International Symposium on Signal Processing and its Applications, ISSPA, Gold Coast, Australia, 25-30 August, 1996.
Organized by the Signal Processing Research Centre, QUT, Brisbane, Australia.

A NEW APPROACH TO UNSUPERVISED MARKOV RANDOM FIELD-BASED SEGMENTATION OF MR IMAGES

Mark W. Morrison, Alison A. Dingle and Yianni Attikiouzel

Centre for Intelligent Information Processing Systems
Department of Electrical and Electronic Engineering
The University of Western Australia
Nedlands, Perth, Western Australia
email: mark@ee.uwa.edu.au

ABSTRACT

This paper describes a new approach to the unsupervised segmentation of images. A Markov random field model is used for prior label field modelling. Unlike conventional stochastic model-based approaches, each image class is not characterised by a parametric model. The algorithm compares the data in local windows around each pixel with the global distribution of data in each class using appropriate distance metrics. A novel method for determining the number of image classes is presented.

1. INTRODUCTION

There has been a large amount of research published on the application of statistical methods to the problem of MR image segmentation. In the most basic case, clustering techniques have been applied in both supervised and unsupervised contexts. In particular the EM algorithm has been applied to the task of unsupervised Gaussian mixture modelling [1, 2, 3, 4]. Information theoretic (IT) measures, such as the Akaike Information criterion (AIC) [5] or minimum description length (MDL) [6] can then be used for cluster validation.

More recently, improvements have resulted through the application of Markov random field (MRF) image models [7, 8, 9, 10]. In particular, MRFs allow prior contextual anatomical information, and our intuitive notions regarding the physical properties of images to be incorporated easily into the segmentation process.

Currently there are two major drawbacks that exist with these approaches. Firstly, in almost all cases, parametric models are utilised to characterise the image formation process. That is, each tissue is assumed to give rise to a characteristic MR signature — Multidimensional Gaussian models are used almost exclusively. There is a need, therefore, to estimate parameters for these models. Supervised methods, which rely on the collection of training data from exemplar images have been shown to be impractical for this task. Unsupervised methods are generally unreliable, and are in most cases very sensitive to the initial estimates used. In addition, there is no inherent reason to suggest *a-priori* that any particular parametric model is a "good" model for the data, and by imputing an incorrect model on the data, sub-optimal segmentations may result.

The second difficulty is in deciding how many tissues are present in the images — the cluster validation problem. Once again, information theoretic criteria have been used. The use of Markov field models, however, introduces dependencies into the data, which tend to make the use of IT criteria both uncertain and computationally expensive (see, for example, [10]).

In this paper, we propose a new method of unsupervised MR image segmentation, based on the recent work of Kervrann and Heitz [11] and Geman & Geman [12], which does not rely on parametric approaches to data modelling. Furthermore, the number of tissues in the image is estimated as an integral part of the segmentation process, rather than as a post-hoc "goodness of fit" measure as provided by AIC or MDL.

2. PROBLEM FORMULATION

Let $\mathcal{L}$ be a rectangular array of sites, $\mathcal{L} = \{(i,j); 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$. Furthermore, for notational convenience, assume that the two-dimensional lattice can be indexed by a single parameter in the manner described by [13], where each site is referenced by a single index, s. Associated with each site $s \in \mathcal{L}$ is a label, x_s, that specifies the image class to which the site belongs. Thus, $x_s = r, r \in \{1, 2, \ldots, K\}$ where K is the number of image classes[1]

It is assumed that there is a true, but unknown, labelling of sites, $\mathbf{X}^* = \{x_s^*; s \in \mathcal{L}\}$. In the context of *unsupervised segmentation*, the problem is to determine the number of image classes, K, and to reconstruct an estimate, $\hat{\mathbf{X}}$, of the true label field from the observed image data, $\mathbf{Y} = \{\mathbf{y}_s; s \in \mathcal{L}\}$; where the image data consists of M observations at each site, $\mathbf{y}_s = [y_s^{(1)}, y_s^{(2)}, \ldots, y_s^{(M)}]$. In addition to the observed data, additional constraints may be imposed on the estimated labelling by the application

[1] Note that each site is considered to belong completely to one class. In general, it may be the case that the true labelling of each site is composed of a mixture of classes - this is the case of so-called "fuzzy" segmentation. This type of model is quite appropriate for MR image segmentation, where partial volume effects can have a significant impact on the performance of "hard" segmentation algorithms. The extension of the algorithm described in this paper to "fuzzy" segmentation is being investigated, but will not be considered here.

of predefined image models for the label field and through assumptions placed on the image formation process. In unsupervised segmentation, the parameters associated with these models either need to be estimated directly from the observed data, or are preset to default values independent of the image data.

3. STOCHASTIC MODEL-BASED IMAGE SEGMENTATION

In terms of the so-called stochastic model-based approach to segmentation [8, 2, 3, 14, 15, 10], the above problem can be stated as

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}\in\Omega} P(\mathbf{X}|\mathbf{Y}) \tag{1}$$

where Ω is the sample space of all possible segmentations. That is, choose the segmentation which has the highest likelihood, given the data that was observed. By use of Bayes theorem, (1) can be reformulated as

$$\begin{aligned}\hat{\mathbf{X}} &= \arg\max_{\mathbf{X}\in\Omega} \frac{P(\mathbf{Y}|\mathbf{X})P(\mathbf{X})}{P(\mathbf{Y})} \\ &= \arg\max_{\mathbf{X}\in\Omega} P(\mathbf{Y}|\mathbf{X})P(\mathbf{X}) \\ &= \arg\max_{\mathbf{X}\in\Omega} [\log P(\mathbf{Y}|\mathbf{X}) + \log P(\mathbf{X})]\end{aligned} \tag{2}$$

Therefore, to obtain an estimate of the label field, it is necessary to specify an observation model, $P(\mathbf{Y}|\mathbf{X})$, and a prior model, $P(\mathbf{X})$. This is known as the maximum *a-posteriori* (MAP) estimate of $\hat{\mathbf{X}}$. Note that in the case of an assumption of a *uniform prior* on the label field[2], (2) reduces to the maximum likelihood (ML) estimate

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}\in\Omega} P(\mathbf{Y}|\mathbf{X}) \tag{3}$$

3.1. Gaussian Mixture Models

The most widely employed assumption in image segmentation is that the observation model can be described by a conditionally independent Gaussian mixture model [3, 16, 17]. In this case, the observations at each site are assumed to be conditionally independent and to have the same conditional probability, $p(\mathbf{y}_s|x_s)$, dependent only on x_s. Therefore, the observation model can be written as

$$P(\mathbf{Y}|\mathbf{X}) = \prod_{s\in\mathcal{L}} p(\mathbf{y}_s|x_s) \tag{4}$$

and with the assumption of a multivariate Gaussian p.d.f.

$$p(\mathbf{y}_s|x_s = k) = \frac{1}{\sqrt{(2\pi)^M|\Lambda_k|}} e^{\left(-\frac{1}{2}(\mathbf{y}_s-\mu_k)^T\Lambda_k^{-1}(\mathbf{y}_s-\mu_k)\right)} \tag{5}$$

where μ_k and Λ_k are the mean and covariance matrix.

[2]That is, all prior distributions are equally probably - generally used in the case when no prior information is available regarding the properties of the label field

3.2. Markov Random Fields

The use of the above model in ML image segmentation has been widely reported [1, 18, 4]. The efficacy of the ML approach, however, has proved to be less than satisfactory. More recently, the MAP approach has become more prevalent. In particular, the use of Markov random field models for the prior, $P(\mathbf{X})$, has been widely reported.

In very general terms, the use of Markov random field priors allows the introduction of a degree of spatial dependence between labels into the segmented image. In its simplest form, this implies a tendency for the label at any particular site to be similar to labels at neighbouring sites. An equivalent formulation[3] of an MRF is to describe the spatial dependency in terms of joint probabilities in the form of a Gibbs Distribution. In this case, the label field prior is of the following form

$$P(\mathbf{X} = \mathbf{x}) = \frac{e^{-U(\mathbf{x})}}{\mathcal{Z}} \tag{6}$$

where $U(\mathbf{x})$ is called the *energy function* and $\mathcal{Z}$ is a normalising constant, known as the *partition function*.

A commonly used energy function is the *multi-level logistic* MRF [19]

$$U(\mathbf{x}) = \beta \sum_{s\in\mathcal{L}} \sum_{t\in\mathcal{N}_s} [1 - 2\delta(x_s - x_t)] \tag{7}$$

where β is a predefine MRF strength constant, $\delta(\cdot)$ is the usual dirac-delta function, and $\mathcal{N}_s$ is a suitably defined neighbourhood of sites about s.

3.3. Current Status

The utility and application of the models described above constitutes a substantial portion of the current literature on statistically-based image segmentation. A number of limitations to this approach have been identified.

Firstly, the determination of the model parameters (mean and covariance matrix) for the observation model is a very difficult problem. A common way of tackling this is to iteratively segment the image and then recalculate the parameters, assuming that the segmentation is correct. This procedure can be shown to converge to a partial optimal solution (POS) [20]. Simultaneous segmentation and parameter estimation can be performed by using an approach akin to the EM algorithm [16]. Zhang *et al* describe such an approach [10], where at each iteration, every site is partially assigned to all classes, with a "hard" segmentation being performed once the parameter values converge. This approach yields significant segmentation improvements, but the EM algorithm is very sensitive to initial conditions, and there is no way to guarantee that a sensible solution will result.

The second problem with mixture modelling approaches is that it is difficult to determine the correct number of image classes. The number of classes is typically estimated by searching, over a range of K, for a value that minimises a

[3]There are certain requirements for this equivalence, which are described in [13] and many of the other papers referenced in this paper.

given model-fitting criterion, $F(K)$, such as Akaike's information criterion (AIC) [5] or minimum description length (MDL) [6]. These measures are basically a trade-off between the fit of the model to the data and the complexity of the model. Experimental results with synthetic images [21], however, clearly demonstrate that they are unable to consistently estimate the correct number of image classes.

4. A NEW APPROACH

Kervrann and Heitz [11] proposed an alternative method to estimate the number of image classes, whereby new image classes are created during the segmentation process. An outlier class, ρ, is introduced and pixels are assigned to this outlier class if they do not conform to any of the existing classes. In a general segmentation task, the distribution of data within each class may not be Gaussian. Rather than imposing a given model on the data, Kervrann and Heitz proposed comparing the local distribution of data around each pixel with the distribution of data within each image class. Their methods were specifically geared toward texture segmentation, where a large number of features are calculated from the original image. The method described here is an extension of these concepts to the case of general image segmentation.

4.1. The Energy Function

In this approach, the MAP segmentation of (2) is restated as a problem of energy minimisation, and so

$$\begin{aligned}\hat{\mathbf{X}} &= \arg\max_{\mathbf{X}\in\Omega}[\log P(\mathbf{Y}|\mathbf{X}) + \log P(\mathbf{X})] \\ &= \arg\min_{\mathbf{X}\in\Omega}[-\log P(\mathbf{Y}|\mathbf{X}) - \log P(\mathbf{X})] \\ &= \arg\min_{\mathbf{X}\in\Omega}[U(\mathbf{Y}|\mathbf{X}) + U(\mathbf{X})]\end{aligned} \tag{8}$$

where $U(\mathbf{Y}|\mathbf{X})$ and $U(\mathbf{X})$ are energy functions.

The second term, $U(\mathbf{X})$, is the energy in the Markov random field. A multi-level logistic MRF is used for this, so that

$$U(\mathbf{X}) = \beta \sum_{s\in\mathcal{L}} \sum_{t\in\mathcal{N}_s} [1 - 2\delta(x_s - x_t)] \tag{9}$$

The first term, $U(\mathbf{Y}|\mathbf{X})$, measures the difference between the local distribution of data near each site s and the distribution of data for the image class k to which s is assigned ($x_s = k$). Thus, this term encourages each site s to be labelled $x_s = k$ so that the local distribution of data matches the regional distribution of data for image class k.

In order to define $U(\mathbf{Y}|\mathbf{X})$ mathematically, let W_s be a window centred on site s over which the local distribution of data is computed and let $\mathbf{Y}_{W_s}$ be the data associated with the subset of sites that lie within the window W_s. Similarly, let $\mathbf{Y}_{R_k}$ be the data associated with sites assigned to image class k. This is illustrated in figure 1. The term $U(\mathbf{Y}|\mathbf{X})$ can then be expressed as :

$$U(\mathbf{Y}|\mathbf{X}) = \sum_{s\in\mathcal{L}} V[f(\mathbf{Y}_{W_s}), f(\mathbf{Y}_{R_{x_s}})] \tag{10}$$

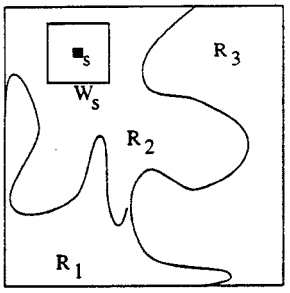

Figure 1: Local and global region statistics

where $f(\mathbf{Y})$ represents the distribution of the data $\mathbf{Y}$, $V[f, g]$ is a measure of the difference between the two distributions f and g, and $\delta(\cdot)$ is the Dirac-Delta function.

The distribution of data within each image class k is used as a model of the conditional probability density function $p(\mathbf{y}_s|x_s = k)$. Thus, $f(\mathbf{Y}_{R_k})$ is simply the histogram of the data in image class k normalised to unit area and $f(\mathbf{Y}_{W_s})$ is the normalised histogram of data within window W_s. Note that if the data $\mathbf{Y}$ is made up of M observations at each site, then $f(\mathbf{Y})$ is an M-dimensional distribution, that is, $f(\mathbf{Y}) : \mathbb{R}^M \rightarrow \mathbb{R}$.

There are many methods available for quantifying the difference between two probability density functions, the most common of which are the L_p distance, the Hellinger distance (H_p), the Kullback-Leibler number, and the Kolmogorov-Smirnov distance. An analysis of the relative merits of each of these measures can be found in [21]. For the remainder of this paper, it will be assumed that the total variation (L_1) metric has been used.

The overall energy function $U(\mathbf{Y}|\mathbf{X}) + U(\mathbf{X})$ is minimised using the iterated conditional modes (ICM) algorithm [7], which selects the label $\hat{x}_s$ at each site s to minimise the local energy

$$\hat{x}_s = \arg\min_k u(x_s = k|\mathbf{Y}, \mathbf{X}_{\mathcal{N}_s}) \tag{11}$$

where the local energy $u(x_s = k|\mathbf{Y}, \hat{\mathbf{X}}_{\mathcal{N}_s})$, is given by :

$$u(x_s = k|\mathbf{Y}, \hat{\mathbf{X}}_{\mathcal{N}_s}) = V[f(\mathbf{Y}_{W_s}), f(\mathbf{Y}_{R_k})] + \beta \sum_{t\in\mathcal{N}_s} [1 - 2\delta(k - \hat{x}_t)] \tag{12}$$

The sites are updated using a partially synchronous scheme based on coding sets [7].

4.2. Region Formation

In order to allow region formation during the segmentation process, an outlier region is also included in the choice of possible image classes during the ICM process. The manner in which this manifests itself is that any site which has an energy greater than some predetermined threshold[4] with the existing image classes is labelled as an outlier, ρ.

[4] In the implementation of the algorithm, this threshold is dynamically adjusted. Space limitations preclude a detailed description in this paper.

The segmentation procedure begins by assuming that the entire image can be represented by a single class. The probability distribution of the data is calculated from the image data, and each pixel is reassigned to the initial class, or to the outlier class, depending on how closely the data around it matches the global distribution of the initial class data. If a sufficiently large *contiguous* outlier region exists (typically 1-2% of the image is used), the pixels within that region are assigned to a new class. The class probability distributions are re-estimated based on the current labelling and the segmentation-estimation process iterates until no outlier regions of a sufficient size exist.

5. RESULTS

We have applied this segmentation algorithm to a number of synthetic images and MR images. Our results show significant improvements over other unsupervised methods for MR image segmentation. In particular, we have found that the estimation of the number of tissues present in MR images using this approach is more accurate than those that rely on information theoretic measures, such as AIC or MDL.

Due to space limitations, only the theoretical aspects of the algorithm are presented here. For a more detailed description and an extensive set of experimental results on a range of images, readers are referred to our technical report [21]. Alternatively, preprints of a full paper submitted to IEEE Trans. Image Processing may also be obtained by contacting the authors.

6. REFERENCES

[1] H.S. Choi, D.R. Haynor, and Y. Kim. Partial volume tissue classification of multichannel magnetic resonance images — a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407, September 1991.

[2] T. Lei and W. Sewchand. Statistical approach to X-ray CT imaging and its applications in image analysis - part II : A new stochastic model-based image segmentation technique for X-ray CT image. *IEEE Transactions on Medical Imaging*, 11(1):62–69, March 1992.

[3] W. Li, M.W. Morrison, and Y. Attikiouzel. Unsupervised segmentation of dual-echo MR images by a sequentially learned Gaussian mixture model. In *Proceedings of the IEEE International Conference on Image Processing*, page Accepted for publication, Washington, USA, October 1995. IEEE.

[4] M.W. Vannier, T.K. Pilgram, C.M. Speidel, L.R. Neumann, D.L. Rickman, and L.D. Schertz. Validation of magnetic resonance imaging (MRI) multispectral tissue classification. *Computerized Medical Imaging and Graphics*, 15(4):217–223, 1991.

[5] H. Akaike. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics A*, 30:9–14, 1978.

[6] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[7] J.E. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3):259–302, 1986.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[9] J.W. Woods. Two-dimensional discrete Markovian fields. *IEEE Transactions on Information Theory*, 18(2):232–240, March 1972.

[10] 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):404–419, July 1994.

[11] C. Kervrann and F. Heitz. A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics. *IEEE Transactions on Image Processing*, 4(6):856–862, 1995.

[12] D. Geman, S. Geman, C. Graffigne, and D. Pong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):609–628, July 1990.

[13] R.C. Dubes and A.K. Jain. Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.

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

[15] C.W. Therrien, T.F. Quatieri, and D.E. Dudgeon. Statistical model-based algorithms for image analysis. *Proceedings of the IEEE*, 74(4):532–551, April 1986.

[16] R.A. Redner and H.F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, April 1984.

[17] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1985.

[18] K.M. Hanson. Introduction to Bayesian image analysis. In *Proceedings of SPIE (Image Processing)*, volume 1898, pages 716–731. SPIE, 1993.

[19] H. Derin and C-S. Won. A parallel image segmentation algorithm using relaxation with varying neighborhoods and its mapping to array processors. *Computer Graphics and Image Processing*, 40:54–78, 1987.

[20] R. E. Wendell and A. P. Hurter. Minimization of a non-separable objective function subject to disjoint constraints. *Operations Research*, 24(4):643–657, 1976.

[21] A. A. Dingle and M. W. Morrison. Unsupervised image segmentation. Technical report, Department of Electrical and Electronic Engineering, The University of Western Australia, The University of Western Australia, Nedlands, Perth, WA, Australia, April 1996.