Atlas-based segmentation of brain magnetic resonance imaging using random walks

Jean-Philippe Morin Christian Desrosiers Luc Duong
École de technologie supérieure
Montreal, Canada
jean-philippe.morin.1@ens.etsmtl.ca, {christian.desrosiers, luc.duong}@etsmtl.ca

Abstract

The segmentation of brain magnetic resonance imaging is a difficult task, essential to several applications in neuroscience. Atlas-based methods are often employed for this task since they provide prior information in the form of labels, without the manual intervention of a trained technician. In this paper, we present a novel and efficient atlas-based segmentation method based on random walks. Unlike most atlas-based approaches, our method combines the registration and label propagation steps in a single efficient framework. Moreover, this method does not depend on a specific deformation model, making it more robust to complex transformations not captured by such models. Experiments on benchmark brain MRI data show the usefulness and efficiency of our method.

1. Introduction

The segmentation of brain magnetic resonance imaging (MRI), essential to several important applications in diagnosis [10], therapy evaluation [6] and human brain mapping [7], is a challenging task due to frequent image artifacts and poor contrast between the structures to segment [3]. Recently, atlas-based segmentation approaches have gained a considerable amount of popularity for this task. Unlike manual segmentation methods, which require a trained technician to manually label regions of interest in target images, atlas-based approaches use pre-labeled images, called *atlases*, to segment target images without human assistance. This allows one to avoid the repetitive and time-consuming task of labeling images manually, and to enhance the segmentation process by providing a priori information on the target regions.

The atlas-based segmentation process is normally composed of three steps: a *registration* step where the target image is aligned to the single atlas or multiple atlases, a *label propagation* or *fusion* step where the labels are transfered from the atlases to the target, and a final *segmentation* step in which transfered labels are used to segment the target image. The registration process is usually carried out in two separate steps: a global registration (affine or rigid transformation) performed to obtain an initial alignment, followed by a local non-rigid registration to take into account target-specific deformations. This local registration is typically performed by applying a non-linear transform to the points of a free-form deformation grid, for instance, based on B-spline curves [15], level sets [17], demons [18] or optical flows [13]. Since it involves a complex discrete optimization problem, the registration process is often computationally expensive. Moreover, because it usually depends on a parametric model, it can fail in the case of complex deformations that do not fit well in this model.

To better capture the variability of target regions, segmentation methods using several atlases have been proposed. Multiatlas-based methods, which have been shown to be more robust and accurate than methods based on a single atlas [9], vary mostly by the way in which atlas labels are combined into the target image. Commonly employed strategies include majority voting [14], (locally) weighted voting [1], or more complex strategies based on the Expectation Maximisation (EM) algorithm [19] and generative probabilistic models [16]. Like most label propagation techniques, the accuracy of these strategies largely depends on the quality of the registration carried out beforehand. To avoid this problem, recent works such as [12] have focused on combining the segmentation and registration processes. While interesting, these approaches are also model-dependent and often computationally complex. For a more extensive survey of the literature on atlas-based segmentation of MRI brain images, refer to [3].

This paper presents a novel atlas-based segmentation method based on random-walks. Like existing probabilistic label fusion approaches, this method uses a generative model to provide location-specific label probabilities and to give high-confidence labels a greater impact on the segmentation. However, unlike most existing approaches,

this method combines the registration and label propagation steps in a single unified framework. Moreover, this method does not depend on a specific deformation model, which makes it robust to complex and highly-variable deformations. Also, our method only requires a few seed labels in each atlas, which can make the segmentation process more user friendly. Finally, as shown by our experiments, this method is highly efficient and can provide accurate segmentations rapidly.

2. The proposed method

The atlas-based segmentation problem can be defined as follows. Suppose we have a set $\mathcal{I} = \{I_1, \ldots, I_M\}$ of M atlases, each of them containing a set $S_j \subseteq I_j$ of labeled pixels known as seeds. The goal is to use these atlases to label every pixel of a target image I_t. Using a probabilistic formulation, we want to find the most probable label $l_k \in L$ for each $x_i \in I_t$. To solve this problem, we use a probabilistic generative model similar to the one proposed in [8]. In this model, illustrated in Figure 1, the label of a pixel x_i is produced by first selecting a label l_k, then by selecting a seed having this label, then transferring this label to x_i through a random-walk process. Following this model, the problem can then be formulated as:

$$\arg\max_{l_k \in L} \; p(l_k \,|\, x_i) \;\propto\; p(l_k) \sum_j \sum_{x_s \in S_j} p(x_i \,|\, x_s)\, p(x_s \,|\, l_k). \quad (1)$$

Considering the a priori label probability $p(l_k)$ and the seed selection probability $p(x_s \,|\, l_k)$ as uniform, and denoting as Q_k the set of seeds with label l_k, we obtain the following problem:

$$\arg\max_{l_k \in L} \; p(l_k \,|\, x_i) \;\propto\; \frac{1}{|Q_k|} \sum_{x_s \in Q_k} p(x_i \,|\, x_s). \quad (2)$$

Thus, we assign to a pixel x_i the label l_k for which the average diffusion probability from the seeds of label l_k to x_i is the highest. As mentioned in [8], considering the *average* probability avoids the bias normally observed when having an uneven distribution of labels among the seeds.

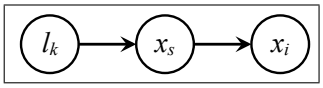


Figure 1. Generative segmentation model.

2.1. Random-walk segmentation in a single image

The diffusion probabilities are modeled using a random-walk with restart (RWR) process. To define this process, let us first consider the simpler setting where an image I is segmented without using atlases. Suppose that I contains a set of labeled pixels $x_s \in S$. In this setting, the diffusion probability $p(x_i \,|\, x_s)$ is computed as the stable-state probability of a random-walker, starting its walk at seed x_s, to visit pixel x_i. At each step t, the walker at a pixel x_i can do one of the two following things: 1) with a probability $\alpha \in]0, 1]$, the walker returns to his starting seed x_s, or 2) with probability $(1-\alpha)$, the walker selects a neighbor pixel $x_j \in \mathcal{N}_i$ with probability p_{ij} and moves to this neighbor. The probability $\pi_{si}^{(t)}$ of a walker starting at seed x_s to visit pixel x_i at step t can then be expressed using the following equation:

$$\pi_{si}^{(t)} \;=\; \alpha\, \delta(i = s) \;+\; (1-\alpha) \sum_{x_j \in \mathcal{N}_i} \pi_{sj}^{(t-1)} p_{ji}, \quad (3)$$

where $\delta(rel) = 1$ when relation rel is true and 0 otherwise. Considering simultaneously every pixel x_i and seed x_s yields a linear system which can be expressed in matrix form as:

$$\Pi^{(t)} \;=\; \alpha I \;+\; (1-\alpha)\Pi^{(t-1)} P, \quad (4)$$

where P is a $|I| \times |I|$ matrix such that $[P]_{ij} = p_{ij}$ is the transition probability from pixel i to pixel j, and $\Pi^{(t)}$ is a $|S| \times |I|$ matrix such that $[\Pi^{(t)}]_{si} = \pi_{si}^{(t)}$. At the stable state of this system, we have that $\Pi^{(t)} = \Pi^{(t-1)} = \Pi$, and thus:

$$\Pi \;=\; \alpha(I - (1-\alpha)P)^{-1}. \quad (5)$$

Once Π has been computed, the label probabilities of individual pixels can then be obtained as follows. Let B be a $|L| \times |I|$ matrix such that

$$[B]_{ki} \;=\; \begin{cases} \frac{1}{|Q_k|}, & \text{if } x_i \in S \text{ and } label(x_i) = l_k \\ 0, & \text{otherwise.} \end{cases}$$

and let $R = B\Pi$. The probability of pixel x_i of having label l_k is given by $p(l_k \,|\, x_i) = [R]_{ki}$.

The transition probabilities can be defined as in most random-walk based segmentation methods. Let W be a $|I| \times |I|$ matrix such that

$$[W]_{ij} \;=\; \begin{cases} \exp\{-\sigma \cdot |x_i - x_j|^2\}, & \text{if } x_j \in \mathcal{N}_i \\ 0, & \text{otherwize.} \end{cases}$$

where $|x_i - x_j|$ is the difference in intensity or any other characteristic of the pixels, and $\sigma \geq 0$ is a user given parameter. Moreover, let D be a diagonal matrix such that $[D]_{ii} = \sum_j [W]_{ij}$. The transition probability matrix corresponds to $P = D^{-1}W$.

2.2. Multiatlas-based random-walk segmentation

Let us now consider the more complex case where we have a set $\mathcal{I} = \{I_1, \ldots, I_M\}$ of atlases and want to segment a target image I_t. In order to diffuse the information from the atlases to the target, we need to establish links between these images. In our approach, this connection is made by connecting specific pixels of the atlas and target images, called *anchors*. Denote respectively by $Z_j \subseteq I_j$ and $Z_t \subseteq I_t$

the sets of anchor pixels in an atlas image I_j and the target image I_t. Furthermore, let $k(x_i, x_q)$ be a function evaluating the similarity between two anchors $x_i \in Z_j$ and $x_q \in Z_t$. For instance, this function could be defined using feature vectors $f(x_i)$ and $f(x_q)$ extracted from pixels x_i and x_q:

$$k(x_i, x_q) \;=\; \exp\{-\gamma \cdot \|f(x_i) - f(x_q)\|^2\}. \tag{6}$$

We then define probabilities for two types of transitions in the random-walk: intra-image and inter-image transitions. For transitions within a single atlas image I_j, we define an $|I_j| \times |I_j|$ matrix P_{jj} such that:

$$[P_{jj}]_{il} \;=\; [1 \;-\; \beta \cdot \delta(x_i \in Z_j)]\frac{w_{il}}{d_i}, \tag{7}$$

where $\beta \in [0, 1]$ is a parameter acting as a trade-off between the intra- and inter-image propagation, $w_{il} = [W_j]_{il}$ and $d_i = \sum_{l \in \mathcal{N}_i} w_{il}$. Essentially, these transition probabilities are computed as in the single image case, except for the anchor pixels where the probabilities are damped by a factor $(1 - \beta)$. Likewise, the transitions probabilities within the target image I_t can be defined as:

$$[P_{tt}]_{ql} \;=\; [1 \;-\; \mu \cdot \delta(x_q \in Z_t)]\frac{w_{ql}}{d_q}, \tag{8}$$

where $w_{ql} = [W_t]_{ql}$, $d_q = \sum_{l \in \mathcal{N}_q} w_{ql}$, and $\mu \in [0, 1]$ plays a similar role as β.

The inter-image transitions are defined in a similar fashion. Thus, the transition probabilities from atlas I_j to target I_t are given in a matrix P_{jt} such that:

$$[P_{jt}]_{iq} \;=\; \begin{cases} \beta\frac{k(x_i, x_q)}{k_i}, & \text{if } x_i \in Z_j \text{ and } x_q \in Z_t \\ 0, & \text{otherwise.} \end{cases}, \tag{9}$$

where $k_i = \sum_{x_q \in Z_t} k(x_i, x_q)$. Correspondingly, the transition probabilities within the target can be defined as:

$$[P_{tj}]_{qi} \;=\; \begin{cases} \mu\frac{k(x_i, x_q)}{k'_q}, & \text{if } x_i \in \bigcup_j Z_j \text{ and } x_q \in Z_t \\ 0, & \text{otherwise.} \end{cases}, \tag{10}$$

where $k'_q = \sum_j \sum_{x_i \in Z_j} k(x_i, x_q)$.

For any anchor pixel x_i in atlas I_j, we can verify that the transition probabilities within I_j and to the target I_t sum to 1:

$$\sum_{l \in \mathcal{N}_i}(1-\beta)\frac{w_{il}}{d_i} + \sum_{q \in Z_t}\beta\frac{k(x_i, x_q)}{k_i} \;=\; (1-\beta)\frac{d_i}{d_i} \;+\; \beta\frac{k_i}{k_i} \;=\; 1.$$

Using a similar process, one can also verify that the transition probabilities of any pixel in I_t sum to 1.

Having defined the transition probabilities, we can then find an expression for the label probabilities of the target image pixels. Let R_t be an $|L| \times |I_t|$ matrix such that $[R_t]_{ki}$ is the probability of $x_i \in I_t$ of having label l_k. Moreover, denote by B_j the label indicating matrix corresponding to atlas I_j and let $C_{jj} = I - (1-\alpha)P_{jj}$. It can be shown that:

$$R_t \;=\; \alpha\Big[\sum_{j=1}^{M} B_j C_{jj}^{-1} P_{jt}\Big]\Big[C_{tt} - \sum_{q=1}^{M} P_{tq} C_{qq}^{-1} P_{qt}\Big]^{-1}. \tag{11}$$

This model can be further simplified by limiting the transitions to only go from the atlases to the target image, not the opposite. To do this, we simply set $\mu = 0$ and obtain:

$$R_t \;=\; \alpha\Big[\sum_{j=1}^{M} B_j C_{jj}^{-1} P_{jt}\Big]C_{tt}^{-1}. \tag{12}$$

2.3. Efficient computation

While matrices C_{jj} are sparse, their inverse is not and, therefore, may require significant computational and memory resources. To avoid this problem, we use an approximation technique limiting the random walk to $T_{\max}$ steps:

$$B_j C_{jj}^{-1} \;\approx\; \sum_{t=0}^{T_{\max}}(1-\alpha)^t B_{jj} P_{jj}^t. \tag{13}$$

This approximation can be computed iteratively, as described in Figure 2. The computational complexity of this technique can be shown to be in $O(T_{\max} \times |I_j| \times |L| \times Q_{\max})$, where $|L|$ is the number of label classes and $Q_{\max}$ is the maximum number of seeds in a label class. Moreover, the space requirements of this techniques are low since the size of matrices A and M is only $|I_j| \times |L|$. A similar approach can be used to compute BC_{tt}^{-1}, where $B = \sum_j B_j C_{jj}^{-1} P_{jt}$. Finally, if the anchor pixels can be obtained independently of the target image, matrices $B_j C_{jj}^{-1}$ can be pre-computed offline, making the complexity of the proposed method constant with respect to the number of atlases.

Approximation scheme

Input: Matrices B, $C = I - (1-\alpha)P$;
Input: The maximum random-walk length $T_{\max}$;
Output: The approximation $A \approx BC^{-1}$;

$A \leftarrow B$;
$M \leftarrow B$;
for $t = 1 \ldots T_{\max}$ **do**
    $M \leftarrow (1-\alpha)MP$;
    $A \leftarrow A + M$;
return A ;

Figure 2. Approximation scheme to compute the label probabilities.

3. Experimental evaluation

In this section, we evaluate our method on the task of segmenting brain MRI images from the BrainWeb Simu-

lated Brain database[1] [4]. For our experiments, we used a set of 7 models of normal brains, corresponding to subjects 04, 05, 06, 18, 20, 38 and 41. Each model is a volume composed of 181 slices (1mm thick) of 256×256 pixels. While the ground truth segmentations provided in the database contain 12 classes, we used only the 4 principal ones: *Background* (BG), *White matter* (WM), *Gray matter* (GM) and *Cerebrospinal fluid* (CSF). To measure the accuracy of the obtained segmentations S with respect to the ground truth R, we used the Dice coefficient:

$$Dice(S,R) = \frac{2\times|S \cap R|}{|S| + |R|}. \quad (14)$$

The proposed segmentation algorithms were implemented in Matlab and the experiments were carried out on a 2.30 Ghz i7 64 bit Intel CPU computer, with 8 Gb of RAM. Unless specified otherwise, the following default parameters were used throughout the experiments: neighborhood $\mathcal{N}_i$ of pixel p_i containing the immediate neighbors of p_i, number of computation steps $T_{\max} = 2500$, $\sigma = 60000$, $\alpha = 0.0001$, $\beta = 1$, and $\mu = 0$.

In a first experiment, we evaluated the robustness of our method to increasing amounts of deformation. To simulate a continuous range of deformations, we selected a central slice S_i, in each of the 7 volumes, and used the pair of slices located at increasing distances of S_i, i.e. slices (S_{i-dist}, S_{i+dist}) for $dist = 1,\ldots,5$, as atlases to segment S_i. Pixels in the corresponding ground truth segmentations were used as atlas seeds. To find the anchor pixels and their transition probabilities, we tested two different feature extraction techniques. In the first technique, we used the *Affine-Scale Invariant Feature Transfom* (ASIFT) algorithm [11] to find matches between the atlases and the target image, and set $k(x_i, x_q) = 1$ for every correspondence (x_i, x_q) found by ASIFT. Likewise, in the second technique, we used the OpenSurf implementation of the *Speeded Up Robust Feature* (SURF) algorithm [2] to find the matches (x_i, x_q), and kept the match strengths returned by the algorithm as values $k(x_i, x_q)$.

Figure 3 summarizes the results of this experiment. As shown in Figures 3(a)-3(b), we can see that the segmentation quality (Dice coefficient, averaged over the 7 models) decreases as slices further away from the target are used as atlases. Moreover, we observe that ASIFT gives better segmentation results than SURF. As presented in Figure 3(c), this is due to the higher number of matches found by ASIFT. Figures 3(d)-3(k) give an example of a target slice S_i, atlas slices (S_{i-1}, S_{i-3} and S_{i-5}), ground truth, and segmentation results obtained for atlas distances ranging from 1 to 5 slices.

In a second experiment, we then evaluated the impact of the number of atlases on the segmentation quality. For this experiment, we segmented a target slice in each volume, using the corresponding slices in the other volumes as atlases. These atlases were randomly selected to form sets of increasing size N_{atlas}, ranging from 1 to 5. In light of the previous experiment, we only tested the ASIFT algorithm to find matches. Figure 4 reports the results of this experiment. As shown in Figure 4(a), the segmentation quality (Dice coefficient, averaged over the 7 models) increases as more atlases are used. An example of a target, ground truth, and segmentation results obtained for 1 to 4 atlases is presented in Figures 4(b)-4(g).

A last experiment was conducted to evaluate the influence of parameter α on the segmentation quality. For this experiment, a central slice S_i was selected in each of the 7 volumes, and the next slice S_{i+1} was used as atlas. Segmentation quality (Dice coefficient, averaged over the 7 models), was measured for α values ranging from 0.00001 to 0.5. As we can see in Figure 5, our method is not so sensitive to this parameter, and a wide range of values work equally well.

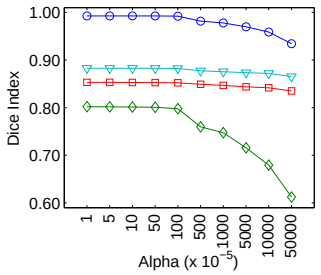
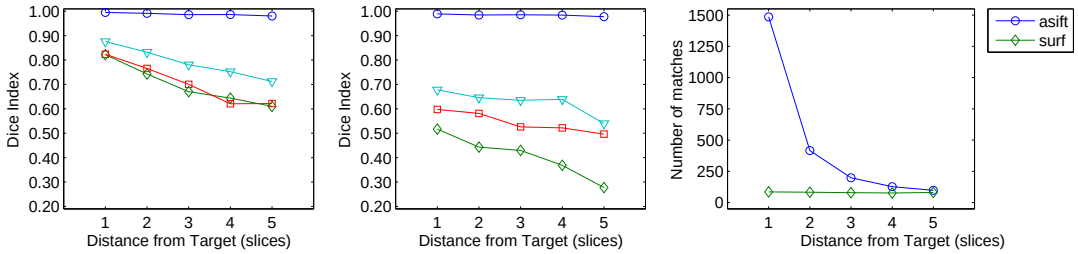


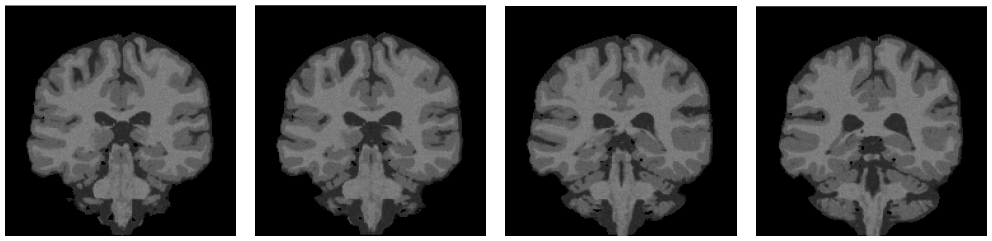
Figure 5. Impact of parameter α on the segmentation quality (Dice coefficient), using the ASIFT algorithm to find atlas-target matches. Dice legend: CSF (◊), GM (□), BG (○), WM (▽).

In terms of efficiency, using the approximation scheme detailed in Section 2.3, our method took about 3-5 seconds to segment the target image, once the matches were obtained. In comparison, ASIFT took 20-30 seconds to compute the matches. Concerning the segmentation results, we point out that no pre-processing step was used to enhance the images. Techniques like Mean-shift [5] could improve the segmentation accuracy by extracting uniform regions in which the labels could diffuse more easily. Also, we note that our method does not directly use the intensity of pixels to infer their class. This could be useful in cases where intensity values are not good indicators of the class, for instance, in functional segmentation. Finally, while this work focused on segmenting brain slices separately, our method could be easily extended to the simultaneous segmentation of all slices in a volume, by connecting the pixels of adjacent slices. Resampling techniques could be used to limit the size of the resulting graph.

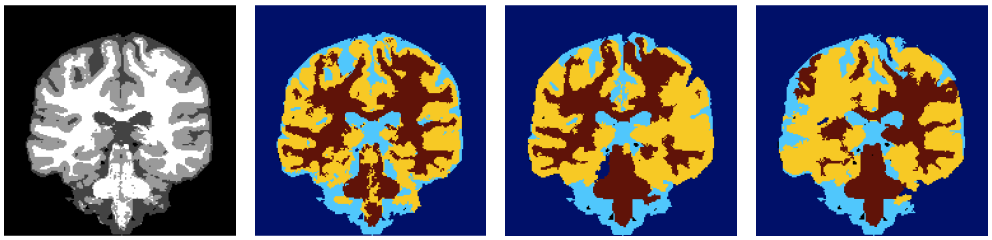
[1] http://www.bic.mni.mcgill.ca/brainweb/



(a) ASIFT (b) SURF (c) #Matches vs atlas distance

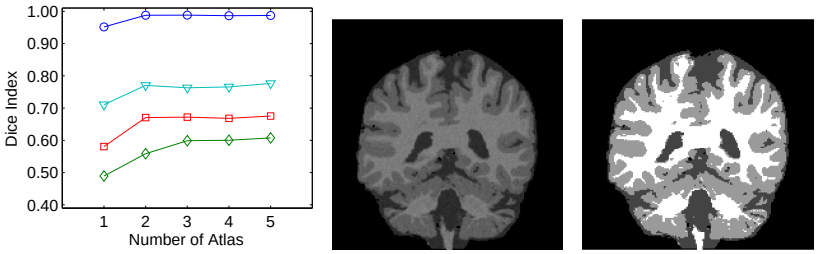


(d) Target (e) Atlas (-1) (f) Atlas (-3) (g) Atlas (-5)

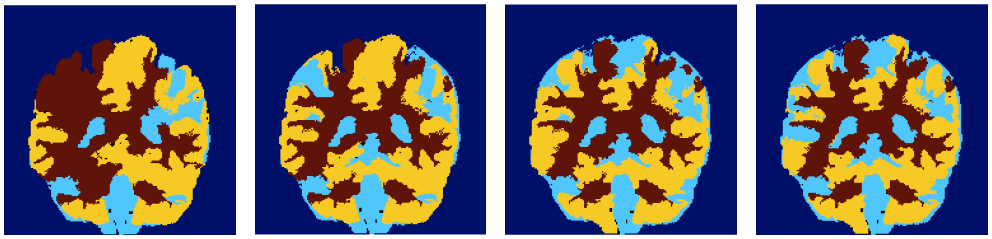


(h) Ground truth (i) Result ($dist$ = 1) (j) Result ($dist$ = 3) (k) Result ($dist$ = 5)

Figure 3. Impact of the deformation level (atlas distance to target slice) on the quality of segmentation (Dice coefficient), using the ASIFT and SURF algorithms to find atlas-target matches. Dice legend: CSF (◊), GM (□), BG (○), WM (▽).



(a) Dice vs #atlases (b) Target (c) Ground truth



(d) Result (N_{atlas} = 1) (e) Result (N_{atlas} = 2) (f) Result (N_{atlas} = 3) (g) Result (N_{atlas} = 4)

Figure 4. Impact of the number of atlases on the segmentation quality (Dice coefficient), using the ASIFT algorithm to find atlas-target matches. Dice legend: CSF (◊), GM (□), BG (○), WM (▽).

4. Conclusion

In summary, this paper presented a novel atlas-based segmentation method using random walks. Unlike most atlas-based approaches, this method combines the registration and label propagation steps in a single efficient framework, and does not depend on a specific deformation model. Experiments on the segmentation of brain MRI have shown the usefulness and efficiency of our method, even without any image pre-processing. As future works, we will investigate the impact of using pre-processing techniques to enhance the diffusion of labels in the target image, test other feature extraction methods than ASIFT and SURF, more suited to medical images such as MRI, and extend our method to the simultaneous segmentation of whole volumes.

References

[1] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *Medical Imaging, IEEE Transactions on*, 28(8):1266–1277, 2009.

[2] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision (ECCV)*, pages 404–417, 2006.

[3] M. Cabezas, A. Oliver, X. Llado, J. Freixenet, and M. Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, 104:158–177, 2011.

[4] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5:425, 1997.

[5] D. Comaniciu and P. Meer. Robust analysis of feature spaces: color image segmentation. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 750–755, 1997.

[6] D. Hill. Neuroimaging to assess safety and efficacy of ad therapies. *Expert Opinion on Investigational Drugs*, 19(1):23–26, 2010.

[7] R. E. Jung, J. M. Segall, H. Jeremy Bockholt, R. A. Flores, S. M. Smith, R. S. Chavez, and R. J. Haier. Neuroanatomy of creativity. *Human Brain Mapping*, 31(3):398–409, 2010.

[8] T. Kim, K. Lee, and S. Lee. Generative image segmentation using random walks with restart. In *European Conference on Computer Vision (ECCV)*, pages 264–275, 2008.

[9] J. Lotjonen, R. Wolz, J. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3):2352–2365, 2010.

[10] K.-O. Lövblad, N. Anzalone, A. Dörfler, M. Essig, B. Hurwitz, L. Kappos, S.-K. Lee, and M. Filippi. MR imaging in multiple sclerosis: review and recommendations for current practice. *American Journal Of Neuroradiology*, 31(6):983–989, 2010.

[11] J.-M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438, 2009.

[12] K. Pohl, J. Fisher, W. Grimson, R. Kikinis, and W. Wells. A bayesian model for joint segmentation and registration. *Neuroimage*, 31(1):228–239, 2006.

[13] G. Postelnicu, L. Zöllei, and B. Fischl. Combined volumetric and surface registration. *IEEE Trans. Med. Imaging*, 28(4):508–522, 2009.

[14] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428–1442, 2004.

[15] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999.

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

[17] B. Vemuri, J. Ye, Y. Chen, and C. Leonard. A level-set based approach to image registration. In *Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on*, pages 86–93, 2000.

[18] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2008.

[19] S. Warfield, K. Zou, and W. Wells. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903–921, 2004.