CROSS MODALITY LABEL FUSION IN MULTI-ATLAS SEGMENTATION

Keyvan Kasiri, Paul Fieguth, David A Clausi

Vision and Image Processing (VIP) Lab
Department of Systems Design Engineering
University of Waterloo
Waterloo, Canada

ABSTRACT

Multi-atlas label fusion is a widely used approach in medical image analysis that has improved the accuracy of segmentation. Majority voting, as the most common combination strategy, weighs each candidate in the atlas database equally. More sophisticated methods rely on the intensity similarity of each atlas to the target volume. However, these methods cannot handle those cases in which the atlases and the target image are in different modalities.

A new method for label fusion is proposed, based on a structural similarity measure, relying on the structural relationships of features extracted from an un-decimated wavelet transform instead of explicit image intensities. The new label fusion method has been tested on simulated and real MR images; segmentation results are promising, and open the door to a wider range of multi-modal approaches.

***Index Terms*—** label fusion, multi-atlas segmentation, similarity measure

1. INTRODUCTION

In traditional atlas-based segmentation, a target scan is labeled by referring to an atlas, where an *atlas* refers to an image, already segmented, and where the target is aligned to the atlas using deformable registration [1]. The major drawback of this approach is that a single deformation cannot represent the range of anatomical variations present in a whole population of potential target cases. The accuracy of the segmentation procedure can be improved using a group of atlases, instead of a single atlas, thus combining labels from some number of registered atlases; this work has led to an active literature on *multi-atlas* approaches [2, 3].

The key challenge associated with the multi-atlas approach is "label fusion" — the strategy by which atlas labels are combined into a single segmentation. Many label fusion methods, such as majority voting (MV) [3] and STAPLE [4], do not consider image intensities after being warped to the target image. If we do consider the image intensities and give higher weights to those more similar atlases, whether globally or locally, we obtain improvements in segmentation accuracy [5, 6, 7].

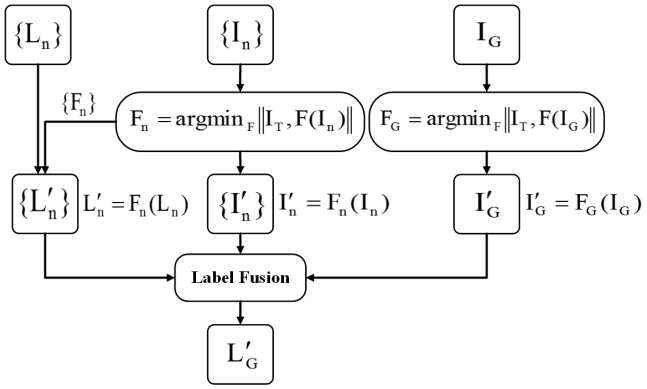


Fig. 1. Block-diagram of the multi-atlas-based segmentation for a multi-modal atlas database.

The multi-atlas approaches are promising, however these methods remain problematic in those cases where the atlases and the target scan are obtained from different sensors or from different acquisition modalities: image-intensity comparisons may no longer be valid, since image brightness can have highly differing meanings and circumstances in different modes [8].

In this paper, we seek to develop an approach for multi-modal atlases, specifically a label fusion approach for multi-atlas segmentation. We propose a similarity measure which uses structural features based on an un-decimated wavelet transform (UDWT), to overcome the problems associated with matching complex intensity relationships.

2. OVERVIEW

2.1. Problem Definition

Fig. 1 shows the block diagram of the general multi-atlas-based segmentation framework. $\{I_n\}$, $\{L_n\}$, and I_G respectively represent the set of N atlases, the labels from these atlases, and the target image. The label alphabet contains L

unique segments:

$$L_i(x) \in \{1, \ldots, L\}, \tag{1}$$

where x denotes the location in the label map L_i corresponding to the i-th atlas. In the first stage, the atlases and the target image are all warped to the template image, I_T, resulting in the inferred transformations $\{F_n\}$ for the atlases and F_G for the target. Given these transformations, each input, whether image or label field, can be transformed to the common reference of the template. Thus I'_G, $\{I'_n\}$ and $\{L'_n\}$ are the target, atlases, and labels in the common reference frame.

The key goal of this paper is to design a label fusion method, a final segmentation result L'_G which will be generated by combining all propagated labels, $\{L'_n\}$ using a label fusion method, with the label fusion weighted on the basis of the similarity of the transformed atlases $\{I'_n\}$ and the transformed reference I'_G.

2.2. Background

Many label fusion methods have been introduced in the medical atlas literature [9]. The simplest and most widely used one is majority voting [3], which asserts an equal contribution for each atlas. As the image intensity is not taken into account during label fusion, a higher accuracy can be achieved by some form of weighting, based on the similarities between the atlases and the target image. Weighting strategies including both global and local forms [5, 10], where local weighted voting (LWV) outperforms global strategies when dealing with high contrast anatomical structures [6, 9, 11], although other authors [12, 13] have proposed a framework for labeling whole brain scans by incorporating a global and stationary Markov random field that ensures the consistency of the neighbourhood relations between structures.

Most label fusion approaches are limited by the assumption that they depend on the consistency of voxel intensities across different MRI scans. In these cases, approaches based on mutual information do help [14], however its inherent nonlocality make it problematic for local weighted label fusion. This issue will be highlighted when atlases and target image are acquired with different modalities [6, 8].

Relying on the similarity between intensity values of the atlases and target scan is often problematic in medical imaging — in particular when the atlases and target image are obtained via different sensor types or imaging protocols. In [15], a generative probabilistic model is proposed that yields an algorithm for solving the atlas-to-target registrations and label fusion steps simultaneously. This model exploits the consistency of voxel intensities within the target scan to drive the registration and label fusion instead of intensity similarity, hence the atlases and target image can be of different modalities. The method is based on exploiting the consistency of voxel intensities within the segmentation regions, as well as their relation with the propagated labels.

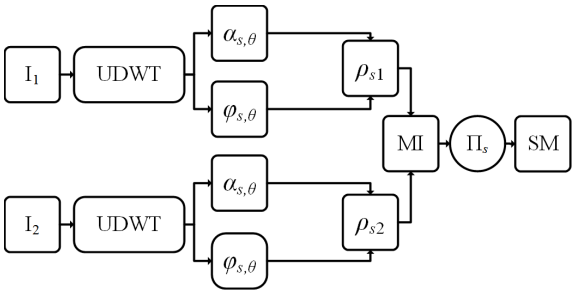


Fig. 2. Similarity measure (SM) for multi-modal images based on structural features. The similarity measure is obtained by computing the mutual information of structural features captured by an un-decimated wavelet transform.

3. METHOD

The aim of the segmentation is to find a label map L'_G associated with I'_G which is the target image already registered to the template image. The label fusion problem in a multi-atlas segmentation can be inferred from a maximum-a-posteriori (MAP) estimation framework [6]:

$$\hat{L}'_G(x) = \underset{L'_G}{\operatorname{argmax}} \sum_n p(L'_G(x)|L'_n) p(I'_G(x)|I'_n) \tag{2}$$

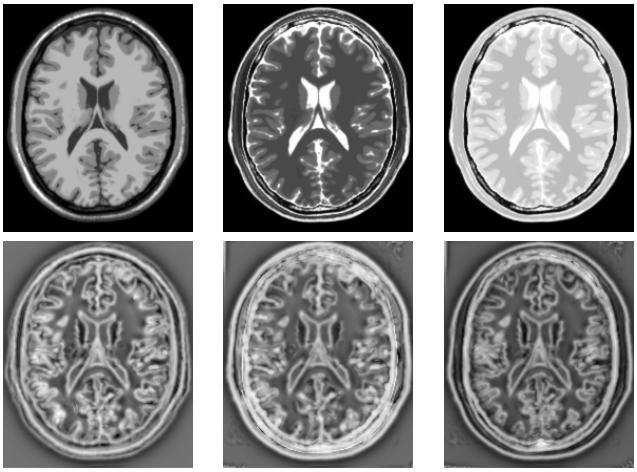
where $p(L'_G(x)|L'_n)$ is the label prior value and $p(I'_G(x)|I'_n)$ is the probability that relates the n-th atlas to the target image which can be interpreted as an assigned weight to the n-th vote [16]. Traditional majority voting produces the final segmentation, L'_G, by assuming that different atlases provide equal registration quality. Typically, for deterministic atlases, discrete values of 0 and 1 are used instead of $p(L'_G(x)|L'_n)$. As mentioned above, $p(I'_G(x)|I'_n)$ gives a hint of the relation between two images which has been interpreted in the literature as image similarity [16, 17].

Since images are obtained from different sensors, therefore the intensity relationship between the images is complex and performing a similarity measure based on image intensity will not yield good results.

The proposed label fusion method is based on defining a structural similarity measure to approximate the similarity of the atlas and the target image, for which the block diagram is depicted in Fig. 2. As shown in the figure, multi-scale complex wavelet representation of the input images are constructed using an un-decimated complex wavelet transform such as dual-tree complex wavelet transform (DT-CWT) [18] and Log-Gabor complex wavelet transform [19]. While the Log-Gabor complex wavelet transform is used in this paper, the DT-WCT may also be employed as well. The resulting wavelet coefficients are noted as $\Upsilon_{s,\theta}(x)$ at location x,

$$\Upsilon_{s,\theta}(x) = \alpha_{s,\theta}(x) \exp[j\phi_{s,\theta}(x)] \tag{3}$$

where $\alpha_{s,\theta}(x)$ and $\phi_{s,\theta}(x)$ are the amplitude and phase of the complex wavelet coefficients, respectively. The phase order,



T1 T2 PD

Fig. 3. Structural features from different MR modes. The first row shows a slice of brain scans in T1, T2, and PD modes. The second row shows the structural features extracted from the second scale of UDWT.

$\rho_s(x)$ at each scale can be defined as the normalised weighted summation of phase deviations from its mean value across all scales:

$$\rho_s(x) = \frac{\sum_\theta \alpha_{s,\theta}(x)\Lambda(x)}{\sum_\theta \alpha_{s,\theta}(x)}, \tag{4}$$

where

$$\Lambda(x) = \cos(\phi_{s,\theta}(x) - \bar{\phi}_\theta(x)). \tag{5}$$

Here, $\Lambda(x)$ is the phase deviation from the mean value of the complex phase $\phi_\theta(x)$. Fig. 3 shows the structural features of different modes of a brain MR slice from the BrainWeb simulated database [20]. As can be seen, the intensity information, which is the problematic part of the label fusion, is no longer present and instead the aspects which remain are the structural features that are almost the same in all modalities.

In order to measure the similarity between each atlas and the target image, the similarity is calculated across all scales based on the structural features represented by ρ_s. Mutual information (MI) based on image intensity entropy is utilised to measure the similarity of structural features at each scale. MI for two images I_1 and I_2 is defined as

$$MI(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) \tag{6}$$

In this equation, $H(I_1)$ and $H(I_2)$ represent the entropy of the intensity in images I_1 and I_2 and $H(I_1, I_2)$ stands for the joint entropy of these two images.

The proposed similarity measure is a function over all scales: the structural features at some scale from the two images are compared using mutual information applied to the phase order from (4):

$$SM(I_1, I_2) = \prod_s MI_s(\rho_{s1}, \rho_{s2}) \tag{7}$$

where s denotes the scale from UDWT. Finally, the resulting similarity measure is applied to (2), contributing to the label fusion paradigm by weighting labels from each atlas based on how similar each atlas image is to the target image:

$$\hat{L}'_G(x) = \underset{L'_G}{\operatorname{argmax}} \sum_n p(L'_G(x)|L'_n) SM(I'_G, I'_n). \tag{8}$$

4. RESULTS AND DISCUSSION

4.1. Data

We have tested our method on the 3D brain MR scans from the BrainWeb simulated database [20], based on the T1, T2, and PD modalities with 3% noise and 20% intensity non-uniformity, and on the T1 images in the LONI real database [21]. The databases provide ground truth of tissue labels for white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF).

4.2. Experimental setup

To assess the proposed method, we compared our approach with conventional majority voting (MV) and mutual information (MI) [17] for segmenting real and simulated MR scans into WM, GM, and CSF tissues. In the first test on simulated data, a set of training data was generated by an artificial deformation using thin-plate spline (TPS). Two different cases are examined: a single mode atlas database and a multi-modal atlas database with a target in a different mode from the atlas set. The registration utilised in this framework is undertaken using statistical parametric mapping (SPM)[22]. For these experiments, 15 different random deformation fields are generated and the whole process of segmentation is run ten times. To validate the method on real data, the second test was performed by using 40 real T1 atlases and a PD target image. A set of ten training scans out of 40 subjects is randomly selected to form the atlas database and this procedure is run ten times to obtain the segmentation results.

To quantitatively assess the accuracy of segmentation, the Dice similarity coefficient [23] is used, defined as

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{9}$$

where A and B are the ground truth and the segmented image.

4.3. Results

Fig. 4 illustrates the advantage of using multi-modal atlases instead of single-mode ones. The effect of adding an atlas in a mode other than the target's mode on the segmentation accuracy is examined using simulated brain data. In this experiment, all atlases are in the same mode as the target image, and a slice of a T1 image is segmented using MV. The experiment is then repeated for the case that additional T2 training data is added.

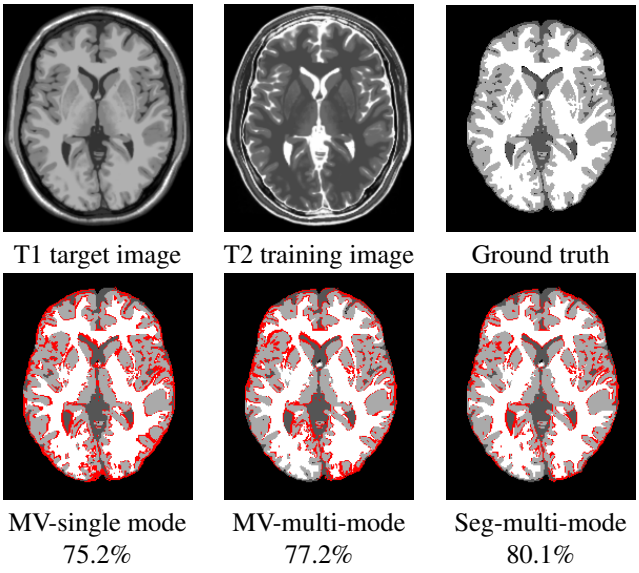


Fig. 4. Multi-modal versus single-mode segmentation: the bottom row shows the results of MV and the proposed method, with the Dice coefficient (9) given. The misclassification error of each case is shown in red. The highest Dice performance is offered by the proposed approach.

The average Dice coefficients by MV method for the WM, GM, and CSF tissues in the two experiments are obtained as 75.2% and 77.2%. Comparatively, the proposed method has the accuracy of 80.1% for the multi-modal case. The misclassification error in each of the segmentation results is shown in red color. One should note that, in the MV method, only label maps are used. However, the proposed method takes advantage of the structural features in the new mode as well as the label map to segment the target image.

The first experiment on simulated data, which is illustrated in Fig. 5, considers the cross-modality segmentation with the single-mode atlas database. For this experiment, first, the target image is assumed to be in T2 mode while the atlas database is in T1. For the second case, the target is changed to PD mode. The atlas database is generated using artificial deformations applied on the simulated images from the BrainWeb database [20]. The segmentation results demonstrate improved performance of the proposed label fusion compared to the traditional MV and MI-based method.

A second experiment is performed to show how the method works for the complex cases with multi-modal atlases and the target image in a mode which does not have any representative in the atlas set. Table 1 reports the segmentation results when the database contains atlases of T1 and T2 mode scans and the target image is in PD mode. Results obtained from the proposed method significantly outperforms MV and shows considerable improvement over MI-based method. The improvement is statistically noticeable as well since the lower standard deviation for the accuracy measurement is achieved.

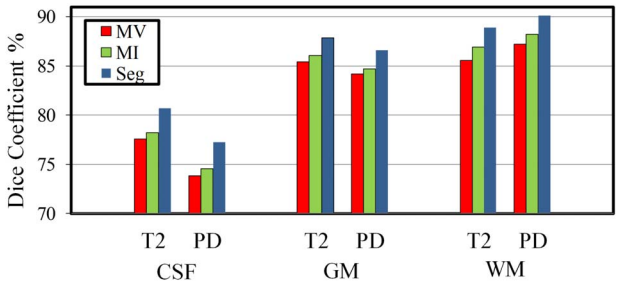


Fig. 5. Single-mode multi-atlas segmentation results in terms of average Dice coefficient for the proposed (Seg), majority voting (MV), and MI-based method (MI). The atlas set is in T1 while the target is in T2 and PD.

Table 1. Average Dice coefficient and its standard deviation for T1 and T2 atlases and PD target mode.

Tissue	WM	GM	CSF
Seg	**88.6±0.2**	**88.2±0.2**	**80.7±0.8**
MI	86.9±0.3	86.1±0.4	78.2±1.2
MV	85.6±0.4	85.4±0.5	77.6±1.3

To evaluate on real data, the method is applied to segment a T2 target image given a set of T1 real normal images randomly selected from LONI database [21]. Table 2 shows the results for this experiment. Although the results of the proposed method does not show any improvement for segmenting the GM, it still does a promising job for delineation of the two other tissues. Furthermore, the method is shown to be robust over different atlas selections compared to other reported methods.

5. CONCLUSION

We proposed a label fusion method based on a structural similarity measure. Unlike most of previous label fusion methods that are working on single-mode multi-atlas segmentation, our method is designed to deal with fusing labels across modalities or utilising single-mode atlas set to segment a target in different mode. For this purpose, we proposed a similarity measure based on structural features which can be extracted from un-decimated wavelet coefficients. To validate our method, experiments for segmenting tissues in the MR brain images were conducted. Based on the results in this paper, the proposed label fusion method outperforms the state of the art and opens the door for future works on multi-modal approaches.

Table 2. Average Dice coefficient and its standard deviation for T2 target given real T1 atlases.

Tissue	WM	GM	CSF
Seg	**80.6±0.4**	75.0±**0.2**	**61.2±0.8**
MI	78.9±0.7	**75.2**±0.4	58.3±1.3
MV	77.6±0.8	72.4±0.4	55.1±1.7

6. REFERENCES

[1] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-D model-based neuroanatomical segmentation," *Human Brain Mapping*, vol. 3, no. 3, pp. 190–208, 1995.

[2] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.

[3] 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*, vol. 33, pp. 115–126, 2006.

[4] S. Warfield, K. Zou, and W. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, pp. 903–921, 2004.

[5] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. Ginneken, "Multi-atlas-based segmentation with local decision fusionapplication to cardiac and aortic segmentation in CT scans," *IEEE Trans. Med. Imag.*, vol. 28, pp. 1000–1010, 2009.

[6] M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Trans. Med. Imag.*, vol. 29, pp. 1714–1729, 2010.

[7] C. Sjöberg and A. Ahnesjö, "Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures," *Comput. Methods and Programs in Biomedicine*, vol. 110, no. 3, pp. 308–319, 2013.

[8] J. E. Iglesias, M. R. Sabuncu, and K. Van Leemput, "A generative model for multi-atlas segmentation accross modalities," in *IEEE Int. Symp. on Biomedical Imaging (ISBI)*, 2012, pp. 888–891.

[9] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz de Solorzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," *IEEE Trans. Med. Imag.*, vol. 28, pp. 1266–1277, 2009.

[10] A. R. Khan, N. Cherbuin, W. Wen, K. J. Anstey, P. Sachdev, and M. F. Beg, "Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): validation on hippocampus segmentation," *NeuroImage*, vol. 56, 2011.

[11] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, and J. P. W. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE Trans. Med. Imag.*, pp. 2000–2008, 2010.

[12] C. Ledig, R. Wolz, P. Aljabar, J. Lotjonen, R. A. Heckemann, A. Hammers, and D. Rueckert, "Multi-class brain segmentation using atlas propagation and EM-based refinement," in *IEEE Int. Symp. on Biomedical Imaging (ISBI)*, 2012, pp. 896–899.

[13] H. Wang, J. W. Suh, J. Pluta, M. Altinay, and P. Yushkevich, "Optimal weights for multi-atlas label fusion," in *Inform. Process. in Medical Imaging (IPMI)*, 2011, pp. 73–84.

[14] M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein, "Optimum template selection for atlas-based segmentation," *NeuroImage*, vol. 34, no. 4, pp. 1612–1618, 2007.

[15] J. E. Iglesias, M. R. Sabuncu, and K. Van Leemput, "A unified framework for cross-modality multi-atlas segmentation of brain MRI," *Medical Image Anal.*, vol. 17, no. 8, pp. 1181–1191, 2013.

[16] H. Wang, J. W. Suh, S. Das, J. Pluta, M. Altinay, and P. Yushkevich, "Regression-based label fusion for multi-atlas segmentation," in *IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR)*, 2011, pp. 1113–1120.

[17] P. Aljabar, R. Heckemann, A. Hammers, J. V Hajnal, and D. Rueckert, "Classifier selection strategies for label fusion using large atlas databases," in *Medical Image Computing and Comput.-Assisted Intervention (MICCAI)*, pp. 523–531. 2007.

[18] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, 2005.

[19] S. Fischer, F. Šroubek, L. Perrinet, R. Redondo, and G. Cristóbal, "Self-invertible 2D log-Gabor wavelets," *Int. J. of Comput. Vision*, vol. 75, no. 2, pp. 231–246, 2007.

[20] McConnell brain imaging center, "BrainWeb: simulated brain database," http://www.bic.mni.mcgill.ca/brainweb/.

[21] LONI, "The UCLA laboratory of Neuro Imaging," http://www.loni.ucla.edu/.

[22] J. Ashburner, "SPM: Statistical parametric mapping," http://www.fil.ion.ucl.ac.uk/spm/.

[23] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.