

**Li, W., deSilver, C. and Attikiouzel, Y. (2004) *A semi-supervised map segmentation of brain tissues.* In: 7th International Conference on Signal Processing ICSP '04, 31 August - 4 September, Beijing, China.**

# A SEMI-SUPERVISED MAP SEGMENTATION OF BRAIN TISSUES

*Wanqing Li*

SITACS, University of Wollongong
email:wanqing@uow.edu.au

*Chris deSilver, Yianni Attikiouzel*

Division of Science and Engineering
Murdoch University

## ABSTRACT

This paper presents a method for semi-supervised MAP (maximum a-posterior probablity) segmentation of brain tissues where labelled data are available for either all types of tissues or only a few types of tissues possibly at different levels of quality. The proposed MAP segmentation takes supervised and unsupervised segmentation as its two special cases where, respectively, quality labelled data is available or there is no labelled data at all. Experiments on real MR images have shown that the proposed method improved the segmentation accuracy substantially with only a few labelled data in comparison with both fully supervised method with the same labelled data set and unsupervised method.

## 1. INTRODUCTION

Segmentation of Magnetic Resonance (MR) images is a process of delineation of regions representing different types of tissues and/or lesions. After more than a decade research, techniques for segmenting MR images are gradually converging to MAP (maximum a-posterior probabliity) segmentation based on Gibbs Random Field (GRF) and Markov Random Field (MRF) [8, 11, 13, 10, 14, 12] and FCM (Fuzzy C-means) based classification [4, 15, 1]. Numerious algorithms based on either MAP or FCM have been developed [6, 3, 5] and most of them were employed in a fully unsupervised manner.

The advantages of unsupervised (or automatic) segmentation over supervised segmentation have been well recognised: less user interaction and high reproductivity. Because almost all automatic techniques are virtually an optimisation process which is governed by an objective function such as total log likelihood in normal mixture modelling and sums of the squared errors in FCM, the techniques inevitably suffer from the problem of local traps (minima or maxima). In consequence, they need to be tuned properly in order to produce satifactory results in a specific application.

On the other hand, supervised techniques usually do not suffer the problem of local traps and often produce accurate results, but they require reliable training data available for every type of tissues. their results will depend entirely on the quality of the training data sets.

Bensaid et. al [2, 9] introduced semi-supervised FCM (ssFCM) which incorporates labelled data into the unsupervised FCM algorithm. The ssFCM relaxes the requirement on the labelled data in comparison with supervised approaches and assumes that

- high quality labelled data are available for every class.
- the labelled data captures the shapes of the clusters.

In this paper, we further relax the requirement and introduce a semi-supervised MAP segmentation (ssMAP) method that is able to utilize any available labelled data. The labelled data, if any, is assumed to have the following characteristics:

- they are available for every type of tissues or for some types of tissues if the number of tissues is known.
- they may capture neither the centers nor the shapes of the clusters.

The paper is organized as follows. Section 2 describes the semi-supervised MAP segmentation of MR images that consists of prior and data models. Section 3 presents the ML (maximum likelihood) estimation of the model parameters that is subject to any labelled data. Some experiments on real dual echo MR images are presented in Section 4. Discussion and conclusions are given in Section 5.

## 2. SEMI-SUPERVISED MAP SEGMENTATION

Let $\{y_t\}_{t=1}^{M}$ be unlabelled pixels in a MR slice to be segmented and they are considered as a realisation of a random field defined on a lattice $\mathcal{L}$, where $t \in \mathcal{L}$. $\{y_t^c : t = 1, 2, \cdots, n_c; c = 1, 2, \cdots, K\}$ denotes all labelled pixels for $K$ types of tissues. The labelled pixels for the $i'th$ tissue are denoted as $\{y_t^i\}_{t=1}^{n_i}$, where $n_i \geq 0$, the number of labelled pixels for the $i'th$ tissue, and $\sum_i n_i = N$.

The true but unknown tissue labels of all pixels are assumed to be a realisation of the random field $X = \{X_t : t \in \mathcal{L}\}$, denoted by $x^* = \{x_t : t \in \mathcal{L}\}$, where $x_t$ labels the tissue type at site $t$.

0-7803-8406-7/04/$20.00 © 2004 IEEE.

Assume $X$ is a local Markov Random Field (MRF) defined in a neighbouhood system and the labelled pixels do not supply any spatial knowledge of their labelled tissues. Using Bayes rule, A maximum *a posteriori* (MAP) estimation of the pixel labells, $\hat{\mathbf{x}}$, is

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\in\Omega} \prod_{t\in\mathcal{L}} f(y_t|x_t)p(x_t|x_{\partial t}) \prod_{s\in\zeta} f(y_s^c|x_s^c) \quad (1)$$

where $f(y_t|x_t)$ is the conditional density of random variables $\{Y_t : 1 \le t \le M\}$ dependent on $\mathbf{x}$, usually known as data model. $p(x_t|x_{\partial t}) = \frac{e^{-v(x_t|\eta_t)}}{Z_t}$ is the prior probability or prior model of $x_t$ given its neighbours. $x_{\partial t}$ is defined in a neighbourhood system $\eta_t$, where $Z_t$ is a partition function and $v(\cdot)$ is usually referred as a energy function.

## 3. MODEL ESTIMATION

MAP segmentation requires both prior and data models. The common way is to parameterise both models and then estimate the parameters from the observations.

### 3.1. Prior model

The prior model, or specifically the energy function $v(\cdot)$, must be defined over cliques in a neighbourhood system. Taking the simple second-order neighbourhood structure containing eight nearest neighbours to the corresponding pixel position $t$, we define the energy function over the posterior probabilities (soft labelling) rather than over discrete (hard) labelling

$$v(x_t = k|\eta_t^2) = -\alpha(k) - \beta G_{\partial t}(k) \quad (2)$$

$$G_{\partial t}(k) = \sum_{r\in\partial t} z_{rk}, \quad (3)$$

where $\alpha(k)$ represents global information about the probability of tissue $k$, $\beta$ is a parameter to be set and $z_{rk}$ is the posterior probability of pixel $r$ belonging to tissue $k$.

The value of $\beta$ controls the degree of spatial clustering observable in the underlying state process. Normally, prior model is very tolerant of small variations of its parameters. Empirical study on simulated MR images having various degrees of noise, partial volume effects and non-uniformity suggests that $\beta = 2.5$ is a good choice [6].

### 3.2. Data model

According to the statistical properties of MR images [11], $f(y_t|x_t)$ can be reasonably approximated as a multivariate Gaussian, i.e.

$$f(y_t|x_t = k) \sim \mathcal{N}(\mu_k, \Sigma_k), \quad (4)$$

where $\mu_k$ and $\Sigma_k$ are the mean vector and covariance matrix respectively of the Gaussian describing the probability density of tissue $k$.

It is appropriate to estimate the tissue parameters $\{\mu_k, \Sigma_k, K\}$ by fitting the parameters to the image data with ML or least squares. However, the estimation must be subject to the labelled data. Assume the independence of the potential labelling over pixels and use a finite normal mixture as the likelihood of realising the multispectral MR images given the underlying tissue types

$$g(Y|\Phi) = \prod_{t=1}^{M}\{\sum_{k=1}^{K} w_k f(y_t|\phi_k)\} \prod_{c=1}^{K}\prod_{t=1}^{n_c} f(y_t^c|\phi_c), \quad (5)$$

where $w_k$ is the ratio of the number of pixels within class k to the total number of pixels $M$, and $\sum_k w_k = 1$. $\Phi = (\phi_1, \phi_2, \cdots, \phi_k)$ and $\phi_k = (w_k, \mu_k, \Sigma_k)$ is the parameter vector for the $k'th$ normal component $f(\cdot)$.

The ML estimation of Equation 5 using the EM algorithm will lead to the $(r+1)'th$ iteration of the parameter estimation

$$w_k^{r+1} = \frac{n_k + \sum_{i=1}^{M} z_{ik}^r}{N+M} \quad (6)$$

$$\mu_k^{r+1} = \frac{\sum_{i=1}^{n_k} y_i^k + \sum_{i=1}^{M} z_{ik} y_i}{n_k + \sum_{i=1}^{M} z_{ik}^r} \quad (7)$$

$$\Sigma_k^{r+1} = \frac{\sum_{i=1}^{n_k}(y_i^k - \mu_k)(y_i^k - \mu_k)^T}{n_k + \sum_{i=1}^{M} z_{ik}^r} + \frac{\sum_{i=1}^{M} z_{ik}(y_i - \mu_i^{r+1})(y_i - \mu_i^{r+1})^T}{n_k + \sum_{i=1}^{M} z_{ik}^r} \quad (8)$$

$$z_{ik}^r = \frac{w_k^r f(y_i|x_i = k)}{\sum_{j=1}^{K} w_j^r f(y_i|x_i = j)} \quad (9)$$

In the case where the labelled pixels provide no information on the mixing proportions,

$$w_k^{r+1} = \frac{1}{M}\sum_{i=1}^{M} z_{ik}^r, \quad (10)$$

It is quite natural in practice that the number of labelled data, $N$, is quite small compared to the number of unlabelled pixels, $M$, i.e. $N \ll M$. In addition, the quality of training data may vary from case to case. To control the contribution of training data to the final estimation, we modify the iteration equations for the mean and covariance (Equations 7 and 8) by weighting the few labelled pixels more heavily than their unlabelled counterparts. This is done by introducing weights $u_i^c, i = 1, 2, \cdots, n_c; c = 1, 2, \cdots, K$

| Tissues | SKIN | FAT | SKULL | GM | WM | CSF |
|---|---|---|---|---|---|---|
| Labelled Pixels | 16 | 16 | 16 | 36 | 24 | 20 |

**Table 1**. Number of labelled pixels used for the semi-supervised segmentation of slice 9 of patient 3.

in Equations 7 and 8 as follows:

$$\mu_k^{r+1} = \frac{\sum_{i=1}^{n_k} u_i^k (y_i^k) + \sum_{i=1}^{M} z_{ik} y_i}{\sum_{i=1}^{n_k} u_i^k + \sum_{i=1}^{M} z_{ik}^r}, \quad (11)$$

$$\Sigma_k^{r+1} = \frac{\sum_{i=1}^{n_k} u_i^k (y_i^k - \mu_k)(y_i^k - \mu_k)^T}{\sum_{i=1}^{n_k} u_i^k + \sum_{i=1}^{M} z_{ik}^r} + \frac{\sum_{i=1}^{M} z_{ik} (y_i - \mu_i^{r+1})(y_i - \mu_i^{r+1})^T}{\sum_{i=1}^{n_k} u_i^k + \sum_{i=1}^{M} z_{ik}^r} \quad (12)$$

where $u_i^c$ is the weight of the $i'th$ labelled pixel for the $c'th$ tissue. In the most general formulation, $u_i^c$ is simply a positive real number. If $u_i^k$ is an integer, $u_i^c$ copies of the corresponding labelled pixel $y_i^c$ are effectively used. The larger the $u_i^c$, the more the $y_i^c$ contributes to the final parameter estimation. The weight allows us to tailor the estimation to agree with any expert knowledge that might be available about the quality or importance of each $y_i^c$ as a training pixel. In the absence of such knowledge, the most rational approach is to take $u_i^c = u$ for all $i$ and $c$, where $u$ is a constant.

## 4. EXPERIMENTAL RESULTS

Slices from the 12 real MRI data sets were used to test the capacity and stability of separating normal brain tissues. All 12 data sets were scanned with a spin echo pulse sequence at repetition times from 1800 msec to 3000 msec. Each data set consisted of about 20 slices covering almost the whole brain and each slice had dual spin echoes: PWD and T2W images. PDW and T2W images were scanned at $T_E = 16$ msec and $T_E = 98$ msec respectively.

Figure 1(a) shows the dual spin echo MR images of slice 9, patient 3, with PDW echo on left and T2W on right. 128 pixels were labelled manually as training data for the six normal tissues: SKIN, FAT, SKULL, GM, WM and CSF. The number of labelled pixels for each tissue is listed in Table 1.

As the labelled pixels occupied only about 0.5% to 1.0% of the true size of the corresponding tissues, it was not possible for those labelled pixels to catch either the cluster centers or the cluster shapes. Therefore, the selection of the confidence weights would influence the segmentation accuracy. For simplicity or in cases without prior knowledge about the expected sizes of tissues, all weights were set equal. A comparison study [6] demonstrated $u = 50.0$ gave the most accurate segmentation in comparion to manual segmentation.

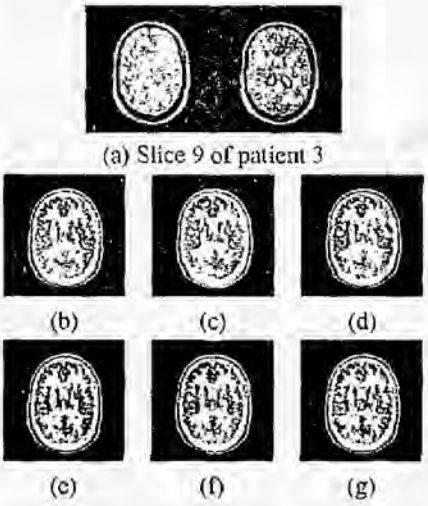

**Fig. 1**. Semi-supervised segmentation of slice 9, patient 3, (a), at $u = 50.0$ given labelled data for various tissues. (a) Labelled data only for WM. (b) Labelled data for WM and GM. (c) Labelled data for WM, GM and CSF. (d) Labelled data for WM, GM, CSF and SKIN. (e) labelled data for all tissues except FAT. (f) Labelled data for all tissues.

Noticeably there were few differences in the intracranial region separation from the slice among the supervised, semi-automatic and automatic approaches. However, there was significant improvement in the separation of CSF, WM, and GM. The segmentation errors were reduced from 13%, 31%, and 22% in supervised segmentation to 3%, 2.6% and 4% in semi-supervised segmentation for these three brain tissues respectively. In comparison with the unsupervised approach, the segmentation errors are also reduced by about 5% on average for CSF, WM and GM.

Figure 1(b) through (g) show the segmentations when labelled pixels were available for various tissues. The segmentations were obtained by setting $u = 50.0$. From (b) to (g), the number of tissues having labelled pixels was gradually increased, where (b) is the case when only WM had training pixels while (g) corresponds to the case when all tissues had labelled pixels. As expected, the segmentation improved gradually as more and more tissues had labelled pixels.

## 5. DISCUSSION

As an unification of fully unsupervised (automatic) and supervised approaches, the proposed semi-supervised MAP segmentation shares their advantages, being less dependent

on the quality of the training data than supervised segmentation and more reliable and accurate than automatic segmentation. It does not require that training data be available for every tissue type. In other words, it makes use of any quantity of training data available to improve the segmentation. Our examples have indicated that the semi-supervised segmentation is superior to both unsupervised (automatic) and supervised ones when a small quantity of reliable training data is available for some important tissues, such as GM, WM, and CSF. When there are no training pixels, the algorithm is the same as the automatic one. If there are many reliably labelled pixels for every tissue, the algorithm can perform like a supervised approach by setting a very large value for the weight $a$.

In the semi-automatic approach, the selection of the confidence weights seems crucial to the final results. In principle, the weights are expected to be set individually according to the ratio of the number of labelled pixels to the total pixel number of the tissues. Choosing same value for all weights is the simplest, but in such a case, it is recommended that number of labelled pixels be roughly proportional to the total pixel number of each tissue. Moreover, caution should be exercised when manually labelling pixels. Since we normally labelled pixels in a $n \times n$ window simultaneously, it is easy to include some noisy pixels. When the weights are set large, the effect of the noise will be automatically magnified. So we would suggest the use of small $n$, say 2 or 3, and discarding the pixels in the window which have the largest intensity or smallest intensity. In this way, the quality of the labelled pixels might be improved. (It should be pointed out that we neither proportionally select the training data nor consider the noise effect on the data in our previously presented examples).

## 6. REFERENCES


[1] S. M. Ahmed, M. N.and Yamany, N. Mohamed, A. A. Farag, and T. Moriarty. A modified fuzzy C-means algorithms for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3):193–199, March 2002.

[2] A.M. Bensaid, L.O. Hall, J.C. Bezdek, and L.P. Clarke. Partially supervised clustering for image segmentation. *Pattern Recognition*, 29(5):859–871, May 1996.

[3] J.C. Bezdek, L.O. Hall, and L.P. Clarke. Review of MR image segmentation techniques using pattern recognition. *Am. Assoc. Phys. Med.*, 20(4):1033–48, Jul/Aug 1993.

[4] M.C. Clark, L.O. Hall, et al. MRI segmentation using fuzzy clustering techniques. *IEEE Engineering in Medicine and Biology*, pages 730–742, Nov/Dec 1994.

[5] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, and M.L. Silbiger. MRI segmentation: Methods and applications. *Magnetic Resonance Imaging*, 13(3):343–368, 1995.

[6] W. Li. *Automatic Segmentation of Multi-spectral Magnetic Resonance Images*. PhD thesis, CIIPS, Department of Electrical and Electronic Engineering, The University of Western Australia, Nedlands, WA 6907, Feburary 1997.

[7] Z. Liang. Tissue classification and segmentation of MR images. *IEEE Engineering in Medicine and Biology*, pages 81–85, March 1993.

[8] Z. Liang, J. R. MacFall, and D. P. Harrington. Parameter estimation and tissue segmentation from multispectral MR images. *IEEE Transactions on Medical Imaging*, 13(3):441–449, Sept 1994.

[9] H. Liu and S. T. Huang. Evolutionary semi-supervised fuzzy clustering. *Pattern Recognition Letter*, 24:3105–3113, 2003.

[10] Y. Wang, T. Adali, J. Xuan, and Z. Szabo. Magnetic resonance image analysis by information theoretic criteria andstochastic site model. *IEEE Trans. Information Technology in Biomedicine*, 5(2):150–158.

[11] Yue Wang and Tianhu Lei. A new stochastic model-based image segmentation technique for MR image. In *Proceedings of ICIP'94*, volume II, pages 182–186. IEEE Singal Processing, IEEE Press, 1994.

[12] P. P. Wyatt and J. A. Noble. MAP MRF joint segmentation and registration of medical images. *Medical Image Analysis*, 7:539–552, 2003.

[13] J. Zhang, J.W. Modestino, and D.A. Langan. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Transactions on Image Processing*, 3(4):404–420, July 1994.

[14] Y. Zhang, Brady. M., and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Medical Imaging*, 20(1):45–57, Jan. 2001.

[15] T. Zhu, C.and Jiang. Multicontext fuzzy clustering forseparation of brain tissues in magnetic resonance images. *NeuroImaging*, 18:685–696, 2003.