Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote

Michaël Sdika

Combining Atlas Based Segmentation and Intensity Classification with Nearest Neighbor Transform and Accuracy Weighted Vote

Michaël Sdika[*,1]

Centre de Résonnance Magnétique Biologique et Médical, CNRS UMR n°6612, Faculté de Médecine de Marseille, Université de la Méditérranée, 27 Bd Jean Moulin, 13005 Marseille, France

Abstract

In this paper, different methods to improve atlas based segmentation are presented. The first technique is a new mapping of the labels of an atlas consistent with a given intensity classification segmentation. This new mapping combines the two segmentations using the nearest neighbor transform and is especially effective for complex and folded regions like the cortex where the registration is difficult. Then, in a multi atlas context, an original weighting is introduced to combine the segmentation of several atlases using a voting procedure. This weighting is derived from statistical classification theory and is computed offline using the atlases as a training dataset. Concretely, the accuracy map of each atlas is computed and the vote is weighted by the accuracy of the atlases. Numerical experiments have been performed on publicly available *in vivo* datasets and show that, when used together, the two techniques provide an important improvement of the segmentation accuracy.

Key words: atlas based segmentation, intensity classification, distance transform, nearest neighbor transform, multi atlas segmentation, weighted vote, accuracy map, cortical segmentation

1. Introduction

Atlas based segmentation is now an established technique in medical image processing. It involves using a non-rigid registration algorithm to find a geometric transformation from the subject image to a pre-labeled atlas. The transformation is then used to map the preexistent segmentation of the atlas on the subject [12][3][22]. When the images to be segmented are all the same up to a spatial deformation, this approach allows to accurately segment all the structures of the subject image in a single pass. Segmentation errors can occur if the registration is trapped in a local minimum. Typically, registration will fail when the structures to match are complex, like in the cortical area of human brain image, or when the contrast-to-noise ratio is low.

To improve the result of atlas based segmentation, one can use an atlas set instead of a single atlas; one can also try to combine the result with other segmentation methods. In this paper, a multi atlas method and a multi segmentation method are proposed and it is shown that, when used together, these methods produce a substantial improvement over state of the art techniques.

Atlas based segmentation is often combined with an intensity classification method. Pixels are, with these methods, clustered based on their intensity. Classes can be found directly on the histogram [13] but models can also be much more complex to account for dependencies between pixels and bias field inhomogeneities [24][27]. Atlases can also be used in this framework to provide spatial priors to the classification [21][15]. Typical use of these intensity classification methods includes segmentation of brain images in white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF). In [6], the intensity classification is performed using a more complete atlas: the atlas not only contains information about the probability of a class at a given pixel but also about the intensity distribution given the label and about the distribution of the class of its neighbors. The same procedure has also been applied to surface parcellation in [7]. More recently, registration and intensity classification have been described by a joint model as in [25],[14] or [26]. Solutions to these models are found by iteratively performing registration and segmentation, each step using the output of the previous one. The registration is improved by the information provided by the intensity classification, and spatial priors provided by the registration are used for intensity classification. The atlas based segmentation can also be combined with the results of intensity classification by subdividing the labels mapped from the atlas with the different classes of the tissues classification. This technique was introduced in [11] for the separation of left and right tissues segmentation. It works well when the structures of the atlas are correcty matched. In [4], an atlas based segmentation coherent with an intensity classification method is proposed. The atlas provides some maps of the most probable label given the tissue type (WM, GM or CSF) and the labeling is made after intensity classification using the label map corresponding to the tissue. This technique has been specifically designed for the cortical region where the registration is likely

Email addresses: michael.sdika@univmed.fr (Michaël Sdika)
URL: http://crmbm.timone.univ-mrs.fr (Michaël Sdika)

[1]This work is supported by CNRS (UMR 6612) and Institut Universitaire de France

to fail but where intensity classification performs well.

To improve atlas based segmentation, some authors have also proposed to use a set of atlases instead of a single atlas. Results are then combined to create the best agreement segmentation. In [23] an EM based procedure is proposed to find the best agreement between the binary segmentations of a group of experts. This method has been extended in [19] to multi class segmentation and applied to the problem of atlas based segmentation with multiple atlases. The best agreement can also be found by shape based averaging as in [18]. For each label, distance maps to the segmentation of each atlas are averaged. The output label is chosen to minimize the mean distance maps. The use of a voting procedure has been investigated in [16] [17] or [8] and although simple, this method produces good results. In [2] and [9], a pixelwise weight is given to the vote of each atlas. Weights are a function of the registration residual in [9] and a distance (e.g. normalized cross correlation or mutual information) between the neighborhood of the pixel in the subject and the atlas in [2].

The present paper contains several contributions. First we propose a mapping of the atlas labels consistent with the intensity classification segmentation of the subject and show how it can be used to segment cortical structures. For the multi atlas part, the concept of accuracy map of an atlas is introduced, an estimator of this quantity is proposed as well as its use as an effective weight in a voting procedure. This new weighting is particularly efficient when different methods are used to label an image and it is shown that, together, the two techniques produce a substantial improvement of the segmentation. For each method, numerical experiments have been conducted on *in vivo* data.

2. Combining Atlas Based Segmentation and Intensity classification with Nearest Neighbor Transform

In this section, it is assumed that a single atlas (a grey level image and its manual labeling) is available. An intensity classification method is also available to segment any grey level image into different classes. The objective is to make the atlas based segmentation of a new subject consistent with its intensity classification as in the ANIMAL+INSECT method described in [4]. However, our method does not require the use of a probabilistic atlas. Using intensity classification will be especially useful in cortical region where local minima of the registration induce frequent defect of the atlas based segmentation (see example on Fig 1). To be clear in our presentation, *label* or *labeling* will be used for the atlas based segmentation and *class* or *classification* will be used for the intensity classification segmentation.

2.1. Standard Atlas based Labeling

Let I_a be the grey level image of the atlas and L_a be the image of the labels of interest. Now, let I_s be the grey level image of the subject to be labeled and T, the transformation from the subject to the atlas found by the registration. The classic way to label the subject image is simply to map the label image L_a with the transformation:

$$L_s(x) = L_a(T(x)). \tag{1}$$

2.2. Atlas Labeling using the Nearest Neighbor Transform

Let's assume now that both I_a and I_s have been segmented into C_a and C_s using the intensity classification method. The labeling proposed here assigns to each pixel of the subject the label of the closest atlas point of the same class. As a consequence, the subject have a labeling coherent with the classification: GM pixels are mapped to GM pixels... To compute the closest point, the Euclidean distance in the subject coordinate system is used.

In concrete terms, the labels and the classes of the atlas are first mapped on the subject:

$$L_s^a(x) = L_a(T(x)),$$

$$C_s^a(x) = C_a(T(x)),$$

then, for each class, the nearest neighbor transform to the current atlas class is computed:

$$p(x, c) = \underset{C_s^a(y)=c}{\operatorname{argmin}}\, d(x, y), \tag{2}$$

where d is the Euclidean distance. Finally, the label assigned to a pixel x is the label of its nearest neighbor in its class:

$$L_s(x) = L_s^a(p(x, C_s(x))).$$

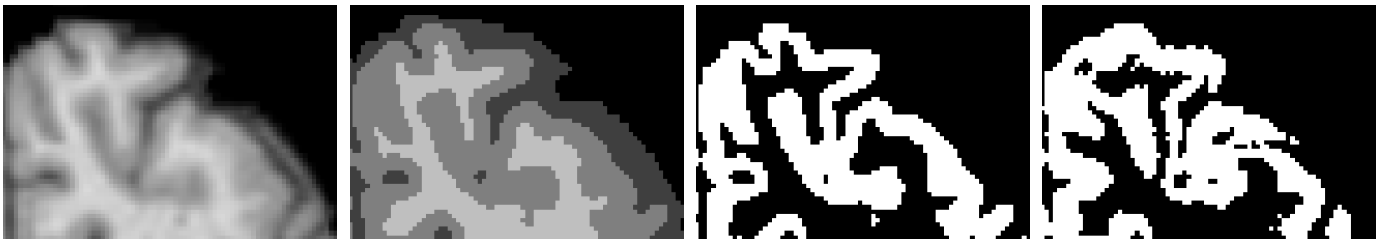
2.3. Improving the Robustness

An important requirement of the new labeling is that borders of L_a and C_a match exactly when they are supposed to. Indeed, as the border pixels are also the nearest neighbors, any mismatch between the two segmentations will have a strong influence on the final labeling. Ideally, L_a and C_a have been produced together with, for example, C_a used as starting point of the manual process to create L_a.

In practice, such a C_a map is not always available or may not be reproduced by the automatic intensity classification that will be used on the subject. A solution is to create it by running the intensity classification on the atlas image. In this case, as L_a and C_a have been produced independently, their borders will mismatch. The labeling can be significantly improved if a binary morphological erosion is applied to each class of the atlas before computing the nearest neighbor transform. The erosion will remove the border of the current class and make the nearest neighbor of each point clearly inside the class. The nearest neighbor computation (eqn. 2) is just replaced by:

$$p_e(x, c) = \underset{y \in C_{s,c}^{e,a}}{\operatorname{argmin}}\, d(x, y),$$

where $C_{s,c}^{e,a}$ is the result of an erosion of the $C_s^a = c$ mask.



(a) (b) (c) (d)

Figure 1: A brain image (a), its automatic intensity classification (b), the manual segmentation of the cortex (c), automatic delineation of the cortex with standard atlas based segmentation (d). A local minimum in the registration makes the segmentation (d) fail while the intensity classification (b) is good for this part of the brain.

3. Multi Atlas Segmentation with Accuracy Weighted Vote

In this section, the concept of accuracy map of an atlas is introduced. A practical estimator for accuracy maps and their use as an effective weighting in a voting procedure are also presented.

The accuracy of a classifier is the likelihood that the output of the classifier is correct. Here, each pixel of the atlas is seen as a classifier. So the value of the accuracy map at pixel x is the probability that pixels segmented with x will be correctly labeled.

Note that, here, atlas refers to the grey level image, its segmentation and the labeling method used to segment other images. If the standard and the nearest neighbor transform based labeling are used, the number of atlases and accuracy maps will be twice the number of pre-segmented images.

Examples of accuracy maps of an atlas with the two labelings are shown on Fig.2. As expected, segmentation with nearest neighbor transform based labeling is better on cortical regions whereas standard labeling is better on deep grey structures. This result is a direct translation to the new labeling of the behavior of the intensity classification algorithm. Indeed, the typical three classes Gaussian mixture model used is not really suited due to the intensity difference between deep grey and cortical grey matter, and consequently deep grey matter are problematic for theses methods. For example, parts of the thalamus has an intensity similar to white matter and consequently, are often classified as such.

3.1. An Estimator for the Accuracy

The accuracy map of an atlas is estimated here using the remaining pre-segmented images as a training dataset. Goodness of segmentation maps are built using each of the other images. These maps are then averaged in the atlas coordinate system. Formally, our dataset is composed of N_a images $(I_i)_i$ along with their pre-existing segmentations $(L_i)_i$. To estimate the accuracy map of the first atlas, the (non-rigid) transformation T_i from I_i to I_1 is computed and used to produce $\tilde{L}_i$, the automatic segmentation of the i^{th} images by the first atlas. The goodness map is then computed as follows:

$$g_i(x) = \begin{cases} 1 & \text{if } L_i(x) = \tilde{L}_i(x) \\ 0 & \text{otherwise.} \end{cases}$$

The accuracy map q_1 of the first atlas is then estimated by averaging the goodness maps in its own coordinate system:

$$q_1(x) = \frac{1}{N_a - 1} \sum_{i \neq 1} T_i^{-1}(g_i(x)). \tag{3}$$

This procedure is then repeated to compute the accuracy map of each atlas.

3.2. Combining Results with Accuracy Weighted Vote

When a set of atlases is available for the segmentation, voting is a simple and efficient way to combine the result of the segmentations. Each atlas is used to segment the subject, then, the final label at a given pixel is the most frequent one. In basic implementation of voting procedure, each atlas has an equal vote. Consequently, the vote does not account for spatial systematic error that an atlas can produce. We propose to weight the vote of each atlas with its accuracy map. When the label image of an atlas is mapped in the subject reference, the accuracy map is mapped as well and used as weight in the vote. Thus more weight is given to more accurate pixels.

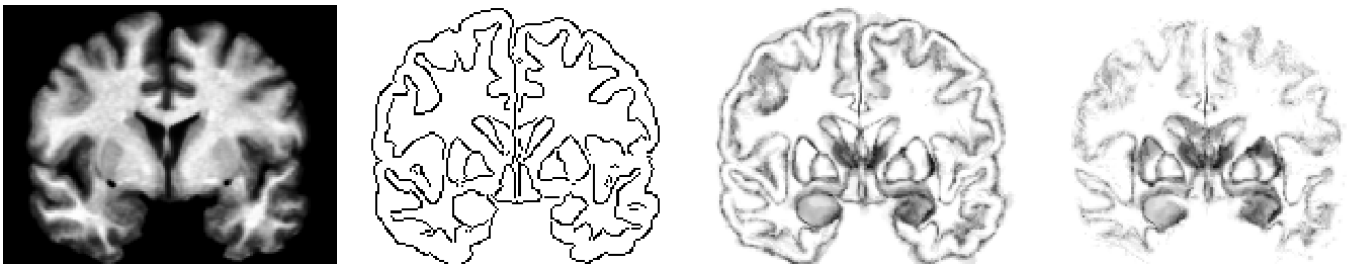
Formally, the label chosen by the vote procedure is:

$$\underset{l}{\text{argmax}} \sum_{L_n(T_n(x))=l} w_n(x),$$

where l index the labels, L_n is the ground truth segmentation of the n^{th} atlas and T_n the transformation from the subject to the n^{th} atlas. When the vote is unweighted, the w_n coefficient is constant ($w_n(x) = 1$). For accuracy weighted vote, w_n is the corresponding accuracy of the n^{th} atlas at this location: $w_n(x) = T_n(q_n(x))$.

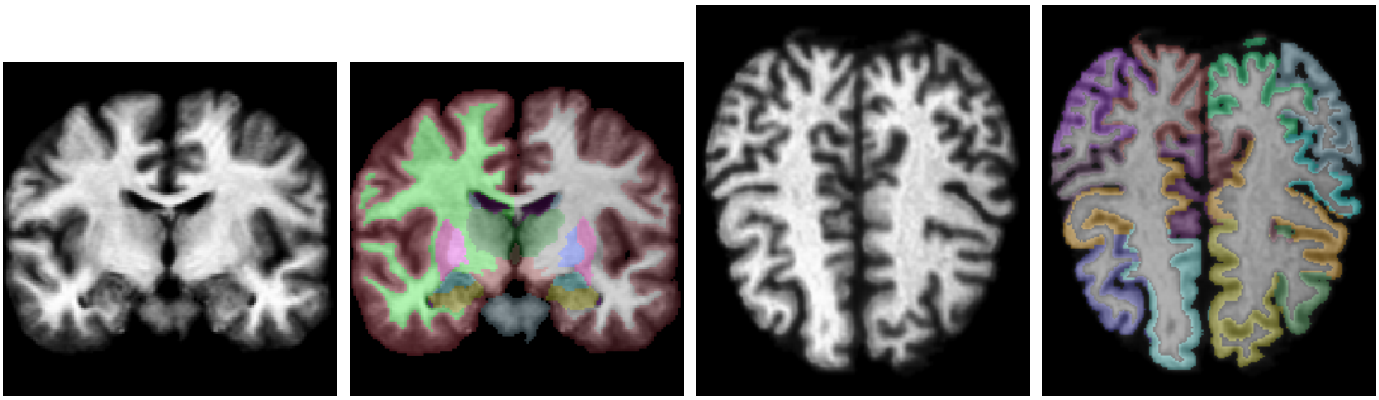
4. Implementation Details

Any non-rigid registration method can be used to compute the transformation from the subject image to the atlas. However, a topology preserving registration would ensure that structures of the atlas are not lost when mapped on the subject. For the experiment of this paper, the method described in [20] has been used. The transformation is modeled by cubic B-spline (with a node spacing of 6 pixels) and non invertibility is penalized by enforcing positivity of the Jacobian on the pixels and



(a) (b) (c) (d)

Figure 2: An atlas (a), the contour of its manual segmentation (b), its accuracy maps for the standard labeling (c), and for the nearest neighbor transform based labeling (d). Bright areas of the accuracy maps are regions where the atlas performs well.



(a) (b) (c) (d)

Figure 3: An atlas from the IBSR dataset (a) and its manual segmentation (b). An atlas from the NA0 dataset (c) and its manual segmentation (d).

penalizing negative Jacobian between the pixels. A least square cost function as been used within a multi-resolution framework.

Invertible transformations are also preferable for the computation of accuracy maps with formula 3. The inverse of a transformation T is computed using a Levenberg-Marquard routine on the following problem:

$$T^{-1}(x) = \underset{y}{\operatorname{argmin}} \|x - T(y)\|_2^2 .$$

The distance and nearest neighbor transform have been computed in linear time using the algorithm of [5].

As it is the case for registration, any method can be used for the intensity classification. For our experiments, the method of [27] has been used. It alternatively estimates the tissue class label of pixels (assumed to be the realizations of a Markov random field), the parameters of the Gaussian mixture modeling the intensity distribution and the bias field inhomogeneities.

The resampling of label images mapped through a geometric transformation has been done using partial volume interpolation (PVI) (see [10]). To label a continuous point, its 8-nearest neighbors vote for their own label with a weight taken as their coefficient in a trilinear interpolation procedure. The label with the highest vote is assigned to the point. In our experience, this resampling technique provides better results than nearest neighbor interpolation. Mapping of goodness of segmentation maps and accuracy maps, have been done using trilinear interpolation to ensure that their values remain in $[0, 1]$.

5. Numerical Experiments

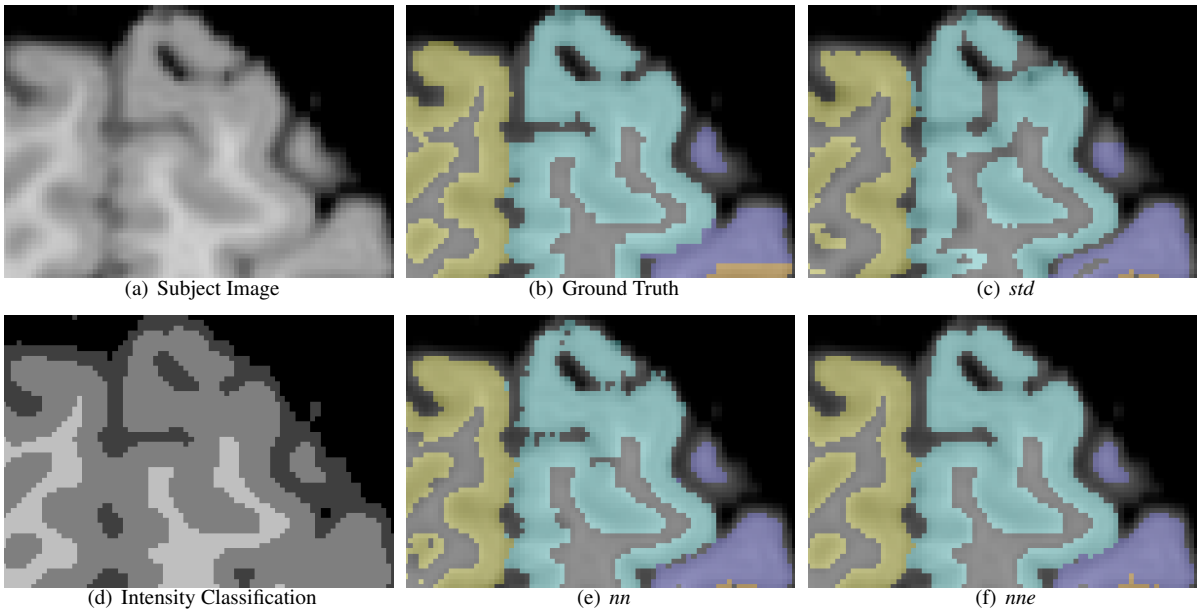
Numerical experiments have been performed using two publicly available *in vivo* datasets. The NA0 dataset provided by the Non-Rigid Image Registration Evaluation Project[2] (NIREP) consists of 16 3D MR images of healthy human brains. On each image, the cortex has been segmented into 32 grey matter regions of interest using a combination of manual and automated procedures. The IBSR dataset provided by the Center for Morphometric Analysis at Massachusetts General Hospital[3] contains 18 images of healthy brains and the corresponding segmentation of the whole brain into 32 structures. Example images from these two dataset are shown on figure 3.

The relative overlap (RO) and the mean accuracy (MA) has been used to assess the quality of an automated segmentation L_a when a ground truth L_g is available. The RO of the structure s is given by the ratio of the intersection of the two segmentations to their union and the MA is the percentage of correctly segmented pixel in the volume of interest (the whole brain):

$$RO(L_g^s, L_a^s) = 100 \frac{V\left(L_g^s \cap L_a^s\right)}{V\left(L_g^s \cup L_a^s\right)},$$

[2] http://www.nirep.org/

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



(a) Subject Image (b) Ground Truth (c) *std*

(d) Intensity Classification (e) *nn* (f) *nne*

Figure 4: Results of the cortical segmentation with the standard atlas based labeling (*std*), the nearest neighbor transform based labeling (*nn*) and the nearest neighbor transform with erosion labeling (*nne*).

$$MA(L_g, L_a) = 100 \frac{V\left(L_g = L_a\right)}{V\left(L_g\right)}.$$

For each experiment, different methods are compared using a leave one out procedure. Each image is used as a subject with a ground truth segmentation and automatically labeled using the others as atlases. The average *RO* or *MA* is collected for each subject and finally, mean and standard deviation over the different subjects are reported.

5.1. Cortical Segmentation using Nearest Neighbor Transform Labeling

The NA0 database has been used to evaluate the nearest neighbor transform based labeling. The leave one out procedure is used to compare the three labelings presented in section 2. Results are reported on Fig. 5(a) for the standard labeling (*std*), the nearest neighbor transform based labeling of section 2.2 (*nn*) and the nearest neighbor transform with erosion labeling of section 2.3 (*nne*). The structures of interest are all in the cortex, in the left part of the brain (similar results are obtained for the right hemisphere). The full names of all cortical structures are given in appendix B.

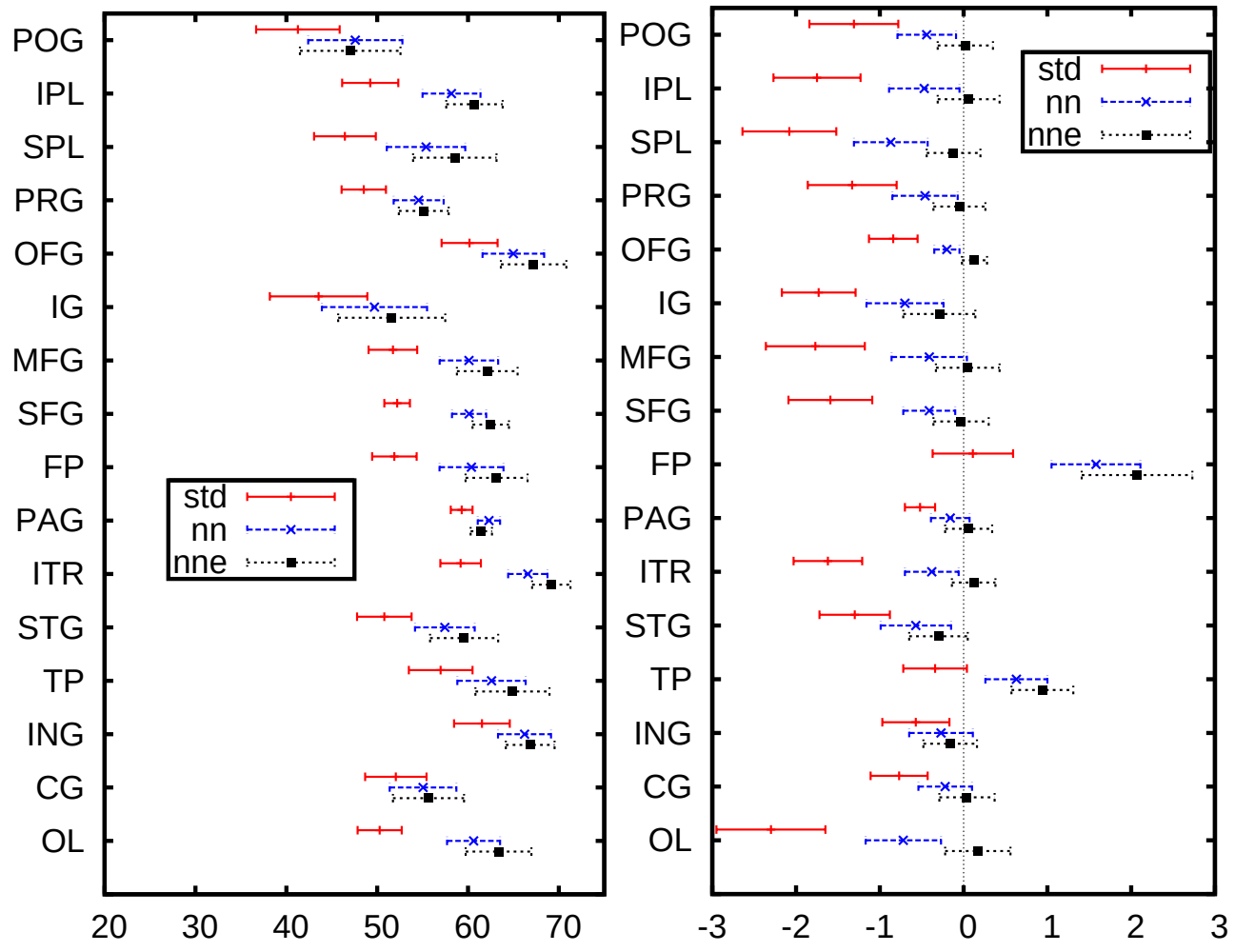
As shown in Fig. 5(a), *nn* and *nne* provide a substantial improvement over standard labeling for all the structures. Indeed, registration is likely to fail in cortical region where structures are complex and folded. It can also be seen that *nne* provide a further improvement over *nn* for most of the structures.

Figure 4 displays the results of the three labelings in a region where the registration fell into a local minimum. The folding of the cortex has not been correctly recovered by the registration leading to the failure of *std*. With *nn*, the labeling respects the tissue classification computed on the subject image and the result is coherent with the topology of the brain as seen by the intensity classification. However, the segmentation produced with *nn* is somewhat rough and seems inaccurate at the border of the structures. The *nne* labeling preserves the advantages observed with *nn* but the segmentation is cleaner and smoother.

During the registration, constraints are imposed to be able to invert the transformation. One can wonder if these constraints are beneficial or not to the to the labeling of cortical structures. The same experiments has been done a second time with an unconstrained registration and the *RO* difference between constrained and unconstrained registration is reported on Fig. 5(b), for each structure and each labeling (difference is positive when constrained registration is better). One can see that, for the *std* labeling, unconstrained registration is preferable for the segmentation of cortical structures. However, with *nn*, the difference is less pronounced and with *nne*, no difference is visible. When used in an accuracy weighted vote context, only constrained registration can be used for the computation of the accuracy maps as the inverse of the transformation is needed. Note that this imposes the use of a constrained registration during the labeling itself for a good representativity of the accuracy maps.

5.2. Multiple Atlas Segmentation with Accuracy Weighted Vote

The accuracy weighted vote has been evaluated on the IBSR dataset. *std* and *nne* labeling (or both) are used with a basic majority vote (*bv*) or the proposed accuracy weighted vote (*awv*). These methods are evaluated and tested using a leave one out procedure for different numbers of atlases. Care has been taken not to use the ground truth labels of an atlas when it is used as the subject. In particular, when a set of N atlases is used to label a subject, the corresponding accuracy maps are built from these N atlases exclusively.



(a) *RO* on cortical structures

(b) *RO* Differences between constrained and unconstrained registration

Figure 5: Results of the leave one out evaluation for cortical segmentation with the standard labeling (*std*), the nearest neighbor transform based labeling (*nn*) and the nearest neighbor transform with erosion labeling (*nne*). Labels abbreviations are given in appendix B (a) relative overlap on each structure with constrained registration and the three labelings, (b) the difference of relative overlap between constrained and unconstrained registration (difference is positive when constrained registration is better).

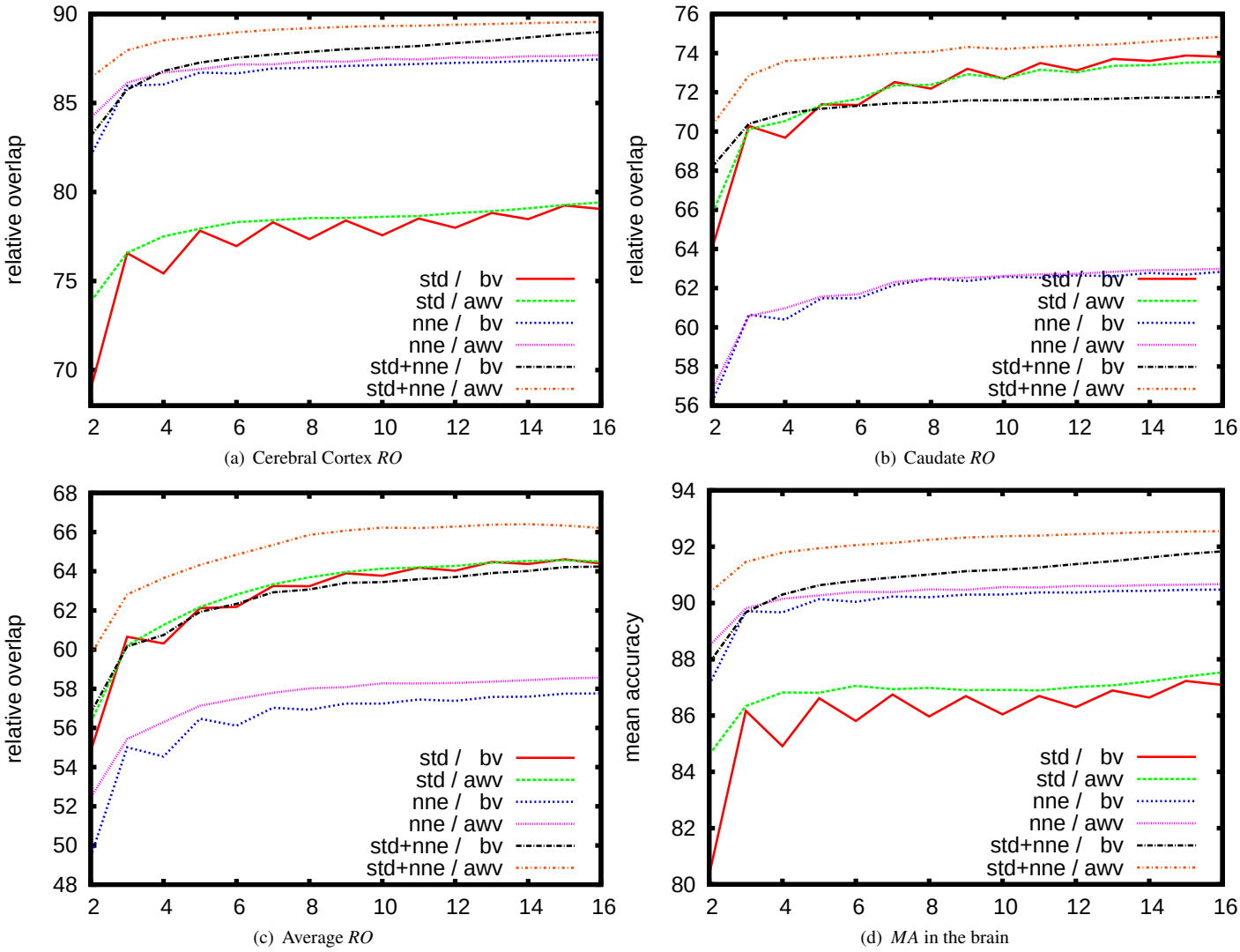
On Fig. 6 are displayed the *RO* for the cerebral cortex and the caudate as well as the mean *RO* over all the structures and the *MA* in the brain. On Fig. 7, the mean and standard deviation of the *RO* are reported for each structure when 3 or 16 atlas are used.

One can first see on all graphs of Fig. 6 that for a given labeling, the mean *RO* is higher with *awv* than with *bv*. On Fig. 7, when both labelings are used, this result is confirmed for all structures with the exception of CuWM when 3 atlases vote.

As the intensity classification used performs poorly on deep grey structures, segmentation with *nne* alone gives bad results on the caudate (Fig. 6(b)). For this structure, all five other methods perform better, especially when *awv* is used. On the contrary, for the left cerebral cortex (Fig. 6(a)), *nne* performs better than *std* and vote with both labeling provide a further important improvement over *nne*. Note that for *std +nne*, a high number of atlases is required by *bv* to reach the performance that *awv* achieve with few atlases. One can also see that, whatever the number of atlases, about 15 points improvement is achieved when both labelings are combined with *awv* compared to a standard atlas based segmentation with a majority vote.

If the average over all the structures is considered (Fig. 6(c)), the methods can be classified in three groups. *nne* with either *bv* or *awv* which perform poorly. Then come *std* with either vote and *std +nne* with basic vote. This group achieves an improvement of about 6 points over *nne*. The best results are obtained by accuracy weighted vote with both labelings for which the mean *RO* is 4 points higher than the second group. Indeed, as shown in Fig. 7 *nne* is better than *std* only on few structures: CuCo, CuWM, CaCo and CaWM. Most of the time, *std +nne* is better or as good as the best of *std* and *nne* with either vote technique. For ventricular structures (LatVe, 3Ve and 4Ve) *bv* does not perform as well as *awv* to the addition of the poorer *nne* segmentation in the vote.

On the mean accuracy graph (Fig. 6(d)), the ranking of *nne* and *std* is exchanged compared to the mean *RO*. Indeed, *std*



(a) Cerebral Cortex *RO* (b) Caudate *RO*

(c) Average *RO* (d) *MA* in the brain

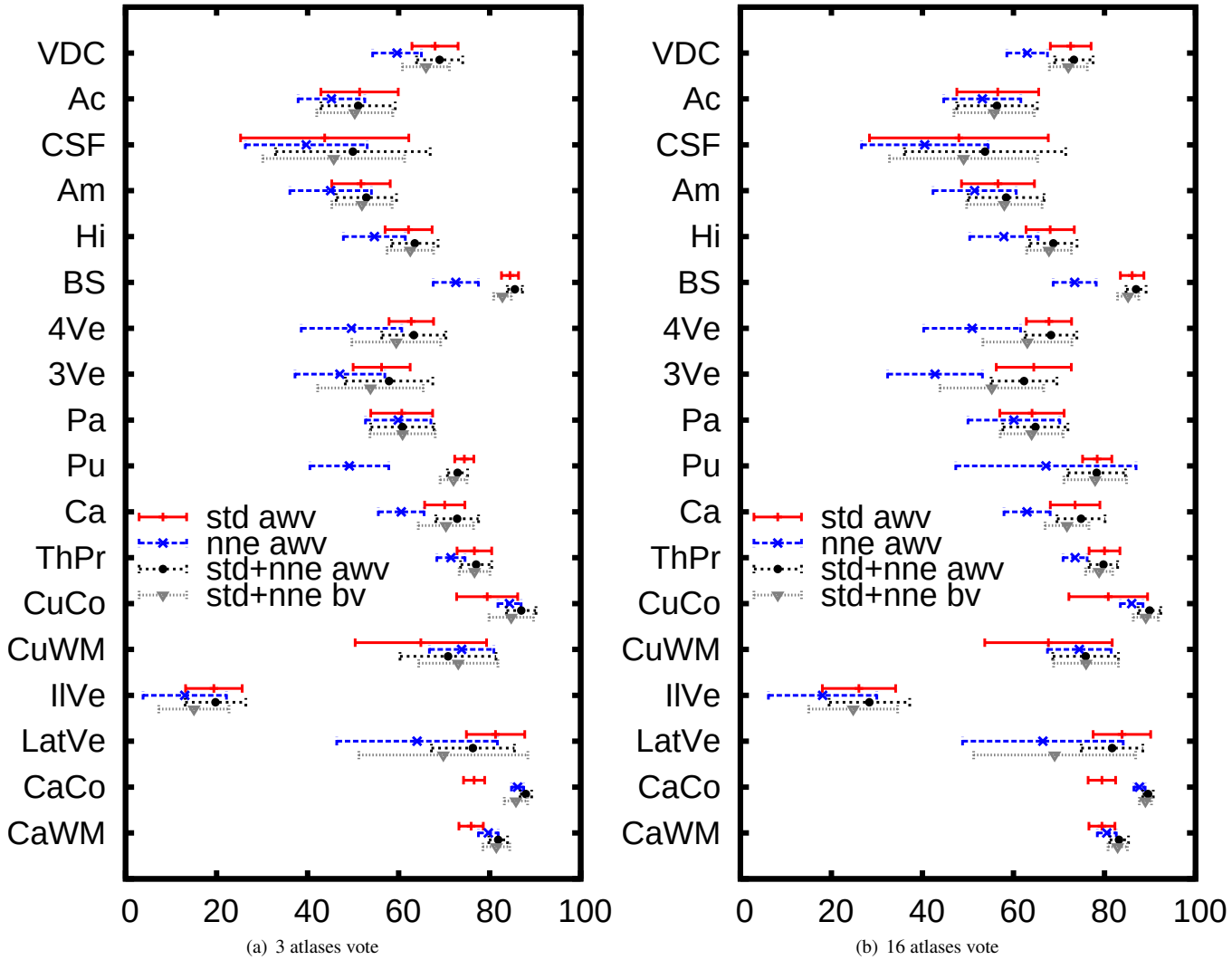
Figure 6: Mean *RO* or *MA* as a function of the number of atlases. The graphs are plotted for the left cerebral cortex *RO* (a), the left caudate *RO*(b), the average *RO* over all the structures (c) and a mean accuracy in the brain (d). Each graph shows the results for combinations of basic vote (*bv*) and accuracy weighted vote (*awv*) with *std*, *nne* or both labelings.

performs better on more structures than *nne* but the few structures for which *nne* is better have larger volumes. However, in agreement with *RO* measures, *awv* always performs better than *bv*. One can finally see that for all the measures we used, accuracy weighted vote using the two labelings together is always the best method and clearly outperforms the other methods.

6. Discussion

In this article, several contributions to atlas based segmentation have been proposed. First, a new labeling technique to combine the output of atlas based segmentation with intensity classification segmentation has been presented. Each pixel is labeled by the closest point of the atlas belonging to the same tissue class. As a consequence, labels are assigned coherently with the intensity classification. This labeling is beneficial for highly folded structures such as the cortex, decently segmented by intensity classification, but where the registration is difficult to achieve. For example, it can be used to segment substructures of the cortex such as Broadmann areas.

Contributions have also been made to multi-atlases segmentation. The concept of accuracy map of an atlas has been introduced to visualize where an atlas performs well. The accuracy map of each atlas is computed offline using the other atlases as a training dataset. The best agreement segmentation of different atlases is then chosen by a vote of each atlas weighted by its accuracy. The principle of this accuracy weighted vote (*awv*), is to give more weight where an atlas "usually" perform well. This weighting is especially useful when the new labeling (*nne*) is used in conjunction with the standard atlas based labeling (*std*). Roughly, *nne* is better on cortex and *std* is better on deep grey structures and *awv* will automatically discover and utilize this information when combining several atlases for segmentation. As a consequence, when *std* and *nne* are combined using accuracy weighted vote, a substantial improvement is achieved over a vote with the best of the two labelings.



(a) 3 atlases vote (b) 16 atlases vote

Figure 7: *RO* mean and standard deviation for all left hemisphere structures. *std* is the standard labeling, *nne* is the new labeling. *bv* is the basic majority vote, *awv* is the new accuracy weighted vote. Results are given for a vote with 3 atlases (a) or 16 atlases (b). Labels abbreviations are given in appendix A

The local weighting proposed in [2] or [9] are both based on local difference between the subject and the atlas. The aim is to weight the vote by a measure of the confidence one can have in the registration of a given pair atlas/subject. The local weighting proposed in the current paper is derived from statistical arguments and the vote is weighted by a measure of the confidence in the atlas itself (and its associated mapping scheme). Future work may include a combination of the two approaches.

It is now established that atlas based segmentation can be improved by using multiple atlas combined by a vote fusion rule. In [17], it is proposed to let the registration parameters vary instead of using new atlases. They found that better results are obtained when several atlases is used. The methods proposed in the current paper allow to let the mapping scheme vary (as well as the atlases). Nearest neighbors based mapping make the mapping scheme change by changing a third party segmentation. The accuracy weighted vote find the best combination of the different mapping schemes used.

In the current paper, only the standard labeling and the neighbors labeling with the method of [27] has been used. No atlas is used for the intensity classification of a new subject. As the registration between the subject and the atlases is performed anyway, more information can be accumulated in the atlas and used during the intensity classification as in [15] or [6]. Note that these methods can be used together with the classification of [27] and not as a replacement. In this case, all theses intensity classifications and the standard labeling would be combined by the accuracy weighted vote.

In term of computation time, the additional cost of *nne* is the cost of the intensity classification. Indeed, the computation of the nearest neighbors transform for few tissues classes is negligible compared to the registration. In the multi atlases perspective, the most time consuming part of the process is also the non-rigid registration to each atlas. If computation time is a limiting constraint, the vote can be performed with a reduced number of atlases. The complete dataset can however be used

offline to learn the accuracy map of each atlas. Note that, atlas selection strategies, as described for example in [16] or [1], can be used to select a subset of the dataset. In this case, accuracy maps can be part of a new criterion in the selection procedure. Indeed, atlases with high accuracy are, by definition, likely to perform better during the segmentation.

A. Abbreviations for the IBSR database

CaWM	cerebral white matter
CaCo	cerebral cortex
LatVe	lateral ventricle
IlVe	inferior lateral ventricle
CuWM	cerebellum white matter
CuCo	cerebellum cortex
ThPr	thalamus proper
Ca	caudate
Pu	putamen
Pa	pallidum
3Ve	3rd ventricle
4Ve	4th ventricle
BS	brain stem
Hi	hippocampus
Am	amygdala
CSF	cerebro spinal fluid
Ac	accumbens area
VDC	ventraldc

B. Abbreviations for the NA0 database

POG	postcentral gyrus
IPL	inferior parietal lobule
SPL	superior parietal lobule
PRG	precentral gyrus
OFG	orbital frontal gyrus
IG	inferior gyrus
MFG	middle frontal gyrus
SFG	superior frontal gyrus
FP	frontal pole
PAG	parahippocampal gyrus
ITR	inferior temporal region
STG	superior temporal gyrus
TP	temporal pole
ING	insula gyrus
CG	cingulate gyrus
OL	occipital lobe

References

[1] Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D., 2007. Classifier selection strategies for label fusion using large atlas databases. In: Medical Image Computing and Computer-Assisted Intervention MICCAI 2007. pp. 523–531.

[2] Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C., Aug 2009. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans Med Imaging 28 (8), 1266–1277.

[3] Collins, D. L., Holmes, C. J., Peters, T. M., Evans, A. C., 1995. Automatic 3d model-based neuroanatomical segmentation. Human Brain Mapping 3 (3), 190–208.

[4] Collins, L. D., Zijdenbos, A. P., Baare, W. F. C., Evans, A. C., 1999. Animal+insect: Improved cortical structure segmentation. In: IPMI '99: Proceedings of the 16th International Conference on Information Processing in Medical Imaging. Springer-Verlag, pp. 210–223.

[5] Felzenszwalb, P. F., Huttenlocher, D. P., September 2004. Distance transforms of sampled functions. Tech. rep., Cornell Computing and Information Science.

[6] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A. M., Jan 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

[7] Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Sgonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A. M., Jan 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.

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

[9] Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B., Jul 2009. Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in CT scans. IEEE Trans on Medical Imaging 28 (7), 1000–1010.

[10] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. Medical Imaging, IEEE Transactions on 16 (2), 187–198.

[11] Maes, F., Leemput, K. V., DeLisi, L. E., Vandermeulen, D., Suetens, P., 1999. Quantification of cerebral grey and white matter asymmetry from MRI. In: MICCAI '99: Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer-Verlag, London, UK, pp. 348–357.

[12] Miller, M., Christensen, G., Amit, Y., Grenander, U., 1993. Mathematical textbook of deformable neuroanatomies. In: Proceedings of the National Academy of Science of the United States of America. Vol. 90. pp. 11944–8.

[13] Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9 (1), 62–66.

[14] Pohl, K. M., Fisher, J., Grimson, W. E. L., Kikinis, R., Wells, W. M., 5 2006. A Bayesian model for joint segmentation and registration. NeuroImage 31 (1), 228–239.

[15] Pohl, K. M., Wells, III, W. M., Guimond, A., Kasai, K., Shenton, M. E., Kikinis, R., Grimson, W. E. L., Warfield, S. K., 2002. Incorporating non-rigid registration into expectation maximization algorithm to segment MR images. In: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part I. Springer-Verlag, pp. 564–571.

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

[17] Rohlfing, T., Maurer, Jr., C. R., 2005. Multi-classifier framework for atlas-based image segmentation. Pattern Recognition Letters 26 (13), 2070–2079.

[18] Rohlfing, T., Maurer, Jr., C. R., Jan. 2007. Shape-based averaging. IEEE Transactions on Image Processing 16 (1), 153–161.

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

[20] Sdika, M., 2008. A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. Medical Imaging, IEEE Transactions on 27 (2), 271–281.

[21] Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., Oct 1999. Automated model-based tissue classification of MR images of the brain. Medical Imaging, IEEE Transactions on 18 (10), 897–908.

[22] Warfield, S., Robatino, A., Dengler, J., Jolesz, F., Kikinis, R., 1998. Nonlinear Registration and Template Driven Segmentation. Brain Warping, Ch. 4, pp. 67–84.

[23] Warfield, S. K., Zou, K. H., William M. Wells, I., 2002. Validation of image segmentation and expert quality with an expectation-maximization algorithm. In: Proceedings of the 5th International Conference on Medical

Image Computing and Computer-Assisted Intervention-Part I. Springer-Verlag, pp. 298–306.

[24] Wells, W., Grimson, W., Kikinis, R., Jolesz, F., 1996. Adaptive segmentation of MRI data. IEEE Trans Med Imaging. 15 (4), 429–442.

[25] Wyatt, P. P., Noble, J. A., 12 2003. Map MRF joint segmentation and registration of medical images. Medical Image Analysis 7 (4), 539–552.

[26] Yeo, B. T., Sabuncu, M. R., Desikan, R., Fischl, B., Golland, P., 2008. Effects of registration regularization and atlas sharpness on segmentation accuracy. Medical image analysis 12 (5), 603–615.

[27] Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. IEEE Transactions on Medical Imaging 20 (1), 45–57.