A MAP based Approach Combining Intensity, Local Prior and Multi-atlas Prior for Brain Tissue Classification

Liping Wang, Frédéric Labrosse, and Reyer Zwiggelaar

Department of Computer Science, Aberystwyth University, SY23 3DB, UK

Abstract. Automated and accurate tissue classification in 3D brain Magnetic Resonance images is essential in volumetric morphometry or as a preprocessing step for diagnosing brain diseases. However, noise, intensity inhomogeneity and partial volume effects limit the classification accuracy of the existing methods. This work performs brain tissue classification using an approach combining three commonly used features: image intensity, local prior and multi-atlas prior. The image intensity is used by K-Means to obtain the initial classification. The local prior is modelled by a Markov Random Field to better deal with the images contaminated by severe artefacts. The multi-atlas prior is derived from applying exhaustive registrations and local label fusion strategy. In this work, we apply the multi-atlas segmentation method for brain segmentation. After that, a maximum a posteriori (MAP) approach combining intensity, local prior and multi-atlas prior is used to segment the brain tissues into three types: white matter, gray matter and cerebral spinal fluid.

Keywords: Tissue classification, Markov Random Field, multi-atlas registration

1 Method

In this work, only the thick-slice T1-weighted scans are used. A multi-atlas segmentation method is applied to remove the skulls from the whole-head images. The image intensity is utilised by K-Means (KM) to obtain a preliminary classification. Considering the continuity of each tissue type in brain images, the central voxel and its neighbours in a local neighbourhood tend to belong to the same tissue class. This spatial information can be modelled by a Markov Random Field model (MRF) and combined in the classification framework as a constraint to improve the accuracy especially in cases where severe image noise and intensity inhomogeneity exist. Recently, multi-atlas segmentation (MAS) has drawn great attention because of its superior performance [1–4]. Atlas selection is carried out based on image similarity measured by mutual information[5, 6] to select the more relevant atlases. And the selected atlases are combined using a local label fusion method, which is probabilistic patch based model[7, 8], to generate the multi-atlas prior. These three image based features are all combined into a

MAP model. Finally, the brain tissues are segmented into three classes: white matter (WM), gray matter (GM) and cerebral spinal fluid (CSF) by applying the expectation-maximisation (EM) algorithm.

1.1 Brain segmentation

Five training images provided by the MRBrainS13[9] database are considered as atlases and nonrigidly registered to each testing image using demons registration[10, 11]. The brain mask of each training image is generated from the label map of tissue segmentation, which is also provided by the database. The voxels of three tissues (WM, GM and CSF) are considered as foreground and the other voxels are considered as background. By applying registration, a transformation between each pair of training and testing images can be obtained. This transformation can then be applied to the brain mask of the training image to generate a candidate brain mask for the testing image. By repeating this process, 5 candidate brain masks are produced for each testing image. Then these candidate brain masks are combined by applying label fusion method to generate the final brain mask for the testing image. In our work, we used majority voting for label fusion. It is the most commonly used label fusion strategy [2, 12]. It treats every atlas equally and assigns to each voxel the class label that most atlases agree on. Finally, the brain segmentation is completed for each testing image by projecting its brain mask to the whole-head image.

1.2 Brain tissue segmentation

In this section, we will formulate the modelling of local prior [13] and multi-atlas prior [8] and the optimisation of the tissue classification model constructed using the three image based features [3, 14].

Problem Formulation
The voxels of a 3D brain MR image are indexed with $i \in \mathcal{S} = \{1, 2, ..., N\}$ where N is the number of voxels. Each voxel at $\mathcal{S}_i$ is associated with a feature vector $y_i \in \mathbb{R}$. In terms of tissue classification in our work, y_i represents the intensity value of the i^{th} voxel. The set of y_i is the observed image denoted by $y = \{y_1, y_2, ..., y_N\} \in \mathbb{R}^N$. The aim of tissue classification in normal brains is to classify each voxel at $\mathcal{S}_i$ into one of the classes labelled by $\mathcal{L} = \{\text{CSF, GM, WM}\}$. The tissue class of the i^{th} voxel is denoted by $x_i \in \mathcal{L}$ and $x = \{x_1, x_2, ..., x_N\} \in \mathcal{L}^N$ is a classification of the image. Our task is to find the best classification $x^* = \{x_1^*, x_2^*, ..., x_N^*\} \in \mathcal{L}^N$ given the image intensity y which can be formulated as a MAP optimisation problem:

$$\begin{aligned} x^* &= \arg\max_{x \in \mathcal{L}^N} P(x|y) = \frac{P(y|x)P(x)}{P(y)} \\ &= \arg\max_{x \in \mathcal{L}^N} P(y|x)P(x) \end{aligned} \tag{1}$$

where $P(x|y)$ is the probability of the classification x given the image intensity y. $P(y)$ is independent from the classification x and hence ignored in the optimisation. By taking the negative logarithm of (1), the probability maximization problem is converted to finding the minimum of an energy functional:

$$\begin{aligned} x^* &= \arg\min_{x\in\mathcal{L}^N}(-lnP(y|x) - lnP(x)) \\ &= \arg\min_{x\in\mathcal{L}^N}(-lnP(y|x) - ln(P_{Lpr}(x)P_{Apr}(x))) \\ &\propto \arg\min_{x\in\mathcal{L}^N}(E_{in}(x) + E_{Lpr}(x) + \gamma E_{Apr}(x)). \end{aligned} \tag{2}$$

E_{in} represents the intensity energy which models the intensity distributions of three tissue classes. It measures how well the current classification x explains the image y. The probability of the prior classification $P(x)$ is determined by both the local and multi-atlas prior probabilities which are represented as P_{Lpr} and P_{Apr}, respectively. Then the two prior probabilities are rewritten as energy terms $E_{Lpr}(x)$ and $E_{Apr}(x)$. The effect of the local prior is balanced with that of the image intensity by the spatial parameter β embodied in $E_{Lpr}(x)$, which will be described later. γ is introduced to balance the effect of the multi-atlas prior with that of the intensity and neighbouring information.

Intensity Energy

The intensity energy E_{in} is calculated by assuming the intensities of all the voxels in images are independently and identically distributed. The likelihood $P(y|x)$ can be expressed as

$$P(y|x) = \prod_{i\in\mathcal{S}} P(y_i|x_i). \tag{3}$$

$P(y_i|x_i)$ is the probability density function of y_i given the tissue class x_i. As in [13], it is usually modelled by a Gaussian function for each tissue class:

$$P(y_i|x_i) = \frac{1}{\sigma_{x_i}\sqrt{2\pi}} exp\left[\frac{-(y_i-\mu_{x_i})^2}{2\sigma_{x_i}^2}\right], x_i \in \mathcal{L} \tag{4}$$

where the model parameters $\theta_{x_i} = \{\mu_{x_i}, \sigma_{x_i}\}$ are the mean and standard deviation of the Gaussian distribution for the tissue class x_i.

E_{in} can be formulated as

$$\begin{aligned} E_{in}(x) &= -\sum_{i\in\mathcal{S}} lnP(y_i|x_i) \\ &= \sum_{i\in\mathcal{S}}\left(\frac{(y_i-\mu_{x_i})^2}{2\sigma_{x_i}^2} + ln\sqrt{2\pi} + ln\sigma_{x_i}\right). \end{aligned} \tag{5}$$

Local Prior Energy

The local prior energy E_{Lpr} can be modelled by a MRF which assumes that

the tissue class of the voxel at $\mathcal{S}_i$ depends only on its local neighbourhood $\mathcal{N}_i$ with $i \notin \mathcal{N}_i$ and $i \in \mathcal{N}_j \leftrightarrow j \in \mathcal{N}_i$ [13]. A Gibbs distribution is often used to characterize a MRF as

$$P(x) = Z^{-1} exp[-U(x, \beta)] \tag{6}$$

where Z is the normalization factor; β is the spatial parameter and $U(x, \beta)$ represents the energy function. A simplified Potts model is generally used to express the energy function $U(x, \beta)$:

$$U(x|\beta) = \sum_{i \in \mathcal{S}} \left(\frac{\beta}{2} \sum_{j \in \mathcal{N}_i} V_{ij}(x_i, x_j) \right). \tag{7}$$

V_{ij} models the local spatial transitions between the voxel at $\mathcal{S}_i$ and its neighbour at $\mathcal{S}_j$ and can be calculated as

$$V_{ij}(x_i, x_j) = \frac{\delta(x_i, x_j)}{d(i, j)} \tag{8}$$

where $d(i, j)$ measures the distance between the two voxels. The weighting function $\delta(x_i, x_j)$ is defined as

$$\delta(x_i, x_j) = \begin{cases} -1 & x_i = x_j \\ +1 & x_i \neq x_j. \end{cases} \tag{9}$$

This weighting function constrains the classification process with a reward and penalty mechanism. Then E_{Lpr} can be formulated as

$$E_{Lpr} = U(x|\beta) + lnZ. \tag{10}$$

Multi-atlas Prior Energy

The multi-atlas prior is derived from the atlas priors (i.e., the segmentations of other images) and applying a local label fusion strategy. Firstly, multiple atlas images (i.e., five skull stripped training images) are nonrigidly registered to the testing image using demons registration [10, 11]. Then the similarities between the registered atlas images and the testing image are measured and only the atlases with high similarities are selected to contribute to the classification. After that, the label maps (i.e., tissue segmentations) of the selected atlases are propagated to the testing image. Finally, a local label fusion methods is applied to the transformed label maps to infer the classification of the testing image.

Atlas Selection. We used mutual information [5, 6] to calculate the similarity between the registered atlas a and the testing image y:

$$I(a; y) = H(a) + H(y) - H(a, y). \tag{11}$$

$I(a; y)$ denotes the mutual information; $H(a)$ and $H(y)$ denote the entropy of a and y, respectively; $H(a, y)$ denotes the joint entropy of a and y. The atlases with high mutual information values are selected for label fusion.

In the following, the set of the selected atlases are denoted by $\mathcal{A}$ with the corresponding label maps $\{x^a|a \in \mathcal{A}\}$. The atlases and the label maps have already been transformed to the coordinate system of the testing image. The similarities between the atlases $\mathcal{A}$ and the testing image y are denoted by $\{I(a;y)|a \in \mathcal{A}\}$.

Local Label Fusion. Local label fusion methods estimate the classification accuracy at each voxel in a local neighbourhood and assign weights accordingly [12]. Probabilistic Patch-Based Model (PPBM) is a local label fusion method, which was firstly proposed by Sabuncu *et al.* [7] and then improved by Bai *et al.* [8]. The improved model accounts for the potential registration error using the following strategy: for each voxel at $\mathcal{S}_i$ in the testing image, the label is depended not only on the corresponding i^{th} voxels in the atlases but a set of candidates in a local neighbourhood $\mathcal{M}_i$ with $i \in \mathcal{M}_i$ in each atlas.

The PPBM consists of two components: the intensity likelihood and the label likelihood. The intensity likelihood models the intensity difference between the target and candidate voxels as a Gaussian distribution. For each voxel at $\mathcal{S}_i$ in the testing image y, given the candidate location $j \in \mathcal{M}_i$ in the atlas a, the probability of y_i is computed as

$$P(y_i|j,a) = \frac{1}{\sqrt{2\pi}\sigma_1} exp\left[-\frac{D(B_i, B_j^a)}{2\sigma_1^2}\right] \tag{12}$$

where σ_1 denotes the standard deviation of the Gaussian distribution. The intensity difference between the two voxels is computed by the difference between their corresponding patches B_i and B_j^a as $D(B_i, B_j^a)$, the mean squared difference between the two patches. The label likelihood models the registration error as a Gaussian distribution with the standard deviation σ_2. Similarly, the probability of the voxel at $\mathcal{S}_i$ with the label l is computed as

$$P(x_i = l|j, x^a) = \frac{1}{\sqrt{2\pi}\sigma_2} exp\left[-\frac{d^2(i,j)}{2\sigma_2^2}\right] \xi_{x_j^a,l} \tag{13}$$

where x_j^a is the label of the candidate and $d(i,j)$ measures the Euclidean distance between the target voxel and the candidate; $\xi_{x_j^a,l}$ is defined as

$$\xi_{x_j^a,l} = \begin{cases} 1 & x_j^a = l \\ 0 & x_j^a \neq l. \end{cases} \tag{14}$$

The multi-atlas prior energy $E_{Apr}(x)$ is formulated as

$$\begin{aligned} E_{Apr}(x) &= -lnP_{Apr}(x) \\ &= -ln \prod_{i \in \mathcal{S}} P_{Apr}(x_i) \\ &= -\sum_{i \in \mathcal{S}} lnP_{Apr}(x_i). \end{aligned} \tag{15}$$

For each tissue class l, taking all the candidates in all the atlases into account, the multi-atlas prior probability $P_{Apr}(x_i = l)$ based on PPBM is calculated as

$$\frac{\sum_{a\in\mathcal{A}}\sum_{j\in\mathcal{M}_i} P(y_i|j,a)P(x_i=l|j,x^a)}{\sum_{u\in\mathcal{L}}\sum_{a\in\mathcal{A}}\sum_{j\in\mathcal{M}_i} P(y_i|j,a)P(x_i=u|j,x^a)}.$$

Then the three energy terms defined by Eq.5, 10 and 15 can be substituted into Eq.2. All the constants can be ignored in minimisation.

Classification Approach
The tissue classification is iterated with parameters estimation in an EM framework which is generalised in Algorithm 1.

Algorithm 1 Tissue Classification Using An EM Approach

1: Initialisation: For each tissue class $l \in \mathcal{L}$, the parameter $\hat{\theta}_l = \{\hat{\mu}_l, \hat{\sigma}_l\}$ is estimated from the classification of KM;
2: Classification: Given the current parameter $\hat{\theta}$, the tissues are classified by minimizing the energy functional stated in (2) with the Iterated Conditional Modes (ICM) algorithm [15];
3: Maximization: The parameter $\hat{\theta}$ is updated according to the new classification;
4: Iterate steps 2 and 3 until $\hat{\theta}$ is stable.

Given a classification, the parameter estimations for the tissue class l in steps 1 and 3 in Algorithm 1 are computed as

$$\begin{aligned}\hat{\mu}_l &= \frac{\sum_{i\in\mathcal{S}} y_i\xi_{x_i,l}}{\sum_{i\in\mathcal{S}}\xi_{x_i,l}}\\ \hat{\sigma_l}^2 &= \frac{\sum_{i\in\mathcal{S}}(y_i-\hat{\mu}_l)^2\xi_{x_i,l}}{\sum_{i\in\mathcal{S}}\xi_{x_i,l}}\end{aligned} \tag{16}$$

where $\xi_{x_i,l}$ is defined as

$$\xi_{x_i,l} = \begin{cases}1 & x_i = l\\ 0 & x_i \neq l.\end{cases} \tag{17}$$

2 Experiment setup and algorithm execution

The method was implemented in MATLAB R2015a on a 64-bit Win7 operating system with Intel 3.00GHz CPU and 32.0GB RAM. For brain tissue classification, we applied the MAP based method described in this document to each skull stripped testing image. The number of clusters K was set to 3 for KM because we aim to segment the brain tissues into three classes: WM, GM and CSF. The spatial parameter β was set to 0.3 and the number of neighbours considered

at each central voxel was 6. In modelling the multi-atlas prior, 3 atlases were selected for tissue classification; for each voxel at $\mathcal{S}_i$ in the testing image, the size of the local neighbourhood $\mathcal{M}_i$ in each atlas a was set to 7; the patch size was also set to 7; the two standard deviations σ_1 and σ_2 were set to 0.1 and 2, respectively. The parameter γ was set to 0.3. These parameters were optimised on the training images. The average execution time was 8.86 ± 0.51 minutes for brain segmentation and 9.63 ± 0.70 minutes for brain tissue segmentation.

3 Conclusion

In this work, A multi-atlas segmentation method was used for skull stripping. Image intensity was used by KM to provide an preliminary brain tissue segmentation; the local prior was modelled by MRF to improve the tissue classification for the brain images with artefacts; The multi-atlas prior was derived by registering multiple atlas images to the testing image and propagating the label maps of the atlases to the testing image. These three image based features were combined into a MAP model and the classification was achieved by applying the EM algorithm.

References

1. Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177, 2011.
2. Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205–219, 2015.
3. Jyrki MP Lötjönen, Robin Wolz, Juha R Koikkalainen, Lennart Thurfjell, Gunhild Waldemar, Hilkka Soininen, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage*, 49(3):2352–2365, 2010.
4. Michaël Sdika. Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote. *Medical Image Analysis*, 14(2):219–226, 2010.
5. Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.
6. Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.
7. Mert R Sabuncu, BT Thomas Yeo, Koen Van Leemput, Bruce Fischl, and Polina Golland. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging*, 29(10):1714–1729, 2010.
8. Wenjia Bai, Wenzhe Shi, Declan P O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, and Daniel Rueckert. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. *IEEE Transactions on Medical Imaging*, 32(7):1302–1315, 2013.

9. Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen de Bresser, Amir Alansary, Marleen de Bruijne, Aaron Carass, Ayman El-Baz, et al. MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. *Computational intelligence and neuroscience*, 2015.
10. J-P. Thirion. Fast non-rigid matching of 3D medical images. *[Research Report] RR-2547*, page 37, 1995.
11. He Wang, Lei Dong, Jennifer O'Daniel, Radhe Mohan, Adam S Garden, K Kian Ang, Deborah A Kuban, Mark Bonnen, Joe Y Chang, and Rex Cheung. Validation of an accelerated'demons' algorithm for deformable image registration in radiation therapy. *Physics in Medicine and Biology*, 50(12):2887, 2005.
12. Xabier Artaechevarria, Arrate Munoz-Barrutia, and Carlos Ortiz-de Solórzano. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Transactions on Medical Imaging*, 28(8):1266–1277, 2009.
13. Meritxell Bach Cuadra, Leila Cammoun, Torsten Butz, Olivier Cuisenaire, and Jean-Philippe Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Transactions on Medical Imaging*, 24(12):1548–1565, 2005.
14. Fedde van der Lijn, Tom den Heijer, Monique MB Breteler, and Wiro J Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *Neuroimage*, 43(4):708–720, 2008.
15. Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 259–302, 1986.