Automatic Generation of Training Data for Brain Tissue Classification from MRI

Chris A. Cocosco, Alex P. Zijdenbos, and Alan C. Evans

McConnell Brain Imaging Centre, Montreal Neurological Institute,
McGill University, Montreal, Canada
`crisco@bic.mni.mcgill.ca`

Abstract. A novel fully automatic procedure for brain tissue classification from 3D magnetic resonance head images (MRI) is described. The procedure uses feature space proximity measures, and does not make any assumptions about the tissue intensity distributions. As opposed to existing methods, which are often sensitive to anatomical variability and pathology (such as atrophy), the proposed procedure is robust against morphological deviations from the model. Starting from a set of samples generated from prior tissue probability maps (the "model") in a standard, brain-based coordinate system ("stereotaxic space"), the method reduces the fraction of incorrectly labeled samples in this set from 25% down to 5%. The corrected set of samples is then used by a supervised classifier for classifying the entire 3D image. Validation experiments were performed on both real and simulated MRI data; the Kappa similarity measure increased from 0.83 to 0.94.

1   Introduction

Fully automatic, accurate, and robust brain tissue classification[1] from anatomical magnetic resonance images (aMRI) is of great importance for research and clinical studies of the normal and diseased human brain. Operator-assisted segmentation methods are impractical for large amounts of data, and also are highly subjective and non-reproducible [1].

Existing methods for fully automatic brain tissue classification typically rely on an existing anatomical model. This makes them sensitive to any deviations from the model due to pathology, due to aging, or simply due to normal anatomical variability between individuals. Also, there may be situations when the only model available was constructed from a different human population than the image to be classified. Moreover, many of the published feature-space classification methods assume multi-variate Normal (Gaussian) tissue intensity distributions. It has been shown that this is a poor assumption for multi-spectral anatomical brain MRI [2, 3].

---

[1] In the context of this paper, "classification" means the labeling of individual image voxels as one of the main tissue classes in the brain: cerebro-spinal fluid (CSF), grey matter, and white matter; a fourth class ("background") denotes everything else.

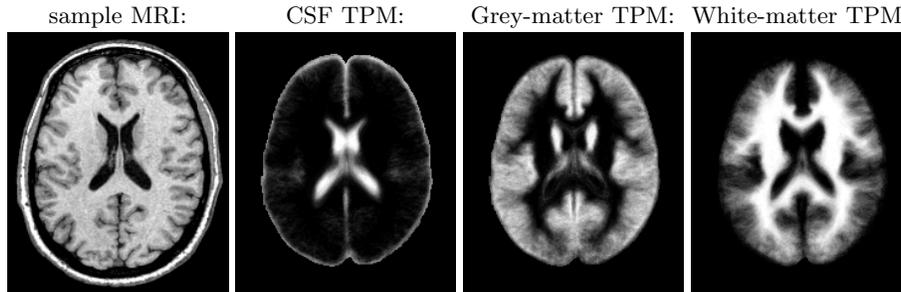sample MRI: CSF TPM: Grey-matter TPM: White-matter TPM:

Fig. 1. All 3D image volumes are spatially registered to the same stereotaxic space. The tissue probability map (TPM) values range from 0 (black) to 1 (white).

The MRI intensity scale has no absolute meaning, and is dependent on the pulse sequence and other variable scanner parameters. An aspect that is ignored by most brain MRI intensity-based classification schemes is the *fully automatic* generation of a *correct* set of training samples for the classifier, when given a never-seen-before MRI brain dataset. Existing approaches to fully automatic classification include:

– EM-style schemes were proposed by Van Leemput [4] and by Ashburner [5]. Both use a probabilistic brain atlas to initialize and constrain the tissue classification. However, both authors report failures on atypical (significantly different than the atlas) brain scans, such as child or pathological brains.
– The use of stereotaxic space tissue probability maps for automating supervised classification algorithms was originally proposed by Kamber [6], and subsequently used by other researchers [1, 7]. The maps are used to select training samples from spatial locations that are very likely to contain a given tissue type. This approach's limitations are described in the next section.

The main contribution of this paper is a novel method for fully automatic generation of correct training samples for tissue classification. The method is non-parametric, hence does not make any assumptions about the feature space distributions. It is based on a prior tissue probability map in stereotaxic space (the "model"), and is designed to accommodate subject anatomies that are significantly different than the model.

2   Problem Statement

A stereotaxic space tissue probability map (TPM) of a given tissue is a spatial probability distribution representing a certain subject population. For each spatial location in a standard, brain-based, coordinate system (stereotaxic space), the TPM value at that location is the probability of the given tissue being observed there, for that particular population.

Once imaging data is spatially registered (normalized) to the stereotaxic space, TPM-s provide an a-priori spatial probability distribution for each tissue (Fig. 1). This distribution can be used to automatically produce a training set for the supervised classifier [6]: for example, choose spatial locations that have a TPM value $\geq \tau = 0.99$ (99%). The lower the τ, the more qualifying spatial locations there will be. However, this simplistic approach has two limitations:

1. Mis-labeled samples ("false positives"): Since the morphology of the human brain is so variable, even among the locations with very high a-priori probability of being a given tissue, some of them will be wrongly labeled as one tissue class when in fact they are from another class. The fraction of false positives in the training set will increase when τ is decreased. Also, for a given τ, this fraction will be larger when the subject is from a different population than the population represented by the TPM.

2. Intensity distribution estimation: For highest τ (where the false positive rate is lowest) the qualifying sample points give a very limited coverage of the brain area, especially for CSF (Fig. 1). Intuitively, this will not give a good estimate of the true tissue intensity distributions (which is needed by a supervised classifier), for two reasons:

- Brain tissue, as seen in aMRI, is not homogeneous throughout the brain [8].
- MRI artifacts, such as intensity non-uniformity (INU), introduce additional spatial variations in the measured tissue signal.

Thus, sampling at a lower τ would be beneficial for the intensity distribution estimation; however, a lower τ also means more false positives.

Our novel contribution is a way to address these two limitations. Specifically, a "pruning" of the raw set of points obtained from the TPM is performed, with the goals of eliminating the false positives caused by anatomical difference, and of allowing for a lower TPM τ. The only requirement for the TPM is that the majority of training points it provides, for a given τ, are correctly labeled.

3   Method

The following presents a fully automatic, non-parametric, brain tissue classification procedure based on feature space proximity measures. Non-parametric classifiers are attractive because they do not make any assumptions about the underlying feature space data density functions. The procedure consists of two stages:

1. A semi-supervised classifier, using a minimum spanning tree graph-theoretic method, and stereotaxic space prior information. It produces a set of training samples customized for the particular individual anatomy subjected to classification. This stage will be referred to as the "pruning" stage.

2. A supervised classifier, using the classic k-nearest-neighbor (kNN) algorithm [9]. It is trained on the set of samples produced by the first stage. In this work: $k = 45$, and the classifier had > 3000 training points per class.

The image features used are only signal intensities of one or more MRI modalities (contrasts). The feature space proximity measure used is the common Euclidean distance in d-dimensional space. However, for $d > 1$ the Euclidean distance is not invariant to independent scaling of the different axes, and the MRI scanner raw output has no absolute nor guaranteed scale. This problem is addressed by a pre-processing step that normalizes the intensities of the input multi-spectral MRI-s. A simple intensity histogram range-matching procedure is used: points located at a small percentile away from the absolute minimum/maximum of the histograms are matched between MRI modalities.

3.1   Pruning stage

The pruning works on a set of input sample points that are selected through random sampling from the qualifying locations in the respective tissue probability map (TPM); an equal number of samples is selected for each tissue class (background, CSF, grey matter, white matter). The qualifying locations are locations where the TPM value (i.e. the prior probability) is $\geq \tau$, where τ is the threshold parameter.

The pruning technique makes use of a minimum spanning tree (MST) in feature space. This method is referred to as "semi-supervised" because, unlike in traditional unsupervised classification ("clustering" techniques), some prior information exists in this application: the number of main clusters, and their relative position in feature space is known. Furthermore, each sample point has an initial labeling suggested by the TPM-based point selection process (section 2). The purpose of the pruning is to reject the points with incorrect labeling.

Here are the three main steps of the pruning method:

1. The minimum spanning tree of the input set of points is constructed in feature space (Fig. 2).
2. Iteratively, the graph is broken into smaller trees (connected components, or clusters) by removing "long", or "inconsistent", edges from the initial MST. At each step, the *main clusters* are identified and labeled by using prior knowledge, and a stop condition is tested on them. If the condition is not satisfied, the graph breaking is continued.
3. At the end, the points that are in the right cluster (i.e. have the same initial labeling as their cluster) are deemed to be true positives and kept; all the other points are deemed to be incorrectly labeled and discarded.

MST breaking: A heuristic method (inspired by [9]) was implemented and experimentally evaluated (section 4). It uses a threshold value T, which is decreased at each iteration of the algorithm and tested on all edges of the graph in parallel:

– an edge (i, j) is removed if $length(i, j) > T \times A(i)$ or if $length(i, j) > T \times A(j)$, where $A(i)$ is the average length of all the other edges incident on node i.
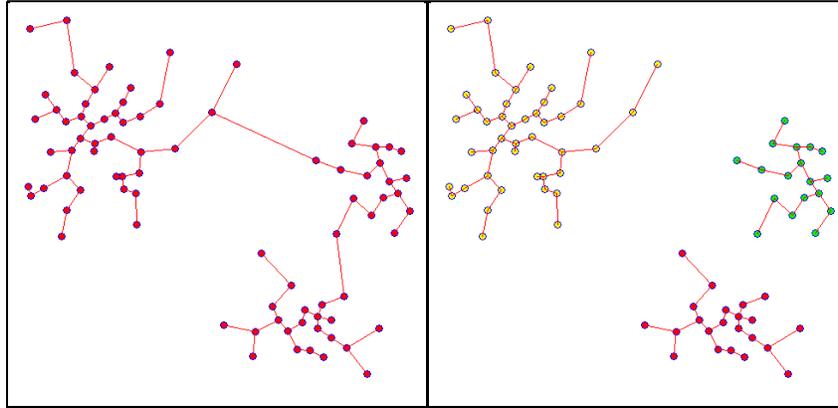
Fig. 2. *Left*: minimum spanning tree (MST) of a set of points in the plane. *Right*: the result of removing the "inconsistent" edges (section 3.1), for a $T = 1.45$.

Main clusters identification: The main clusters are the best guesses for the true background, CSF, grey matter, and white matter clusters in feature space. Under the assumption that the majority of points have correct initial labels, the best guess for each class is the cluster which contains the largest number of points labeled as that class.

Stop condition: If the above determined main clusters are found to be four distinct clusters, and the relative cluster locations in feature space correspond to prior knowledge, the iterative graph breaking stops.

4   Experiments and Results

Experiments were performed in order to validate the training set pruning method, and also the entire brain tissue classification scheme proposed here. All experiments were performed with repetitions, for assessing the statistical significance of each resulting data point. The performance on both subject brains similar to the TPM used and, more importantly, on brains with significant morphological differences from the TPM, was explored using the following MRI datasets:

1. Realistic simulations [10] driven by a new custom set of "phantoms" (digital anatomical models) resembling elderly brains. These phantoms were produced using a non-linear spatial registration procedure between real MRI-s of elderly subjects (aged 60-70) and a standard anatomical model [10]; the resulting deformation field was inverted and used for deforming the standard phantom. Multi-spectral MRI-s (T1, T2, PD) were simulated as $1\,mm^3$ isotropic voxel acquisitions, with 3% noise and 20% INU (intensity non-uniformity).
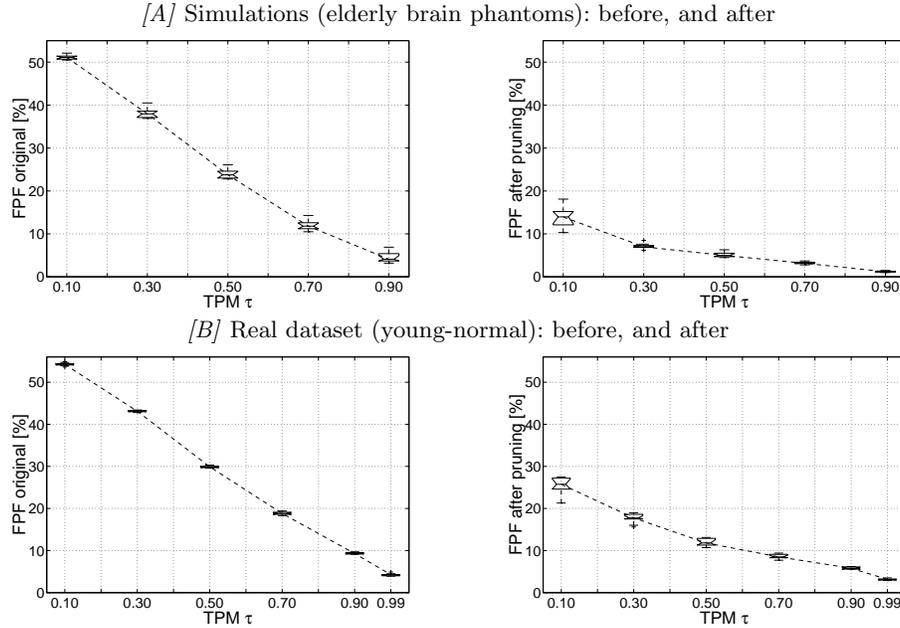
[A] Simulations (elderly brain phantoms): before, and after

[B] Real dataset (young-normal): before, and after

Fig. 3. Multi-spectral aMRI (T1-T2-PD) operation: false positives fraction (FPF) in the point set. *[A]:* 10 repetitions, each with a different phantom (anatomical model). *[B]:* 10 repetitions, each with a different initial point set. It can observed that for all experiments the pruning significantly reduces the FPF. The results for T1-only (single-feature) operation are similar.

2. Real T1-T2-PD scans of a young normal individual (aged 36), $1\,mm^3$ resolution. INU correction was performed using MNI-N3 [11]; the different acquisitions were spatially normalized to each other using a linear (affine) registration procedure. The head T1 scan was completely manually segmented (except the cerebellum) by a human expert – a trained neuroanatomist.
3. Real multi-spectral scans of 31 ischemia patients, who exhibit brain atrophy. T1: $1\,mm^3$ resolution; T2/PD: 1x1x3.5 mm resolution; same processing as for 2 above. Only a qualitative evaluation was performed on these data.

For the quantitative measurements (on 1 and 2 above), the "gold standard" was the anatomical model used for the simulations, and the manual classification for the real dataset. For each pruning experiment, 7500 candidate points per class were selected based on the TPM-s (section 3.1); the TPM-s used were produced [7] from a group of 53 young normal subjects (aged 18-35).

An intuitive figure of merit for the pruning method is the rate of false positives (mis-labeled samples) left in the point set. A low such rate is desired in the pruned point set, as it corresponds to a "mostly correct" training set for the final supervised tissue classification stage (Fig. 3).

Simulations (elderly brain phantoms):   Real dataset (young-normal):
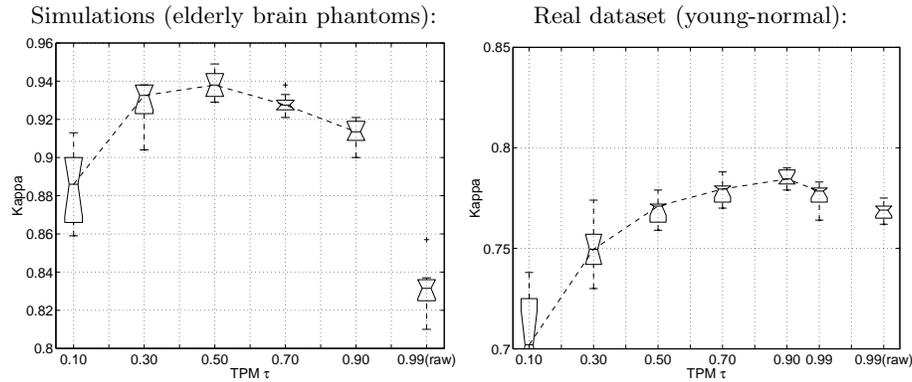

Fig. 4. Multi-spectral aMRI (T1-T2-PD) operation: final classification Kappa (repetitions same as Fig. 3). *Left:* the pruning gives a clear improvement over the "raw" (no pruning) operation. *Right:* for $\tau \geq 0.7$, the pruning produces a small, but statistically significant, improvement over "raw" (box notches do not overlap: $p < 0.05$).

However, it is more important to study how the pruning influences the final tissue classification result. For a quantitative measure of performance, the *Kappa* measure (a chance-corrected similarity measure between two labelings [12]) was computed against the gold standard, over the intra-cranial area. For comparison, Figs. 4-5 also show the result for an experiment with $\tau = 0.99$ and no pruning ("raw") [7]: the final kNN classifier was simply trained with the raw samples extracted from the TPM at $\tau = 0.99$.

5   Discussion and Conclusion

Based on the experimental results presented above, it can be concluded that the MST-based pruning method achieves its goal of reducing the rate of mislabeled samples in the point set selected using the TPM-s. Moreover, the pruning improves the final tissue classification compared to the "raw" method. This improvement is substantial for some ischemia patients (Fig. 5).

Limitations: Not all of the false positives are "pruned", and part of the true positives in the original training set are discarded as well. The cause of this is the partial overlap of the tissue class distributions in feature space – an inherent limitation of anatomical brain MRI.

This paper describes a fully automatic procedure for brain tissue classification from MR anatomical images. As opposed to existing methods, it does not make any assumptions about the image intensity distributions, or about the close morphological similarity between the subject's brain and the anatomical model. This procedure (which can operate on both single-spectral and multi-spectral aMRI) will provide improved tissue segmentation for research and clinical studies of the development, functioning, and pathology of the human brain.
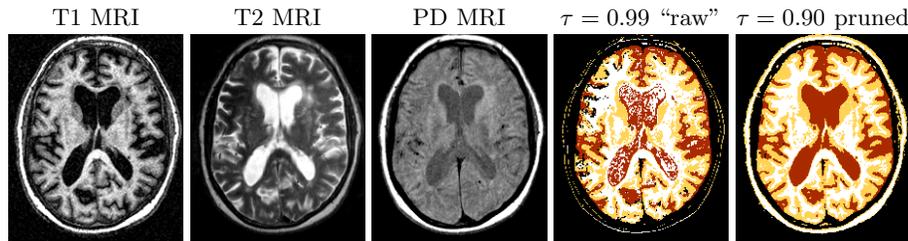
| T1 MRI | T2 MRI | PD MRI | $\tau = 0.99$ "raw" | $\tau = 0.90$ pruned |

Fig. 5. Real dataset (ischemia patient): multi-spectral aMRI, and classified images (the different gray levels correspond to the different tissues). This subject exhibits severe brain atrophy (i.e. significant morphological difference from the "young normal" model), which is the likely cause for the poor performance of the "raw" method (no pruning). The pruning method, for $\tau = 0.90$, gives significantly better classification.

References

1. Zijdenbos, A., et al. : Automatic quantification of MS lesions in 3D MRI brain data sets: Validation of INSECT. In: Proc. of MICCAI, Cambridge (1998) 439–448
2. Clarke, L. P., et al. : MRI: stability of three supervised segmentation techniques. Magnetic Resonance Imaging **11** (1993) 95–106
3. DeCarli, C., et al. : Method for quantification of brain, ventricular and subarachnoid CSF volumes from MR images. Journal of Computer Assisted Tomography **16** (1992) 274–284
4. Van Leemput, K., et al. : Automated model-based tissue classification of MR images of the brain. IEEE Trans. on Medical Imaging **18** (1999) 897–908
5. Ashburner, J., Friston, K.J.: Voxel-based morphometry — the methods. NeuroImage **11** (2000) 805–821
6. Kamber, M., et al. : Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. IEEE Trans. on Medical Imaging **14** (1995) 442–53
7. Kollokian, V.: Performance analysis of automatic techniques for tissue classification in magnetic resonance images of the human brain. Master's thesis, Concordia University, Montreal, Canada (1996)
8. Kandel, E.R., et al. : Principles of Neural Science. fourth edn. McGraw Hill (2000)
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. 2nd edn. Wiley (2001)
10. Brainweb – simulated brain database. (http://www.bic.mni.mcgill.ca/brainweb/)
11. Sled, J.G., et al. : A non-parametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Trans. on Medical Imaging **17** (1998) 87–97
12. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurements **20** (1960) 37–46