Pictorial Multi-atlas Segmentation of Brain MRI

Cheng-Yi Liu[1], Juan Eugenio Iglesias[1,2], Zhuowen Tu[1]
[1]University of California, Los Angeles, Laboratory of Neuro Imaging
[2]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, MGH, Harvard Medical School
chengyiliu@ucla.edu, iglesias@nmr.mgh.harvard.edu, zhuowen.tu@loni.ucla.edu

Abstract

The use of a single labeled volume ("atlas") is limited in registration-based segmentation because it is hard for one atlas to represent the whole data population, especially if input images observe large variation. Moreover, the choice of volume to label biases the algorithm. Multi-atlas segmentation has emerged as an alternative but it has a similar drawback due to combinatory combinations of different anatomical structures; and in addition, the computation time grows linearly with the number of atlases. In this paper, a pictorial-structure-based approach to achieving both the high performance and the efficiency of registration-based segmentation is proposed. Our method performs segmentation via registering each structure of the atlas in an exemplar-based graphical model. We compared the proposed approach with multi-atlas segmentation and show the advantage of our method in both effectiveness and efficiency.

1. Introduction

Atlas-based registration has been ubiquitous in medical image analysis in the last decade [15, 2, 8]. Given a labeled three-dimensional scan (henceforth an "atlas") and a test volume, the spatial correspondence between the two can be obtained with a registration process. Registration is typically cast as an optimization problem in which a predefined transform is adjusted to maximize a similarity measure. Once the transform has been determined, it can be used to transfer the labels in the atlas to the test scan and obtain its segmentation, which can be used for further clinical analysis e.g. identifying the pathological changes of structures, or comparing the condition of a subject with control cohorts.

In practice, the use of single-atlas segmentation is limited for two reasons: (1) the choice of atlas biases the results; and (2) a single atlas cannot in general represent a whole population, especially on datasets with large variation. Multi-atlas registration (MAR) [12, 7, 11, 13] has been widely used to ameliorate these problems. In MAR, a set of atlases are registered to the target image. By decision rules such as averaging, the final segmentation can be computed from the transferred labels. The main disadvantage of MAR is the heavy computational burden of registering a number of atlases to the test case.

To reduce the computational cost of MAR, several studies have attempted to predict which atlases will yield the most accurate segmentation for each region when registered to the test case. In [15], van Rikxoort et al. proposed an adaptive selection procedure based on the amount of information gained from registering each candidate atlas. Another way of decreasing the processing time of atlas-based registration is carrying out the analysis at a structural (as opposed to global) level. Wu et al. [16] used a criterion based on mutual information to choose the best atlas for each structure.

In this paper, we propose a method inspired by Pictorial Structures [6] for multi-atlas 3D brain image segmentation. Pictorials structures were originally proposed for object recognition; our application differs in model and implementation. In our framework, a collection of anatomical structures are arranged in a deformable configuration such that each structure is modeled separately and the global configuration is represented by spring-like connections between pairs of structures. Each individual structure is modeled by local registration (exemplar-based) and the global configuration is captured through the overlap of the propagated labels. Our algorithm is also related to [4], but is not restricted to just two neighboring structures of interest. The proposed method is evaluated through the segmentation of 12 subcortical brain structures (left and right thalamus proper, L/R caudate, L/R putamen, L/R pallidum, L/R hippocampus, and L/R amygdala) in a publicly available MRI dataset.

2. Method

Registration-based segmentation of brain subcortical structures is intrinsically inefficient because a whole MRI volume is registered when only small parts of the deforma-

tion field will be used. To segment the structures of interest in a target scan efficiently and precisely, our method exploits the advantages of MAR and pictorial structures throughout the process:

- MAR: using a set of atlases helps capture the variability of the population in both appearance and shape. Another advantage is that the output is probabilistic, i.e., each voxel is assigned a probability of belonging to the structure of interest.
- Pictorial structures: the target scan is decomposed into smaller structures. Matching the model to the target is formulated as the optimization of a composite metric that combines individual similarity terms for each structure and a global configuration term.

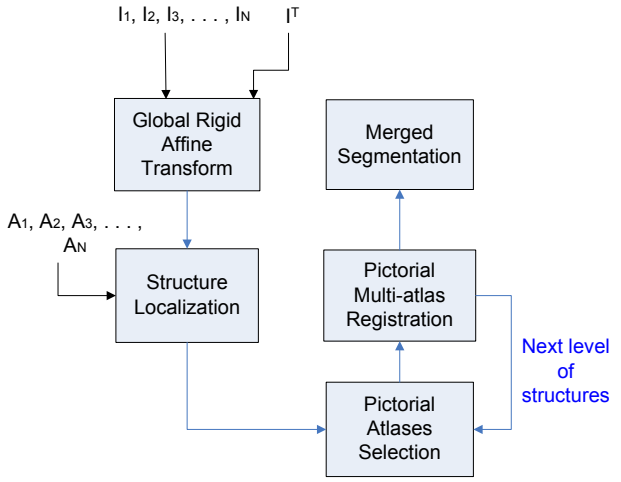

Figure 1. Process of pictorial multi-atlas structural registration. I_i's and A_i's are the intensity and label images from the atlas set. I^T is the target image. After the global image transform, the following processes are performed on structures. All structures are assigned some level numbers to indicate their processing orders, which forms the pictorial atlas selection-registration loop.

Fig. (1) gives an overview of our pictorial multi-atlas segmentation. After a global affine transform, each structure in the target image will be roughly localized. This enables further local atlas selection and registration. The final stage is to integrate all the local MAR results, into the final segmentation in the global image space.

2.1. Pictorial atlas localization and selection

Our algorithm assumes the availability of N atlases with annotations of J structures: $\{\{I_n, A_n\}, n = 1..N\}$, where I_n's and A_n's denote the intensity and label images, respectively. For an atlas image I_n, the sub-image (bounding box) containing the complete jth structure is a pictorial atlas and denoted as I_{nj}. In order to capture the context and reduce the boundary effect, I_{nj} is defined such that its size is slightly larger (4mm extension from the structural boundary) than the structure itself.

Given a target image I^T, a global affine transform is first optimized between I^T and each I_n. The sub-image in I^T corresponding to I_{nj} is then roughly located; we denote it as I^T_{nj}. However, the disparity between the two sub-images is still large and the initialization of the structures can affect the performance of the following registration-based segmentation [16]. Hence, we apply a local affine transform between each I_{nj} and I^T_{nj} to refine the alignment. This local affine transform is driven by jointly minimizing the following energy functions:

$$E_j = \sum_{n=1}^{N} \left[S(I^T_{nj}, a_{nj}(I_{nj})) + \lambda \sum_{i=1}^{N} \|c_{nj} - c_{ij}\| \right] \quad (1)$$

where S is an energy function based on some similarity measure between the target patch and the pictorial atlas, $a_{nj}(\cdot)$ is the objective affine transform, which can be applied to both I_{nj} and A_{nj}, and c_{ij} is the centroid of structure j in $a_{ij}(A_{ij})$ (mapped to the global image space). The second term of Eqn. 1 provides a localization constraint for the N affine-transformed pictorial structures so that they do not drift away from one another. λ is a weight to combine the two terms. The total engergy, $\sum_j E_j$, can be jointly minimized through the coordinate descent algorithm.

It has been shown that increasing the number of atlases in MAR does not always improve (and sometimes even decreases) the segmentation accuracy [2]. Therefore, before applying the local registration, we select a set of best matching local components (structures) in the pictorial atlases for structure j, i.e., R_j atlases among $I_{nj}, n = 1..N$ and $R_j \leq N$. The target measure μ_{nj} for the pictorial atlas selection can be directly made based on Eqn. 1:

$$\mu_{nj} = S(I^T_{nj}, a_{nj}(I_{nj})) + \lambda \sum_{i=1}^{N} \|c_{nj} - c_{ij}\| \quad (2)$$

Applying Eqn. 2 allows us to select the best R_j number of structures from N atlases. Note that these structures may come from different sets of R_j atlases for different structure j. Please note that both Eqn. 1 and. 2 utilize the information from MAR. This reduces the bias from each pictorial atlas at the initialization stage.

2.2. Pictorial multi-atlas registration

The selected pictorial atlases will be refined by transforms more complex than affine at the local (structural) level. The goal is to jointly optimize the following energy function:

$$E = \sum_{j=1}^{J} \sum_{l=1}^{R_j} \left[S(I^T_{lj}, t_{lj}(I_{lj})) - \beta \sum_{k \neq j, k=1}^{J} P(lj, k) \right]. \quad (3)$$

where β is a constant. To simplify the notation, we use image index l to denote the index in the selected pictorial atlases ranging from 1 to R_j. The transform $t_{lj}(\cdot)$ is not restricted to any specific type; it can be a single non-linear transform or a cascade thereof. In our experiments, $t_{lj}(\cdot)$ is a cascade of an affine transform followed by a diffeomorphic transform (SyN [3]). P is a pairwise energy function to model the relation between the different structures in the pictorial structure framework; for example, we can simply define $P(lj,k) \propto \exp\{-||d(lj,k) - \hat{d}(j,k)||\}$, where $d(j,k)$ is the average distance between structure j and k.

Instead of registering all the pictorial structures at once in Eqn. 3, we employ a hierarchical processing approach similar to Wu [16] to decompose the objective function. The structures of interest are divided into several levels; the easier a structure is to segment, the earlier it is processed (low level number). Meanwhile, its less stable neighbors will be assigned larger level numbers. Such information can be obtained from the set of atlases, and the detailed algorithm is shown in Fig. (2) which automatically assigns a level number to each structure. The hierarchical levels of structures form the iterative selection-registration cycle in Fig. (1) which registers one level of structures per iteration. Level zero corresponds to the ventricles, which are full of cerebrospinal fluid (CSF) and therefore dark in the image and easy to segment. The ventricles help guide the segmentation of the surrounding structures[16, 8]. The levels of the remaining structures are listed in Table. 1, and Fig. (3) shows part of the structural connections in our method.

Level 0	Left/Right inferior horn of the lateral ventricle, Left/Right lateral ventricles, 3rd ventricle, 4th ventricle
Level 1	Left/Right caudate nuclei, Left/Right putamens, Left/Right thalami
Level 2	Left/Right pallida, Left/Right hippocampi
Level 3	Left/Right amygdalum

Table 1. Levels of subcortical structures for pictorial segmentation.

At the beginning of each iteration, Eqn. 2 is used to select R_j candidate local atlases in the pictorial model. At the end of each iteration, R_j segmentations corresponding to structure j will be mapped back to the global image space for generating a joint MAR segmentation. The probability of a voxel belonging to structure j is computed by combining the propagated labels and the segmentation s_j^* is obtained by thresholding at 0.5.

The merging of all s_j^*'s in the global image space follows the level ordering. Therefore, structures with a larger level number will be rendered later and have less probability to be occluded by other structures. In addition, thanks to the use of $P(lj,k)$ in Eqn, 3, serious overlapping (i.e., large change of centroid distance) can be avoided in our method.

Given: N atlas images with K labeled structures: $s_1..s_K$.
Initialize: $Struct = \{s_1..s_K\}$, $Assigned = \emptyset$, $L(i) = 0, i = 1..K$.
Determine the set of neighboring structures for each structure $i : N(s_i)$.
Perform cross-registration between each image , i.e., $N \times (N-1)$ whole image registrations. Calculate the average DICE coefficients of the K structures from all the registration results: $\overline{D} = \left\{\overline{d}(i); i = 1..K\right\}$.
Sort the structures in $Struct$ according to $\overline{D}$ in a descending order.
for $i = 1$ to K **do**
  if structure $i \in CSF$ **then**
    $Struct = Struct - \{s_i\}$
    $Assigned = Assigned + \{s_i\}$
  end if
end for
while $Struct \neq \emptyset$ **do**
  Pick the first s_i from $Struct$
  if $N(s_i) \neq \emptyset$ **then**
    $L(i) = 1 + \underset{L(j)}{\text{argmax}}(j; s_j \in (Assigned \cap N(s_i)))$
  else
    $L(i) = 1$
  end if
  $Struct = Struct - \{i\}$
  $Assigned = Assigned + \{i\}$
end while

Figure 2. Structural level assignment algorithm. The CSF structures are always assigned to level 0. The other structures are assigned according to their neighboring structures with higher average DICE coefficients, which can be determined by the N atlases.

3. Experimental Results

3.1. Datasets

The proposed algorithm was tested on the 18 scans from the Internet Brain Segmentation Repository (IBSR) [9], which have 84 manually annotated structures. All the scans were first skull-stripped by BET [14] and then cropped and rotated to 160x192x160 with $1mm$ spacing in each direction, preserving the whole brain. We chose 18 structures as listed in Table. 1, but the ventricles were only used for identifying the CSF regions and not for evaluation.

3.2. Experimental Setup

To implement our method, we used ITK 3.16 [10] to perform the affine optimization in Eqn. 1. Different methods can be used as the underlying pictorial registration algorithm in Eqn. 3 (this will be detailed in Section 3.3). When the number of atlases was large enough ($N >= 10$), we performed the selection in Section 2.1 with all $R_j = 0.8N$ and $\lambda = 0.2$, to exclude inappropriate atlases. For small

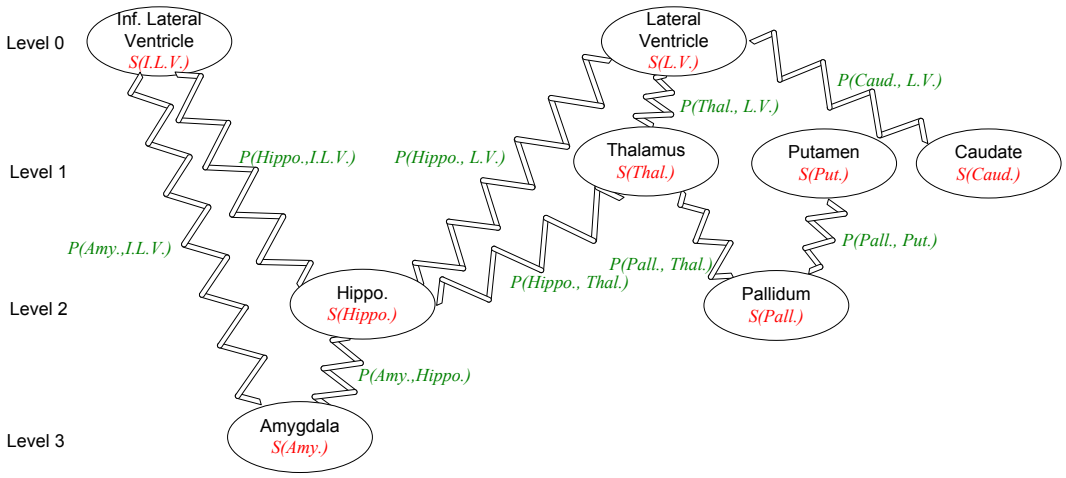

Figure 3. A graph of the pictorial structures in our method. Each node represents a structure, and the spring-like connections maintain the relations between neighboring structures. We omit the 3rd and 4th ventricles from the graph for clarity. Details of the energy terms, $S(\cdot)$ and $P(\cdot)$, are described in Section 2.2.

number of atlases ($N < 10$), all atlases will be included. β in Eqn. 3 was empirically set to 0.2 to balance the appearance similarity and the forces from other structures. In the following comparisons, we use the DICE overlap [5] as performance measure, and PMS stands for "pictorial multi-atlas segmentation." A postfix number after the name of a method indicates the number of atlases included.

3.3. Improvement with Respect to Registration Methods

To show the improvement of our algorithm to typical registration methods, we included two non-linear methods in our comparisons. One is the B-spline nonrigid registration implemented in ITK 3.16 and the other is a diffeomorphic method, SyN, in ANTS 1.9 [3]. SyN compares favorably with other state-of-the-art algorithms [1].

For ITK, the spline order was 3 with 8 grid nodes in each dimension and the metric was Mattes mutual information. For SyN, we used cross correlation (CC) as the similarity metric and the radius of region was 5mm. The more sophisticated metric in SyN, probability mapping (PR), was not chosen because it performed similar to CC in our preliminary experiments but required more computation time. The deformation fields are regularized by Gaussian with sigma=2. Based on these configurations, we performed two tests with different number of iterations: (1) 3-level multi-resolution with 30x50x5 as the number of iterations, which can be executed fast and denoted as MAR_SyN; (2) 5-level multi-resolution with the number of iterations as 10x10x100x100x100, which provides excellent performance in whole image registration and is denoted as MAR_SyN*.

SyN was also the non-linear transform of pictorial registration in our implementation. It was part of the transform t_{lj} in Eqn. 3. The setting of SyN was the same as above and the number of iterations was set as MAR_SyN, 30x50x5, which was sufficient for registering the small structural images.

The test dataset was also IBSR. To avoid biased measures of PMS and MAR, the 18 images were randomly separated into three folds for cross-validation. For each fold, the test set consists of 6 scans, whereas the remaining 12 played the role of atlases. Each measure of the baseline single-atlas registration method is an average of 12 tests.

We list the detailed DICE overlaps of MAR_SyN, MAR_SyN*, and PMS in Table. 3 and their curves in Fig. (4). The computation times of each method respective to the number of atlases were listed in Table. 2. PMS outperforms MAR_SyN which has the same registration parameters. PMS with four atlases (PMS4) was already comparable to MAR_SyN with ten or twelve atlases (MAR_SyN10, MAR_SyN12). However, to achieve the same performance, PMS4 used only 1.4 hours while MAR_SyN10 needed 3.1 hours to process. When both methods used 12 atlases, PMS was significantly better than MAR_SyN on putamens and pallida. Besides, Fig. (4) shows that MAR_SyN required more than five atlases to reach stable results, which implied a fixed cost of 1.6 hours. Table. 2 also shows that based on the setting of 18 structures, PMS and MAR_SyN had similar execution time respective to the same number of atlases.

Number of atlases	3	4	5	6	8	10	12
MAR_SyN	0.9	1.2	1.6	1.9	2.5	3.1	3.7
MAR_SyN*	4.9	6.5	8.2	9.8	13	16.3	19.5
PMS	1.0	1.4	1.7	2.0	2.7	3.1	3.7

Table 2. Computation time of MAR_SyN, MAR_SyN*, and PMS for different number of atlases (in hours). Both the atlases and the target images were 160x192x160 voxels. The test machine was Intel® Xeon® 2.80GHz with 16GB RAM, and the code was single threaded.

Table 3. (a) DICE overlaps of fast SyN-based multi-atlas (MAR_SyN), precise SyN-based multi-atlas (MAR_SyN*), and our method (PMS) on 18 IBSR images, which are shown in three blocks. A postfix number after the name of each method indicates the number of atlases included. The values are the averages of the left and right structures. In each block, the improvement due to the increased number of atlases for a single method is observed. Across the blocks, comparisons between these methods using different numbers of atlases can be made. (b) p-values between MAR_SyN, MAR_SyN*, and PMS, for #atlases=3 and 12.

Method & #Atlases	**Thalamus**	**Caudate**	**Putamen**	**Pallidum**	**Hippocampus**	**Amygdala**	**Average**
MAR_SyN3	0.847	0.701	0.792	0.674	0.669	0.573	0.709
MAR_SyN4	0.835	0.694	0.771	0.657	0.670	0.568	0.699
MAR_SyN5	0.857	0.737	0.812	0.716	0.714	0.650	0.748
MAR_SyN6	0.863	0.762	0.814	0.720	0.729	0.661	0.758
MAR_SyN8	0.872	0.788	0.839	0.757	0.751	0.692	0.783
MAR_SyN10	0.875	0.796	0.852	0.766	0.770	0.709	0.794
MAR_SyN12	0.874	0.793	0.849	0.759	0.762	0.703	0.790
MAR_SyN3*	0.862	0.741	0.846	0.764	0.752	0.670	0.772
MAR_SyN4*	0.865	0.733	0.840	0.761	0.749	0.676	0.770
MAR_SyN5*	0.874	0.779	0.857	0.787	0.769	0.710	0.796
MAR_SyN6*	0.873	0.772	0.853	0.780	0.768	0.715	0.793
MAR_SyN8*	0.876	0.788	0.868	0.789	0.773	0.714	0.799
MAR_SyN10*	0.879	0.792	0.863	0.790	0.783	0.721	0.804
MAR_SyN12*	0.882	0.795	0.866	0.795	0.786	0.723	0.808
PMS3	0.867	0.795	0.850	0.748	0.716	0.687	0.777
PMS4	0.872	0.796	0.848	0.782	0.744	0.714	0.794
PMS5	0.879	0.798	0.857	0.781	0.759	0.719	0.799
PMS6	0.880	0.810	0.859	0.793	0.771	0.730	0.807
PMS8	0.884	0.818	0.860	0.790	0.771	0.720	0.808
PMS10	0.882	0.816	0.860	0.793	0.776	0.721	0.808
PMS12	0.884	0.817	0.863	0.797	0.778	0.724	0.810

(a) DICE overlaps

Label Name	**Thalamus**	**Caudate**	**Putamen**	**Pallidum**	**Hippocampus**	**Amygdala**
MAR_SyN3 v.s. MAR_SyN3*	0.01849	0.06631	0.00170	0.00171	0.00241	0.00133
MAR_SyN3 v.s. PMS3	0.00497	0.00010	0.00081	0.00700	0.04815	0.00016
MAR_SyN3* v.s. PMS3	0.21768	0.00723	0.04197	0.05428	0.00017	0.16101
MAR_SyN12 v.s. MAR_SyN12*	0.10777	0.45051	0.00694	0.01401	0.02534	0.17580
MAR_SyN12 v.s. PMS12	0.06224	0.06457	0.01915	0.00871	0.08770	0.16412
MAR_SyN12* v.s. PMS12	0.37644	0.09107	0.16478	0.37520	0.20543	0.48401

(b) One-sided p-values between MAR_SyN, MAR_SyN*, and PMS, with #atlases=3 and 12, respective to each structure

MAR_SyN* required longer time to execute and produced better results than MAR_SyN. In Table. 3, if the number of atlases was small ($3 \sim 6$ atlases) the improvement of DICE overlaps by PMS were larger than MAR_SyN*. As the number of atlases reached 12, PMS performed better on the caudate nuclei and worse on the hippocampi than MAR_SyN*. The average DICE of PMS12 ($N = 12$, $R = 10$) was 0.810 and the highest in Table. 3, which was equivalent to MAR_SyN12* (0.808). The one-sided p-values also suggested that PMS3 performed significantly better than MAR_SyN3* on caudate nuclei, putamens, and hippocampi, while MAR_SyN12* and PMS12 had no significant differences. The obvious difference between PMS and MAR_SyN* was their computation time. Table. 2 shows that our method required under four hours to completed a PMS12 process, which was **19%** as MAR_SyN12* and more favored in applications.

We also showed the curves of ITK-based MAR (MAR_ITK) at the right of Fig. (4). Although the MAR_ITK required more execution time than MAR_SyN*, its performance was worse than MAR_SyN* in all structures. Comparing ITK with our method, using four atlases in PMS already had all structures better than ITK12.

4. Conclusions

A method of efficient atlas-based segmentation has been presented in this paper. The method benefits from the unbiased statistics of multi-atlas segmentation and the structural analysis of pictorial structures. No tailored rules were used for specific structures so this framework is general and applicable to other multi-part segmentation problems. These advantages make our method more applicable in practical segmentation scenarios.

As shown in our experiments, the performance of PMS is comparable to multi-atlas registration but its computational burden is much smaller. We directly used the energy term of registration for atlas selection; more sophisticate methods, such as information gain in [15], may further reduce the number of atlases without drop in performance. The computational advantage then can be kept for larger number of structures.

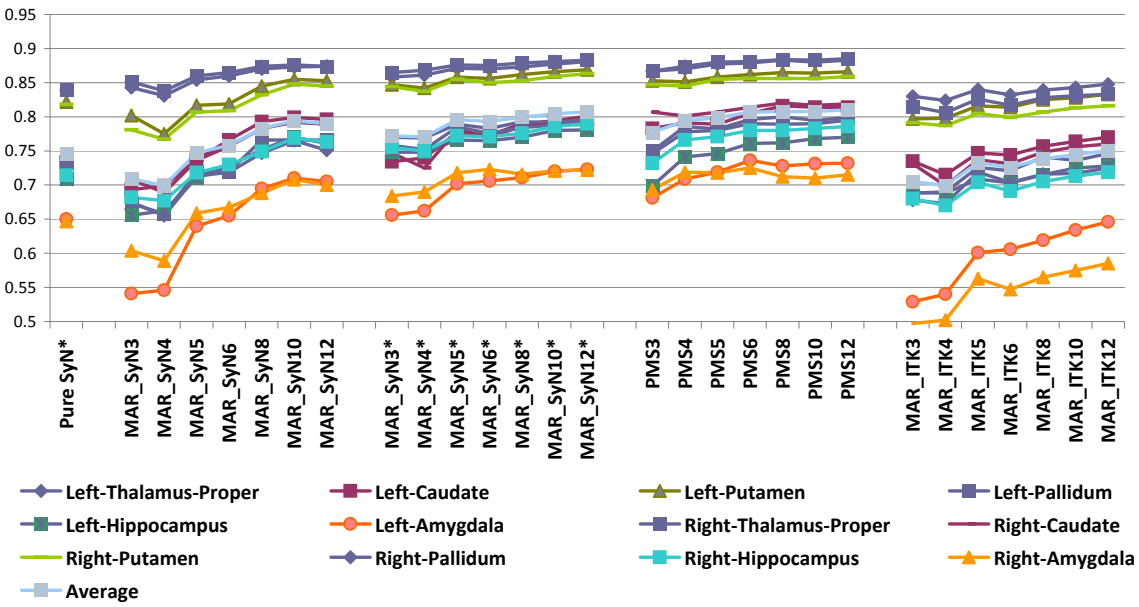

Figure 4. Comparison between fast SyN-based multi-atlas (MAR_SyN), precise SyN-based multi-atlas (MAR_SyN*), PMS, and ITK-based multi-atlas (MAR_ITK). We also show SyN-based single-atlas registration (Pure SyN) as the common registration baseline. Pure SyN can be considered as MAR_SyN1* as it used the same parameters and iterations as MAR_SyN*. The test data were 18 IBSR images.

Acknowledgments

This work was funded by NSF IIS-0844566 and also in part by the Office of Naval Research (ONR) N000140910099.

References

[1] A. Klein et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46:786–802, 2009.

[2] P. Aljabar, R. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Classifier selection strategies for label fusion using large atlas databases. In *Proc. of MICCAI 2007*, volume 4791, pages 523–531, 2007.

[3] B. B. Avants, C. L. Epstein, M. Grossman, and . J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12:26–41, 2008.

[4] M. Chupin, A. Hammers, R. Liu, O. Colliot, J. Burdett, E. Bardinet, J. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation. *NeuroImage*, 46:749–761, 2009.

[5] L. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297–302, 1945.

[6] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *Intl. Journal of Computer Vision*, 61:55–79, 2005.

[7] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, Oct. 2006.

[8] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, A. Hammers, and T. A. D. N. Initiative. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage*, 51:221–227, 2010.

[9] http://www.cma.mgh.harvard.edu/ibsr.

[10] L. Ibanez, W. Schroeder, L. Ng, J. Cates, and the Insight Software Consortium. *The ITK Software Guide Second Edition*. Kitware Inc, 2005.

[11] J. E. Iglesias and N. Karssemeijer. Robust initial detection of landmarks in film-screen mammograms using multiple ffdm atlases. *IEEE Trans. on Medical Imaging*, 11:1815–1824, 2009.

[12] T. Rohlfing, D. B. Russakoff, and C. R. Maurer. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans. on Medical Imaging*, 23(8):983–994, Aug. 2004.

[13] M. Sabuncu, B. Yeo, K. V. Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *IEEE Trans. on Medical Imaging*, 29(10):1714–1729, 2010.

[14] S. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17:143–155, 2002.

[15] E. M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. A. Viergever, J. P. Pluim, and B. van Ginneken. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis*, 14:39–49, 2010.

[16] J. Wu and A. C. Chung. A novel framework for segmentation of deep brain structures based on markov dependence tree. *NeuroImage*, 49:1027–1036, 2009.