P^{(k)}(X_p = l|\, I_p,\, \theta_l^{(k-1)})}{\sum\limits_{\forall p\,\in\,\nu} P^{(k)}(X_p = l|\, I_p,\, \theta_l^{(k-1)})}.$$

The σ value for each tissue class is fixed to the typical value 5 [10]. β is fixed empirically to 1.0 similar to [1, 10], based on the visual inspection of the segmentation results.

In order to compare the energy convergence results from all the methods under identical conditions, final μ values obtained by iteratively updating them in one of the methods (GC-Expansion) are used for the remaining methods. For instance, final μ values with GC-Expansion, obtained for CSF, GM and WM, for the image with 5% noise and 20% INU are, 46.3, 92.9 and 126.4 respectively, and the same values are used with other MRF optimization methods. However, we noticed that, even allowing each method to independently update the μ values has resulted in similar mean values with small variations in their first decimal.

All the experiments are run on the same machine (2.26 GHz Intel Xeon Processor, 12 GB RAM). The results from all 12 brain volumes have shown consistently similar behavior. Because of space limitations, we show here energy results for just one image with 5% noise and 20% INU, and visual results for one of the axial slices extracted from the above brain volume. However, the quantitative results presented here are computed over the entire dataset of 12 volumes.

TRW-S has the ability to compute a lower bound on the energy of the optimal solution. Similar to [1], we take advantage of this lower bound; instead of comparing the absolute energies, we normalize the energies by dividing them with the best lower bound computed by TRW-S, and then compare the energies.

Fig. 1 shows the energy convergence results for the brain volume with 5% noise and 20% INU. The energy differences among these methods, including ICM, are found to be marginal. In terms of time taken for convergence, ICM is obviously faster compared to global optimization methods since, its convergence is based on local optimization criteria. Among the global optimization methods, expansion-move version of graph cuts is the fastest one. The energy convergence results for the remaining brain volumes in the dataset are also similar to the above mentioned results.

Fig. 2 shows an axial slice extracted from the image brain volume with 5% noise and 20% INU, and its ground truth segmentation. Fig. 3 shows automated tissue classification results obtained from all methods. Quantitative evaluation is performed using the commonly used metrics: (i) sensitivity (a measure of true-positive fraction), (ii) specificity (a measure of true-negative fraction), (iii) dice similarity metric (DSM) (a measure of overlap between ground truth and automated segmentation), and (iv) % of error in volume; the results are shown in Table 1 and Table 2. The differences among the methods for all these metrics are found to be marginal. The results from ICM in terms of these metrics are, surprisingly, quite close to the best ones.

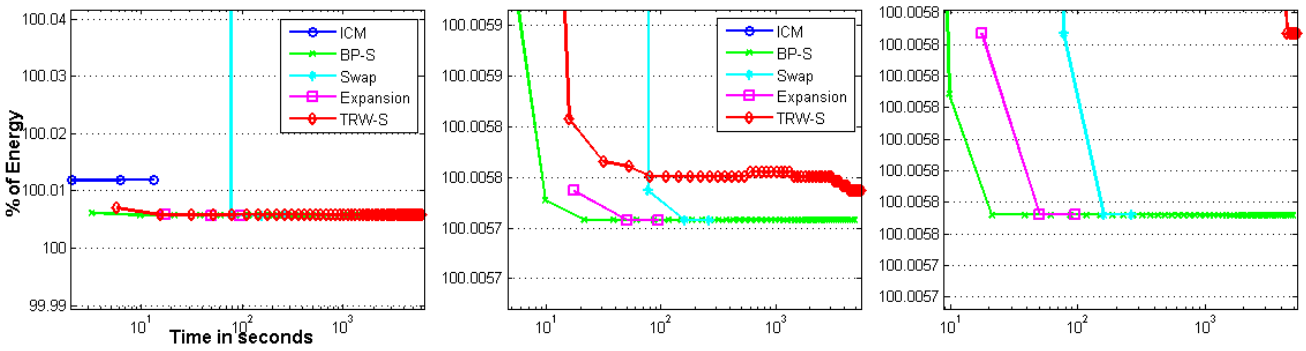
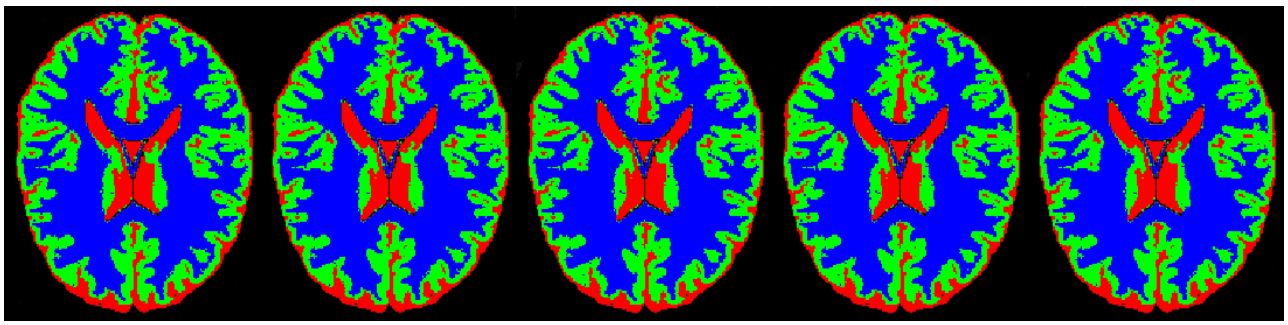


Fig. 1: Comparison of energy results for the MRF minimization methods. These are the results obtained from FGMMRF model, for the simulated brain volume with 5% noise and 20% INU. The second and third plots are the zoomed versions of the first plot.



(a) ICM (b) BP-S (c) GC-Swap (d) GC-Expansion (e) TRW-S

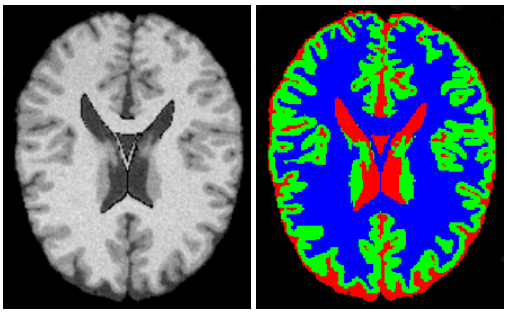
Fig. 3: Qualitative comparison of segmentation results in one of the slices in the axial direction, extracted from the brain volume with 5% noise and 20% INU. Results from all the methods are very close, and are quite similar to the ground truth segmentation shown in Fig. 2b.

Table 1: Mean and standard deviations of sensitivity and specificity of CSF, GM and WM tissues, obtained from different MRF optimization methods. These values are computed over a dataset of 12 brain volumes with varying noise (0-9%) and INU (0-20%).

Method	Sensitivity(%)			Specificity(%)		
	CSF	GM	WM	CSF	GM	WM
ICM	93.42 ± 4.41	92.23 ± 3.41	94.55 ± 2.81	99.28 ± 0.85	99.21 ± 0.47	99.21 ± 0.39
BP-S	93.43 ± 4.39	92.21 ± 3.39	94.62 ± 2.77	99.26 ± 0.88	99.22 ± 0.46	99.20 ± 0.38
GC-Swap	93.23±4.45	91.74 ± 3.59	95.14 ± 2.46	99.16 ± 1.11	99.25 ± 0.45	99.10 ± 0.42
GC-Exp	93.26 ± 4.43	92.01 ± 3.46	94.84 ± 2.61	99.27 ± 0.89	99.23 ± 0.45	99.16 ± 0.40
TRW-S	93.43 ± 4.39	92.21 ± 3.39	94.62 ± 2.76	99.26 ± 0.88	99.22 ± 0.46	99.20 ± 0.38

Table 2: Mean and standard deviations of dice similarity metric and % error in volume of CSF, GM and WM tissues, obtained from different MRF optimization methods. These values are computed over a dataset of 12 brain volumes with varying noise (0-9%) and INU (0-20%).

Method	Dice Similarity Metric(%)			%Error in Volume		
	CSF	GM	WM	CSF	GM	WM
ICM	90.89 ± 8.30	93.32 ± 3.34	93.60 ± 3.07	6.49 ± 11.43	-2.35 ± 0.75	2.05 ± 1.67
BP-S	90.79 ± 8.48	93.33 ± 3.32	93.62 ± 3.04	6.81 ± 12.05	-2.42 ± 0.74	2.16 ± 1.64
GC-Swap	90.11 ± 9.75	93.17 ± 3.39	93.48 ± 3.04	8.46 ± 15.98	-3.09 ± 0.86	3.60 ± 2.00
GC-Exp	90.81 ± 8.48	93.27 ± 3.33	93.55 ± 3.01	6.39 ± 12.16	-2.70 ± 0.82	2.77 ± 1.76
TRW-S	90.79 ± 8.48	93.33 ± 3.32	93.62 ± 3.04	6.81 ± 12.06	-2.42 ± 0.74	2.17 ± 1.64



(a) (b)

Fig. 2: (a) One of the axial slices extracted from the brain volume with 5% noise and 20% INU. (b) Ground truth segmentation of tissues for the same slice; CSF, GM and WM tissue classes are represented with labels of red, green and blue colors respectively.

4. CONCLUSIONS

In this paper, we have presented tissue classification of MR brain volumes with different MRF optimization algorithms, and compared them using identical parameters. The evaluated algorithms are: expansion-move and swap-move versions of the graph cuts method, loopy belief propagation, sequential tree reweighted message passing, and the older iterated conditional modes algorithm. The evaluation is performed on the widely used FGMMRF model, using a dataset of 12 brain volumes with varying noise and INU.

In order to draw definite conclusions about the accuracy of the widely used FGMMRF model (or any other model in general), it is essential to make sure that the solution has converged to a global optimum. Otherwise, it will be ambiguous whether the resulting errors are due to the model itself, or, whether the errors are related to the convergence of optimization method used for solving the model. The current study is important in this perspective also.

It is well known that since ICM converges to a local optimum, its results depend very much on the initialization of the labels. In this paper, for ICM, each voxel is initialized with that tissue-class label with whose μ value the voxel's intensity difference is minimum. From the results, we notice that although ICM is a local optimization method, the above mentioned initialization resulted in an accurate classification. However, if a dataset to be segmented is quite different from the above tested volumes from Brainweb, it may be safe to use GC-Expansion rather than ICM so that convergence to a global minimum is guaranteed. Note that among the global optimization methods, GC-Expansion converged quickly. In the future work, we want to study in more detail, the effects of different initializations on the convergence of ICM. We would like to extend this evaluation to real brain volumes. Further, we would like to perform similar investigation on other tissue classification models like partial volume models [14, 10]. It would be also interesting to evaluate other fast optimization methods like [15].

5. REFERENCES

[1] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[2] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[3] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[4] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[5] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[6] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proceedings of the IEEE International Conference on Computer Vision*, 2003, vol. 2, pp. 900–907.

[7] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Map estimation via agreement on trees: Message-passing and linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3697–3717, 2005.

[8] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568–1583, 2006.

[9] S. Birchfield and C. Tomasi, "On the statistical analysis of dirty pictures (with discussion)," *J. Royal Statistical Soc., Series B*, vol. 20, no. 4, pp. 259–302, 1986.

[10] M. Bach Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J.-Ph. Thiran, "Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1548– 1565, 2005.

[11] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[12] Z. Song, N. Tustison, B. Avants, and J. Gee, "Adaptive graph cuts with tissue priors for brain MRI segmentation," in *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, 2006, pp. 762–765.

[13] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sied, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 463–468, 1998.

[14] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *NeuroImage*, vol. 13, no. 5, pp. 856–876, 2001.

[15] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1436–1453, 2007.