Finding Seeds for Segmentation Using Statistical Fusion

Fangxu Xing[*a], Andrew J. Asman[b], Jerry L. Prince[a,c], Bennett A. Landman[b,c,d]

[a]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA 21218
[b]Department of Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235
[c]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA 21218
[d]Department of Radiology and Radiological Sciences, Vanderbilt University, Nashville, TN, USA 37235

ABSTRACT

Image labeling is an essential step for quantitative analysis of medical images. Many image labeling algorithms require seed identification in order to initialize segmentation algorithms such as region growing, graph cuts, and the random walker. Seeds are usually placed manually by human raters, which makes these algorithms semi-automatic and can be prohibitive for very large datasets. In this paper an automatic algorithm for placing seeds using multi-atlas registration and statistical fusion is proposed. Atlases containing the centers of mass of a collection of neuroanatomical objects are deformably registered in a training set to determine where these centers of mass go after labels transformed by registration. The biases of these transformations are determined and incorporated in a continuous form of Simultaneous Truth And Performance Level Estimation (STAPLE) fusion, thereby improving the estimates (on average) over a single registration strategy that does not incorporate bias or fusion. We evaluate this technique using real 3D brain MR image atlases and demonstrate its efficacy on correcting the data bias and reducing the fusion error.

Keywords: Labeling, seed, continuous, STAPLE, atlas, statistics, fusion, bias, registration

1. INTRODUCTION

Image labeling is an important step in automated analysis of medical images. Several segmentation algorithms that can be used for labeling, such as region growing, the random walker[1], and graph cuts[2], require the identification of one or more initial seeds. Typically, seeds are manually picked by human raters, which means that these algorithms are not fully-automatic, and the burden of providing human raters to pick seeds on a very large database of images can become prohibitive due to either time or cost reasons. In some cases, multiple raters may be desirable for robustness in face of human error, and this becomes even more costly. It has become increasingly common to use human raters to establish an initial atlas of validated training data from which a fully-automated, multi-atlas algorithm is established to mimic the human raters to carry out the desired task on an otherwise prohibitively large dataset. The statistical fusion of object labels carried over from multiple registrations of a set of labeled atlases has been well explored and is being actively investigated and improved[3, 4, 5, 6]. In this paper, we explore the statistical fusion of seeds that are intended to initialize a segmentation algorithm that would run in the native subject space. At this stage of development, we consider seeds that are defined as object *centers of mass*, but the basic technique we present can be easily modified to estimate seeds that are optimized for a particular segmentation algorithm or to establish multiple seeds within each object.

2. METHODS

Suppose an atlas of labeled objects in a standardized space (such as affine-registered MNI space) is available. Then a subject image is registered into this standardized space and the centers of mass on each of its objects are desired. A "standard" approach might be to register all objects into the subject space, fuse the labels using (for example) classic discrete Simultaneous Truth And Performance Level Estimation (STAPLE) algorithm[3], and then recompute the centers of mass. In this work, we are looking for a general approach in which centers of mass that are known in the atlas spaces might be directly fused after registration without requiring a recomputation based on an estimated label fusion.

Therefore, first we deformably register[7] each atlas from the training set into the subject, establishing a collection of estimates of centers of mass for each object. Before these data can be fused, in correspondence to our previously

---

* fxing1@jhu.edu; http://iacl.ece.jhu.edu; Image Analysis and Communications Laboratory, Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA 21218

reported method called MAP-CSTAPLE[6] where both the continuous STAPLE algorithm (CSTAPLE)[4, 5] and prior information on rater (training atlas) performance (bias and variance) are needed, we still need to establish a prior model of rater performance. So now we only consider the training set having the centers of mass of all labeled objects. Every atlas in it is deformably registered to all other atlases and the transformed centers of mass are compared to the true positions, generating a set of error vectors. Some atlases will perform well on some datasets and not so well on others, and in general each object within each atlas will have an *a priori* performance[8] characterized statistically by its bias and variance. With this prior information also incorporated, continuous data fusion is established by MAP-CSTAPLE resulting in reduced error due to the expected bias and variance of each training atlas taken into account.

To explain mathematically, let the subject atlas be denoted as A_t, and R training atlases are denoted as $\{A_1, A_2, \dots, A_R\}$ with corresponding labels $\{I_1(\boldsymbol{x}), I_2(\boldsymbol{x}), \dots, I_R(\boldsymbol{x})\}$. Each label set contains same number of object labels, i.e. the jth label of atlas A_m is donated with indicator function $I_m^j(\boldsymbol{x})$ where it is 1 when coordinate $\boldsymbol{x}$ belongs to the label and 0 elsewhere. The center of mass of this object label is define as

$$\boldsymbol{c}_m^j = \frac{\int \boldsymbol{x} \cdot I_m^j(\boldsymbol{x})d\boldsymbol{x}}{\int I_m^j(\boldsymbol{x})d\boldsymbol{x}} \tag{1}$$

Our approach has the following steps, as illustrated using a flowchart in Figure 1.

1. Register every training atlas to the subject atlas. Deformable registration is required to properly align different atlases. In our experiments we apply FNIRT algorithm[7] for the registration. The deformation fields from atlas A_m to A_t are calculated and stored as $\{\varphi_{1t}, \varphi_{2t}, \dots, \varphi_{Rt}\}$.

2. Deform the labels of all training atlases with the deformation field, which will result in R registered label sets in the subject space $\{I_{1t}(\boldsymbol{x}), I_{2t}(\boldsymbol{x}), \dots, I_{Rt}(\boldsymbol{x})\}$, where

$$I_{mt}(\boldsymbol{x}) = I_m(\varphi_{mt}^{-1}(\boldsymbol{x})) \tag{2}$$

3. For every label j, calculate the centers of mass $\{\boldsymbol{c}_{1t}^j, \boldsymbol{c}_{2t}^j, \dots, \boldsymbol{c}_{Rt}^j\}$ after labels are transformed by registration, i.e., the centers of mass of $\{I_{1t}(\boldsymbol{x}), I_{2t}(\boldsymbol{x}), \dots, I_{Rt}(\boldsymbol{x})\}$, by

$$\boldsymbol{c}_{mt}^j = \frac{\int \boldsymbol{x} \cdot I_{mt}^j(\boldsymbol{x})d\boldsymbol{x}}{\int I_{mt}^j(\boldsymbol{x})d\boldsymbol{x}} \tag{3}$$

4. The bias of a center of mass must be learned as prior information to be applied within MAP-CSTAPLE. Both forward and backward registrations are performed on every pair of the training atlases, so that we know the location $\boldsymbol{c}_m^j$ of every center of mass j in its own space A_m as well as the location $\boldsymbol{c}_{mn}^j$ in the space of any other training atlas A_n. Under the assumption that the biases of all atlases lie in the same standardized space after registration, the bias $\boldsymbol{b}_m^j$ of center of mass j is computed by the equation below.

$$\boldsymbol{b}_m^j = \frac{1}{R-1} \sum_{n \neq m} (\boldsymbol{c}_{mn}^j - \boldsymbol{c}_n^j) \tag{4}$$

5. Now we can apply the MAP-CSTAPLE algorithm as described in our previous report. In contrast to STAPLE, which estimates a fusion of discrete labels as well as rater performance parameters (sensitivity and specificity) simultaneously, CSTAPLE deals with multi-dimensional continuous variables. The labeled point location is usually written in the form of vector $\boldsymbol{d}_m^j$, which is the decision of rater m for point j. A fusion of these decisions denoted as $\boldsymbol{t}^j$ is to be found for each point j. And also for each rater m, the performance parameter, i.e., bias $\boldsymbol{\mu}_m$, and variance $\boldsymbol{\Sigma}_m$ is to be estimated simultaneously. Mathematically, it can be proved that under the same maximum likelihood estimation framework as discrete STAPLE, CSTAPLE yields equal likelihood for arbitrary rater bias parameter $\boldsymbol{\mu}_m$ so that bias is indeterminate[6]. If a rater has a systematic bias, the resulting fusion may not be accurate if this bias is not accounted. Therefore we previously proposed the MAP-CSTAPLE algorithm which utilizes prior information to restrict the rater bias and guide the iterations to converge to a proper bias estimate through a maximum *a posteriori* estimation framework. The proposed EM iteration[9, 10] equations are

E-Step:

$$
\begin{cases}
\boldsymbol{A}^{(n)} = \left(\sum_{\mathrm{m}} \boldsymbol{\Sigma}_m^{-1(n)} \right)^{-1} \\
\boldsymbol{t}^{j(n)} = \boldsymbol{A}^{(n)} \cdot \sum_{m} \boldsymbol{\Sigma}_m^{-1(n)} \left(\boldsymbol{d}_m^j - \boldsymbol{\mu}_m^{(n)} \right)
\end{cases}
\tag{5}
$$

M-Step:

$$
\begin{cases}
\boldsymbol{\mu}_m^{(n+1)} = \left(\boldsymbol{I} + \dfrac{\boldsymbol{\Sigma}_m^{(n+1)}}{J\sigma_{\mu_m}^2} \right)^{-1} \left(\dfrac{1}{J} \sum_{j} \left(\boldsymbol{d}_m^j - \boldsymbol{t}^{j(n)} \right) + \dfrac{\boldsymbol{\Sigma}_m^{(n+1)}}{J\sigma_{\mu_m}^2} \boldsymbol{\mu}_{\mu_m} \right) \\
\boldsymbol{\Sigma}_m^{(n+1)} = \dfrac{1}{J} \sum_{j} \left[\boldsymbol{A}^{(n)} + \left(\boldsymbol{d}_m^j - \boldsymbol{\mu}_m^{(n+1)} - \boldsymbol{t}^{j(n)} \right) \left(\boldsymbol{d}_m^j - \boldsymbol{\mu}_m^{(n+1)} - \boldsymbol{t}^{j(n)} \right)^T \right]
\end{cases}
\tag{6}
$$

where $\boldsymbol{\mu}_{\mu_m}$ and $\sigma_{\mu_m}^2$ are the prior mean and prior variance of bias $\boldsymbol{\mu}_m$, and J is the total number of points. Notice that in this method, the bias of a specific rater is assumed to be unchanged at different points. Comparison against discrete STAPLE shows that this is not always true and might introduce a systematic deviation. Also in our experiments with the real brain label centroid dataset, we have observed that the same rater (registered atlas) has different biases at different label points. Therefore we modified the algorithm to consider spatially-varying bias by assuming different biases at different points, i.e., mathematically, change $\boldsymbol{\mu}_m$ to $\boldsymbol{\mu}_m^j$. And the new EM iteration equations are

E-Step:

$$
\begin{cases}
\boldsymbol{A}^{(n)} = \left(\sum_{\mathrm{m}} \boldsymbol{\Sigma}_m^{-1(n)} \right)^{-1} \\
\boldsymbol{t}^{j(n)} = \boldsymbol{A}^{(n)} \cdot \sum_{m} \boldsymbol{\Sigma}_m^{-1(n)} \left(\boldsymbol{d}_m^j - \boldsymbol{\mu}_m^{j(n)} \right)
\end{cases}
\tag{7}
$$

M-Step:

$$
\begin{cases}
\boldsymbol{\mu}_m^{j(n+1)} = \left(\boldsymbol{I} + \dfrac{\boldsymbol{\Sigma}_m^{(n+1)}}{(\sigma_{\mu_m}^j)^2} \right)^{-1} \left(\boldsymbol{d}_m^j - \boldsymbol{t}^{j(n)} + \dfrac{\boldsymbol{\Sigma}_m^{(n+1)}}{(\sigma_{\mu_m}^j)^2} \boldsymbol{\mu}_{\mu_m}^j \right) \\
\boldsymbol{\Sigma}_m^{(n+1)} = \dfrac{1}{J} \sum_{j} \left[\boldsymbol{A}^{(n)} + \left(\boldsymbol{d}_m^j - \boldsymbol{\mu}_m^{(n+1)} - \boldsymbol{t}^{j(n)} \right) \left(\boldsymbol{d}_m^j - \boldsymbol{\mu}_m^{(n+1)} - \boldsymbol{t}^{j(n)} \right)^T \right]
\end{cases}
\tag{8}
$$

Note that the new bias updating equation avoids the averaging through all the points and now treat each point differently with a specific bias, even for the same rater. Using symbols from previous steps, for any target atlas t, we substitute decision $\boldsymbol{d}_m^j$ with $\boldsymbol{c}_{mt}^j$ and bias prior mean $\boldsymbol{\mu}_{\mu_m}^j$ with $\boldsymbol{b}_m^j$. Also for bias prior variance $(\sigma_{\mu_m}^j)^2$ we use the scalar variance of $(\boldsymbol{c}_{mn}^j - \boldsymbol{c}_n^j)$ over all targets n. After equations (7) and (8) converge[11], the estimated fusion $\boldsymbol{t}^{j(\infty)}$ can be used as seed for label point j.

3. RESULTS

The dataset of 148 complete 3D real brain atlases were evaluated in the experiment. On each atlas, 38 labels were identified and validated which segmented the entire brain volume. In correspondence with our model, for any target atlas, the total number of points J was 38. First, 147 of the atlases were used as training dataset to get the bias prior for raters in order to evaluate the one rest subject atlas. We explored the fusion of different rater numbers from 1 to 147. For any fixed rater number, the raters were randomly drawn from the training dataset to be used to calculate the fusion. Every one of the atlases was regarded as the subject atlas with the rest as training. The entire setup was repeated in 50 Monte-Carlos with random raters and random subject atlas. We then computed the mean squared error (MSE) of fusion from the truth for every subject atlas, and got an average MSE of 1.9613 ± 1.0565 voxel^2. At the same time, we also did the evaluation on each of the same dataset with the fusion technique replaced with normal averaging and MAP-CSTAPLE with zero prior, and got an average MSE of 2.2881 ± 1.0230 voxel^2 and 2.3126 ± 1.1153 voxel^2 respectively. The MSE of one subject atlas is shown in Figure 2(d).

We also used 50 of the atlases as training data and the remaining 98 atlases as testing data. 50 Monte-Carlos were performed with the same setup as in previous experiment. For fixed training and testing datasets, the fusion on every single subject atlas was also evaluated with MAP-CSTAPLE and normal averaging. The average MSE was 1.6734 ± 1.3962 voxel^2 for MAP-CSTAPLE and 2.0130 ± 1.4050 voxel^2 for averaging. Figure 2(c) shows the number of fusion points with respect to distance from the truth. Generally speaking, at a certain distance from the truth (a ball centered at truth with certain radius), MAP-CSTAPLE is more likely to capture points than normal averaging technique.

4. CONCLUSION AND DISCUSSION

In this work we described a new approach to automatically determine seed locations for segmentation algorithms that require initial seeds. The approach was based on the STAPLE framework used for fusion approximation. We designed our model based on a previously proposed method – MAP-CSTAPLE – for continuous vector fusion problems and modified the equations and assumptions to fit the real brain atlas data we are evaluating, i.e., augmented the algorithm to handle the spatially bias-varying case. Experiment results demonstrated improved performance as compared to simple averaging. By using this approach, semi-automatic segmentation algorithms can become fully automatic, and this provides an alternative multi-atlas approach to image labeling over the conventional label fusion approaches.

By directly modifying the bias parameters to spatially vary at different label points, we may have introduced another problem of insufficiency. In particular, since each rater performs only once at each label point, it is difficult to reliably estimate a spatially-varying bias for the rater at that point. In the future, therefore, we will develop and investigate a model that incorporates measurement repetitions.

ACKNOWLEDGEMENTS

REFERENCES

[1] Grady, L., "Random walks for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1768 - 1783 (2006).

[2] Freedman, D., Zhang, T., "Interactive graph cut based segmentation with shape priors," Computer Vision and Pattern Recognition 2005, Vol 1, pp. 755-62 (2005).

[3] Warfield, S. K., Zou, K. H., and Wells, W. M., "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," IEEE Trans Med Imaging, 23(7), 903-21 (2004).

[4] Warfield, S., Zou, K., Wells, W., "Validation of image segmentation by estimating rater bias and variance," Medical Image Computing and Computer-Assisted Intervention, 4190:839–47 (2006).

[5] Commowick, O., and Warfield, S.K., "A Continuous STAPLE for Scalar, Vector, and Tensor Images: An Application to DTI Analysis," IEEE Trans Med Imaging, 28(6), 838-46 (2009).

[6] Xing, F., Soleimanifarda, S., Prince, J. L., Landman, B. A., "Statistical Fusion of Continuous Labels: Identification of Cardiac Landmarks," Proceedings of SPIE, Vol. 7962, 796206 (2011).

[7] Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S. M., "Bayesian analysis of neuroimaging data in FSL," NeuroImage, 45:S173-186, (2009).

[8] Gauvain, J. L., Lee, C. H., "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," IEEE Transactions on Speech and Audio Processing 2, 291–298 (1994).

[9] McLachlan, G. J., Krishnan, T., "The EM algorithm and extensions," New York: John Wiley and Sons Ltd (1997).

[10] Rohlfing, T., Russakoff, D. B., Maurer, C. R., "Expectation maximization strategies for multi-atlas multi-label segmentation," Proc. Int. Conf. Information Processing in Medical Imaging, pp. 210–221 (2003).

[11] Wu, C. F. J., "On the convergence properties of the EM algorithm," The Annals of Statistics, 11(1):95–103 (1983).
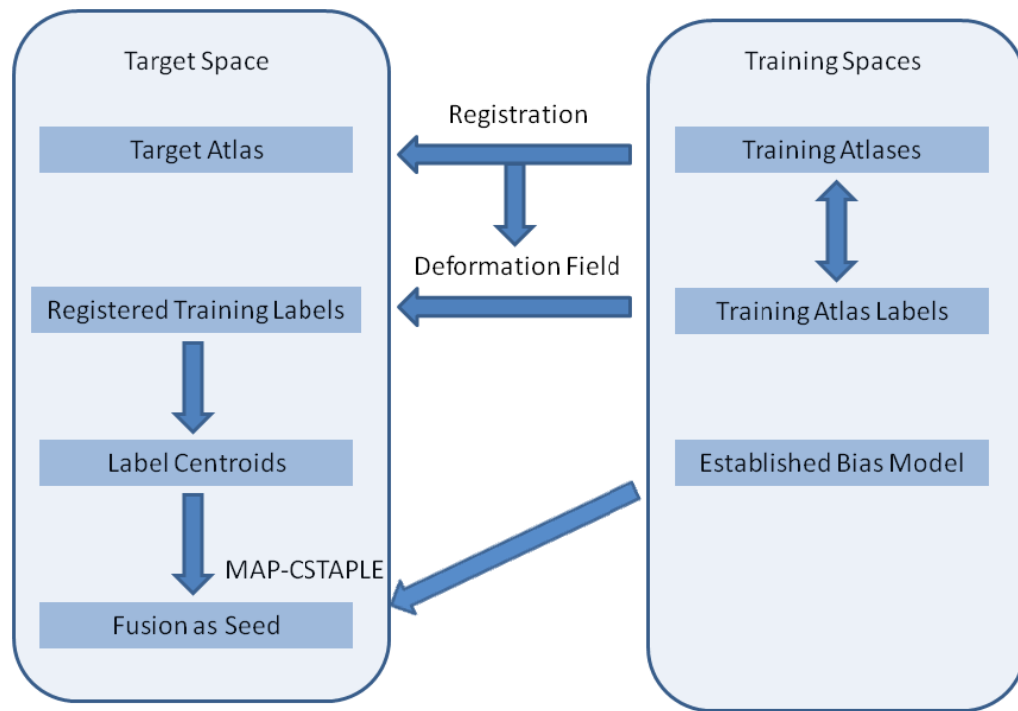
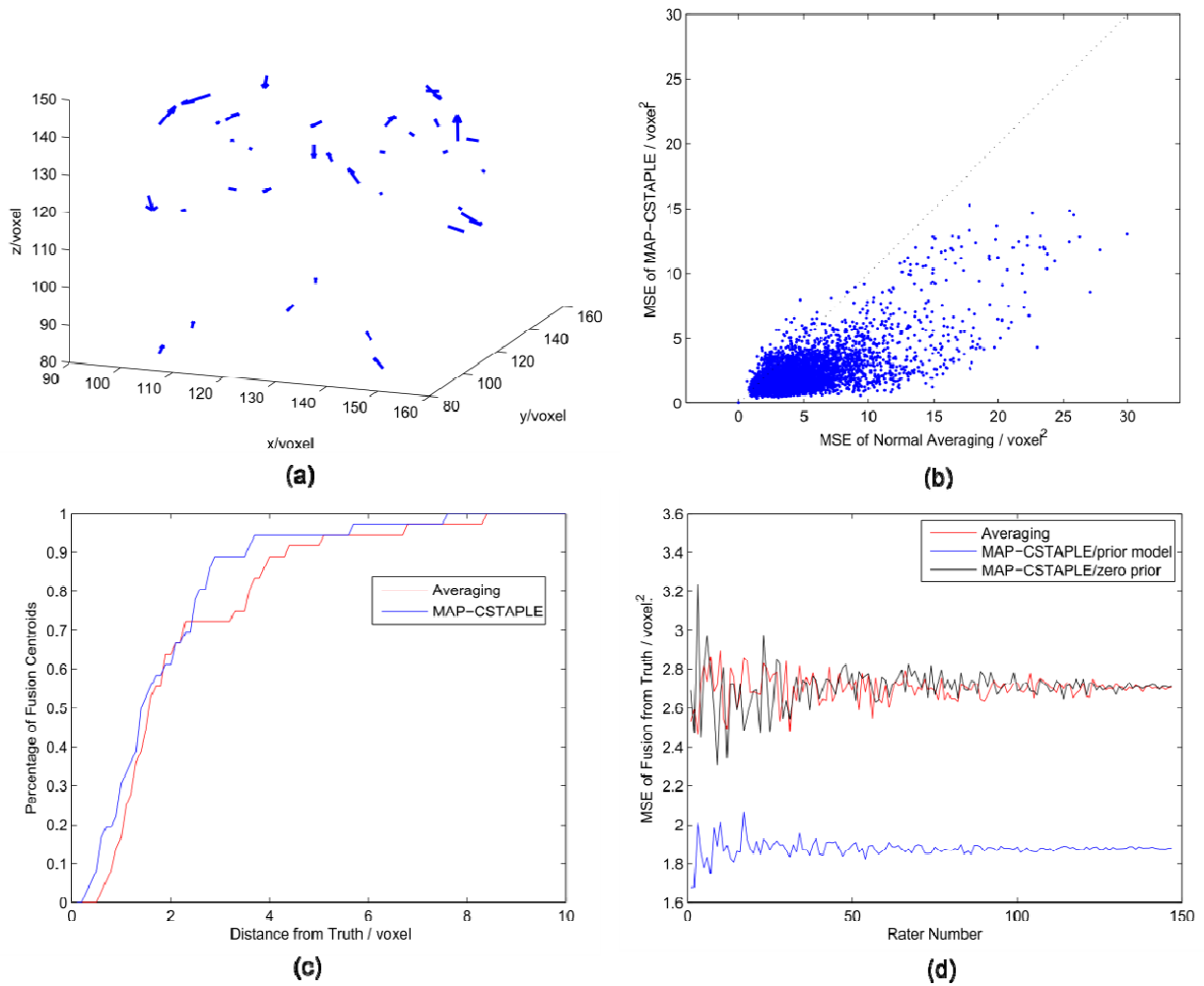Figure 1: A flowchart for the proposed seed identification technique.

Figure 2: (a) The bias model of one training atlas that has a symmetrical spatial pattern. (b) The MSE of every fusion point from the truth, which demonstrates MAP-CSTAPLE has an ability to reduce the fusion errors with the bias prior model. (c) The number of fusion points close to the truth. (d) The MSE of the fusion of one subject with respect to rater numbers. Training data has 147 atlases.