Automatic Framework for Medical Image Registration, Segmentation and Modeling

V Petrović*and T Cootes and C Twining and C Taylor[a]

[a]Imaging Science and Biomedical Engineering, University of Manchester, Oxford Road, Manchester, M13 9PT

Abstract. We describe an framework that simultaneously segments and registers a set of medical images in an automatic manner, incrementally constructing a model of the structure and shape deformations of the set. The framework extends existing groupwise registration and modeling approaches by explicitly modeling the fraction of each tissue type in each voxel, rather than the expected intensity in each voxel. This decouples the model from the effects of the imaging sequence and thus imaging modality. When estimating the optimal deformation field between examples in the set each image is compared to its reconstruction generated from the model segment fractions and the current estimate of its intensity distributions for each tissue type (i.e. an estimate of how the model would appear given the imaging conditions for that image). We also present a method to determine the optimal number of tissue types and fully automate the approach as well as model construction methods that ensure efficient convergeance. We describe the algorithm in detail and present results of applying it to a set of MR images of the brain.

1 Introduction

This paper proposes a fully automatic approach for analysing, understanding and representing structure in groups of medical images. Ideally, given a set of images of different examples of a structure, one would like to derive in an efficient and robust (converging) manner:

- a set of deformation fields defining dense correspondences between the images (registration)
- a classification of the voxels in each image into different tissue types (segmentation)
- a statistical representation of the variability of shape and appearance across the image set (modeling)

There has already been considerable research into techniques that could provide each of these requirements independently. Groupwise non-rigid image registration methods derive a dense, spatial correspondence across sets of images [1–3], segmentation of the brain into tissue types in MR images has been considered in [4, 5], while Statistical Appearance Models [1] capture and describe the appearance (shape and texture) variation of the modelled structure.

A variety of methods have been proposed combining pairs of techniques such as segmentation and registration with active countours [6] and maximum a posteriori segmentation using hidden markov random fields and B-spline non-rigid registration [7]. Models of deformation have been constructed from correspondences estimated by non-rigid registration [8, 9], but it was also shown that by integrating modeling and registration more tightly, by performing them in parallel [1, 10] quantitatively better models result.

In this paper we describe an algorithm which combines segmentation, registration and model building in a single iterative framework that incrementally improves the structure analysis to satisfy all the requirements identified above. The method is described in detail in Section 2 while results of applying it to a set of MR images of the brain are shown in Section 3.

2 Method

An overview of the algorithm is illustrated in Figure 1. A set of N images T_i, $i = 1...N$, (the training set) is assumed to have been roughly aligned (either the imaging protocol is sufficient, or for instance by affine registration to one example using a Mutual Information metric). It is further assumed that the structures in the images consist of M distinct tissue types whose intensity probability density functions have parameters θ_i. In the following, $W_i()$ represents a spatial deformation defining correspondence from a reference frame to the image T_i. The segmentation, registration and modeling proceeds sequentially through the following steps (but see below for more details):

1. Warp each training image T_i into the reference frame using the current estimate of the deformation field. $T'_i = W_i^{-1}(T_i)$.

* v.petrovic@manchester.ac.uk

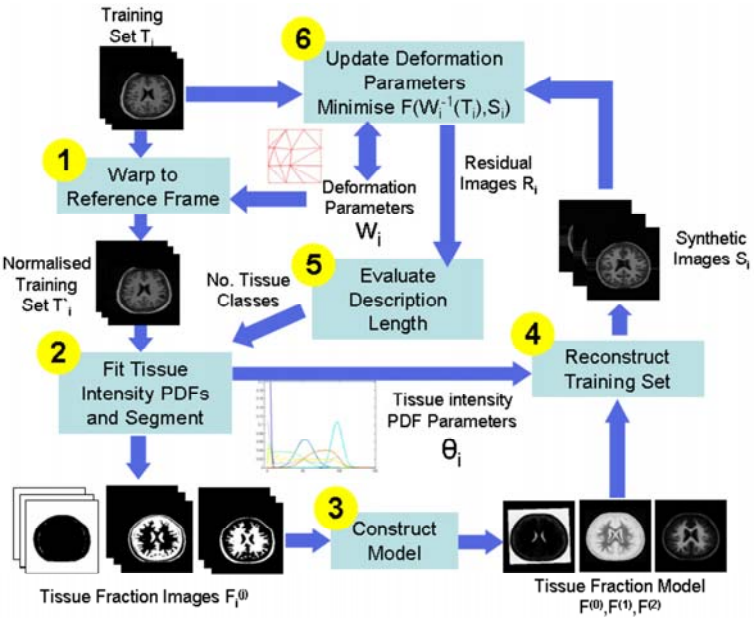

Figure 1. Outline of the algorithm.

2. Compute the intensity histogram of T_i', and fit a mixture model to estimate the distributions and proportions of each pure tissue class (means, SDs and weights encoded in $\theta_i = \{\mu_{ij}, \sigma_{ij}, w_{ik}\}$). Use the resulting distributions to estimate the most probable fraction of each tissue in each voxel, encoded in a set of tissue fraction images $F_i^{(j)}$, $j = 1...M$ for each training example.
3. Combine the tissue fraction images from all examples to construct a model, $\{\hat{F}^{(1)}...\hat{F}^{(M)}\}$.
4. Synthesize a reconstruction of each training set image using the current estimates of the mean pure tissue class distributions, μ_{ij}, and the current tissue fraction model, $S_i = \sum_{j=1}^{M} \mu_{ij}\hat{F}^{(j)}$
5. Find optimal number of tissue classes M by evaluating the Description Length (DL) of the modeled training set representation as a function of M, see [1], and using the one that results in minimum DL (MDL).
6. Update the current estimate of W_i to best register S_i onto T_i, minimising a suitable similarity measure, $D_{im}(T_i, W_i(S_i))$.

During the first pass of the algorithm we apply steps 1 and 2 to all images, then combine them (step 3) into an initial model of the tissue fractions at each voxel in the reference frame. Steps 2 to 5 are then repeated iteratively until an optimum M is found. We then apply steps 1 to 6 (exc. 5) to each image in turn multiple times, performing registration of coarse details in the early stages and progressively finer details as the iterations progresses. The process of shape model building is excluded from the Figure 1 for the sake of clarity, but is an integral part of step 3 and can further be used to constrain the deformation field optimisation in step 6 [10].

2.1 Segmentation

We follow Pokric et al. [4] in assuming that the objects in MR images of the brain are constructed from a small number (M) of different tissue types, and that each voxel contains either a pure tissue or a mixture of at most two different tissues (partial volumes). If we know the distributions of intensities for these classes, we can construct the distribution for a particular fractional distribution by convolution. In the experiments described below we assume that the pure tissues distributions are Normals, $p_i(g) = N(g : \mu_i, \sigma_i^2)$, then the distribution for a partial volume with fraction f of tissue type i and $1 - f$ of type j is given by

$$p_{ij}(g|f) = N(g : f\mu_i + (1-f)\mu_j, f\sigma_i^2 + (1-f)\sigma_j^2) \tag{1}$$

The distribution over all partial volumes containing i and j is given by (2) where we assume all values of f in the range $[0, 1]$ are equally likely ($p(f) = 1$).

$$p_{ij}(g) = \int_{f=0}^{f=1} p_{ij}(g|f)p(f)df = \int_{f=0}^{f=1} p_{ij}(g|f)df \tag{2}$$

If we assume that any voxel contains at most 2 different tissue types, then we need only consider M pure tissue classes with distributions $p_k(g)$, $k = 1..M$, and $M(M-1)/2$ partial tissue classes (enumerated $p_k(g)$,$k = (M+1)..M_t = M(M+1)/2$). Thus the measured image intensity distribution, $h(g)$, can be approximated as a weighted sum where $\theta = \{\mu_i, \sigma_i, w_k\}$ ($i = 1..M, k = 1..M_t$):

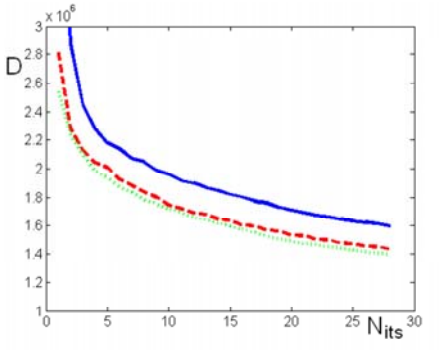

Figure 2. Cummulative obj. function D vs. registration iteration for simple mean (solid), half trimmed mean (dashed) and median model estimation (dotted line)

$$p(g:\theta) = \sum_{i=1}^{M_t} w_i p_i(g) \tag{3}$$

We thus perform an optimisation to estimate the parameters θ which optimise $D_p(p(g:\theta), h(g))$, where $D_p(p,q)$ is a suitable measure of divergence between distributions. The probability that a voxel with intensity g belongs to class k is given by $P_k(g) = w_k p_k(g) / \left(\sum w_k p_k(g)\right)$ (see Figure 4). That voxel can then be classified as belonging to class $k_c = \arg\max_k P_k(g)$. However, we are actually interested in the estimate of the fraction of each pure class tissue (f_i,$i = 1..M$), in the voxel, not the probability of each class. If $k_c \leq M$ then the voxel is a pure tissue, so we define $f_{k_c} = 1$ and $f_{i \neq k_c} = 0$. If $k_c > M$ then the voxel is classified as a partial volume, containing two tissues, say of type i and type j. In this case we wish to find the most likely value of the fractions for each tissue. We define

$$f_i \quad = \quad \arg\max_f p_{ij}(f|g) = \arg\max_f p_{ij}(g|f)p(f)/p(g) = \arg\max_f p_{ij}(g|f) \tag{4}$$

where $p_{ij}(g|f)$ is defined above in Equation 1. We then set $f_j = 1 - f_i$ and $f_{k \neq i,j} = 0$. Figure 4 shows an example of this, demonstrating that tissue probabilities are not the same as estimates of pure tissue fractions. Using this approach we compute M images, $\{F_i^{(1)}, ...F_i^{(M)}\}$, recording the fraction of each tissue type at each voxel in the normalised version of image i (that projected into the reference frame).

2.2 Model Construction

We train a model from the M fractional images from each of the N images. [1] Constraints can be imposed on model construction that would say limit any voxel to a mixture of at most two tissue types but although this directly supports convergence, we found that even the simple mean is powerful enough to drive the process to convergeance.

$$\{\hat{F}^{(1)}...\hat{F}^{(M)}\} = \frac{1}{N}\sum_i \{F_i^{(1)}...F_i^{(M)}\} \tag{5}$$

Greater convergence speed is achieved by using a more robust trimmed mean estimation (evaluated by eliminiting outliers). On a separate data set of 50 synthetic, half light - half dark images that vary randomly in the position intensity tranistion and levels, we found that using only half the samples in mean estimation considerably improves the speed of registration convergence. Convergence for this data set is shown as cumulative objective funcion (D) vs. the registration iteration for the simple mean, half trimmed mean and median on Figure 2.

2.3 Image Reconstruction

In order to align the training set a deformation field is optimised between each T_i and the model (reference frame) embodied in a reconstruction, S_i produced using the model tissue fractions and the estimates of intensity distributions for each tissue type in the current image (i.e. an estimate of how the model would appear given the imaging conditions for that image). Pure tissue types exhibiting Normal distributions are optimally represented by their mean ($\mu_i j$) while partial volume voxels are represented by a sum of pure tissue means weighted by the current model tissue fractions;

$$S_i = \sum_{j=1}^{M} \mu_{ij} \hat{F}^{(j)} \tag{6}$$

For an example, see Figure 5. Ideally S_i is a noise free version of T_i but in practice it starts blurred due to missalignments and gets progressively sharper and closer to T_i as alignment and model improve. Deformation parameters W_i

[1] In practice, when working on image i we construct the model from all $N-1$ other images. Such a leave-one-out approach tends to give more generalisable models and lead to faster convergence.

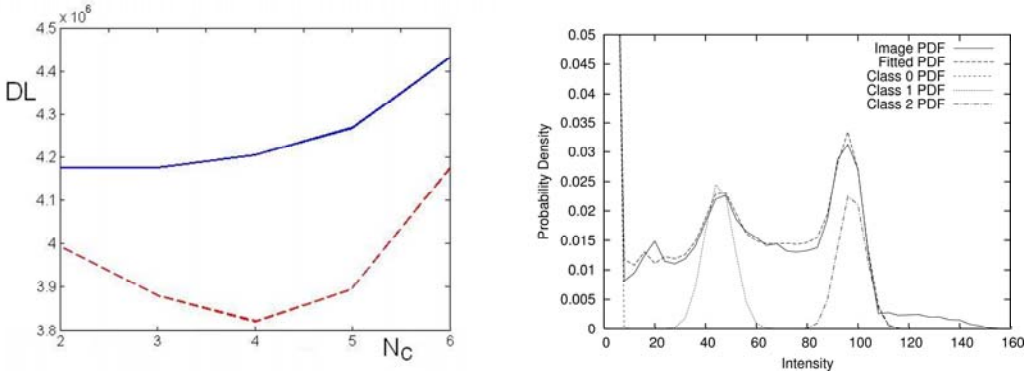

Figure 3. Description Length of modeled training set vs number of tissue classes (N_c) for non-aligned (solid line) and aligned (dashed line) brain images (left); Distribution of intensities for image in Fig. 4 and result of fitting pure and partial tissue class PDFs

are optimised with respect to an objective function measuring similarity between T_i and S_i in the training image frame - $D_{im}(T_i, W_i^{-1}(S_i))$. The deformation is optimised in several stages and at several resolution levels (for the sake of efficiency) by moving individual or groups of control points across the scene, see [10] for details.

2.4 Number of Segment Classes

To initialise the image analysis automaticaly and avoid explicit prior assumptions about image structure we search for an optimal number of tissue types M automatically. For this purpose we use the proposed segmentation-registration algorithm in a Minimum Description Lenght framework, analogously to [1]. We start with the minimum $M = 2$ and perform steps 2 to 4 (Fig. 1 for increasing values of M. For each we evaluate the Decription Lenght of transmiting the training set using the resulting model, given as the cost of the model, model parameters and the residual images ($R_i = S_i - T_i$), $C_{total} = C_{model} + C_{parameters} + C_{residuals}$.

In practice the residual term dominates the cost and improving the model and reconstructed images explicitly improves the Decription Length. If adding another class improves and reconstructed images more than it increases the cost of the model the Description Length decreases. The optimal number of classes is the one that results in a Minimum Description Length (MDL). Note that as we improve the model by aligning images we can realisticaly expect that the optimal value for M might change during registration, see Section 3. For details on how to evaluate C, see [1].

3 Experiments and Results

We applied the method to a set of 32 2D slices of MR brain images (choosing equivalent slices from affine aligned 3D datasets) (Figure 4). We used sum of square differences for both the image similarity, $D_{im}()$, and the divergence between intensity distributions, $D_p()$. Figure 4 shows one of the images from the training set. By fitting pure and partial tissue class distributions to the image histogram (Figure 3), we can estimate the probability that each voxel belongs to each pure tissue class ($P_i(x, y)$ in column 2) and to a partial tissue class ($P_{partial}(x, y)$). We can then estimate the fractions of each tissue type in each voxel ($F_i(x, y)$ in column 3). Finally, combining these fractions with the means of the pure class distributions, we can generate the resconstruction, $S(x, y)$.

Automatic estimation of the number of tissue classes in the data is shown on Figure 3. While the images are not aligned fewer types are sufficient, in this case $M = 3$ is optimal. As the images are brought into alignment, the structures are more coherent accross the examples and further patterns emerge, Figure 3 indicates that four different, consistent tissue types exist in the registered set. As we have not performed any "skull-stripping", this region also contributes to the intensity distributions, in addition to three types (white and grey matter and CSF + background) identified initially.

Figure 5 shows the mean of the model tissue fractions for the three tissue types at the beginning and end of the registration, together with an example of the resulting reconstruction using the means of each pure tissue class. As registration progresses the alignment becomes more accurate, resulting in crisper estimates of the tissue fractions.

4 Discussion

We have demonstrated that it is possible to integrate segmentation, modeling and registration into a single framework and perform complex object structure analysis in a fully automatic manner. By constructing a model of segment

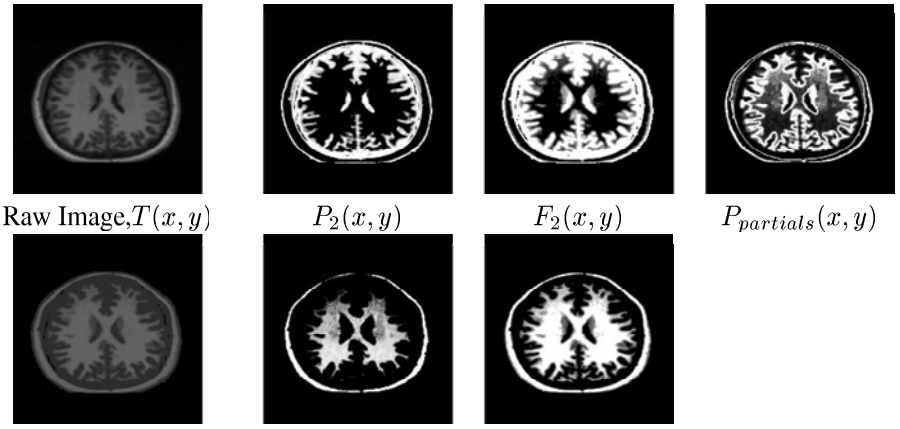

Figure 4. MR brain slice and resulting segmentations. Column 2 shows the pure tissue probabilities, column 3 the estimates of pure tissue fractions.

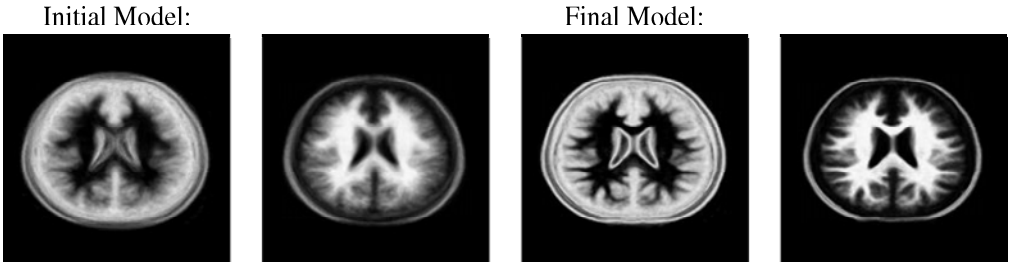

Figure 5. Mean tissue fraction estimates $\hat{F}_2(x,y)$ and $\hat{F}_3(x,y)$ at the beginning and end of registration.

fractions, rather than intensities, we decouple the model from details of the imaging process, and concentrate on explicitly learning tissue structure in medical images. The final model includes both tissue class information and statistical models of shape variation. It should be possible to match such a model to new images taken with different modalities, and we anticipate that such models would have wide application.

We are currently extending the implementation to full 3D data (the extension is natural). Further work will also concentrate on including spatial consistency in estimating the optimal number of tissue types as well as their intensity distributions. Further improvements could also be made by including the current estimates of the mean fraction as a prior when segmenting each image.

References

1. C. J. Twining, T. F. Cootes, S. Marsland et al. "A unified information-theoretic approach to groupwise non-rigid registration and model building." In *Proceedings of Information Processing in Medical Imaging (IPMI)*, volume 3565 of *Lecture Notes in Computer Science*, pp. 1–14. Springer, 2005.
2. K. K. Bhatia, J. V. Hajnal, B. K. Puri et al. "Consistent groupwise non-rigid registration for atlas construction." *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)* pp. 908–911, 2004.
3. B. Zitová & J. Flusser. "Image registration methods: A survey." *Image and Vision Computing* **21**, pp. 977 – 1000, 2003.
4. N. A. T. M. Pokric & A. Jackson. "The importance of partial voluming in multi-dimensional medical image segmentation." In *Proceedings of Information Processing in Medical Imaging (IPMI)*, volume 2208 of *Lecture Notes in Computer Science*, pp. 1293–1294. Springer, 2001.
5. Y. Zhang, M. Brady & S. Smith. "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm." *IEEE Transactions on Medical Imaging* **20**, pp. 45 – 57, 2001.
6. A. Yezzi, L. Zollei & T. Kapur. "A variational framework for inregrating segmentation and registration through active contours." *Medical Image Analysis* **7**, pp. 171–185, 2003.
7. D. R. C. Xiaohua, M. Brady. "Simultaneous segmentation and registration for medical image." In *Proceedings of MICCAI 2004*, number 3216 in Lecture Notes in Computer Science, pp. 663 – 670. 2004.
8. A. F. Frangi, D. Rueckert, J. A. Schnabel et al. "Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling." *IEEE Transactions on Medical Imaging* **21(9)**, pp. 1151–1166, 2002.
9. D. Rueckert, A. F. Frangi & J. A. Schnabel. "Automatic construction of 3D statistical deformation models using non-rigid registration." *Lecture Notes in Computer Science* **2208**, pp. 77–84, 2001.
10. T. Cootes, C. Twining, V. Petrovic et al. "Groupwise construction of appearance models using piece-wise affine deformations." In *Proceedings of the* 16^{th} *British Machine Vision Conference (BMVC)*, volume 2, pp. 879–888. 2005.