A multi-atlas approach to automatic segmentation of the caudate nucleus in MR brain images

Eva van Rikxoort, Yulia Arzhaeva, and Bram van Ginneken

Image Sciences Institute, University Medical Center Utrecht, the Netherlands
{eva,yulia,bram}@isi.uu.nl

Abstract. Automatic segmentation of brain structures is an important prerequisite for many applications in neuroscience. In this paper a fully automatic method for segmenting the caudate from 3D MRI brain images is presented. The method is based on multi-atlas registration which has shown to be a powerful concept for segmentation. The results show that the automatic segmentation is similar to segmentations by human observers for routine data and slightly worse but still acceptable for non-routine data.

1 Introduction

The identification and delineation of brain structures from magnetic resonance imaging (MRI) brain images is an important task and has many applications in neuroscience, such as the study of brain development, the mapping of functional activation onto brain anatomy, and the analysis of neuroanatomical variability among normal brains [1].

In this paper we focus on the segmentation of the caudate nucleus (CN), a subcortical component of the basal ganglia that is involved in sensory-motor control, cognition, language, emotion and other important brain functions. The caudate is a periventricular gray matter structure that shows up lighter compared to the majority of cortical gray structures in T1-weighted MRI images. The caudate has a rather homogeneous intensity but is difficult to segment because it attaches to other gray structures at multiple locations. It consists of a relatively large head, a cone-shaped body and a thin tail.

Due to the large amount of data and the monotony of the task, manual segmentation of the caudate is difficult and time consuming. To facilitate the task, various automatic and computer-assisted segmentation methods have been developed including atlas-based registration techniques, deformable models, knowledge-driven and histogram-driven approaches, and statistical modelling (see e.g. [2]). However, none of these methods is routinely used in clinical practice at the moment.

In this paper we present a fully automatic multi-atlas based segmentation of the caudate nuclei. Atlas-based segmentation uses registration to achieve segmentation and this has proven to be a powerful concept [3]. Atlas-based methods

T. Heimann, M. Styner, B. van Ginneken (Eds.): 3D Segmentation in The Clinic:
A Grand Challenge, pp. 29-36, 2007.

start by registering an atlas image with a target image that is to be segmented. To obtain a segmentation of the target image, a manual labelling of the atlas is transformed using the mapping determined during the registration. This process is called label propagation. Major advantages of this segmentation approach, especially for 3D applications, are simplicity (only a registration framework and a number of pre-segmented data sets are required, no need for landmarking and complex training procedures), general applicability (a wide range of segmentation tasks can be solved by this method) and robustness (fairly stable volumetric registration methods exist).

The basic underlying assumption of atlas-based segmentation is that it is possible to find a deformation to the atlas that aligns it with the target such that the objects of interest line up perfectly. However, insufficient similarity between the atlas and the target image often results in local mismatches, which in turn leads to segmentation errors [4]. This fundamental problem of single atlas-based segmentation has been recognized and an obvious solution has been attempted: instead of a single atlas, multiple atlases are registered and the propagated labels are fused, for example by averaging. Local errors of a single atlas can now be corrected, as long as the majority of atlases register correctly. Several studies have shown that multi-atlas segmentation outperforms methods that use a single atlas [3, 5, 6].

The method is described in Section 2. The data and experiments are given in Section 3 followed by the results in Section 3.3. Finally Section 4 provides a discussion and a conclusion.

2 Method

Registration. The method we propose is generally applicable and any registration method can be plugged in. The particular method used in this paper is a freely available package [7], that formulates the registration problem as an optimization problem in which the similarity between the target and atlas image is maximized. For the cost function negative mutual information was used, following the implementation in [8]. The transformation is initialized with an affine transformation followed by a non-rigid transformation modeled by a B-spline. For the optimization of the cost function, an iterative stochastic gradient optimizer is used. To avoid local minima, a multi-resolution strategy was taken. Parameter settings are provided in Section 3.

Multi-atlas segmentation. All training atlases are registered to the target image, followed by propagation of the labels. The probability of a label in the segmentation of the target image is determined by averaging the propagated labels. This method was used in e.g. [5].

Post-processing. The result of the multi-atlas segmentation is for each voxel a probability that it belongs to a caudate. To obtain the final segmentation result, the probabilities were first post-processed by Gaussian blurring with scale σ, this has the effect of pooling local evidence and smoothing the borders. The resulting probabilities were thresholded to obtain a binary segmentation.

In the resulting binary segmentation, often a part of the ventricles was included, therefore optimal thresholding [9] was applied to the gray-values in the resulting binary segmentation as a final post-processing step.

3 Experiments and Results

3.1 Data

In this paper 66 MRI images of the brain were used, 32 for training the system and 34 for testing. All MRI images were scanned with an Inversion Recovery Prepped Spoiled Grass sequence on a variety of scanners (GE, Siemens, Phillips, mostly 1.5 Tesla). Some data sets were acquired in axial direction, whereas others in coronal direction. All data sets have been re-oriented to axial RAI-orientation, but have not been aligned in any fashion. The training data was taken from two major sources:

- 18 MRIs and structural segmentations from the internet brain segmentation repository (ISBR) at Mass General Hospital, Boston
- 15 MRIs and caudate segmentations from the Psychiatry Neuroimaging Laboratory at the Brigham and Women's Hospital Boston (BWH)

There is a slight difference in the definition of the caudate in both training data sets in regard to the border with the nucleus accumbens anteriorly and the vanishing tail posteriorly. The test data consisted of data from four different sources.

- 14 MRIs from the Psychiatry Neuroimaging Laboratory at the Brigham and Women's Hospital, Boston (BWH PNL). This data is from the same study as the BWH training set.
- 5 pediatric (no older than 2 years) MRI images acquired at the UNC Neuro Image Analysis Laboratory, Chapel Hill (UNC Ped)
- 5 MRIs from a Parkinson's Disease study at the UNC Neuro Image Analysis Laboratory, Chapel Hill (UNC Eld)
- 10 MRIs from a test/re-test study acquired within 60 days on 5 different scanners at the UNC Neuro Image Analysis Laboratory, Chapel Hill and the Duke Image Analysis Laboratory, DIAL

3.2 Experiments

The segmentation method was applied to all 34 testing images. For the 14 scans from the BWH PNL test data, the 15 training scans from the same study were used as atlases. For the UNC test data, the 18 ISBR training scans were used as atlases.

For the affine registration 4 resolutions were used, in each of which 512 iterations of the stochastic gradient descent optimizer were performed. For the non-rigid B-spline registration 5 resolutions were used. The B-spline grid spacing used in these resolutions was 64, 64, 32, 16, and 8 voxels respectively. The

Correl	UNC Ped	UNC Eld	BWH PNL	Total
Left	0.1170	0.5665	0.6877	0.4571
Right	0.3772	0.5247	0.7370	0.5463
Average	0.2471	0.5456	0.7124	0.5017

Table 1. 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	1894	1856	1859	1892	1830	1733	1694	1747	1710	1780	1799	76	4.2
Right	2247	2271	2199	2155	2169	2072	2127	2155	2077	2118	2159	66	3.0
Total	-										-	-	3.6

Table 2. 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.

optimizer performed 256 iterations in each resolution. For both affine and non-rigid registration 32 histogram bins were used.

Due to the different definitions of the caudate in the two training sets, we found it advantageous to apply different settings for the post-processing of the results. For the UNC test data the probabilities were blurred with $\sigma = 1$ and thresholded at 0.5 before applying optimal thresholding. For the BWH PNL test data different post-processing settings were applied to the body and the tail of the caudate. For the body the same post-processing was applied as for the UNC data, the blurring however caused the tail to vanish, therefore the tail was blurred with $\sigma = 0.5$ and thresholded at 0.25 before applying optimal thresholding. Since the current method does not contain any information about the shape of the object, the difference between the body and tail was determined on 2D slices in the transversal direction; on each transversal slice connected component analysis was performed. If two components were present the upper component was defined as the body and the lower component as the tail. If only one component was present it was identified as body. Although the identification of the body and tail uses only 2D information, the blurring is applied in 3D.

3.3 Results

Figure 1 shows the segmentation results in three slices from three different subjects. To quantify the performance of the automatic segmentation methods the results were compared to manual segmentations for the BWH PNL, UNC Ped, and UNC Eld test scans. Five metrics were computed (see Table 3), volumetric

overlap error, relative volume difference, mean, RMS and maximum surface distance. For each metric and each case, a score was computed, in such a way that 100 points is a perfect result and 90 points roughly corresponds to the score that can be expected of a human observer. A detailed description of the data and the scoring system can be found on [10].

In addition to the metrics mentioned above, Table 1 shows the Pearson correlation coefficient in total as well as per testing group. This coefficient captures the correlation between the automatic and the reference segmentation. Manual segmentations of human experts for the caudate show in average a Pearson correlation of 0.71.

Table 2 shows the coefficient of variation (COV) of the volumetric measurements on the 10 scans of the same young healthy person acquired within 60 days on 5 different scanners. This measure indicates how stable the automatic segmentation is on a test/re-test situation including scanner variability. Manual segmentation of human experts for the caudate show on average a COV of 3.1

4 Discussion

In this paper a fully automatic segmentation by registration method was presented and applied to segmenting the caudates from MRI brain scans. The results show that the automatic segmentation performs similar to human observers on a routine data set (BWH PNL) and doesn't break down on harder data sets (UNC ELd and UNC Ped).

The different types of testing data that were supplied allowed for evaluation of the method not only on routine data but also with respect to flexibility towards pathology, age group and signal-to-noise aspects. The results show that our method performs substantially better on the BWH PNL test data than on the other, non-routine data. This is not only caused by the fact that the UNC data contained pathology and pediatric data; a registration-based approach can be expected to work best when the unseen test images are similar to the training data. This was the case for the BWH PNL test data but not for the UNC test data. When the atlases used for the UNC test data would also have contained pediatric and pathologic cases, we expect our method would have performed better.

The advantage of the proposed method is that it is generally applicable; it can be used for many different structures provided the appropriate examples are used for training. The only task specific part of the method is the post-processing, which consists of basic image processing techniques.

The results show that the caudate is in general under segmented by our method (see Figure 1 and Table 3). This is probably due to the settings of the threshold applied to the probabilities resulting from the registrations and the settings of the optimal thresholding.

The method as presented in this paper is relatively slow. Segmenting one scan takes on average 92 minutes. The main bottleneck is the registration of the atlases, registering one atlas to a test scan takes on average 6 minutes. The

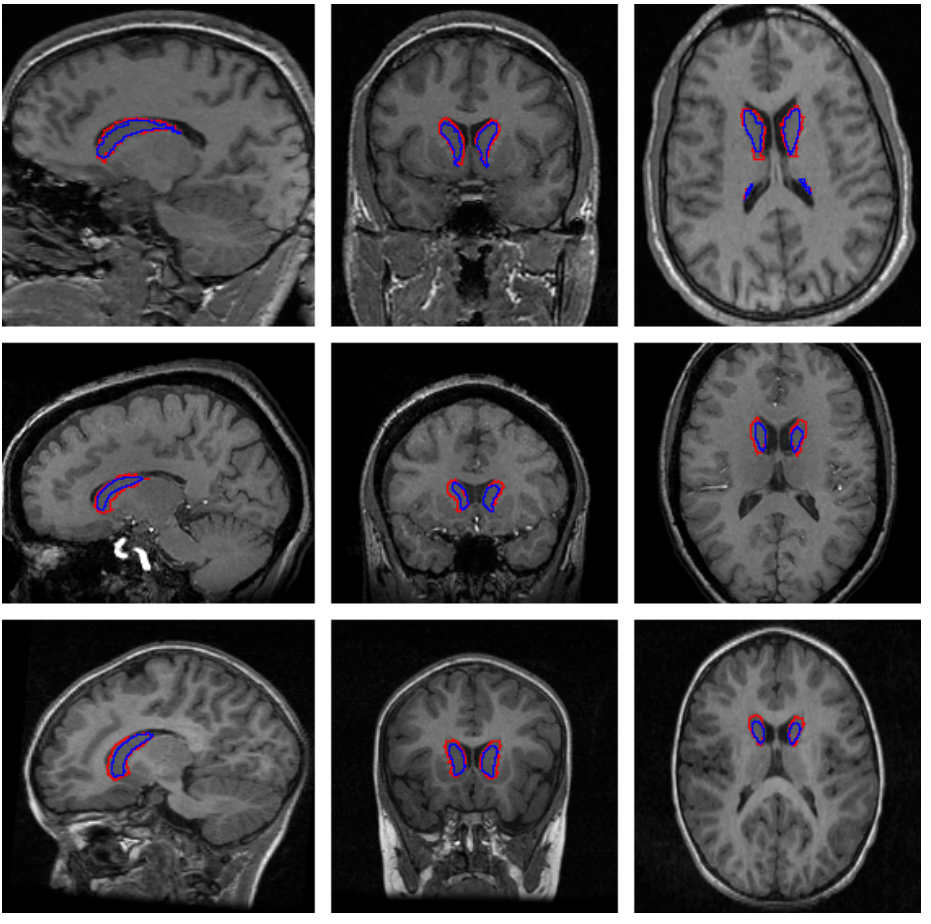

Fig. 1. From left to right, a sagittal, coronal and transversal slice from a a subject in the adults BWH group (top), one in the elderly UNC group (middle) and one in the pediatric UNC group (bottom). The outline of the reference standard segmentation is in red, the outline of the segmentation of the method described in this paper is in blue.

method can be sped up by using less atlases (e.g. selecting the best ones) and optimizing the registration.

To conclude, a general method was applied to the task of caudate segmentation from brain MRI scans. The results show that the method is able to perform comparable to human observers and is quite robust against scanner and patient variability.

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	50.1	68	-41.7	27	1.5	44	2.7	52	18.1	46	48
UNC Ped 14	54.5	66	-47.6	16	1.3	50	1.9	66	11.6	66	53
UNC Ped 15	47.4	70	-44.7	22	1.2	58	1.6	72	9.9	71	58
UNC Ped 19	58.2	64	-54.9	6	1.6	41	2.3	60	8.9	74	48
UNC Ped 30	50.9	68	-50.0	12	1.2	54	1.7	69	10.4	70	54
UNC Eld 01	52.4	67	12.1	78	1.2	54	1.5	73	4.0	88	72
UNC Eld 12	47.2	70	-6.2	69	1.0	62	1.3	77	3.6	90	73
UNC Eld 13	51.6	68	-45.9	20	1.3	52	1.9	66	11.2	67	54
UNC Eld 20	41.1	74	-19.6	66	0.9	67	1.2	78	7.1	79	73
UNC Eld 26	55.5	65	-41.9	26	1.2	56	1.5	72	6.5	80	60
BWH PNL 16	31.1	80	-11.2	80	0.5	80	1.1	81	10.6	69	78
BWH PNL 17	23.8	85	-10.4	82	0.3	87	0.7	88	5.1	84	86
BWH PNL 18	30.3	81	-8.0	86	0.6	78	1.3	77	13.4	61	76
BWH PNL 19	31.3	80	-18.2	68	0.5	80	0.9	85	4.5	87	80
BWH PNL 20	25.8	84	-5.0	92	0.4	86	0.7	87	4.1	88	87
BWH PNL 21	29.2	82	-3.1	93	0.5	81	0.9	85	4.4	87	86
BWH PNL 22	33.5	79	-19.3	66	0.6	78	0.9	84	4.5	87	79
BWH PNL 23	25.3	84	-4.0	93	0.5	82	1.2	78	16.0	52	78
BWH PNL 24	24.5	84	-14.4	74	0.5	82	1.6	72	18.7	45	72
BWH PNL 25	26.7	84	1.1	90	0.4	84	0.8	86	8.1	76	84
BWH PNL 26	24.3	84	-0.3	98	0.5	84	1.2	78	14.8	56	80
BWH PNL 27	38.8	76	-35.9	37	0.7	73	1.0	82	3.7	89	72
BWH PNL 28	30.9	80	-21.4	62	0.5	82	0.8	86	4.1	88	80
BWH PNL 29	47.6	70	-31.1	45	1.0	64	1.7	70	17.0	50	60
Average All	38.8	76	-21.7	59	0.8	69	1.3	76	9.2	73	70
Average UNC Ped	52.2	67	-47.8	16	1.4	49	2.0	63	11.8	65	52
Average UNC Eld	49.6	69	-20.3	52	1.1	58	1.5	73	6.5	81	67
Average BWH PNL	30.2	81	-12.9	76	0.5	80	1.0	81	9.2	73	78

Table 3. 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.

References

1. Han, X., Fischl, B.: Atlas renormalization for improved brain MR image segmentation across scanner platforms. IEEE Transactions on Medical Imaging **26**(4) (2007) 479–486
2. Clarke, L., Velthuizen, R., Camacho, M., Heine, J., Vaydianathan, M., Hall, L., Thatcher, R., Silbiger, M.: MRI segmentation: Methods and applications. Magnetic Resonance Imaging **13** (1995) 343–368
3. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D.B., Maurer, Jr., C.R.: Quo vadis, atlas-based segmentation? In: The Handbook of Medical Image Analysis – Volume III: Registration Models, New York, NY, Kluwer Academic / Plenum Publishers (2005) 435–486

4. Crum, W., Griffin, L., Hill, D., Hawkes, D.: Zen and the art of medical image registration: correspondence, homology, and quality. NeuroImage **20** (2003) 1425–1437
5. Rohlfing, T., Brandt, R., Menzel, R., Maurer, Jr., C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage **21**(4) (2004) 1428–1442
6. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage **33**(1) (2006) 115–26
7. Klein, S., Staring, M.: Elastix. http://www.isi.uu.nl/Elastix/ (Accessed July 17, 2007)
8. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. IEEE Transactions on Image Processing **9**(12) (2000) 2083 – 2099
9. Otsu, N.: A threshold selection method from gray level histograms. IEEE Trans. Systems, Man and Cybernetics **9** (1979) 62–66
10. Heimann, T., van Ginneken, B., Styner, M.: Workshop on 3D segmentation in the clinic: A grand challenge. http://mbi.dkfz-heidelberg.de/grand-challenge2007/ (Accessed July 18, 2007)