Simultaneous Alignment and Central Tendency Estimation for Brain Atlas Construction

Neil I. Weisenfeld[1,2] and Simon K. Warfield [1]

[1] Computational Radiology Laboratory
Brigham and Women's and Children's Hospitals
Harvard Medical School, Boston, MA
[2] Department of Cognitive and Neural Systems
Boston University, Boston, MA

Abstract. Two key questions for investigators seeking to compare images across subjects are how to define a common coordinate system and how best to relate each subject to that coordinate system. We have developed a generalized expectation-maximization (EM) algorithm which iterates between estimating the most typical anatomy, assessing the typicality of each individual subject's anatomy, and registering each individual subject to the typical anatomy estimate. Unlike methods which seek a coordinate system unbiased from the average, we weight the contribution of each tissue class from each subject based on the estimated typicality of that subject's features. We demonstrate our algorithm using segmented brain images from a group of 14 adults and 14 preterm infants.

1 Introduction

Atlases play several key roles in brain image analysis. Atlases can be used as statistical prior probabilities for segmentation algorithms, as a representation of a population norm against which subjects are graded, or to provide a common coordinate system for inter-subject studies. Atlases are distinguished from a single-subject image by the encapsulation of some measure of population statistics. Two key issues are the representation of group statistics and how a new individual is related to the group. Different applications and even subject populations require different atlases. For instance, Talairach and Tournoux created an atlas of a single subject along with a landmark-based approach for aligning individual subjects with the atlas[1]. This atlas is not suitable, however, for pediatric images [2, 3].

A number of registration-based approaches for atlas construction have been proposed and a key issue is how the target coordinate system is defined. A simple strategy is to choose a single subject and then align each other subject in the cohort to this target. Rueckert [4] offered a dramatic demonstration of how the particular choice of target image can bias the atlas, and several authors have proposed schemes to either avoid choosing a single subject target or to minimize the impact of such a choice. Guimond proposed an iterative scheme in which a

single reference is chosen, but the average transformation is factored out and a new average target is developed at each iteration [5]. This minimizes the impact of which subject *within* a cohort is chosen as the reference standard, as long as no subject differs from the desired group average by a transformation that can't be expressed by the registration scheme being employed. Rueckert performed statistics on the parameters of a B-spline deformation model in order to define a reference anatomy that is a minimal distance from each subject in the cohort [4]. Zöllei developed a joint registration scheme which naturally avoids bias toward a particular reference subject through a joint entropy-based objective function evaluated simultaneously across all subjects [6].

Such registration based schemes depend critically upon the alignment metric used (e.g. mutual information, correlation) and upon the capability of the underlying transform model. The use of alignment metrics based on signal intensity may not distinguish between structures with similar intensities, such as primary sensory and motor gyri. Furthermore, in the typical situation where subjects are aligned and statistics are gathered about the post-alignment residual variability, information about a given subject's anatomy is partitioned somewhat arbitrarily between the registration transform and the residual error.

In this paper we propose a new registration metric that extends previous work on segmentation validation [7] and atlas construction [8]. The goal is to generate an unbiased group registration which also takes into account the typicality of a subject's anatomy. Such a scheme is not biased to any particular subject and also minimizes the contribution of outliers. The proposed method operates directly on segmentations which, in principle, allows arbitrarily complicated anatomy to be utilized, and can operate on grayscale data by first employing automatic segmentation techniques. Our method operates by considering a cohort of subject segmentations to each be a representation of the unknown typical anatomy, altered by an unknown transformation, and solves iteratively for both the typical anatomy and the transform that best aligns each individual segmentation to the typical.

2 Materials and Methods

2.1 Algorithm

The Simultaneous Truth and Performance Level Estimation (STAPLE) [7] algorithm was originally described for the assessment of repeated segmentations of the same image data by a group of raters. In the presence of a reference standard, assessing the performance of individual raters (e.g. sensitivity and specificity) is straightforward, but typically no reference standard is available. STAPLE is an Expectation-Maximization algorithm [9] that treats the true, underlying segmentation as hidden or missing data with a known distribution and iterates between estimating the true segmentation and estimating each of the raters' performance scores. The resulting true segmentation estimate is a linear combination of each rater's segmentation weighted by that rater's performance.

The result is a reference standard in which poorly performing raters are given less weight than well-performing ones.

For the purpose of atlas construction, we can take a group of segmentations of *different* subjects and consider them drawn distorted from a typical anatomy. We have developed a new algorithm which estimates a registration between each subject and the unknown reference standard and also a set of typicality parameters for each tissue class in each subject's segmentation. Using the typicality parameters, we construct an atlas in a group coordinate system made by emphasizing typicality and reducing the influence of subjects who do not match the typical anatomy.

The algorithm proceeds as follows:

1. *Transform Initialization:* the center of mass of each segmentation is computed and each subject's registration is initialized with a translation to the centroid of the centers of mass.
2. *Typicality Parameters Initialization:* the typicality parameters matrix is set with values very close to 1 on-diagonal and the residual is divided evenly among the off-diagonal entries.
3. *Expectation:* The probability density function for the typical anatomy is computed, and this provides a probabilistic estimate of the typical anatomy. This is used to compute the expectation over the typical anatomy.
4. *Maximization:* each parameter is estimated in order to maximize the complete-data log likelihood. Since the registration depends on the typicality parameters and vice-versa, no closed form solution exists. Instead we employ a coordinate ascent scheme alternately fixing the typicality parameters and the transforms while maximizing the complete-data log likelihood.
5. The E and M steps are iterated until convergence.

The Expectation-Maximization algorithm is designed for the estimation of parameters in the face of missing data, if a probability density function for the missing data is known or can be estimated. In that case, the desired quantity, referred to as "parameters" in EM literature, is estimated in a manner that maximizes the expectation of the "complete-data" log-likelihood over the missing data. The complete-data comprises the observed data as well as the missing or "hidden" observations.

In our case, the observed data $\mathbf{D}$ consists of the segmentations or label maps D_j for our $j = 1 \ldots N$ subjects. The hidden data is the group central tendency or typical segmentation $\mathbf{T}$ and the complete data consists of $(\mathbf{D}, \mathbf{T})$. The parameters to be estimated are both the typicality parameters θ_j and a registration transformation A_j for each subject. θ_j is defined as a matrix with rows s and columns s' such that

$$\theta_{jss'} \equiv f(D_j = s | T = s')$$

Each of the diagonal elements represents agreement for a given label between the subject and the central tendency and the off-diagonal elements represent the probability of each possible combination of disagreement. The transformation

parameter A_j represents a registration between subject j and the current estimate of the group central tendency T. In the present work we model A_j as an affine transformation, however other transformations (e.g. non-linear) could be substituted.

Given these definitions, the complete data log likelihood is $\ln f(\mathbf{D}, \mathbf{T}|\mathbf{\Theta}, \mathbf{A})$.

2.2 E-Step: Central Tendency Estimation

Two key components of this algorithm are estimation of the "weights" or the probability density function of the central tendency conditioned upon each subject's segmentation and the previous iteration's estimate of the typicality parameters and registration:

$$W_{si}^{(k)} \equiv f(T_i = s|\mathbf{D_i}, \mathbf{\Theta}^{(k)}, \mathbf{A}^{(k)}) \tag{1}$$

$$= \frac{f(T_i = s) \prod_j f(D_{(A_j i)j}|T_i = s, \theta_j^{(k)}, A_j^{(k)})}{\sum_{s'} f(T_i = s') \prod_j f(D_{(A_j i)j}|T_i = s', \theta_j^{(k)}, A_j^{(k)})} \tag{2}$$

At each voxel i, the probability that the central tendency segmentation label is s, conditioned upon the estimates available so far is the prior probability of a label $f(T_i = s)$ multiplied by the product of the probability of the j independent subject decisions conditioned on the true typical segmentation and the typicality parameter estimates. Our notation $(A_j i)$ signifies the transform A_j applied to voxel index i.

With this probability density function, the expectation can be computed:

$$Q(\mathbf{\Theta}^{(k)}, \mathbf{A}^{(k)}|\mathbf{\Theta}^{(k-1)}, \mathbf{A}^{(k-1)}) = E[\ln f(\mathbf{D}, \mathbf{T}|\mathbf{\Theta}, \mathbf{A})\big|\mathbf{D}, \mathbf{\Theta}^{(k-1)}, \mathbf{A}^{(k-1)}] \tag{3}$$

$$= \sum_{\mathbf{T}} W^{(k-1)} \ln f(\mathbf{D}, \mathbf{T}|\mathbf{\Theta}^{(k-1)}, \mathbf{A}^{(k-1)}) \tag{4}$$

2.3 M-Step: Typicality Parameter Estimation and Registration

Our goal for the M-Step is computation of

$$(\hat{\mathbf{\Theta}}, \hat{\mathbf{A}}) = \arg\max_{(\mathbf{\Theta}, \mathbf{A})} Q(\mathbf{\Theta}, \mathbf{A}|\mathbf{\Theta}^{(k-1)}, \mathbf{A}^{(k-1)}) \tag{5}$$

Since $\mathbf{\Theta}$ and $\mathbf{A}$ are interdependent, we use a coordinate ascent scheme alternatively fixing one parameter while maximizing Q over the other until convergence. Considering the problem with $\mathbf{A}$ fixed, the typicality parameters that maximize Q are:

$$\theta_{jss'}^{(k)} = \frac{\sum_{i:D_{(A_j i)j}=s} W_{s'i}^{(k-1)}}{\sum_i W_{s'i}^{(k-1)}} \tag{6}$$

Careful derivations of this result, minus the coordinate transform, is available in [7].

Now for a fixed set of typicality parameters, we seek the registration transform A_j which maximizes the complete data log-likelihood:

$$\mathbf{A}^{(k)} = \arg\max_{\mathbf{A}} \ln f(\mathbf{D}, \mathbf{T}|\mathbf{\Theta}^{(k-1)}, \mathbf{A}^{(k-1)}) \tag{7}$$

Since a given transformation A_j aligns only a single subject j:

$$A_j^{(k)} = \arg\max_{A_j} \sum_i E\big[\ln f(D_{(A_j i)j}|T_i, \theta_j, A_j\big] \tag{8}$$

$$= \arg\max_{A_j} \sum_s \sum_{i:D_{(A_j i)j}=s} \sum_{s'} W_{s'i}^{(k-1)} \ln \theta_{jss'}^{(k-1)} \tag{9}$$

Equation 9 provides our metric for registration of an individual segmentation D_j to our current estimate of the central tendency T. The current weights estimate W provides the conditional probability of the central tendency used in the conditional expectation, while the performance parameters are $f(\mathbf{D}|\mathbf{T}, \mathbf{\Theta}, \mathbf{A})$. The metric is such that each voxel's contribution is the log probability of the segmentation label given the possibility that the central tendency is each possible labeling, multiplied by the probability that each possible labeling is true.

Since no closed form solution to equation 9 exists, we use numerical optimization to maximize the metric with respect to the transforms.

3 Results

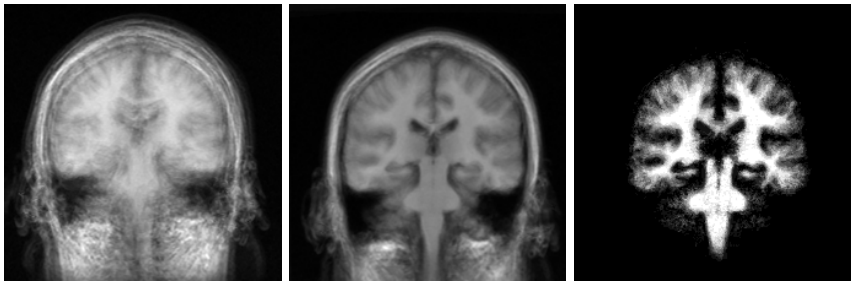

Fig. 1. Average grayscale images from adult subjects after a single iteration of translation correction (left) and after full affine alignment (center). The corresponding white matter probability for the resulting atlas is shown at (right).

For the purpose of investigating the algorithm, previously generated segmentations were used from other studies. 14 adult images were segmented using a previously published method and the segmentations were registered using the

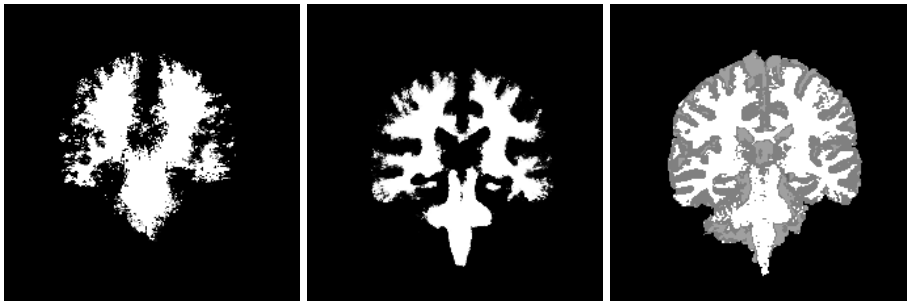

Fig. 2. Typical anatomy for white matter after a single iteration (left) and at the end of the algorithm (center). An example subject's segmentation is shown (right) aligned with the typical anatomy.

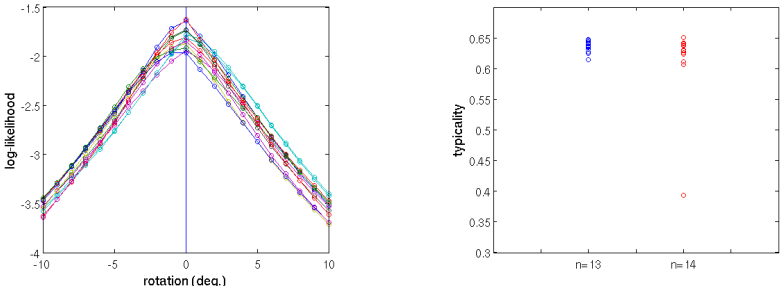

Fig. 3. Plot of the new metric in the neighborhood of the solution for each of 13 subjects compared with the typical anatomy (left). The peak shows a local maximum at the proper alignment. On the (right) is a plot of mean (across tissue) typicality scores (mean of the diagonal elements) for a group of 13 subjects (left column; n=13) and for that same group with one grossly mismatched subject (right column; n=14). The mismatched subject is scored very low (0.39) and its effect on the typical anatomy estimate is weighted accordingly.

proposed metric. The resulting affine transformations taking each subject's segmentation to the space of the typical anatomy was applied, solely for visualization purposes, to the associated T1-weighted MRI and an average grayscale image was formed. Figure 1 shows an initial registration (left) and the final resulting registration (center). Despite gross initial misalignment, our algorithm identifies common structure in a typical anatomy estimate which is refined across iterations as the data becomes better aligned. Figure 2 shows the initial estimate of the central tendency (left) for a tissue class representing white matter. The final white matter estimate central tendency is shown at center.

Figure 1 (right) shows the spatially varying white-matter probability in the generated atlas. The central tendency shown in Figure 2 (center) is used to define the atlas coordinate frame, while simply counting the voxelwise prevalence of a given tissue class, in this space, is used as the atlas itself. Figure 2 (right) shows

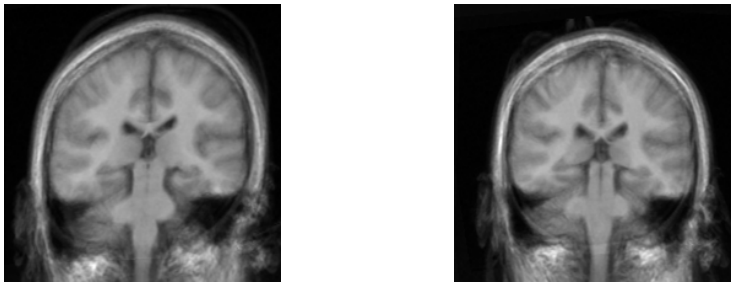

Fig. 4. Mean grayscale images from the same group of subjects as in Figure 1, but with the addition of a single grossly mismatched subject. The figure on the (left) shows the distorted average image which results when typicality parameters estimates are ignored and each subject, including the outlier, is weighted equally. On the (right) is the generated mean grayscale image when the outlier is weighted poorly based on the algorithm's typicality score for the subject. This more closely resembles the registration achieved without the outlier, shown in Figure 1.

a coronal slice from an individual's segmentation after registration to the central tendency.

Figure 3 (left) shows the results of plotting the metric for each subject compared to the derived typical anatomy for different rotations around the algorithm-derived solution indicated by the zero value line in the center of the graph. The function has a clear local maximum at proper alignment with the typical anatomy. In order to demonstrate the robustness to outliers, a single individual was rotated 90 degrees around each of two axes. When the estimation of typicality parameters was disabled, leading to an equal weighting of each subject, the atlas was sheared in an attempt to match the outlier image (Figure 4 (left)). With normal operation of the algorithm, the outlier is scored as atypical and has a reduced influence on the resulting atlas (Figure 4 (right)). The mean (across tissue types) typicality parameters are plotted in Figure 3 (right) for the original group of 13 (left column) and the same group with the added outlier image (right column). The well-matched images garner typicality scores above 0.6 while the outlier scores 0.39.

We performed a similar study on pre-term infants with a cohort of 13 chidren born pre-term, but scanned at 40 weeks post-menstrual age (PMA) and two additional children scanned at 30 weeks PMA. The affine transform for each subject was decomposed into the rigid and affine component. The trace of the affine component is an indicator of the amount of scaling necessary to align each subject to the atlas. We found that the mean and standard deviation of the trace of the 40-week PMA children was 3.28 ± 0.12 and the 30-week values were 2.71 and 2.72. This indicates that the algorithm correctly scaled the smaller brains of the 30-week PMA children to match that of the atlas.

4 Discussion and Conclusion

We have developed a novel method for the construction of a statistical atlas from the segmentations of a group of subjects. Our algorithm automatically determines how typical each subject is and a group common coordinate system is influenced by each individual's anatomy in a manner proportional to their typicality. We have developed a new registration metric which incorporates typicality parameters into a properly weighted comparison of segmentations with a probabilistic estimate of the group's central tendency and have demonstrated an EM algorithm which simultaneously estimates individual registrations, typicality parameters, and the group central tendency which defines the common coordinate system. Future work will include investigation of non-linear transform models which the method naturally supports.

5 Acknowledgments

This investigation was supported in part by a research grant from CIMIT, grant RG 3478A2/2 from the NMSS, by NSF ITR 0426558, and by NIH grants R03 CA126466, R01 HL074942, P30 HD018655, R01 RR021885 and R01 HD046855.

References

1. Talairach, J., Tournoux, P.: Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers, Inc., New York (1988)
2. Mewes, A., Zollei, L., Huppi, P., Als, H., McAnulty, G., Inder, T., Wells, W., Warfield, S.: Displacement of brain regions in preterm infants with non-synostotic dolichocephaly investigated by MRI. Neuroimage (2007)
3. Wilke, M., Schmithorst, V.J., Holland, S.K.: Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data. Magn Reson Med **50**(4) (2003) 749–757
4. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3D statistical deformation models of the brain using non-rigid registration. IEEE Trans. Med. Imaging **22**(8) (2003) 1014–1025
5. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: a convergence study. Comput. Vis. Image Underst. **77**(9) (2000) 192–210
6. Zöllei, L.: A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images. PhD thesis, MIT, Cambridge, MA, USA (2006) available at http://people.csail.mit.edu/lzollei.
7. Warfield, S.K., Zou, K.H., III, W.M.W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging **23**(7) (2004) 903–921
8. Craene, M.D., du Bois d'Aische, A., Macq, B., Warfield, S.K.: : Multi-subject registration for unbiased statistical atlas construction. In Barillot, C., Haynor, D.R., Hellier, P., eds.: MICCAI (1). Volume 3216 of Lecture Notes in Computer Science., Springer (2004) 655–662
9. Dempster, A., Laird, N., Rubin, D.: Maximum-likelihood from incomplete data via the EM algorithm. J. Royal Statistical Soc. Ser. B **39** (1977) 34–37