Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2017 February ; 10137: . doi:10.1117/12.2256147.

Efficient Multi-Atlas Registration using an Intermediate Template Image

Blake E. Dewey^a, Aaron Carass^{a,b}, Ari M. Blitz^c, and Jerry L. Prince^a

^aDept. of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

^bDept. of Computer Science, The Johns Hopkins University, Baltimore, MD 21218, USA

^cDept. of Radiology and Radiological Sciences, The Johns Hopkins School of Medicine, Baltimore, MD 21287, USA

Abstract

Multi-atlas label fusion is an accurate but time-consuming method of labeling the human brain. Using an intermediate image as a registration target can allow researchers to reduce time constraints by storing the deformations required of the atlas images. In this paper, we investigate the effect of registration through an intermediate template image on multi-atlas label fusion and propose a novel registration technique to counteract the negative effects of through-template registration. We show that overall computation time can be decreased dramatically with minimal impact on final label accuracy and time can be exchanged for improved results in a predictable manner. We see almost complete recovery of Dice similarity over a simple through-template registration using the corrected method and still maintain a 3–4 times speed increase. Further, we evaluate the effectiveness of this method on brains of patients with normal-pressure hydrocephalus, where abnormal brain shape presents labeling difficulties, specifically the ventricular labels. Our correction method creates substantially better ventricular labeling than traditional methods and maintains the speed increase seen in healthy subjects.

Keywords

deformable registration; multi-atlas label fusion; normal-pressure hydrocephalus

1. INTRODUCTION

Multi-atlas label fusion (MALF) is a class of automated methods used to determine brain labels of an unlabeled subject image. MALF techniques rely on a set of previously labeled atlas images, which are spatially aligned to the subject image, to estimate the appropriate label for each voxel in the subject image. A large body of MALF literature focuses on the exact method for both alignment and estimation.¹ Traditionally, MALF utilizes pairwise deformable registration to align the atlas images and the unlabeled subject image (see Figure 1). The associated atlas labels are then propagated to the subject image using an inverse

transform. These propagated labels are then fused using a statistical method that can range from a simple majority vote to more complicated methods that may use intensity information or other feature sets.¹

While MALF may present an accurate automated labeling approach, many clinical research applications still choose to use other labeling methods due to the time and high computational cost required by conventional MALF. This large cost is mainly concentrated in the registration step, during which conventional MALF perform pairwise deformable registration between each atlas image and the subject image. In recent years, multiple researchers have attempted to reduce the computational burden by using an intermediate image as a common target for subject and atlas images alike (Figure 1). For example, Depa et al.² used an intermediate average template created from the atlas set, demonstrating comparable label fusion results in cardiac labeling. However, it has been shown that successive transforms performed in this manner have a negative effect on the outcome of fusion algorithms.³

In this paper, we propose a novel registration technique (illustrated in Figure 1) to recover losses associated with an intermediate image by introducing a second, computationally inexpensive registration step between the subject and atlas images (after initial alignment with the intermediate template). By performing this additional step, we trade a portion of the gains in computation time in the first step for improved final label accuracy. This method is evaluated against pairwise registration and the method outlined in Depa et al.² in the human brain with two common similarity metrics (Demons and Cross Correlation) to determine an optimal registration scheme that balances time with accuracy. In addition, we compare the results to MABMIS, a MALF technique that uses PCA-based analysis to determine the proper registration transforms.⁴ Finally, we further validate our results qualitatively on a set of brains of pathologic patients with normal-pressure hydrocephalus (NPH). We aim to improve label fusion results on brains with pathological differences, which have been shown to be particularly difficult to accurately label.⁵

2. METHODS

2.1 Processing Tools and Data

All image processing was performed using a combination of the ANTs image processing toolkit and in-house Python code. Pipelining and support for parallel processing was facilitated using Nipype, an open source pipeline creation tool based in Python.⁶ Comparisons to MABMIS were run using the native code available from algorithm developers and scripted using Nipype. All experiments were run on an Intel Xeon E7-4870 CPU (2.40 GHz), but restricted to 12 threads per process for consistency.

To perform quantitative analysis of the final label fusion results, we used the T1-weighted images manually labeled by Neuromorphometrics (Neuromorphometrics Inc.), specifically the 50 adult brains from the OASIS dataset that were labeled as a part of the 2012 Academic Subscription. This set was split randomly into a training set of 30 images (NMM30) and a testing set of 20 images (NMM20). For a qualitative analysis of images from NPH patients, T1-weighted images were acquired for 20 patients (NPH20) randomly selected from a set of

192 preoperative clinical MRI scans gathered as a part of a retroactive study. All research relating to clinical patients was approved by the local Institutional Review Board.

Two unbiased, average templates were constructed from the NMM30 training set using the template construction scripts included in the ANTs toolkit,⁷ one using the cross-correlation (CC) metric and the other using the “Demons” metric (mean-squared error with Demons gradient update) as the cost function for deformable registration. These two commonly used metrics for deformable registration allow for consistency in the registration process when using either of the two metrics in subsequent registrations. The transformations and transformed images used to create each template from the images in the atlas set were saved to facilitate label propagation.

2.2 Preprocessing

All images were first converted to the NIFTI1 file format, if required, and resampled to a 0.8mm isotropic voxel size using cubic B-Splines. The N4 implementation in ANTs⁸ was used to correct for image inhomogeneity and the images were pre-aligned in an affine manner to the MNI152 ICBM 2009c Nonlinear Asymmetric template (upsampled to 0.8mm isotropic voxel resolution).⁹ The skull was then removed from each image using a mask fusion technique that incorporates multiple skull removal algorithms including: BET,¹⁰ BSE,¹¹ mri watershed,¹² 3dSkullStrip,¹³ MASS,¹⁴ and ROBEX.¹⁵ Finally, the brain-only images were intensity normalized to a randomly chosen atlas image using histogram matching.

2.3 Registration

Registration was performed using the linear and SyN deformable frameworks implemented in ANTs.¹⁶ Two registration schemes were used to implement the various registration experiments. A COMPLETE REGISTRATION used rigid-body and affine registration (using mutual information) to initialize the deformable registration step. Each step was performed at four levels using the suggested parameters in `antsRegistrationSyN` in the ANTs package. A QUICK REGISTRATION used the deformable registration step using the same parameters as the full registration, and was only performed at the first three levels, excluding the full resolution level. Each registration was performed twice, once using the CC similarity metric and once with the Demons metric.

In order to generate baseline data for comparison, PAIRWISE REGISTRATION (PR) was performed using COMPLETE REGISTRATION between each subject image in the subject group (NMM20 or NPH20) and each of the atlas images (NMM30). Final label propagation was performed using the inverse of the composed linear (S_{aff_i}) and deformable (S_{def_i}) transforms for each atlas image.

$$T_{\text{PR}_i} = S_{\text{def}_i}^{-1} \circ S_{\text{aff}_i}^{-1} \quad (1)$$

In the proposed TEMPLATE-CORRECTED REGISTRATION (TCR) method, an initial COMPLETE REGISTRATION is performed between each image in the subject group and

the corresponding template image depending on the metric used. The subject image is then transformed using the initial transformation and a QUICK REGISTRATION is performed between the transformed subject image and each transformed atlas image that was saved during the creation of the template. The label propagation is then performed by using the composition of the *stored* forward transforms of each atlas image (A_{aff_i} and A_{def_i}) and the inverses of the correction transform for that atlas ($C_{\text{def}_i}^{-1}$) and the inverse of the initial transforms for the subject image (S_{aff} and S_{def}),

$$T_{\text{TCR}_i} = A_{\text{aff}_i} \circ A_{\text{def}_i} \circ C_{\text{def}_i}^{-1} \circ S_{\text{def}}^{-1} \circ S_{\text{aff}}^{-1}. \quad (2)$$

The labels were also propagated without the correction transform to create a TEMPLATE-ONLY REGISTRATION (TOR) in the manner of Depa et al.²

$$T_{\text{TOR}_i} = A_{\text{aff}_i} \circ A_{\text{def}_i} \circ S_{\text{def}}^{-1} \circ S_{\text{aff}}^{-1}. \quad (3)$$

2.4 Evaluation and Comparison with MABMIS

Label fusion was performed for all registration experiments using Joint Label Fusion (JLF),¹⁷ implemented in ANTs, with the exception of when testing the number of atlases to use, where majority voting was used for quicker analysis. For each subject, the propagated label images were randomly sorted and label fusion was performed on the first n images ($n = 1 \dots 30$). All evaluations were performed using the Dice coefficient between binary masks of the compared regions.

The MABMIS method was trained on the NMM30 training set and then labeled simultaneously and in a groupwise fashion on each complete subject set. Both registration and label fusion were included in the MABMIS algorithm. MABMIS was only performed on the full NMM30 set (30 atlas images) due to the time required for training.

3. RESULTS

The TEMPLATE-CORRECTED REGISTRATION and TEMPLATE-ONLY REGISTRATION methods performed as expected by providing a substantial increase in computational speed (3.9 times and 30 times, respectively), while the Demons similarity metric provided approximately a 4.5 times increase in speed on average. Median computation times for each registration experiment can be found in Table 1 (all times represent the use of the full atlas set).

Qualitative analysis of the final fusion results (using 30 labels) shows little difference between the methods, see Figure 2(a). Statistical results were calculated using the Wilcoxon Signed Rank test. Quantitative results are outlined in Table 2.

Figure 3 illustrates increasing mean Dice coefficient corresponding to increasing the number of atlas images (and thus increasing time) in each registration scheme. This allows the direct comparison of fusion accuracy and registration time. The results of the TEMPLATE-ONLY REGISTRATION methods are plotted in a vertical line as the only difference in time would be due to label propagation and fusion, which were expressly not included in this study to focus on front-end registration results.

Qualitative analysis of the NPH subjects showed substantial improvement in the labeling of the lateral ventricles and sulcal CSF using the TEMPLATE-CORRECTED REGISTRATION method and similar accuracy in gray and white matter labels, as illustrated in Figure 2(b). We can see almost complete covering of the ventricle areas with a ventricle label (green or orange), whereas the other methods included thalamus or other deep gray structures in the ventricular labels.

4. DISCUSSION AND CONCLUSION

This paper presents an approach to decrease the effects of decreasing the computational burden of multi-atlas label fusion methods. However, before determining an optimal registration scheme, we must discuss the caveats to increasing the speed of our methods. First, we must consider the end purpose of this multi-atlas fusion. If this fused image (or its associated probabilities) are to be used in an intensity based algorithm such as MALP-EM,⁵ there may be no need to spend the extra time at the MALF step of the algorithm, as the expectation-maximization may correct for many errors that may be caused by slightly less accurate multi-atlas labeling. Similarly, a refinement step based on another segmentation technique could be used to refine the gray matter-white matter boundary, which can be seen to degrade in faster methods. We can appreciate the compromises that are made in Figure 3. In NPH brains, however, the difference could be more substantial. Future investigation could compare various refinement methods to determine if time is better spent in registration or in the following fusion and refinement steps.

Another clear advantage to a template-based method is the avenues available for atlas selection, where a subset of atlases are chosen from a larger atlas space by relation to the subject image. Often this is done via affine alignment to a template space, but, by using our TEMPLATE-CORRECTED REGISTRATION framework, we could align to the template image using a deformable method and compare to very similar transformed atlas images. Since we will continue with our QUICK REGISTRATIONS from that point on, we can use a much more accurate method of alignment before doing our atlas selection. Further investigation into the impact of atlas selection in these methods is warranted.

We evaluated the impact of an intermediate template image on registration for the purposes of multi-atlas label fusion. We also presented a novel technique for overcoming losses in registration accuracy associated with TEMPLATE-ONLY REGISTRATION. Our results show that the decrease in computation time eclipses any (small) decreases in accuracy in both quantitative analysis for healthy subjects and demonstrate qualitative increases in accuracy for pathologic brains of subjects with NPH, while showing identical decreases in computation time.

References

1. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*. Aug.2015 24:205–219. [PubMed: 26201875]
2. Depa, M., Holmvang, G., Schmidt, EJ., Golland, P., Sabuncu, MR. Towards Efficient Label Fusion by Pre-Alignment of Training Data. 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2011); 2011. p. 38-46.
3. Sjöberg C, Johansson S, Ahnesjö A. How much will linked deformable registrations decrease the quality of multi-atlas segmentation fusions? *Radiation Oncology*. Dec.2014 9:1. [PubMed: 24382205]
4. Jia H, Yap P-T, Shen D. Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage*. Jan.2012 59:422–430. [PubMed: 21807102]
5. Ledig C, Heckemann RA, Hammers A, Lopez JC, Newcombe VFJ, Makropoulos A, Lötjönen J, Menon DK, Rueckert D. Robust whole-brain segmentation: Application to traumatic brain injury. *Medical Image Analysis*. Apr.2015 21:40–58. [PubMed: 25596765]
6. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. - PubMed - NCBI. *Frontiers in neuroinformatics*. 2011; 5:13. [PubMed: 21897815]
7. Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, Gee JC. The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*. Feb.2010 49:2457–2466. [PubMed: 19818860]
8. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. *IEEE Trans Med Imag*. Jun.2010 29:1310–1320.
9. Fonov V, Evans AC, Botteron K, Almli CR, McKinsty RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*. Jan.2011 54:313–327. [PubMed: 20656036]
10. Smith SM. Fast robust automated brain extraction. *Human Brain Mapping*. Nov.2002 17:143–155. [PubMed: 12391568]
11. Dogdas B, Shattuck DW, Leahy RM. Segmentation of skull and scalp in 3-D human MRI using mathematical morphology. *Human Brain Mapping*. Dec.2005 26:273–285. [PubMed: 15966000]
12. Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. A hybrid approach to the skull stripping problem in MRI. *NeuroImage*. Jul.2004 22:1060–1075. [PubMed: 15219578]
13. Cox RW. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*. Jun.1996 29:162–173. [PubMed: 8812068]
14. Doshi J, Erus G, Ou Y, Gaonkar B, Davatzikos C. Multi-Atlas Skull-Stripping. *Academic Radiology*. Dec.2013 20:1566–1576. [PubMed: 24200484]
15. Iglesias JE, Liu CY, Thompson PM, Tu Z. Robust Brain Extraction Across Datasets and Comparison With Publicly Available Methods. *IEEE Trans Med Imag*. Aug.2011 30:1617–1634.
16. Avants B, Epstein C, Grossman M, Gee J. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. Feb.2008 12:26–41. [PubMed: 17659998]
17. Wang H, Suh JW, Das SR, Pluta J, Craige C, Yushkevich PA. Multi-Atlas Segmentation with Joint Label Fusion. *Mar.2013 35:611–623*.

Figure 1.

Solid black lines represent COMPLETE REGISTRATION between images and dashed black lines represent a stored transform created during the atlas creation process. Solid gray lines represent QUICK REGISTRATION.

(a) NMM20

(b) NPH20

Figure 2.
Final fusion results for representative subjects using each registration method (30 atlases)

Figure 3.

Mean Dice coefficient of all labels across all subjects versus of the registration time in seconds for increasing number of atlas registrations. Shaded regions represent the standard deviation between subjects of the mean Dice for all labels. MABMIS results are represented in a single point as only one training set was created (30 atlases used).

Table 1

Computation time for each registration (REG) experiment (using all 30 atlas images). Times and numbers in parentheses indicate QUICK REGISTRATIONS done during the correction step. Relative speed-up is given first against pairwise registration (same metric), then pairwise registration using CC. For MABMIS, we report a mean registration time (total time/N) as all registration and labeling was done in a groupwise manner.

Method	Mean Time Per Reg. (sec.)	# REG	Total Time (hrs)	Relative Speed-up
Pairwise (CC)	4030	30	33.58	1
Pairwise (Demons)	888	30	7.40	1 (4.54)
MABMIS	—	—	4.39	7.65
Template-Only (CC)	4030	1	1.12	30.0
Template-Only (Demons)	888	1	0.25	30.0 (134.3)
Template-Corrected (CC)	4030(880)	1(30)	8.45	3.97
Template-Corrected (Demons)	888(202)	1(30)	1.92	3.85 (17.49)

Table 2

Quantitative comparison of median Dice scores for each registration method on the cerebral white matter (WM), cerebral gray matter (GM), and ventricles (Vent). Standard deviation of the Dice scores is given in parentheses. Numbers in bold were statistically significant ($p < 0.01$) with respect to PAIRWISE REGISTRATION (CC), based on a Wilcoxon Signed Rank test without multiple comparison correction.

Method	WM	GM	Vent
Pairwise (CC)	0.921 (0.019)	0.897 (0.032)	0.910 (0.031)
Pairwise (Demons)	0.924 (0.023)	0.904 (0.036)	0.913 (0.027)
MABMIS	0.915 (0.025)	0.886 (0.039)	0.894 (0.032)
Template-Only (CC)	0.909 (0.018)	0.887 (0.030)	0.915 (0.035)
Template-Only (Demons)	0.910 (0.023)	0.883 (0.037)	0.913 (0.029)
Template-Corrected (CC)	0.917 (0.019)	0.894 (0.031)	0.919 (0.035)
Template-Corrected (Demons)	0.918 (0.024)	0.893 (0.031)	0.916 (0.029)