Subcortical Structure Segmentation using Probabilistic Atlas Priors

Sylvain Gouttard[1], Martin Styner[1,2], Sarang Joshi[3], Brad Davis[2], Rachel G. Smith[1], Heather Cody Hazlett[1], Guido Gerig[1,2]

[1] Department of Psychiatry, University of North Carolina at Chapel Hill, NC
[2] Department of Computer Science, University of North Carolina at Chapel Hill, NC
[3] Department of Biomedical Engineering, University of Utah, Salt Lake City, Utah

Abstract. The segmentation of the subcortical structures of the brain is required for many forms of quantitative neuroanatomic analysis. The volumetric and shape parameters of structures such as caudate are employed to characterize a disease or its evolution. This paper presents our fully automatic segmentation of the caudate. The segmentation is based on an unbiased diffeomorphic atlas with probabilistic spatial priors built from a training set of MR images with corresponding manual segmentations. When applying the atlas for automatic structural segmentation, an MR image is first intensity inhomogeneity corrected, skull stripped and intensity calibrated to the atlas. Then the atlas image is registered to the image using an affine followed by a deformable registration matching the gray level intensity. Finally, the registration transformation is applied to the caudate probabilistic maps, which are then thresholded at 0.5 probability.
Our method has been tested on all datasets provided by workshop as our atlas was build on a separate training population. The results show intermediate overlap results (76% Dice) and high correlation with the IBSR data (93%) and moderate correlation with the BWH data (64%). This indicates that our manual segmentation procedure is more similar to the procedure used for the IBSR than for the BWH dataset.

1 Introduction

Magnetic resonance imaging (MRI) is able to provide a detailed information of normal and diseased anatomy for medical research and has become a significant imaging modality in clinical diagnosis and brain studies. Segmentation of subcortical structures such as the caudate from MR brain scans is a critical task that has many applications such as volume assessments and shape analysis. Even though manual delineation by experts is still common practice for high quality segmentation, it is time-consuming and subjective. Neuroimaging studies tend to become ever larger and manual segmentation with its time requirement and low reproducibility is ill-suited for such large imaging studies.

Many methods exist to perform automatic segmentations of the caudate and subcortical structures in general [1]. Deformable models have been employed in

numerous medical imaging applications [2]. Leventon et al. [3] and Tsai et al. [4] use a shape-based approach to curve evolution for the segmentation of medical images containing known object types. A hierarchical, atlas based expectation-Maximization segmentation algorithm [5] is used by Pohl et al. Of course, many more algorithms are available and will be presented at this workshop.

In this paper we describe our automatic segmentation method presented in detail in [6]. Our approach is based on an unbiased diffeomorphic atlas with probabilistic spatial priors built from a training set of images with corresponding manual segmentations. Using the transformation fields from the previous step we generate probabilistic maps for the caudate. After several preprocessing steps, the atlas is registered to the image using an affine followed by a deformable registration matching the gray level intensities. Lastly, the registration transformations are applied to the probabilistic maps of the caudate, which is then thresholded at 0.5 probability.

2 Method

The method described in this paper is based on the registration of an atlas with probabilistic maps onto the case to be segmented. The whole process can be separated into an atlas building step, which is performed only once, plus a segmentation step performed for each dataset. This section summarizes this methodology, which is described in more detail in [6].

2.1 Atlas computation

Our caudate segmentation method is based on an atlas registration. Prior to any segmentation we need to create the atlas itself and the caudate probabilistic maps in the atlas coordinate space. The atlas is generated in two parts: first a template image is computed as the unbiased average of a training population, then the probabilistic maps are computed in the template image space. The atlas used in this study was created from 10 separate healthy adult control subjects (20 to 55 year old, IRprepped SPGR, GE 1.5) that were not part of the training population provided by the workshop organizers. The probabilstic maps have been generated from manual segmentations performed by a single expert. As we chose a separate training atlas, the provided training datasets in the workshop were used as unbiased testing datasets.

Our experience with the atlas creation for our segmentation shows that the probabilistic maps provide a better segmentation if they come from the same kind of manual segmentation. If the range of variation of the manual segmentation used in the atlas creation process is too wide, the automatic segmentation does not give satisfying results. The atlas used in this study is the one we created from a very consistent manual segmentation study, and it showed good results on numbers of different datasets. In the training dataset provided for this workshop, the manual segmentations have different sources and the age range is broad. This atlas would not have given a better segmentation than the one we used.

UU IR Author Manuscript

As a preprocessing to the atlas building, we select randomly a single training image and registered initially all datasets affinely to that image [7, 8]. The datasets are then skull stripped, intensity inhomogeneity corrected, and intensity calibrated (see next section for more details). After this preprocessing, we use fluid, deformable registration to compute the unbiased average from the whole training population, along with deformation fields containing the information to transform each image to the atlas [9]. The deformable registration refines iteratively an average image and computes a fluid-model based deformation field via voxel-by-voxel diffeomorphic mapping to that average image.

2.2 Atlas based segmentation

Preprocessing The first preprocessing step puts the rather heterogeneous set of images provided by the workshop into the same image space using a rigid, normalized mutual information registration approach [8]. We choose one of the datasets (UNC_03) as the registration target and all images are rigidly registered and resliced to an isotropic 1mm resolution image grid. We choose this particular case because it is quite well-centered, its unspecific appearance and the intactness of the whole skull in the image.

The second step computes a brain tissue classification using a probabilistic atlas driven automatic segmentation called itkEMS [10]. In addition to the tissue segmentation, itkEMS performs an intensity inhomogeneity correction that removes gradual variations in the image intensities mainly due to RF coil imperfections. The atlas used for itkEMS is different from the one used in the registration process. The tissue segmentation atlas is a specific one that contains prior knowledge of the tissue distribution. Additionally skull stripping is performed using the hard tissue segmentation. This itkEMS step is used only for the inhomogeneity correction and the skull stripping.

In the last step of the preprocessing an intensity calibration of the image to the atlas image. This intensity calibration of the skull stripped images is computed using a histogram matching only at a specified number of quantile values (ITK filter, 100 quantiles).

After this preprocessing, all images are aligned in the same anatomical space, skull stripped, their intensity inhomogeneity is corrected and normalized. All of these steps are computed fully automatically via shell scripting. The resulting images are used as inputs in the atlas computation and the subcortical structure segmentation.

Segmentation step Our segmentation consists of registration of the atlas to the preprocessed image and the propagation of the probabilistic atlas caudate definition. The registration is computed in two steps. First, the atlas image is affinely registered to the preprocessed image with 15 parameters [8]. Then, the affinely registered template is warped to the case using the fluid, diffeomorphic, deformable process described in the atlas building step [9].

Both transformations (affine and warping) are then applied to the caudate probabilistic definitions to have them in the coordinate space of the case to

UU IR Author Manuscript

be segmented. In order to have the final segmentations in the original image space, and thus to allow the comparison with the manual expert segmentations in this workshop, we apply the inverse of the prior rigid registration computed in the first preprocessing step to the caudate probabilistic maps. In the final step these probabilistic maps are thresholded at 50% probability to yield the binary segmentation of the caudate. We use a probabilistic map along the process because it makes the segmentation more robust when we apply the transformations (affine and warping).

The quality of this segmentation is strongly related to the appropriateness of the atlas to represent the cases to be segmented. The workshop dataset provided has a really large range of age and we expect that our method will not perform well for cases that are highly different from the atlas training population.

3 Results

We present here in this section both results we generated comparing our segmentation to the workshop's training dataset, as well as to the workshop's testing dataset. The first was computed by us, while the latter was computed by the workshop organizers.

3.1 Training data results

The training dataset has 33 cases provided with their manual caudate segmentations: 15 from the BWH dataset and 18 from the IBSR dataset. We applied our segmentation method as described above on these 33 cases.

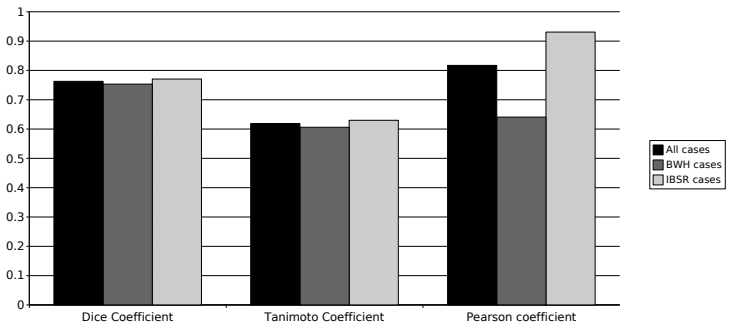

Fig. 1. Dice, Tanimoto and Pearson coefficients for the training dataset.

We calculated the two mainly used volumetric overlap definitions as well as a correlation coefficient:

1. Dice overlap coefficient which is the intersection of the two volumes divided by the average volume
2. Tanimoto overlap coefficient which is the intersection divided by the union of the two volumes
3. Pearson correlation coefficient which measures the correlation of the volumetric measurements across all cases.

In figure 1 the results for both right and left caudate have been averaged, and we present the values for all the data along with the results for the two dataset separated.

The Tanimoto coefficient, which is the hardest test for volume comparison, presents an average value of 63% and the Dice coefficient is at 76%. We notice that the values are quite constant across the two datasets. The Pearson coefficient is much more different between the two groups: while the BWH dataset shows a quite poor result with 0.64, the IBSR dataset has a very good correlation of 0.93.

This big difference in the Pearson coefficient is explained by the nature of our segmentation. The agreement of our segmentation and the manual segmentation depends on the way the manual segmentation procedure in our training population agrees with the one employed in the workshop images. This indicates that the segmentation procedure of our training expert segmentations is more similar to the procedure used for the IBSR dataset than for the BWH dataset.

3.2 Testing data results

Our segmentations on the testing data set are evaluated with the workshop organizer automatic program. The testing group is composed of 4 different populations: scans from 2 year old children, scans from elderly people (over 55 years), scans from adults and finally 10 scans of the same adult person to test the reliability.

Table 1 presents the scores of the segmentations of the 3 first groups. The computation of these scores is based on several values described in the table 1. The table shows that our segmentation gives much better results on the pediatric and elderly dataset (averaging at 80) than on the adult dataset (with 62). The volumetric overlap, the relative absolute volume difference, the symmetric RMS surface distance and the maximum symmetric absolute surface distance scores are better for the pediatric and elderly group whereas the volumetric difference shows surprisingly better results for the adult group.

Table 2 shows the Pearson correlation coefficient for the three groups. The table presents good results for the elderly (.87) and the adult group (.83) but they are not as good with the pediatric dataset (.59). When compared with the average Pearson correlation for the manual segmentation of human experts, which is .71, one can see that our method correlates for the 2 UNC datasets at least as much with the manual raters as they with each other.

Table 3 presents the coefficient of variations (COV) among the 10 scans of the same person. It averages at 1.2% whereas the variablility of the human expert

UU IR Author Manuscript

All Dataset	Overlap Err		Volume Diff.		Abs. Dist.		RMS Dist.		Max. Dist.		Total
	[%]	Score	[%]	Score	[mm]	Score	[mm]	Score	[mm]	Score	Score
UNC Ped 10	29.3	82	-14.1	75	0.7	76	1.2	78	10.7	68	76
UNC Ped 14	26.3	84	-12.4	78	0.5	82	0.8	86	3.9	88	84
UNC Ped 15	22.6	86	-13.2	77	0.4	84	0.7	87	4.8	86	84
UNC Ped 19	29.7	82	-23.2	60	0.6	78	0.9	84	5.2	85	78
UNC Ped 30	28.3	82	-12.7	78	0.6	78	0.9	85	4.7	86	82
UNC Eld 01	39.0	76	22.1	61	0.9	68	1.5	74	6.0	82	72
UNC Eld 12	30.5	81	12.4	78	0.6	78	1.0	82	4.2	88	82
UNC Eld 13	26.9	83	10.3	82	0.5	81	0.9	85	3.6	90	84
UNC Eld 20	27.0	83	14.9	74	0.5	80	0.9	84	4.5	86	82
UNC Eld 26	33.7	79	23.9	58	0.6	76	1.1	80	4.7	86	76
BWH PNL 16	41.6	74	-9.8	82	1.2	58	2.7	52	23.6	30	59
BWH PNL 17	38.6	76	-7.9	86	1.1	60	2.8	50	27.5	19	58
BWH PNL 18	47.2	70	-21.3	62	1.2	56	2.0	64	10.8	68	64
BWH PNL 19	44.5	72	-15.3	74	1.3	52	2.8	49	25.6	25	54
BWH PNL 20	38.4	76	-5.3	91	1.0	64	2.4	57	26.2	23	62
BWH PNL 21	47.6	70	-16.2	72	1.6	40	3.1	44	25.9	24	50
BWH PNL 22	42.7	74	-13.4	76	1.3	51	2.9	48	24.4	28	56
BWH PNL 23	37.7	76	3.2	94	0.8	69	1.4	76	10.1	70	77
BWH PNL 24	37.3	76	-12.6	78	0.8	68	1.5	74	8.4	75	74
BWH PNL 25	40.6	74	-8.0	86	1.2	55	2.8	50	24.9	27	58
BWH PNL 26	37.3	76	-2.7	84	0.8	70	1.6	72	10.0	70	75
BWH PNL 27	39.9	75	-16.9	70	1.4	46	3.5	38	26.2	23	50
BWH PNL 28	43.4	72	-9.2	84	1.4	48	3.2	42	24.9	27	55
BWH PNL 29	44.8	72	5.7	90	1.0	61	1.5	72	9.0	74	74
Average All	36.5	77	-5.1	77	0.9	66	1.8	67	13.7	60	69
Average UNC Ped	27.2	83	-15.1	74	0.6	80	0.9	84	5.9	83	80
Average UNC Eld	31.4	80	16.7	71	0.6	77	1.1	81	4.6	86	79
Average BWH PNL	41.5	74	-9.3	81	1.2	57	2.4	56	19.8	42	62

Table 1. Results of the comparison metrics and corresponding scores for all test cases averaged for the left and right segmentation. The summary rows at the end of the table display the overall average across all test cases, as well as grouped for the three testing groups.

Correl	UNC Ped	UNC Eld	BWH PNL	Total
Left	0.6730	0.8780	0.8339	0.7949
Right	0.5121	0.8767	0.8357	0.7415
Average	0.5925	0.8773	0.8348	0.7682

Table 2. Pearson correlation for the volume measurements in the three testing groups as well as in total. This coefficient captures how well the volumetric measurements correlate with those of the reference segmentations.

Test/Re-Test	UNC 03 [mm^3]	UNC 04 [mm^3]	UNC 09 [mm^3]	UNC 11 [mm^3]	UNC 17 [mm^3]	UNC 18 [mm^3]	UNC 21 [mm^3]	UNC 22 [mm^3]	UNC 24 [mm^3]	UNC 25 [mm^3]	Mean [mm^3]	Stdev [mm^3]	COV [%]
Left	3284	3180	3239	3299	3242	3249	3242	3265	3253	3237	3249	32	1.0
Right	3429	3350	3389	3534	3399	3356	3381	3410	3395	3424	3407	52	1.5
Total	-										-	-	1.2

Table 3. The volumetric measurements of the 10 data sets of the same young adult acquired on 5 different scanners within 60 days. The coefficient of variation (COV = standard deviation / average, last column) indicates the stability of the algorithm in a test/re-test situation including scanner variability.

segmentation is at 3.1%. This shows that our segmentation is more repeatable and thus reliable than the one of human expert raters.

Finally the last figure (Fig. 2) is the comparison between our segmentation (in blue) and the manual one (in red) on selected cases of each of the three groups. The adult case presented shows that our segmentation does not work properly at the posterior tail area of the caudate. On the two other cases (pediatric and elderly) the automatic segmentation agrees pretty well with the general shape of the manual one and there is no general pattern in the differences.

4 Discussion

The results obtained with the testing dataset are interesting. All the different errors computed to assess the scores in each group show that our segmentation is better on the pediatric and the elderly scans than on the adult brains. As our automatic segmentation results are strongly related to the atlas used, these results are somewhat unexpected since we used an atlas based on adult brains. However the quality of our results depends on the correlation between the way the manual segmentation used in the atlas have been done and the actual "true" segmentation our results are compared to. It is thus not surprising that the elderly and pediatric manual segmentation protocol has a higher similarity with the one used in our separate training population compared to the protocol used for the adult BWH dataset. We can note that even for the fairly well segmented cases, the final score is not near 90 (which means a good segmentation according to the workshop evaluation process). Looking closely at the table we notice that the two main scores that lower the overall score for the pediatric and elderly dataset are the volume difference and the absolute distance. These two measures emphasize differences over the whole volume. Although the automatic segmentation is close to the manual one, if there are small errors all around the volume, they add up to increase these two measurements. The difference might come from the registration step of our segmentation. It could introduce a consistent extra layer around the caudate which would give these low scores. It might also be due to the thresholding part of our method. All the registrations are applied on the probabilistic maps, and finally to get the volume, these maps are

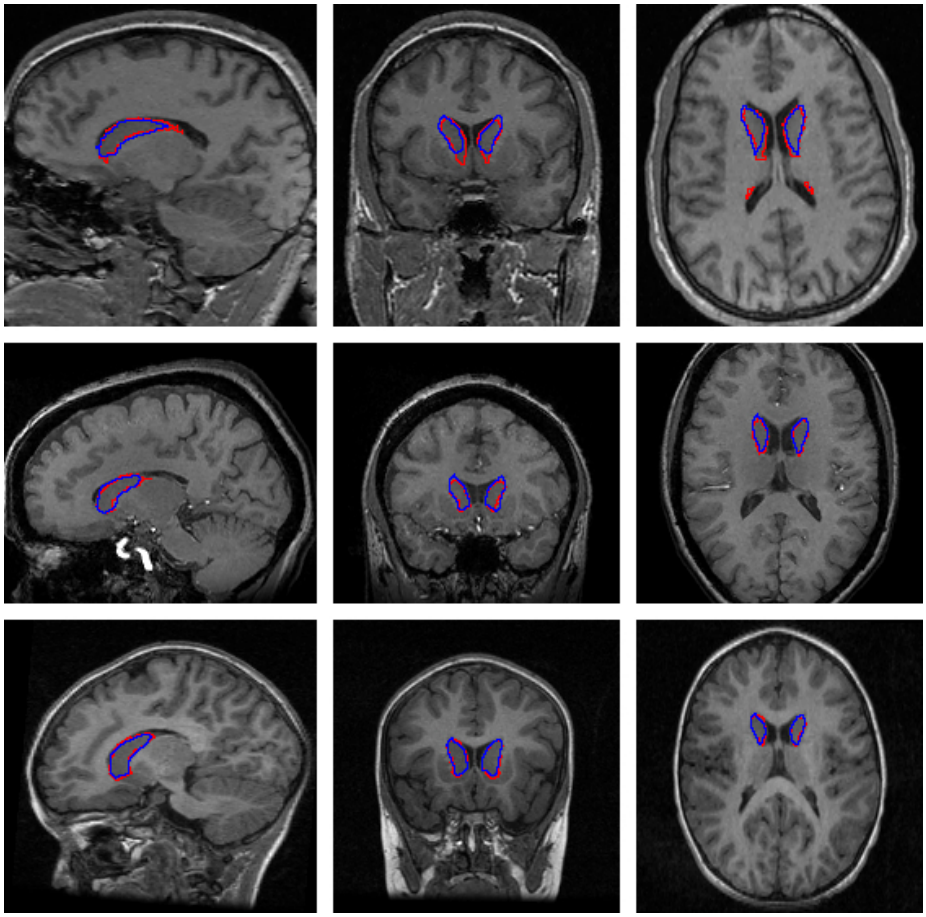

Fig. 2. From left to right, a sagittal, coronal and transversal slice from a a subject in the adults BWH group (top), one in the elderly UNC group (middle) and one in the pediatric UNC group (bottom). The outline of the reference standard segmentation is in red, the outline of the segmentation of the method described in this paper is in blue.

thresholded at .5. The fifty percent probability might not be the more accurate one.

Concerning the Pearson correlation coefficient, the results are in the same order of magnitude as what we had presented in the original paper [6]. The coefficients computed on two datasets were .89 and .73. On the testing dataset the elderly and adult segmentations present good values (0.87 and 0.83) but the coefficient drops for the pediatric segmentation (0.59). This can be explained by the caudate shape variation and the small size of the structures within the pediatric dataset which are not handled properly with our method.

UU IR Author Manuscript

Finally, the reliability assessment over the 10 scans of the same person averages at 1.2% for both caudates. This is a very good result regarding the stability of the method considering that, on this same dataset, manual raters have a COV of 3.1%.

The overall results of our segmentation are quite satisfying considering that with our method we obtained a segmentation for the whole worshop dataset. Some general parameters had to be set up, but we found the proper tuning such that all the caudates have been fully automatically segmented.

5 Conclusion

In this paper we presented our fully automatic segmentation method for the caudate which has been previously described in [6]. We applied this method on a wide range of brain datasets and successfully segmented all of them. Even though we have not trained our model to all different caudate segmentation protocols used by manual expert raters for the workshop testing dataset, the results are quite satisfying. Using a non-trained atlas gave us fairly good results, which could be enhanced with an appropriately retrained atlas, but there seems limited necessity as our results correlated well with the human expert segmentation. As we are also more reliable in a test/retest situation than manual raters, we can conclude that our caudate segmentation allows an efficient and reliable processing of large scale studies.

6 Acknowledgment

We would like to thank Daniel Rueckert for provided the affine registration tookit, also called The Image Registration Toolkit. Matthieu Jomier has also contributed software developments in several of the segmentation steps. We also want to thank Cecile Lepin for her help. This research has been supported by the UNC Neurodevelopmental Disorders Research Center HD 03110 (J. Piven), the NIH Conte Center MH064065, NIH RO1 MH61696, NIMH MH64580 and Lilly sponsored HGDH morphology study.

7 Updated method

After the results from the first assessment, we enhanced our method to try to get better results.

Instead of using our own atlas created from UNC dataset, we used the 15 BWH_PNL cases from the training dataset. As we have seen with our first segmentation the results were not consistent across the different datasets. We discussed in section 4 about the lower segmentation quality for the BWH_PNL dataset compared to the UNC Ped and Eld dataset. In order to improve our segmentation we created the new atlas using the process described in section 2.1.

UU IR Author Manuscript

References

1. Pham, D., Xu, C., Prince, J.: Current methods in medical image segmentation. Annu Rev Biomed Eng **2** (2000) 315–337
2. Sonka, M., Fitzpatrick, J. In: Handbook of medical imaging. Med. Image Process. Anal., SPIE Press. Volume 1 and 2. (2000) 69–211
3. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: IEEE CVPR. Volume 1. (2000) 316–323
4. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., , Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. Volume 22. (2003) 137–154
5. Pohl, K., Bouix, S., Kikinis, R., Grimson, W.: Anatomical guided segmentation with non-stationary tissue class distributions in an expectation-maximization framework. In: ISBI 2004, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2004) pp. 81–84
6. Gouttard, S., Styner, M., Joshi, S., Smith, R., Cody Hazlett, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. In: SPIE Medical Imaging. Volume Vol 6512. (2007) 65122J–1 – 65122J–11
7. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application tobreast mr images. IEEE Transactions on Medical Imaging **18**(8) (1999) 712–721
8. C.Studholme, D.L.G.Hill, D.J.Hawkes: An overlap invariant entropy measure of 3d medical image alignment. In: Pattern Recognition. Volume 32. (1999) 71–86
9. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23** (2004) S151–S160
10. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of mr images of the developing newborn brain. Medical Image Analysis **9** (2005) 457–466

UU IR Author Manuscript