

Segmentation Of MR Images Using Neural Nets

D.L.Toulson

J.F.Boyce

Wheatstone Laboratory

Kings College

London WC2R 2LS

Abstract

A two tier system for MR image segmentation is proposed. The first stage involves probabilistically classifying the pixels of the input image based on second order grey level statistics obtained from the co-occurrence matrix. These probabilities then form the input to a multi-layer perceptron, allowing contextual constraints to be applied to local areas of the image thereby improving the spatial coherence of the classification.

1 Introduction

The objective of the work presented in this paper is to produce a data driven segmentation of cranial MR images. Such segmentation is a pre-requisite for three dimensional surface rendering algorithms, allowing three dimensional data to be graphically displayed.

It is also needed to facilitate the automated extraction of clinical data useful to a clinician in planning radiotherapy.

When segmenting MR images it is important to find regions that have contextual meaning. For this reason the segmentation process deliberately incorporates *prior knowledge* in the form of a set of manually segmented images produced by the majority decision of three radiologists.

An overview of the complete segmentation scheme is shown in Figure 1. The first step requires a probabilistic segmentation of the image.

This is achieved by forming the co-occurrence matrix of the image and then parameterising the component distributions of the matrix associated with the various image classes. Once parametrised, the distributions may be used to calculate the class membership probabilities for each pixel based on the position of its contribution to the co-occurrence matrix.

These probabilities are then input to a multi-layer perceptron to produce the final pixel classifications. This stage of the segmentation is performed in a pixel by pixel manner, the input to the neural net being the class membership probabilities of pixels from a specified neighbourhood around the pixel being classified.

2 The Image

Consider an image consisting of $N \times N$ pixels. Let us denote each of these pixels by a vector $\vec{x} = (x, y)$, $x, y = 1 \dots N$. Let the intensity of each pixel be $i(\vec{x})$ and the set of all pixels within the image be P . For the purposes of computer vision it is also useful to associate with each pixel a representative label $l(\vec{x})$ describing that part of the image of which the pixel is constituent. These labels will in general be from a small finite set, for instance in the case of MR images a possible set may be Background, Scalp, Skull, Grey Matter and White Matter. The task of classification is to provide a mapping from the pixels to the correct labels.

3 Co-occurrence Matrices

The Co-occurrence Matrix, as defined by Haralick *et al* [1] was first used as a tool for developing textural descriptors based on the second order grey level statistics of an image. More recently it has been utilised by Haddon and Boyce [2] as a feature space suitable for image segmentation. Before formally defining the matrix it is useful to make two further definitions. Let us define the Δ -Neighbour of a pixel \vec{x} to be the pixel $\vec{x} + \vec{\Delta}$ and a Δ -Pair be the pixel pair $(\vec{x}, \vec{x} + \vec{\Delta})$.

We may now define the co-occurrence matrix to be a sample of the joint intensities of all Δ -Pairs within an image for a given vector Δ . This may be written,

$$S_{\Delta}(i, j) = \sum_{\vec{x} \in P} \sum_{\vec{x}' \in P} \delta(\vec{x}; \vec{x}' + \vec{\Delta}) \delta(i; i(\vec{x})) \delta(j; i(\vec{x}')). \quad (1)$$

where δ is the Kronecker-Delta function.

One property of the co-occurrence matrix is that unlike the one dimensional grey level histogram it separates *interior* and *boundary* Δ -Pairs into on- (i.e. $i = j$) and off-diagonal contributions allowing each to be examined separately.

4 Probabilistic Segmentation

In order to probabilistically segment an image it is first necessary to represent the intensity of a pixel with a given label as a *random variable*. The distribution of intensities across the entire image $p(i)$ may then be written as a finite mixture of the class conditional distributions,

$$p(i) = \sum_{a=1}^A \pi_a p_a(i) \quad (2)$$

where A is the number of distinct classes in the image, π_a is the *a priori* probability for a pixel to belong to class a and $p_a(i)$ is the class conditional

intensity distributions. In a similar way the distribution of the joint intensities of $\Delta - Pairs$, $p(i, j)$ within an image may be written

$$p(i, j) = \sum_{a,b=1}^A \pi_{a;b} p_{a;b}(i, j) \quad (3)$$

where $\pi_{a;b}$ is the *a priori* probability that a pixel and its $\Delta - Neighbour$ belong to classes a and b respectively and $p_{a;b}(i, j)$ is the corresponding joint intensity distribution for such $\Delta - Pairs$.

As noted in the previous section, it is possible to split the distribution of joint intensities into two separate terms representing the *interior* and *boundary* contributions respectively.

$$p(i, j) = \sum_{a=1}^A \pi_{a;a} p_{a;a}(i, j) + \sum_{a \neq b} \pi_{a;b} p_{a;b}(i, j) \quad (4)$$

Now, the purpose of probabilistic labelling is to assign class membership probabilities for each pixel of the image. To simplify this, we shall ignore the relatively smaller off-diagonal boundary distributions, allowing us to write [using Bayes theorem]

$$p(a | i, a | j) = \frac{\pi_{a;a} p_{a;a}(i, j)}{p(i, j)} \quad (5)$$

where $p(a | i, a | j)$ is the probability of a pixel belonging to class a given that it's intensity is i and its $\Delta - Neighbour's$ intensity is j .

The denominator $p(i, j)$ is a class independent normalising term that need not be explicitly calculated. This leaves the problem of finding the densities $\pi_{a;a}$ and $p_{a;a}(i, j)$. These will not usually be known but may be estimated from the co-occurrence matrix.

The exact methodology for finding the distributions will be dependent on such factors as the availability of training data and the assumed parametric form (if any) of the distributions.

5 Parameterisation Of The Co-occurrence Matrix

Most techniques for parameterising a sample distribution assume prior knowledge of the parametric family to which the distribution belongs. For this reason we assume that the class conditional distributions $p_{a;a}(i, j)$ being sought in the co-occurrence matrix were multivariate normals centered roughly on the leading diagonal. It was found that this assumption is made more valid if smaller local areas of the image are considered independently, reducing the effect of large scale intensity variations across the image. It is then possible to write the Likelihood Function of the parameters in terms of the co-occurrence matrix,

$$\mathcal{L}_0(\psi) = \prod_{i,j=1}^L \left[\sum_{a=1}^A \pi_{a;a} p_{a;a}(i, j) \right]^{S_{\Delta}(i,j)} \quad (6)$$

where,

$$p_{a;a}(i, j) = \frac{1}{2\pi |\Sigma_a|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_a)^T \Sigma_a^{-1} (\vec{x} - \vec{\mu}_a)\right\}. \quad (7)$$

We wish to find the set of parameters $\psi = (\pi_a, \mu_a, \Sigma_a)$ which maximises $\mathcal{L}_0(\psi)$. This can be done by locating the stationary point

$$\frac{\partial \mathcal{L}_0(\psi)}{\partial \psi} = 0. \quad (8)$$

Unfortunately this cannot be solved explicitly and moreover $\mathcal{L}_0(\psi)$ contains a number of singularities degrading the performance of *gradient descent* methods. Fortunately there is an alternative method for parameterising a finite mixture of distributions using the EM algorithm [4].

The EM algorithm operates by starting with an initial estimate of the parameters and then iteratively updating them based on the actual sample data. For the case of a finite mixture of multivariate normals we can write the update equations as,

$$\begin{aligned} \pi_{a;a} &= n^{-1} \sum_{i,j} \omega_{i,j,a} S_{\vec{\Delta}}(i, j), \\ \mu_{a;a} &= n_a^{-1} \sum_{i,j} \omega_{i,j,a} S_{\vec{\Delta}}(i, j) \vec{x}, \quad \vec{x} = (i, j), \\ \Sigma_{a;a} &= n_a^{-1} \sum_{i,j} \omega_{i,j,a} S_{\vec{\Delta}}(i, j) (\vec{x} - \vec{\mu}_a)(\vec{x} - \vec{\mu}_a)^T \end{aligned}$$

where

$$\begin{aligned} n &= \sum_{i,j} S_{\vec{\Delta}} \\ n_a &= n \pi_{a;a} \\ \omega_{i,j,a} &= \frac{\pi_{a;a} p_{a;a}(i, j)}{p(i, j)} \end{aligned}$$

In practice the initial estimates for the parameters were found using the manually segmented training data and the scope of the parameters limited by this data to avoid 'contextually meaningless' classes being found.

6 Classification By Neural Net

One of the main problems associated with statistical classifiers relying solely on grey level information is their inability to take into account spatial relationships between pixels, resulting in characteristically *noisy* segmentations. Methods such as relaxation labelling have been applied to reduce such effects [3] although their iterative nature leads to a somewhat slower segmentation process. As an

alternative the use of a multi-layer perceptron is proposed. The pixels are classified individually by the perceptron, taking as input the class probabilities of pixels over a pre-specified mask and outputting the label of the central pixel. This is shown diagrammatically in figure 2.

The training data for the net is formed using the manual segmentations as the *Ground Truth*. The net is then trained using Back Error Propagation [5].

The purpose of using the neural net is to allow contextual rules to be learnt over local areas, improving the spatial consistency of the segmentation. It also adds *expert knowledge* into the segmentation where boundaries may not be justifiable based solely on the grey level of the pixels alone.

7 Conclusion

Figure 3 shows the results obtained using the techniques described in this paper. As a comparison a net was also trained using the original *raw* grey levels of the image instead of the probabilistic classification of the cooccurrence segmentation. It was found that using mask sizes up to 9×9 the net's performance was seriously degraded, confirming the importance of the initial probabilistic segmentation.

The two main conclusions drawn from the results were that

- Employing cooccurrence mapping as a preprocessor to a neural net improves the performance significantly.
- The composite net was extremely successful at classifying the skull/scalp regions but less successful with the grey/white matter.

At present the process does not explicitly incorporate edge information. This will be the next step and the structure of the process will facilitate the easy integration of edge/region information by simply adding edge information as input to the neural net.

8 Acknowledgment

Darren Toulson wishes to thank the IBM Scientific Center, Winchester for their support of his CASE studentship. We would also like to thank the Gregorio Marañon Hospital, Madrid for the test data produced as part of the AIM project COVIRA.

References

- [1] R.M.Haralick, K.Shanmugan, Its'Hak Dinstein Texture Features for Image Classification, 1973, IEEE Trans. Sys. Cyber., SMC-3 , 610-621.
- [2] J.F.Haddon, J.F.Boyce Image Segmentation by Unifying Region and Boundary Information. PAMI-12 921-948. 1990.
- [3] S.Peleg A New Probabilistic Relaxation Scheme IEEE Trans. Patt. Anal. Mach. Int. PAMI-2, 362-369, 1980.

- [4] A.P.Dempster, N.M.Laird, D.B.Rubin. Maximum Likelihood Estimation From Incomplete Data via the EM Algorithm (with discussion). *J.R. Statis. Soc. B* 39, 1-38.
- [5] D.E.Rummelhart, G.E. Hinton, R.J.Williams Learning Internal Representations By Error Propagation. In *Parallel Distributed Processing*, Chapter 8. MIT Press, Cambridge, Mass., 1986

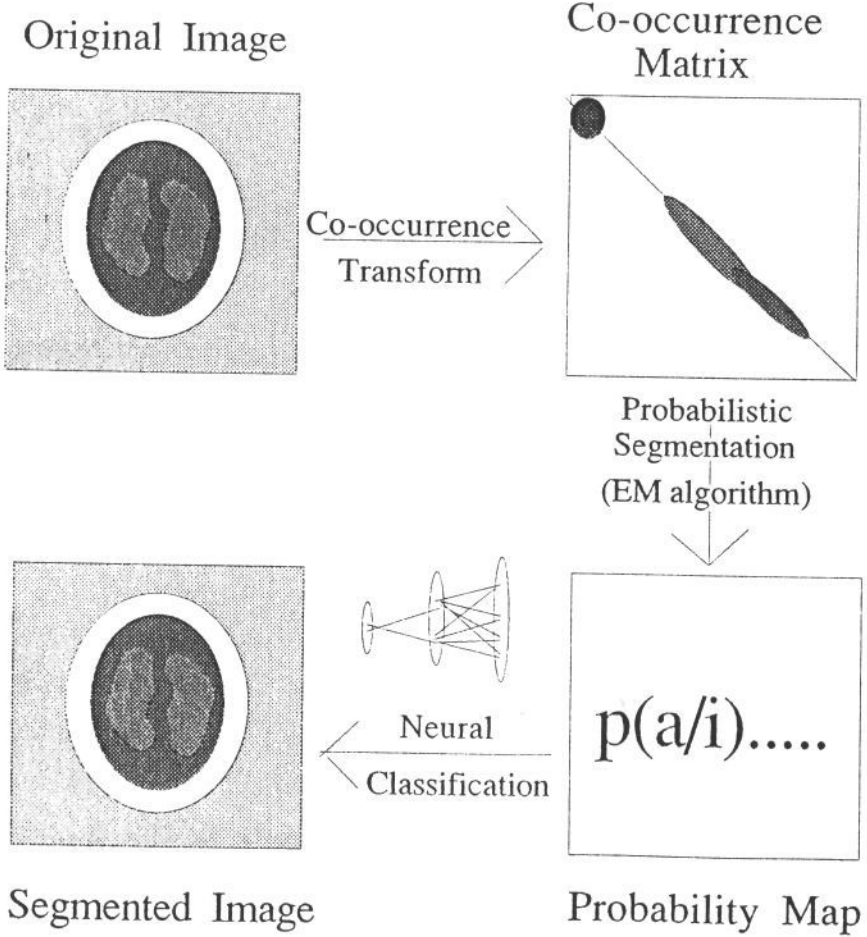


Figure 1: Overview

Neural Implementation

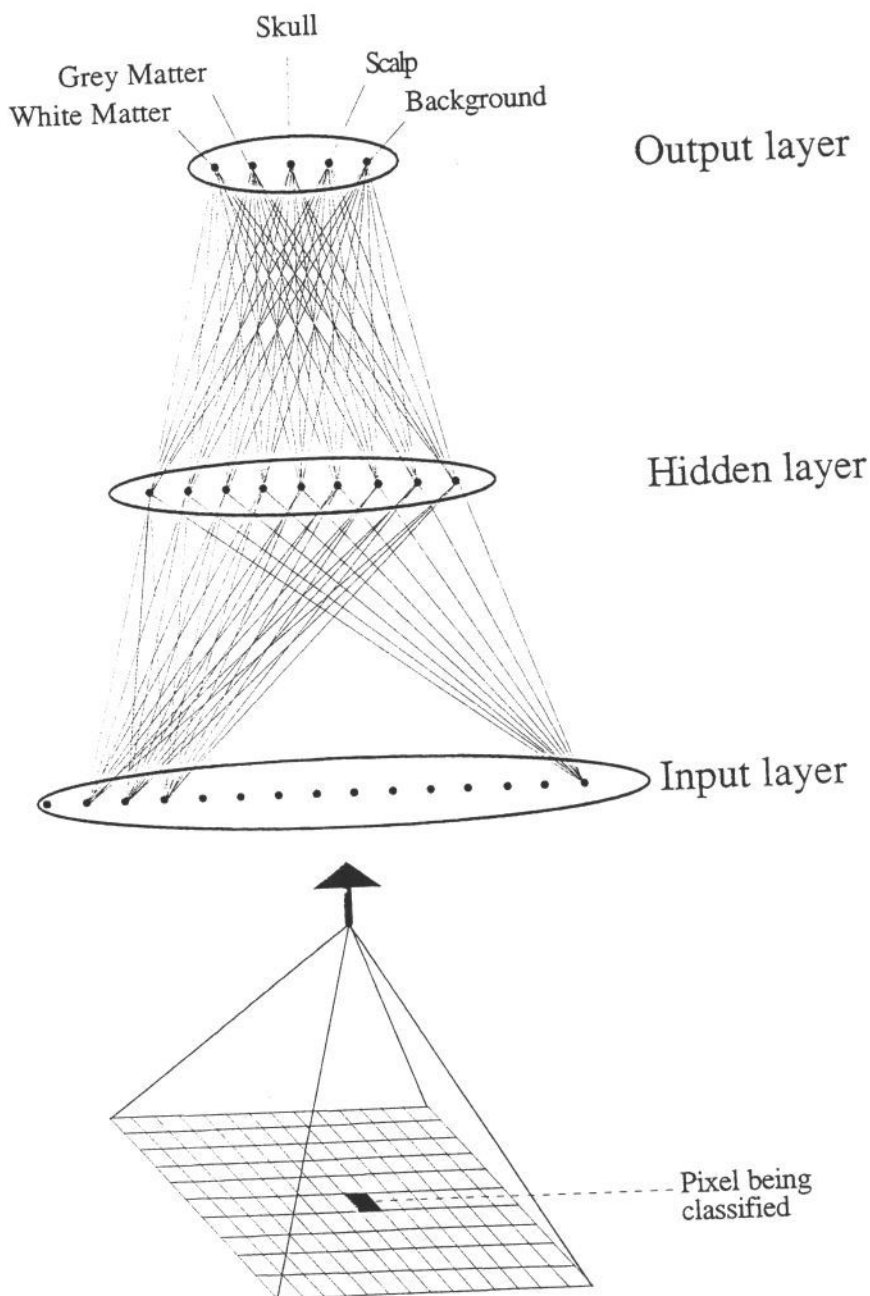


Figure 2: Neural representation of data

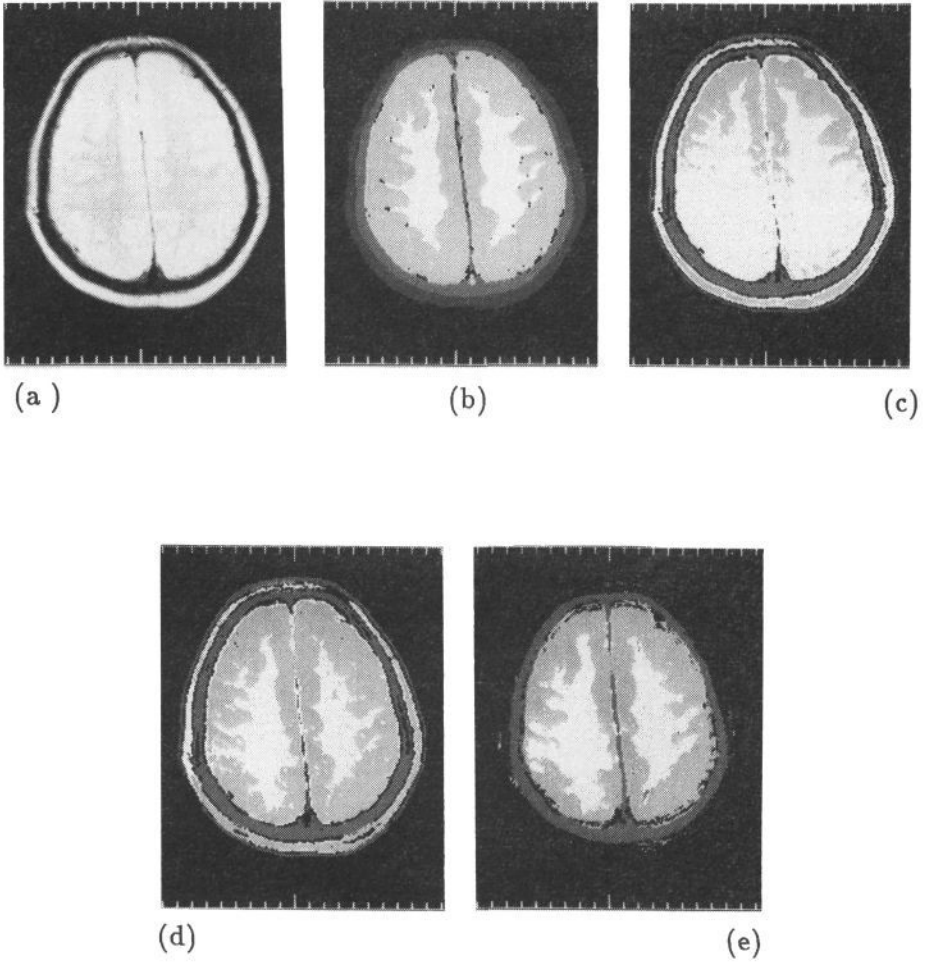


Figure 3. (a) The original image (b) Manual Segmentation (c) Segmentation Using Co-occurrence Matrices , whole image (82.1%) (d) Segmentation Using Co-occurrence Matrices ,64 x 64 sub-images (85.6%) (e) Enhancement of (d) using neural net with 7x7 input mask (92.4%)