

# Nonparametric Mixture Models for Supervised Image Parcellation

Mert R. Sabuncu<sup>1</sup>, B.T. Thomas Yeo<sup>1</sup>, Koen Van Leemput<sup>1,2,3</sup>, Bruce Fischl<sup>1,2</sup>, and Polina Golland<sup>1</sup>

<sup>1</sup> Computer Science and Artificial Intelligence Lab, MIT, USA

<sup>2</sup> Department of Radiology, Harvard Medical School, USA

<sup>3</sup> Dep. of Information and Comp. Sci., Helsinki University of Technology, Finland

**Abstract.** We present a nonparametric, probabilistic mixture model for the supervised parcellation of images. The proposed model yields segmentation algorithms conceptually similar to the recently developed label fusion methods, which register a new image with each training image *separately*. Segmentation is achieved via the fusion of transferred manual labels. We show that in our framework various settings of a model parameter yield algorithms that use image intensity information differently in determining the weight of a training subject during fusion. One particular setting computes a single, global weight per training subject, whereas another setting uses locally varying weights when fusing the training data. The proposed nonparametric parcellation approach capitalizes on recently developed fast and robust pairwise image alignment tools. The use of multiple registrations allows the algorithm to be robust to occasional registration failures. We report experiments on 39 volumetric brain MRI scans with expert manual labels for the white matter, cerebral cortex, ventricles and subcortical structures. The results demonstrate that the proposed nonparametric segmentation framework yields significantly better segmentation than state-of-the-art algorithms.

## 1 Introduction

Supervised image parcellation (segmentation) tools traditionally use atlases, which are parametric models that summarize the training data in a single coordinate system [1–9]. Yet, recent work has shown that more accurate segmentation can be achieved by utilizing the entire training data [10–16], by mapping each training subject into the coordinates of the new image via a pairwise registration algorithm. The transferred manual labels are then fused to generate a segmentation of the new subject. There are at least two advantages of this approach: (1) across-subject anatomical variability is better captured than in a parametric model, and (2) multiple registrations improve robustness against occasional registration failures. The main drawback of the label fusion (multi-atlas) approach is the computational burden introduced by the multiple registrations and the manipulation of the entire training data.

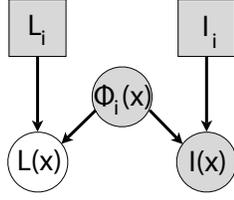
Early label fusion methods proposed to transfer the manual labels to the test image via nearest neighbor interpolation after pairwise registration [12, 14].

Segmentation labels of the test image were then estimated via majority voting. Empirical results suggested that errors in the manual labeling and in registration are averaged out during label fusion, resulting in accurate segmentation. More recent work has shown that a weighted averaging strategy can be used to improve segmentation quality [11]. The basic idea is that training subjects more similar to the test subject should carry more weight during label fusion. The practical advantages of various strategies based on this idea have lately been demonstrated [11, 13, 16]. Some of these strategies use the whole image to determine a single, global weight for each training subject [11, 15, 16], whereas others use local image intensities for locally adapting the weights [11, 13].

This paper presents a novel unified probabilistic model that enables local and global weighting strategies within a label fusion-like segmentation framework. We formulate segmentation using MAP, where a nonparametric approach is used to estimate the joint density on the image intensities and segmentation labels of the new subject. The proposed framework generalizes a model we recently presented at MICCAI 2009 [16], which is based on the assumption that the test subject is generated from a single, unknown training subject. That specific model leads to a segmentation algorithm that assigns greater importance to the training subjects that are globally more similar to the test subject and can be viewed as a particular instantiation of the more general approach presented in this paper. In addition, the proposed approach further enables two possible variants within the same framework. First, we present a local mixture model that assumes each voxel in the test image is generated from some training subject with a uniform prior, independently of other voxels. This local model yields a pixel-wise weighting strategy in segmentation. Second, we develop a semi-local mixture model that relaxes the independence assumption of the local model with a Markov Random Field prior. This model leads to a weighting strategy where intensity information in a local neighborhood is pooled in a principled manner.

In related literature, soft weighting of training subjects was recently used for shape regression [17], where the weights depended on the subjects' age. The proposed nonparametric parcellation framework is also parallel to STAPLE [18], which fuses multiple segmentations of a single subject. In contrast, our framework handles multiple subjects and accounts for inter-subject variability through registration.

The paper is organized as follows. The next section presents the non-parametric generative model for image segmentation. In section 3, we discuss three instantiations of the framework. In section 4, we present inference algorithms for these instantiations. We conclude with experiments in section 5. We report experiments on 39 brain MRI scans that have corresponding manual labels, including the cerebral cortex, white matter, and sub-cortical structures. Experimental results suggest that the proposed nonparametric parcellation framework achieves better segmentation than the existing state-of-the-art algorithms.



**Fig. 1.** Generative model for  $(L(x), I(x))$  given  $M(x) = i$  and  $(L_i, I_i, \Phi_i)$ .  $\Phi_i$  is the mapping from the image coordinates to the template coordinates. Squares indicate non-random parameters, while circles indicate random variables. Shaded variables are assumed to be observed.

## 2 Theory

Let  $\{I_i\}$  be  $N$  training images with corresponding label maps  $\{L_i\}$ ,  $i = 1, \dots, N$ . We assume the label maps take discrete values from 1 to  $\mathcal{L}$  that indicate the label identity at each spatial location. We treat these training images as spatially continuous functions on  $\mathbb{R}^3$  by assuming a suitable interpolator. Let  $I : \Omega \mapsto \mathbb{R}$  denote a new, previously unseen test image defined on a discrete grid  $\Omega \subset \mathbb{R}^3$ . Let  $\Phi_i : \Omega \mapsto \mathbb{R}^3$  denote the spatial mapping (warp) from the test image coordinates to the coordinates of the training image  $i$ . We assume that  $\{\Phi_i\}$  have been pre-computed using a pairwise registration procedure, such as the one described in Section 4.1.

Our objective is to estimate the label map  $\hat{L}$  associated with the test image  $I$ . One common formulation to compute  $\hat{L}$  is via MAP:

$$\hat{L} = \underset{L}{\operatorname{argmax}} p(L|I, \{L_i, I_i, \Phi_i\}) = \underset{L}{\operatorname{argmax}} p(L, I|\{L_i, I_i, \Phi_i\}), \quad (1)$$

where  $p(L, I|\{L_i, I_i, \Phi_i\})$  denotes the joint probability of the label map  $L$  and image  $I$  given the training data.

Rather than using a parametric model for  $p(L, I|\{L_i, I_i, \Phi_i\})$ , we employ a non-parametric estimator, which is an explicit function of the entire training data, not a summary of it. Let  $M : \Omega \mapsto \{1, \dots, N\}$  denote an unknown (hidden) random field that, for each voxel in test image  $I$ , specifies the training image  $I_i$  that generated that voxel. Given  $M$ , the training data, and warps, and assuming the factorization depicted in the graphical model of Fig. 1, we can construct the conditional probability of generating the test image and label map:

$$\begin{aligned} p(L, I|M, \{L_i, I_i, \Phi_i\}) \\ = \prod_{x \in \Omega} p(L(x), I(x) | M(x), \{L_i, I_i, \Phi_i\}) \end{aligned} \quad (2)$$

$$= \prod_{x \in \Omega} p_{M(x)}(L(x), I(x) | L_{M(x)}, I_{M(x)}, \Phi_{M(x)}(x)) \quad (3)$$

$$= \prod_{x \in \Omega} p_{M(x)}\left(L(x) | L_{M(x)}, \Phi_{M(x)}(x)\right) p_{M(x)}\left(I(x) | I_{M(x)}, \Phi_{M(x)}(x)\right), \quad (4)$$

where  $p_{M(x)}(L(x), I(x) | L_{M(x)}, I_{M(x)}, \Phi_{M(x)}(x))$  is the conditional probability of  $(L(x), I(x))$  given that voxel  $x \in \Omega$  of the test subject was generated from training subject  $M(x)$ . Note that Eq. (4) assumes that  $(L(x), I(x))$  are conditionally

independent given the membership  $M(x)$ , corresponding warp  $\Phi_{M(x)}$  and training data. Given a prior on  $M$ , we can view  $p(L, I|\{L_i, I_i, \Phi_i\})$  as a mixture:

$$p(L, I|\{L_i, I_i, \Phi_i\}) = \sum_{\mathcal{M}} p(M)p(L, I|M, \{L_i, I_i, \Phi_i\}), \quad (5)$$

where  $\sum_{\mathcal{M}}$  denotes the marginalization over the unknown random field  $M$ . Substituting Eq. (4) into Eq. (5) yields:

$$\begin{aligned} \hat{L} = \operatorname{argmax}_L \sum_{\mathcal{M}} p(M) \prod_{x \in \Omega} p_{M(x)}(L(x) | L_{M(x)}, \Phi_{M(x)}(x)) \\ \times p_{M(x)}(I(x) | I_{M(x)}, \Phi_{M(x)}(x)). \end{aligned} \quad (6)$$

In the next section, we present instantiations of the individual terms in Eq. (6).

### 3 Model Instantiation

#### 3.1 Image Likelihood

We adopt a Gaussian distribution with a stationary variance  $\sigma^2$  as the image likelihood:

$$p_i(I(x)|I_i, \Phi_i(x)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (I(x) - I_i(\Phi_i(x)))^2\right]. \quad (7)$$

#### 3.2 Label Likelihood

We use the distance transform representation to encode the label prior information, cf. [19]. Let  $D_i^l$  denote the signed distance transform of label  $l$  in training subject  $i$ , assumed to be positive inside the structure of interest. We define the label likelihood as:

$$p_i(L(x) = l | L_i, \Phi_i(x)) \propto \exp(\rho D_i^l(\Phi_i(x))), \quad (8)$$

where  $\rho > 0$  is the slope constant and  $\sum_{l=1}^{\mathcal{L}} p_i(L(x) = l | L_i, \Phi_i(x)) = 1$ , where  $\mathcal{L}$  is the total number of labels including a background label.  $p_i(L(x) = l | L_i, \Phi_i(x))$  encodes the conditional probability of observing label  $l$  at voxel  $x \in \Omega$  of the test image, given that it was generated from training image  $i$ .

#### 3.3 Membership Prior

The latent random field  $M : \Omega \mapsto \{1, \dots, N\}$  encodes the local association between the test image and training data. We place a Markov Random Field (MRF) prior on  $M$ :

$$p(M) = \frac{1}{Z_\beta} \prod_{x \in \Omega} \exp\left(\beta \sum_{y \in \mathcal{N}_x} \delta(M(x), M(y))\right), \quad (9)$$

where  $\beta \geq 0$  is a scalar parameter,  $\mathcal{N}_x$  is a spatial neighborhood of voxel  $x$ ,  $Z_\beta$  is the partition function that only depends on  $\beta$ , and  $\delta(M(x), M(y)) = 1$ , if  $M(x) = M(y)$  and zero otherwise. In our implementation,  $\mathcal{N}_x$  includes the immediate 8 neighbors of each voxel. Similar models have been used in the segmentation literature, e.g. [5, 9], mainly as priors on label maps to encourage spatially contiguous segmentations. In contrast, we use the MRF prior to pool local intensity information in determining the association between the test subject and training data.

The parameter  $\beta$  influences the average size of the local patches of the test subject that are generated from a particular training subject. In this work, we consider three settings of the parameter  $\beta$ . For  $\beta = 0$ , the model effectively assumes that each test image voxel is independently generated from a training subject, drawn with a uniform prior.  $\beta \rightarrow +\infty$  forces the membership of all voxels to be the same and corresponds to assuming that the whole test subject is generated from a single unknown training subject, drawn from a uniform prior. A positive, finite  $\beta$  favors local patches of voxels to have the same membership.

The  $\beta \rightarrow +\infty$  case reduces to a model similar to the one we presented in [16], except now we make the simplifying assumption that the training data is apriori mapped to the test subject's coordinate frame as a preprocessing step. Due to this simplification, the warp cost in registration plays no role in the segmentation algorithms we present in this paper. Without this simplification, however, inference for finite values of  $\beta$  becomes intractable. As demonstrated in the next section, the resulting inference algorithms allow us to determine the association between the training data and test subject using local intensity information.

## 4 Algorithms

### 4.1 Efficient Pairwise Registration

To perform pairwise registration, we employ an efficient algorithm [20, 21] that uses a one-parameter subgroup of diffeomorphisms, where a warp  $\Phi$  is parameterized with a smooth, stationary velocity field  $v : \mathbb{R}^3 \mapsto \mathbb{R}^3$  via an ODE [22]:  $\frac{\partial \Phi(x, t)}{\partial t} = v(\Phi(x, t))$  and initial condition  $\Phi(x, 0) = x$ . The warp  $\Phi(x) = \exp(v)(x)$  can be computed efficiently using scaling and squaring and inverted by using the negative of the velocity field:  $\Phi^{-1} = \exp(-v)$  [22].

We impose an elastic-like regularization on the stationary velocity field:

$$p(\Phi = \exp(v)) = \frac{1}{Z_\lambda} \exp \left[ -\lambda \sum_{y \in \Omega} \sum_{j, k=1, 2, 3} \left( \frac{\partial^2}{\partial x_j^2} v_k(x) \Big|_{x=y} \right)^2 \right], \quad (10)$$

where  $\lambda > 0$  is the warp stiffness parameter,  $Z_\lambda$  is a partition function that depends only on  $\lambda$ , and  $x_j$  and  $v_k$  denote the  $j$ 'th and  $k$ 'th component (dimension) of position  $x$  and velocity  $v$ , respectively. A higher warp stiffness parameter  $\lambda$  yields more rigid warps.

To derive the registration objective function, we assume a simple additive Gaussian noise model, consistent with the image likelihood term described in Section 3.1. This model leads to the following optimization problem for registering the  $i$ -th training image to the test subject:

$$\hat{v}^i = \operatorname{argmin}_v \sum_{y \in \Omega} \left[ (I(y) - I_i(\exp(v)(y)))^2 + 2\lambda\sigma^2 \sum_{j,k=1,2,3} \left( \frac{\partial^2}{\partial x_j^2} v_k(x) \Big|_{x=y} \right)^2 \right], \quad (11)$$

where  $\sigma^2$  is the stationary image noise variance, and  $\Phi_i \triangleq \exp(\hat{v}^i)$ . To solve Eq. (11), we use the bidirectional log-domain Demons framework [20], which decouples the optimization of the first and second terms by introducing an auxiliary transformation. The update warp is first computed using the Gauss-Newton method. The regularization is achieved by smoothing the updated warp field. It can be shown that the smoothing kernel corresponding to Eq. (10) can be approximated with a Gaussian;  $K(x) \propto \exp(-\alpha \sum_{i=1,2,3} x_i^2)$ , where  $\alpha = \frac{\gamma}{8\lambda\sigma^2}$  and  $\gamma > 0$  controls the size of the Gauss-Newton step.

## 4.2 Segmentation Algorithms

Here, we present algorithms to solve the optimization problem of Eq. (6) for the three cases of  $\beta$  in the model presented in Section 3.

## 4.3 Global Mixture

First, we consider  $\beta \rightarrow +\infty$ , which is equivalent to a global mixture model, where the test subject is assumed to be generated from a single, unknown training subject. In this case, the segmentation problem in Eq. (6) reduces to

$$\hat{L} = \operatorname{argmax}_L \sum_{i=1}^N \prod_{x \in \Omega} p_i(L(x)|L_i, \Phi_i(x)) p_i(I(x)|I_i, \Phi_i(x)). \quad (12)$$

Eq. (12) cannot be solved in closed form. However, an efficient solution to this MAP formulation can be obtained via Expectation Maximization (EM). Here, we present a summary.

The E-step updates the posterior of the membership associated with each training image:

$$m_i^{(n)} \propto \prod_{x \in \Omega} p_i(I(x)|I_i, \Phi_i(x)) p_i(\hat{L}^{(n-1)}(x)|L_i, \Phi_i(x)), \quad (13)$$

where  $\hat{L}^{(n-1)}(x)$  is the segmentation estimate of the test image from the previous iteration and  $\sum_i m_i^{(n)} = 1$ . The M-step updates the segmentation estimate:

$$\hat{L}^{(n)}(x) = \operatorname{argmax}_{l \in \{1, \dots, \mathcal{L}\}} \sum_{i=1}^N m_i^{(n)} \log(p_i(L(x) = l|L_i, \Phi_i(x))). \quad (14)$$

The E-step in Eq. (13) determines a single membership index for the entire training image, based on all the voxels. The M-step in Eq. (14) performs an independent optimization at each voxel  $x \in \Omega$ ; it determines the mode of a length- $\mathcal{L}$  vector, where  $\mathcal{L}$  is the number of labels. The EM algorithm is initialized with  $m_i^{(1)} \propto \prod_{x \in \Omega} p_i(I(x)|I_i, \Phi_i(x))$  and iterates between Equations (14) and (13), until convergence.

#### 4.4 Local Mixture: Independent Prior

The second case we consider is  $\beta = 0$ , which corresponds to assuming a voxel-wise independent mixture model with a uniform prior on  $M$ :

$$p(M) = \frac{1}{N^{|\Omega|}}, \quad (15)$$

where  $|\Omega|$  is the cardinality of the image domain, i.e., the number of voxels. It is easy to show that the segmentation problem reduces to

$$\hat{L}(x) = \operatorname{argmax}_{l \in \{1, \dots, \mathcal{L}\}} \sum_{i=1}^N p_i(L(x) = l | L_i, \Phi_i(x)) p_i(I(x) | I_i, \Phi_i(x)), \quad (16)$$

where the image and label likelihood terms in the summation can be computed using Eqs. (7) and (8). The optimization problem can be solved by simply comparing  $\mathcal{L}$  numbers at each voxel.

#### 4.5 Semi-local mixture: MRF Prior

Finally, we consider a finite, positive  $\beta$ . This leads to an MRF prior, which couples neighboring voxels and thus the exact marginalization of Eq. (6) becomes computationally intractable. An efficient approximate solution can be obtained using variational mean field [23]. The main idea of variational mean field is to approximate the posterior distribution of the membership  $p(M|I, L, \{I_i, L_i, \Phi_i\})$ , with a simple distribution  $q$  that is fully factorized:

$$q(M) = \prod_{x \in \Omega} q_x(M(x)). \quad (17)$$

The objective function of Eq. (6) can then be approximated by an easier-to-optimize lower bound, which is a function of  $q$ . This approximate problem can be solved via coordinate-ascent, where the segmentation  $L$  and approximate posterior  $q$  are updated sequentially, by solving the optimization for each variable while fixing the other. One particular formulation leads to a straightforward update rule for  $L$ :

$$\hat{L}^{(n)}(x) = \operatorname{argmax}_{l \in \{1, \dots, \mathcal{L}\}} \sum_{i=1}^N q_x^{(n-1)}(M(x) = i) \log p_i(L(x) = l | L_i, \Phi_i(x)), \quad (18)$$

where  $q^{(n-1)}$  is an estimate of the posterior at the  $(n-1)$ 'th iteration. Eq. (18) is independent for each voxel and entails determining the mode of a length- $\mathcal{L}$  vector. For a fixed segmentation estimate  $\hat{L}^{(n)}(x)$  the optimal  $q$  is the solution of the following fixed-point equation:

$$q_x^{(n)}(M(x)) \propto p_{M(x)}(I(x)|L_{M(x)}, \Phi_{M(x)}(x)) \times p_{M(x)}(\hat{L}^{(n)}(x)|L_{M(x)}, \Phi_{M(x)}(x)) \exp\left(\beta \sum_{y \in \mathcal{N}_x} q_y^{(n)}(M(x))\right), \quad (19)$$

and  $\sum_i q_x^{(n)}(M(x) = i) = 1$ . We solve Eq. (19) iteratively. The variational mean field algorithm alternates between Eqs. (19) and (18), until convergence.

## 5 Experiments

We validate the proposed framework on 39 T1-weighted brain MRI scans of dimensions  $256 \times 256 \times 256$ , 1mm isotropic. Each MRI volume is an average of 3-4 scans and was gain-field corrected and skull-stripped. These volumes were then manually delineated by an expert anatomist into left and right White Matter (WM), Cerebral Cortex (CT), Lateral Ventricle (LV), Hippocampus (HP), Thalamus (TH), Caudate (CA), Putamen (PU), Pallidum (PA) and Amygdala (AM). We use volume overlap with manual labels, as measured by the Dice score [24], to quantify segmentation quality. The Dice score ranges from 0 to 1, with higher values indicating improved segmentation.

### 5.1 Setting Parameters Through Training

The proposed nonparametric parcellation framework has two stages with several input parameters. The registration stage has two independent parameters:  $\gamma$  that controls the step size in the Gauss-Newton optimization and  $\alpha$  that determines the smoothness of the final warp. The segmentation stage has two additional input parameters:  $\sigma^2$ , which is the intensity variance of the image likelihood in Eq. (7) and the slope  $\rho$  of the distance transform used to compute the label prior in Eq. (8). Furthermore, the semi-local model of Section 4.5 has a non-zero, finite  $\beta$  parameter.

Nine subjects were used to determine the optimal values of these parameters. First, 20 random pairs of these nine subjects were registered for a range of values of  $\gamma$  and  $\alpha$ . Registration quality was assessed by the amount of pairwise label overlap and used to select the optimal  $(\gamma^*, \alpha^*)$  pair.

We used the optimal  $(\gamma^*, \alpha^*)$  pair to register all 72 ordered pairs of the nine training subjects. We performed nine leave-one-out segmentations using both the global and local models to determine the corresponding optimal pairs of  $\sigma^2$  and  $\rho$ . The optimal pair for the local model was then used to determine the optimal value for  $\beta$  in the semi-local model. The optimal parameter values were finally used to segment the remaining 30 subjects.

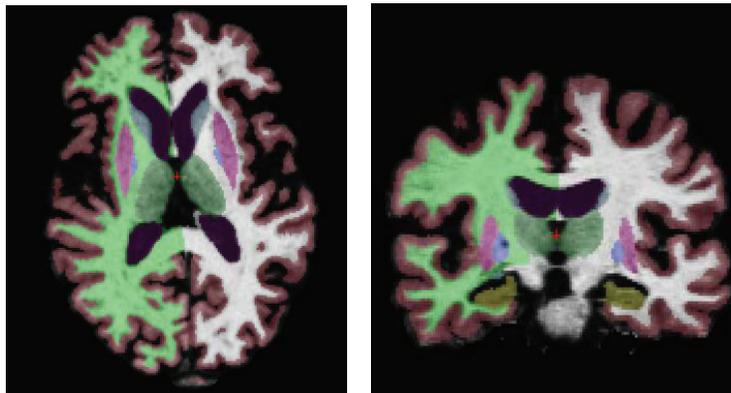


Fig. 2. A typical segmentation obtained with the local mixture model.

## 5.2 Benchmarks

The first benchmark we consider is the whole-brain parcellation tool available in the Freesurfer software package [25]. The Freesurfer parcellation tool uses a unified registration-segmentation procedure that models across-scanner intensity variation [2, 3]. We consider this as a state-of-the-art benchmark, since numerous imaging studies across multiple centers have shown Freesurfer's robustness and accuracy as a segmentation tool.

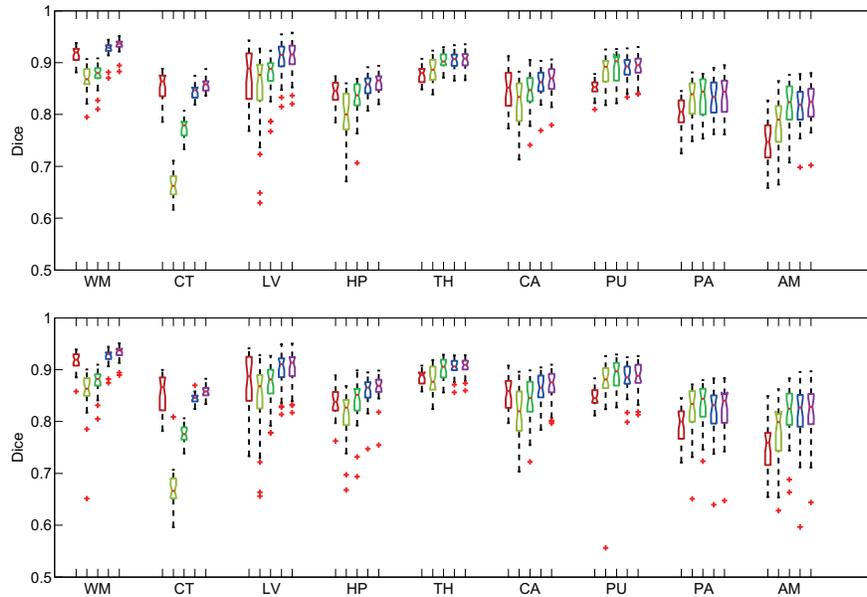
As a second benchmark, we use our implementation of the Label Fusion algorithm [12, 14]. We employ the pairwise registrations obtained with  $(\gamma^*, \alpha^*)$  to transfer the labels of the training subjects via the trilinear interpolation of the probability maps, obtained by assigning 1 to entries corresponding to the manual labels and zero elsewhere. Segmentation is then computed through majority voting at each voxel. We use trilinear interpolation instead of nearest neighbor interpolation because we find that trilinear interpolation yields better results.

## 5.3 Results

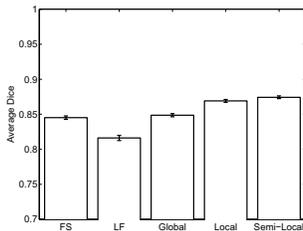
We report test results for the 30 subjects not included in the group used for setting the algorithm parameters  $\gamma$ ,  $\alpha$ ,  $\sigma^2$ ,  $\rho$ , and  $\beta$ . For each test subject, we treated the remaining subjects as training data in a cross-validation evaluation.

Fig. 2 illustrates a typical automatic segmentation result obtained with the local mixture model and overlaid on the MRI volume. Fig. 3 shows box-plots of Dice scores for the two benchmarks and the proposed non-parametric parcellation algorithms. Table 1 provides the mean Dice scores averaged over all subjects and both hemispheres. Fig. 4 provides an overall comparison between the average dice scores achieved by the algorithms.

On average, the local and semi-local mixture models yield better segmentations than the global mixture model, mainly due to the large improvement in the white matter, cerebral cortex and lateral ventricles, the segmentation of which



**Fig. 3.** Boxplots of Dice scores for Freesurfer (red), Label Fusion (yellow), the global mixture model (green), the local mixture model (blue) and the semi-local mixture model (purple). Top row is left hemisphere. Bottom row is right hemisphere. Medians are indicated by horizontal bars. Boxes indicate the lower and upper quartiles and vertical lines extend to 1.5 inter-quartile spacing. ‘+’s indicate outliers.



**Fig. 4.** Average Dice scores for each algorithm (FS: Freesurfer, LF: Label Fusion, Global: Global Mixture, Local: Local Mixture, and Semi-Local: MRF-based model). Error bars show standard error. Each subject and ROI was treated as an independent sample with an equal weight.

clearly benefits from the additional use of local intensity information. A paired t-test between the local and semi-local models reveals that a statistically significant improvement is achieved with the MRF model that pools local intensity information. Yet, this improvement is overall quite modest: about 1% per ROI.

As discussed earlier, the global mixture model is similar to that of [16], except that [16] incorporates registration into the model. Despite this, we find that both algorithms achieve similar segmentation accuracy (results not shown).

A paired sample t-test implies that the difference in accuracy between the proposed semi-local mixture model and Freesurfer is statistically significant ( $p < 0.05$ , Bonferroni corrected) for all ROIs, except the cerebral cortex and right

**Table 1.** Comparison of average dice scores. Boldface font indicates best scores for each structure. As a reference, the last row lists approximate average volumes.

	WM	CT	LV	HP	TH	CA	PU	PA	AM
Freesurfer	0.92	0.85	0.87	0.84	0.88	0.85	0.85	0.80	0.75
Label Fusion	0.85	0.66	0.84	0.77	0.86	0.80	0.86	0.81	0.75
Global Mixture	0.88	0.77	0.87	0.83	0.90	0.84	<b>0.89</b>	<b>0.83</b>	0.81
Local Mixture	<b>0.93</b>	0.84	0.90	0.86	0.90	0.86	0.88	0.82	<b>0.82</b>
Semi-local Mixture	<b>0.93</b>	<b>0.86</b>	<b>0.91</b>	<b>0.87</b>	<b>0.91</b>	<b>0.87</b>	<b>0.89</b>	<b>0.83</b>	<b>0.82</b>
Volumes ( $\times 10^3 mm^3$ )	450	448	25	7	14	7	10	3	3

**Table 2.** Approximate average run-time to segment one test subject (in CPU hours).

Freesurfer	Label Fusion	Global Mixture	Local Mixture	Semi-local Mixture
10	24	32	24	40

Caudate, where the two methods yield comparable results. The same results are obtained when comparing the local mixture model and Freesurfer.

Compared to the Label Fusion benchmark, the nonparametric parcellation algorithms (global, local and semi-local) yield significantly better segmentation (paired sample t-test,  $p < 0.05$ , Bonferroni corrected) in all regions, except Pallidum and Putamen, where the improvement over Label Fusion does not reach statistical significance. We note, however, that the results we report for our Label Fusion implementation are lower than the ones reported in [12]. This might be due to differences in the data and/or registration algorithm. Specifically, normalized mutual information (NMI) was used as the registration cost function in [12]. Entropy-based measures such as NMI are known to yield more robust alignment results. We leave a careful analysis of this issue to future work.

Table 2 lists the average run-times for all five algorithms. The parametric atlas-based Freesurfer algorithm is the fastest, mainly because it needs to compute only a single registration. The remaining algorithms take up more than 20 hours of CPU time on a modern machine, most of which is dedicated to the many registrations performed with the training data. The two iterative algorithms that solve the global mixture and semi-local mixture models (EM and variational mean field, respectively) require significantly longer run-times. The local-mixture model, on the other hand, requires minimal computation time once the registrations are complete, since it simply performs a voxelwise weighted averaging. Its run-time is similar to that required by Label Fusion.

## 6 Conclusion

This paper presents a novel, nonparametric mixture model of images and label maps, that yields accurate image segmentation algorithms. The resulting algorithms are conceptually similar to recent label fusion (or multi-atlas) methods that utilize the entire training data, rather than a summary of it, and register the test subject to each training subject separately. Segmentation is then achieved

by fusing the transferred manual labels. In the proposed framework, similarities between the test image and training data determine how the transferred labels are weighed during fusion. As we discuss in this paper, different settings of a model parameter yields various weighting strategies. Our experiments suggests that a semi-local strategy that is derived from an MRF model that encourages local image patches to be associated with the same training data provides the best segmentation results. We also show that a computationally less expensive local strategy that treats each voxel independently leads to accurate segmentations that are better than the current state-of-the-art.

We leave an investigation of various registration algorithms within the proposed framework to future work. It is clear that alternative strategies can be used to improve the alignment between the training data and test subject. For example, one could use a richer representation of diffeomorphic warps, cf. [4], or a more sophisticated registration cost function, cf. [12]. Since any multi-atlas segmentation algorithm will be robust against occasional registration failures, whether a better alignment algorithm will lead to more accurate segmentation remains an open question.

**Acknowledgements:** Support for this research is provided in part by: NAMIC (NIH NIBIB NIMIC U54-EB005149), the NAC (NIH NCRR NAC P41-RR13218), the mBIRN (NIH NCRR mBIRN U24-RR021382), the NIH NINDS R01-NS051826 grant, the NSF CAREER 0642971 grant, NCRR (P41-RR14075, R01 RR16594-01A1), the NIBIB (R01 EB001550, R01EB006758), the NINDS (R01 NS052585-01), the MIND Institute, and the Autism & Dyslexia Project funded by the Ellison Medical Foundation. B.T. Thomas Yeo is funded by the A\*STAR, Singapore.

## References

1. Ashburner, J., Friston, K.: Unified segmentation. *Neuroimage* **26** (2005) 839–851
2. Fischl, B., et al.: Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23** (2004) 69–84
3. Han, X., Fischl, B.: Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE TMI* **26**(4) (2007) 479–486
4. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphism atlas construction for computational anatomy. *Neuroimage* **23** (2004) 151–160
5. Van Leemput, K., et al.: Automated model-based bias field correction of MR images of the brain. **18**(10) (1999) 885–896
6. Pohl, K., et al.: A bayesian model for joint segmentation and registration. *Neuroimage* **31** (2006) 228–239
7. Twining, C., et al.: A unified information-theoretic approach to groupwise non-rigid registration and model building. In Proc. IPMI **LNCS 3565** (2005) 1–14
8. Tu, Z., et al.: Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE TMI* **27**(4) (2008) 495–508
9. Yeo, B.T.T., et al.: Effects of registration regularization and atlas sharpness on segmentation accuracy. *Medical Image Analysis* **12**(5) (2008) 603–615
10. Aljabar, P., et al.: Classifier selection strategies for label fusion using large atlas databases. In Proc. of MICCAI **LNCS 4791** (2007) 523–531

11. Artaechevarria, X., et al.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE TMI* **28**(8) (2009) 1266 – 1277
12. Heckemann, R., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* **33**(1) (2006) 115–126
13. Isgum, I., et al.: Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. *IEEE TMI* **28**(7) (2009)
14. Rohlfing, T., et al.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* **21**(4) (2004) 1428–1442
15. Rohlfing, T., et al.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE TMI* **23**(8) (2004) 983–994
16. Sabuncu, M., et al.: Supervised nonparametric image parcellation. *In Proc. of MICCAI LNCS 5762* (2009)
17. Davis, B., et al.: Population shape regression from random design data. *In Proc. of ICCV* (2007) 1–7
18. Warfield, S., et al.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for validation of image segmentation. *TMI* **23**(7) (2004) 903–921
19. Pohl, K., et al.: Logarithm odds maps for shape representation. *MICCAI 2006 LNCS 4191* (2006) 955–963
20. Vercauteren, T., et al.: Symmetric log-domain diffeomorphic registration: A demons-based approach. *In Proc of MICCAI LNCS 5241* (2008) 754–761
21. Sabuncu, M., et al.: Asymmetric image template registration. *In Proc. of MICCAI LNCS 5761* (2009)
22. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-Euclidean framework for statistics on diffeomorphisms. *In Proc of MICCAI LNCS 4190* (2006) 924–931
23. Jaakkola, T.: Tutorial on variational approximation methods. *Advanced Mean Field Methods: Theory and Practice* (2000)
24. Dice, L.: Measures of the amount of ecologic association between species. *Ecology* **26**(3) (1945) 297–302
25. Freesurfer: <http://surfer.nmr.mgh.harvard.edu>