Joint Registration and Clustering of Images

Mert R. Sabuncu[1], Martha E. Shenton[2], and Polina Golland[1]

1. Computer Science and Artificial Intelligence Lab,
Massachusetts Institute of Technology, Cambridge MA 02139, USA
2. Psychiatry Neuroimaging Labortory, Department of Psychiatry, Brigham and
Women's Hospital, Harvard Medical School, Boston, MA 02215, USA

Abstract. We demonstrate an EM-based algorithm that jointly registers and clusters a group of images using an affine transformation model. The output is a small number of prototype images that represent the different modes of the population. The proposed algorithm can be viewed as a generalization of other well-known atlas construction algorithms, where the collection of prototypes represent multiple atlases for that population. Our experiments indicate that the employment of multiple atlases improves the localization of the underlying structure in a new subject.

Key words: Image Registration, Image Clustering, Multiple Atlases

1 Introduction

This paper introduces a general probabilistic framework to jointly co-register and cluster a group of images. This method is applicable in a wide range of applications, where a database of images needs to be summarized concisely, e.g. with a small number of prototypes. In medical imaging, atlases are used for various purposes, including structure/function localization, morphometry, segmentation, and parcellation. Unlike the traditional approach that uses one atlas (i.e., one mean image or one probabilistic image) to represent the whole population, we employ the proposed algorithm to compute multiple atlases to capture the different modes in a population. The same framework then can be used to register a new image and determine its cluster membership. We demonstrate the utility of having multiple atlases for the application of localizing medial temporal brain structures in a pool of subjects that consists of healthy controls and schizophrenics. Our experiments indicate that the best results are achieved if the actual group memberships (schizophrenic vs. healthy) are used and two different atlases are computed. Our clustering algorithm, on the other hand, achieves comparably good results without the ground truth membership information. The alignment quality of the underlying structures, as measured by the Dice measure [4], is around 5% better than the results obtained with a single atlas.

1.1 Prior Work

An important problem in medical imaging is the construction of an atlas from a group of subjects. The term atlas usually refers to a probabilistic model, of which

the parameters are learned from a training data set [14]. In its simplest form, an atlas is a mean intensity image. Yet, richer statistics, e.g. intensity variance, segmentation label counts, etc., can also be included in the atlas model [5].

Atlas construction requires a dense correspondence across subjects. Earlier techniques used a template image (either a universal template, such as the MNI template brain [3], or an arbitrary subject from the training data set [6]) to initially align the training subjects using a pairwise registration algorithm. Other techniques have focused on determining the least biased template from the training set [11, 9, 10]. The drawback of such algorithms is that they represent the whole population using a single template. This can be sub-optimal in situations where there are more than one "mode" in the population. To circumvent this, more recent approaches have proposed to co-register the group of images simultaneously without computing a group template [12, 16]. These algorithms, however, don't yield the multiple modes of the population. In [2], Blezek and Miller have examined a method to automatically identify the modes of a population using a mean-shift algorithm. Rather than integrating image registration into their framework, the authors treat the transformation as a degree of freedom over which the algorithm optimizes when computing pairwise image distances. This resulted in an explosion of pairwise image registration instances, each of which can be computationally expensive. An alternative strategy, employed in [7], is to use all training images as the atlas. A new subject is registered with each training image and segmentation label inference, for example, will be based on a fusion of the manual labels in the training data. This approach is not suitable for anatomical variability studies, where the subjects should be in a common coordinate frame.

In this paper, we investigate a probabilistic framework for joint registration of a set of images into a common coordinate frame, while clustering them into a small number of groups, each represented by a prototype image. We employ a simple mixture of Gaussians model and a maximum likelihood framework which we solve using Expectation Maximization (EM). Our implementation can be viewed as an extension of the approach of [13], which solves the registration problem as an initial, seperate step. We demonstrate the algorithm using 3D MR data and an affine transformation model. A recent study [1], provides a statistical analysis of a MAP formulation based on a similar model and proves asymptotic consistency of the final algorithm[1].

2 Theory

Let $\{P_k\}_{k=1}^K$ be a small number of prototype images that summarize the group of images $\{I_n\}_{n=1}^N$. Our model is, for all $n \in \{1, \ldots, N\}$, there exists a $k \in \{1, \ldots, K\}$ such that:

$$I_n(\boldsymbol{\Phi}_n(\boldsymbol{x})) = P_k(\boldsymbol{x}) + \epsilon(\boldsymbol{x}),\ \forall \boldsymbol{x} \in \Omega \subset \mathbb{R}^3, \tag{1}$$

where $\boldsymbol{\Phi}_n : \Omega \mapsto \mathbb{R}^3$ is an admissible spatial warp, e.g., an affine transformation, $\epsilon(\boldsymbol{x})$ is an independent, non-stationary Gaussian random variable with zero mean

[1] Thanks to the anonymous reviewer who pointed us to this excellent paper.

and a variance of σ_x^2. Our goal is to find all prototypes P_k and variance estimates σ_x^2's, while simultaneously solving for the Φ_n's.

We can view the observed images as a group of spatially transformed samples from a mixture of Gaussians. Cluster k has the mean image $P_k(\boldsymbol{x})$. The generative process begins with a random choice for the cluster k. Each cluster may have a different prior probability π_k. Next, a zero mean non-stationary independent Gaussian noise is added to the mean image of the cluster. This image is then transformed by applying a randomly selected spatial transformation Φ^{-1} to *generate* the observed image I. Thus, given the prototype images $\{P_k\}$, variance image $\Sigma(\boldsymbol{x}) = \sigma_x^2$, spatial transformation Φ, and prototype priors $\{\pi_k\}$ ($\sum_k \pi_k = 1$), the probability of observing the image I is[2]:

$$p(I|\{P_k\}, \{\pi_k\}, \Sigma, \Phi) = \sum_k \pi_k p(I|C = k, P_k, \Sigma, \Phi) \tag{2}$$

where C denotes the cluster that generates I. Using Eq. (1), we obtain:

$$p(I|C = k, P_k, \Sigma, \Phi) = \prod_{\boldsymbol{x}\in\Omega} \mathcal{N}(I(\Phi(\boldsymbol{x})); P_k(\boldsymbol{x}), \Sigma(\boldsymbol{x})), \tag{3}$$

where $\mathcal{N}(x; \mu, \sigma)$ is the Gaussian density with mean μ and standard deviation σ. We formulate the problem of atlas construction as a maximum likelihood estimation:

$$\theta^* = \arg\max_\theta \sum_n \log p(I_n|\theta), \tag{4}$$

where $\theta = \{\{P_k\}, \{\pi_k\}, \Sigma, \{\Phi_n\}\}$ are the parameters and $p(I_n|\theta)$ is defined in Eq. (2). In this paper, we use the Expectation Maximization (EM) algorithm to solve Eq. (4). In this context (and with some abuse of notation), the EM algorithm can be derived in the following manner:

$$L(\theta) = \sum_n \log p(I_n|\theta) \geq \mathbb{E}_{q_n(k)} \log p(I_n, C_n = k|\theta) + c, \tag{5}$$

where $q_n(k)$ is any probability distribution, c is a constant that doesn't depend on θ and $\mathbb{E}_q$ denotes expectation with respect to q. The lower bound is a direct consequence of Jensen's inequality. For a fixed θ_0 value, the equality holds if and only if $q_n(k) = p(C_n = k|I_i, \theta_0)$. Let's define a function Q by inserting this $q_n(k)$ into Eq (5). Then, we have: $L(\theta) \geq Q(\theta; \theta_0)$, and $L(\theta_0) = Q(\theta_0; \theta_0)$. The EM algorithm can be viewed as iteratively maximizing this lower bound. Let $\theta^{(i)}$ be the guess of θ at the (i)th iteration. Computing $Q(\theta; \theta^{(i)})$ is the E-step of the $(i+1)$th iteration. The M-step updates θ to maximize $Q(\theta; \theta^{(i)})$.

[2] An extension of our model would be to compute a separate variance image for each cluster. In our experiments, the effect of this on the final result was minimal. So, to save computational resources, we opted to have a common variance image, Σ.

2.1 E-step:

In the E-step, the algorithm computes the posterior cluster membership probabilities $\tilde{p}_n^{(i)}(k) \triangleq p(C_n = k|I_n, \theta^{(i)})$ for each image given the model parameters from the previous iteration:

$$\tilde{p}_n^{(i)}(k) \propto \pi_k^{(i)} \prod_{\boldsymbol{x}\in\Omega} \mathcal{N}(I_n(\Phi_n^{(i)}(\boldsymbol{x})); P_k^{(i)}(\boldsymbol{x}), \Sigma(\boldsymbol{x})) \tag{6}$$

and $\sum_k \tilde{p}_n^{(i)}(k) = 1$ for all i. These membership probabilities can be seen as "fuzzy membership", where $\tilde{p}_n(k) = 1$ for some k would indicate a "hard membership" in cluster k.

2.2 M-step:

In the M-step, the algorithm updates the model parameters, $\theta = \{\{P_k\}, \{\pi_k\}, \Sigma, \{\Phi_n\}\}$, to maximize the expected log-likelihood as defined in Eq. (5). This entails updating all prototype images P_k, the prior probabilities π_k, the variance image Σ and all image transformations Φ_n. For a fixed set of $\{\Phi_n^{(i)}\}$, there are closed form expressions for the first three parameters that maximize the objective. These can be derived by taking the derivative of the expected log-likelihood (with the Lagrange multiplier for $\sum_k \pi_k = 1$) and equating it to zero:

$$P_k^{(i+1)}(\boldsymbol{x}) = \frac{\sum_n \tilde{p}_n^{(i)}(k) I_n(\Phi_n^{(i)}(\boldsymbol{x}))}{\sum_n \tilde{p}_n^{(i)}(k)} \tag{7}$$

$$\Sigma^{(i+1)}(\boldsymbol{x}) = \frac{1}{N} \sum_{n,k} \tilde{p}_n^{(i)}(k)(I_n(\Phi_n^{(i)}(\boldsymbol{x})) - P_k^{(i+1)}(\boldsymbol{x}))^2 \tag{8}$$

$$\pi_k^{(i+1)} = \frac{\sum_n \tilde{p}_n^{(i)}(k)}{\sum_{n,k} \tilde{p}_n^{(i)}(k)}. \tag{9}$$

Given these updated model parameters, the new transformations can be computed by optimizing:

$$\Phi_n^{(i+1)} = \arg\min_{\Phi} \sum_{\boldsymbol{x}\in\Omega} \frac{(I_n(\Phi(\boldsymbol{x})) - \bar{P}_n^{(i)}(\boldsymbol{x}))^2}{\Sigma^{(i+1)}(\boldsymbol{x})}, \tag{10}$$

where $\bar{P}_n^{(i+1)} = \sum_k \tilde{p}_n^{(i)}(k) P_k^{(i+1)}$ is the "effective prototype" (i.e., target image in registration) for image I_n at iteration i. This is simply a weighted average of the current prototypes and the weights are membership probabilities.

Equations (7,8,9), and (10) implement a generalized EM, where the optimization is done with coordinate-ascent and the convergence of the log-likelihood to a local optimum is guaranteed. We solve (10) using an iterative gradient-descent type optimizer, an affine transformation model and a multi-resolution pyramid strategy. In contrast to traditional approaches, each image is registered to a different target image: a unique average of the current prototype images, where the averaging is done in a weighted fashion and the weights are corresponding membership probabilities. Note that the registration of each image

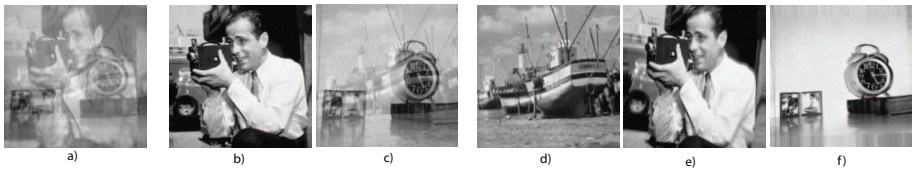

Fig. 1. Synthetic 2D results. Prototype images for three different numbers of clusters: 1 (a), 2 (b,c), and 3 (d,e,f).

can be done in parallel. As discussed in other group-wise registration papers, e.g. [12, 16], we need to anchor the registration parameters to avoid global drifts across subjects. A suitable constraint is that each point in the atlas coordinate frame lies at the average location of corresponding points in the subjects, i.e., $\frac{1}{N}\sum_n \boldsymbol{\Phi}_n^{(i)}(\boldsymbol{x}) = \boldsymbol{x}, \forall \boldsymbol{x} \in \Omega \subset \mathbb{R}^3$ at each iteration i. With the affine transformation, this constraint can be satisfied by applying the inverse of the average affine transformation to all subject transformations after the M-step.

3 Empirical Results

3.1 2D Simulations

As a proof of concept, we implemented and tested our algorithm on a group of synthetic 2D images of size 256×256 and a 9-parameter affine transformation model. Three original images were used to generate a group of 34 images simulating the process described in Section 2. The number of images from each prototype, the affine transformation parameters[3] and the added Gaussian noise[4] were randomly generated. Figure 1 shows the prototype images for three values of k: $1, 2$, and the correct choice of 3. For the last choice, the algorithm could separate out the three original images used to generate the whole group.

3.2 3D MR data

We used MR brain images of 16 patients with first episode schizophrenia and 17 healthy subjects to compute atlases. Because first episode patients are relatively free of confounds such as the long-term effects of medication, there are only subtle structural differences between the two groups, which are difficult to identify by looking at individual scans. A detailed description of the data and related findings are reported in [8]. The images also contained manual labels of medial temporal lobe structures: the Superior Temporal Gyrus, Hippocampus, Amygdala and Parahippocampal Gyrus. We used these segmentation labels and the group membership information to explore the proposed clustering approach. As input, we provided our algorithm with the 33 MR volumes (with no membership or label information). To normalize for global differences in scale and orientation, we ran the algorithm with $k = 1$ and the registration parameters from this step was used to initialize the algorithm with $k = 2$. Figure 2 shows slices from the two output prototype volumes. The first prototype is formed by

[3] All zero mean, standard dev: translations 10 pixels, rotation 0.3 radian, log scale 0.1
[4] zero mean, 0.1*(max intensity value) standard dev.

17 brains, 11 of which are healthy controls. The second prototype is an average of 16 subjects, 10 of which are schizophrenics. We call these atlases C1 and C2, respectively. As benchmarks we computed two types of atlases: we employed the same algorithm with $k = 1$ on healthy-controls (CON) and schizophrenic patients (SZ) separately to compute two atlases (CON and SZ), and all subjects together to compute one atlas (POOL) for the whole population.

Next, we used the manual labels to explore the alignment of the ROI's across the subjects. To quantify this, we used a label entropy measure, defined as: $E = -\sum_{\boldsymbol{x}} \sum_l f(\boldsymbol{x}; l) \log f(\boldsymbol{x}; l)$, where $f(\boldsymbol{x}; l)$ denotes the frequency (or, prior probability) of structure l at location $\boldsymbol{x}$ in the common atlas space. A small entropy is an indication of *overall* good label alignment and a *sharper* label frequency image. If we have two atlases for a population, the combined label entropy for that population can be computed as a weighted sum: $\pi_1 E_1 + \pi_2 E_2$, where π_i and E_i are the prior probability and marginal entropy of atlas i. For CON and SZ, the prior probabilities were computed as: 17/33 and 16/33. Table 1 lists the label entropy measures. Based on these results we conclude that, in the individual CON and SZ atlases, the label maps are aligned much better than in the POOL atlas space. The proposed clustering algorithm also yields two atlases where label maps are significantly better aligned than POOL.

A sharper label frequency image suggests a better localization of the underlying structure for a new subject [15]. To test the predictive power of the atlases, we computed a pairwise Dice measure [4] between each subject and the corresponding 50% probability volume[5] for the three different approaches. For POOL, we registered all subjects to an atlas computed with all other subjects. For CON/SZ, the ground truth membership was used to determine the atlas the new subject was registered with. A schizophrenic, for example, was registered into SZ space. The Dice overlap was computed with the 50% probability volume of all other schizophrenics in SZ space. For C1 and C2, we employed Equations (6) and (10) to iteratively compute the cluster membership and register with the effective prototype. The final membership probability was used to assign the new subject to a cluster. All other subjects assigned to that cluster were used to compute the 50% probability volume. Label-specific dice values were averaged over all labels to get one measurement per subject. Figure 3 shows a box-plot of these values for the three atlas spaces. CON/SZ and C1/C2 achieve significantly better predictive power than POOL (one-sided t-test, $p < 0.05$).

These results suggest that the schizophrenics vs. controls partitioning of the data set captures the dominant anatomical variability in structures we have manual labels for. This is not surprising, given the involvement of these ROI's in schizophrenia development. Note that the sharper frequency images obtained using a multiple atlas strategy can be a consequence of having a smaller number of subjects in each atlas. However, the improved overlap for the new subjects, which were excluded in the atlas, is encouraging and supports the usefulness of the proposed approach for segmentation.

[5] For each label, the 50% probability volume is the region in the corresponding atlas where that label occurred in at least half of the subjects, excluding the test subject.

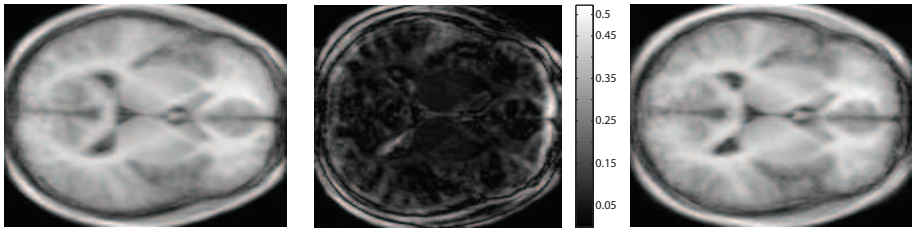

Fig. 2. Axial slices of two prototypes (healthy-dominant, left, and schizophrenic dominant, right) and the absolute difference image (middle).

Table 1. Label Entropies

Atlas	Pooled	Healthy Controls	Schizophrenics	C1	C2
Marginal Entropy	45,814	45,239	42,045	44,842	43,367
Combined Entropy	45,814	43,987		44,139	

4 Discussion

The proposed algorithm is a generalization of well-known unbiased atlas registration algorithms: $k = 1$ is equivalent to the implementations that repeatedly register to a dynamic mean image, e.g. [9]; $k = N$ corresponds to keeping all subjects as atlases, similar to [7]; and the entropy-based group-wise registration approach of [16] can be viewed as a non-parametric version of the proposed algorithm. A novelty of our framework is that for a new subject it computes membership probabilities, which, for example, can be used as weights for a decision fusion-type analysis where inferences from each atlas are combined, e.g. [7].

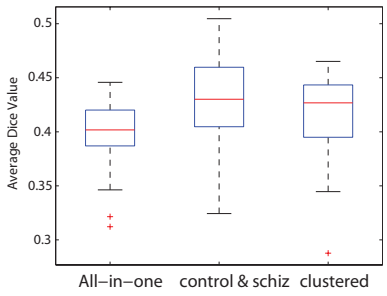

Fig. 3. Predictive power of the three atlases: For each subject, the average dice with the 50-percent atlas volume was computed. Blue boxes indicate the lower and upper quartiles, red lines are the medians. The lines extend to 1.5 times the inter-quartile spacing. Data points outside of the lines are outliers.

The proposed framework can be extended in various ways. The EM method yields an algorithm where the cluster assignments are "soft". An alternative approach can be to perform "hard clustering" at each iteration. Additionally, one can employ a richer, nonlinear transformation model, with a prior on the transformations. Also, more liberal image-to-image distance metrics, such as Mutual Information, can be motivated using more flexible models than a simple additive Gaussian. This should produce better results in cases where inter-image intensity variations are significant.

Acknowledgments

Support for this research was provided in part by the following grants and programs: NSF CAREER 0642971, NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NINDS R01-NS051826, NIH NCRR mBIRN U24-RR021382, NSF JHU ERC CISST.

References

1. A. Allassonniere, Y. Amit, and A. Trouve. Towards a coherent statistical framework for dense deformable template estimation. *J.R. Statist. Soc.B*, 69:3–29, 2007.
2. D. Blezek and J. Miller. Atlas stratification. *MICCAI*, pages 712–719, 2006.
3. D. Collins et al. Automatic 3d intersubject registration of mr volumetric data in standardized talariach space. *J Comp Assist Tom*, 18(2):192–205, 1994.
4. L. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
5. B. Fischl et al. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14:11–22, 2004.
6. A. Guimond, F. Meunier, and J. Thirion. Average brain models: A convergence study. *Technicel Report 3731, INRIA*, 1999.
7. R. Heckemann et al. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.
8. Y. Hirayasu et al. Lower left temporal lobe mri volumes in patients with first-episode schizophrenia compared with psychotic patients with first-episode affective disorder and normal subjects. *Am J Psychiatry*, 155(10):1384–1391, 1998.
9. S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeormophism atlas construction for computational anatomy. *NeuroImage*, 23:151–160, 2004.
10. P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi. Multi-modal image set registration and atlas formation. *Medical Image Analysis*, 10:440–451, 2006.
11. H. Park, P. Bland, A. Hero, and C. Meyer. Least biased target selection in probabilistic atlas construction. *Proc of MICCAI'05*, 3750:419–426, 2005.
12. C. Studholme and V. Cardenas. A template free approach to volumetric spatial normalization of brain anatomy. 25:1191–1202, 2004.
13. A. Tsai et al. An em algorithm for shape classification based on level sets. *Medical Image Analysis*, 9:491–502, 2005.
14. C. Twining et al. A unified information-theoretic approach to groupwise non-rigid registration and model building. *Proc IPMI'05*, pages 1–14, 2005.
15. B. Yeo et al. What data to co-register for computing atlases. *MMBIA*, 2007.
16. L. Zöllei et al. Efficient population registration of 3d data. *Proc of ICCV*, 2005.