

Fully Automated Knowledge-Based Segmentation of the Caudate Nuclei in 3-D MRI

Michael Wels¹, Martin Huber², and Joachim Hornegger¹

¹ Institute of Pattern Recognition, University Erlangen-Nuremberg, Germany

michael.wels@informatik.uni-erlangen.de

² Siemens AG, CT SE SCR 2, Erlangen, Germany

Abstract. In this paper we present a fully automated approach to segmentation of the caudate nuclei in 3-D magnetic resonance images. It is based on the technique of probabilistic boosting trees. As a strategy for supervised learning it is capable to derive a discriminative model for the distinction of object and non-object voxels from expert annotated imaging data and rather rough anatomical prior knowledge provided by a probabilistic anatomical atlas. Training the model involves successively selecting and combining features that best separate the available training samples, i.e., image voxels, and grouping the resulting boosted classifiers in a tree structure. Most of the features used are taken from an intra-axial 2-D context surrounding the voxel of interest and its transformation to a particular set of Haar-like features. The final segmentation is obtained after post-processing the preliminary result by a fast marching approach whose two speed images are seeded at the inner and outer bounds of the object detected. This allows for adaptation to local edges close to the initially detected object's boundary. A detailed quantitative evaluation critically reveals strengths and weaknesses of the proposed method.

1 Introduction

Fully automated segmentation tools both for pathologic as well as non-pathologic structures within the human brain serve as valuable components of clinical decision support systems. Proper and reliable segmentation results are considered critical for the automatic extraction of quantitative or more abstract findings that are diagnostically relevant like, for example, the segmented objects' volume or their relative location, respectively. Those semantic features are, in a second step, easily combined with other findings emerging from different examinations a patient undergoes during diagnostics to form the input to integrated decision support systems.

In the context of brain tissue classification within magnetic resonance (MR) volume sequences segmentation of particular anatomical entities states a challenging problem for fully automated approaches: the distinction of deep gray matter structures, like the caudate nuclei, and cortical gray matter based on observed intensities only, is virtually impossible. Prior knowledge about the anatomical composition of the human brain has to be integrated to guide the

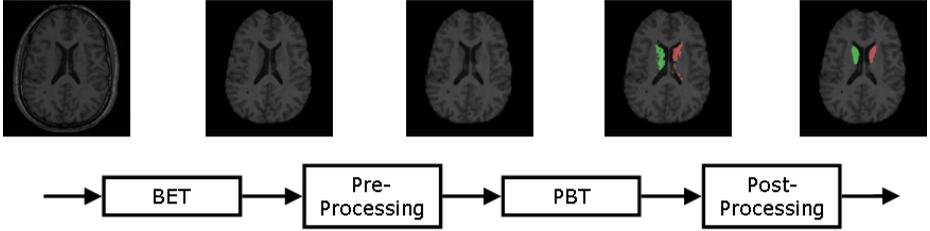


Fig. 1. The proposed segmentation framework.

segmentation process. Furthermore, segmentation methods need to be robust with regard to the characteristic artifacts of the MR imaging modality: Rician noise, partial volume effects, and intensity inhomogeneities.

The contribution of this work is a fully automated knowledge-based method for delineation of both the right as well as the left caudate nuclei in 3-D brain MR images that allows for fast and convenient segmentation of those two anatomical structures. Our method consists of four major steps: First, the whole brain is extracted from its surroundings with the Brain Extraction Tool (BET) [1]. Then, in a pre-processing step, the observed intensities within the area that is detected to be part of the brain itself are normalized to have zero mean and unit variance. Subsequently, two probabilistic boosting trees (PBT) [2] are used in a similar manner to [3] to classify individual voxels of the input data into foreground and background with respect to the right and the left caudate nucleus. This step also incorporates spatial knowledge about the typical location of the caudate nuclei in brain MR scans taken from a probabilistic anatomical atlas. Finally, a twofold fast marching approach that is seeded at the inner and outer bounds of the detected objects is adopted as a post-processing step to better adapt to local object edges apparent in the image. The inner and outer bound of the object necessary for this operation are determined by means of mathematical morphology. Experimental results on several previously unseen real-world data sets demonstrate strengths and weaknesses of the proposed method. The overall system block diagram is depicted in Fig. 1.

2 Methods

2.1 Probabilistic Boosting Tree

Our method’s main component PBT recursively groups boosted ensembles of weak classifiers to a tree structure during learning from expert annotated data. We have chosen Real AdaBoost [4] to be the boosting strategy within each tree node. In this case learning a PBT resembles building a multivariate binary regression tree as the final strong classifier

$$H(\mathbf{x}) = \sum_{t=1}^T h_t(\mathbf{x}) \quad (1)$$

generated within each inner node for a feature vector \mathbf{x} through a combination of real-valued contributions $h_t(\mathbf{x})$ of $T \in \mathbb{N}$ weak classifiers asymptotically approaches the additive logistic regression model [4]:

$$H(\mathbf{x}) \approx \frac{1}{2} \ln \frac{p(y = +1|\mathbf{x})}{p(y = -1|\mathbf{x})} \quad (2)$$

where $y \in \{+1, -1\}$ denotes the classification outcome. Therefore, at each inner node v of the resulting PBT with boosted strong classifier H_v there are current approximations of the posterior probabilities

$$q_v(y = +1|\mathbf{x}) = \frac{e^{2H_v(\mathbf{x})}}{1 + e^{2H_v(\mathbf{x})}} \quad (3)$$

and

$$q_v(y = -1|\mathbf{x}) = \frac{e^{-2H_v(\mathbf{x})}}{1 + e^{-2H_v(\mathbf{x})}}. \quad (4)$$

While training the classifier, those probabilities are used to successively split the set of training data relative to the prior probability $p_v(y = +1)$ associated with the current training (sub-)set into two new subsets. The soft thresholding parameter $\epsilon > 0$ sees to pass on training samples \mathbf{x} that are close to the current node’s decision boundary, i.e., if $q_v(y = +1|\mathbf{x}) \in [(1-\epsilon)p_v(y = +1); (1+\epsilon)p_v(y = +1)]$, to both of the resulting subsets and associated subtrees. This differs from the original formulation in [2] that did not explicitly take into account disparities of positive and negative examples within the training (sub-)sets available at each tree node. Apart from that [2] contains a detailed discussion of the overall learning procedure.

During classification the values for $q_v(y = +1|\mathbf{x})$ are used to guide tree traversing and combined propagation of posteriors in order to get final approximations $\tilde{p}_v(y|\mathbf{x})$ of the true posterior probabilities $p_v(y|\mathbf{x})$ at each tree node v : For outgoing arrows r_v^{-1} and r_v^1 associated with the possible classifications the approximation $\tilde{p}_v(y|\mathbf{x})$ can be computed via the recursive formula

$$\tilde{p}_v(y|\mathbf{x}) = \begin{cases} \tilde{p}_{\beta(r_v^{-1})}(y|\mathbf{x}) & : q_v(+1) < (1-\epsilon)p_v(+1) \\ \tilde{p}_{\beta(r_v^1)}(y|\mathbf{x}) & : q_v(+1) \geq (1-\epsilon)p_v(+1) \\ \sum_i \tilde{p}_{\beta(r_v^i)}(y|\mathbf{x}) \cdot q_v(i|\mathbf{x}) & : \text{otherwise} \end{cases} \quad (5)$$

where $\beta(r)$ denotes the vertex where arrow r ends.

2.2 Discriminative Features

In order to adequately capture discriminative knowledge about image voxels the features used in our method origin from two categories:

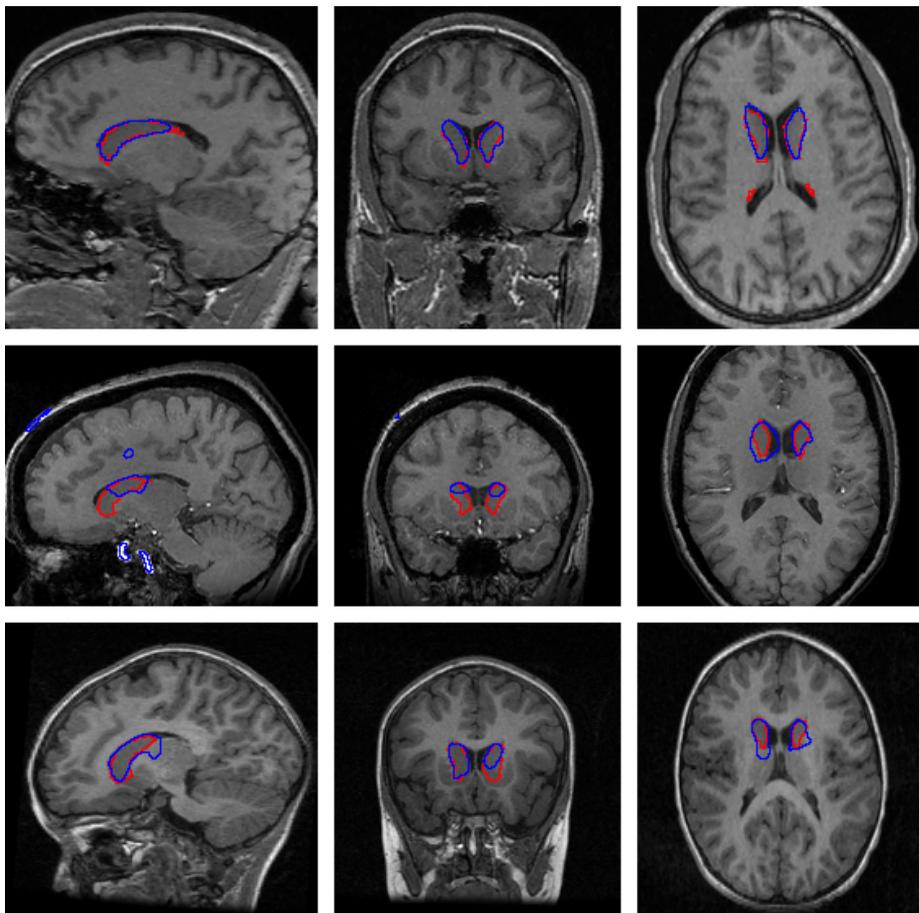


Fig. 2. From left to right, a sagittal, coronal and axial slice from a subject in the adults BWH group (top), one in the elderly UNC group (middle) and one in the pediatric UNC group (bottom). The outline of the reference standard segmentation is in red, the outline of the segmentation of the method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

The by far largest fraction of feature values in the feature vector \mathbf{x}_i associated with a given image voxel i state 2-D Haar-like features [5] computed on an intra-axial 2-D context surrounding the voxel of interest. They are derived from a subset of the extended set of Haar-like feature prototypes by [6] and are represented only implicitly in memory by so-called (Rotated) Integral Images. This allows for fast re-computation of the features with respect to a given voxel when they are actually assessed. Mainly due to combinatorial complexity we did not consider 3-D Haar-like features and furthermore restricted on 2-D Haar-like features whose centers are aligned with the center of the context considered.

All Dataset	Overlap Err		Volume Diff.		Abs. Dist.		RMS Dist.		Max. Dist.		Total Score
	[%]	Score	[%]	Score	[mm]	Score	[mm]	Score	[mm]	Score	
UNC Ped 10	70.8	56	8.4	0	4.2	0	6.6	6	23.6	30	18
UNC Ped 14	89.8	44	-52.2	42	7.1	0	9.0	0	22.6	34	24
UNC Ped 15	47.3	70	-7.5	63	2.0	32	3.2	42	12.2	64	54
UNC Ped 19	69.9	56	-51.9	26	4.7	0	9.9	0	46.6	14	20
UNC Ped 30	64.4	60	-11.8	8	4.8	0	10.5	0	43.8	13	16
UNC Eld 01	83.2	48	183.8	0	15.5	0	24.0	0	66.4	0	10
UNC Eld 12	83.7	47	175.2	0	17.4	0	25.6	0	70.6	0	9
UNC Eld 13	82.4	48	130.3	0	29.7	0	39.3	0	77.3	0	10
UNC Eld 20	88.5	44	178.1	0	27.5	0	36.8	0	87.1	0	9
UNC Eld 26	84.5	47	235.0	0	17.3	0	25.5	0	71.6	0	10
BWH PNL 16	46.3	71	-10.4	70	1.9	30	4.6	18	33.6	2	38
BWH PNL 17	31.6	80	-8.3	86	1.3	54	4.1	26	34.6	6	50
BWH PNL 18	58.8	63	-37.1	39	3.2	0	6.1	0	26.1	24	25
BWH PNL 19	48.2	70	-41.4	28	2.5	20	5.5	13	34.2	6	28
BWH PNL 20	47.3	70	41.8	26	1.7	38	3.6	34	33.3	5	35
BWH PNL 21	36.5	77	-21.8	62	2.3	14	6.5	0	38.9	0	30
BWH PNL 22	50.7	68	-22.3	61	2.3	14	5.2	6	35.2	2	30
BWH PNL 23	33.5	79	10.2	67	1.0	64	2.3	60	17.0	50	64
BWH PNL 24	27.2	83	-4.1	75	0.6	78	1.3	77	11.2	67	76
BWH PNL 25	58.7	63	-5.5	79	4.3	0	8.3	0	37.0	0	28
BWH PNL 26	49.0	69	-38.2	39	2.2	31	4.2	28	23.0	32	40
BWH PNL 27	26.0	84	-8.4	86	1.5	44	5.0	11	34.1	0	45
BWH PNL 28	37.6	76	2.5	74	2.6	5	6.7	0	31.3	8	33
BWH PNL 29	43.6	72	-6.3	50	1.9	30	3.8	34	21.8	36	44
Average All	56.6	64	26.6	41	6.6	19	10.7	15	38.9	16	31
Average UNC Ped	68.4	57	-23.0	28	4.6	6	7.8	10	29.7	31	26
Average UNC Eld	84.5	47	180.5	0	21.5	0	30.2	0	74.6	0	9
Average BWH PNL	42.5	73	-10.7	60	2.1	30	4.8	22	29.4	17	40

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.

The second category of features consists of the three voxel indices in patient space and the voxel’s probability to be part of the caudate nuclei. The latter is taken from a probabilistic anatomical atlas [7] which is roughly registered with the current patient data set by aligning orientations and image centers. This approach might seem brute force on the first sight, yet boosting asks for weak classifiers, i.e., individual features in our case. Further, there are certain quality standards for medical image acquisition such that the image region where the anatomical object that is about to be segmented appears in the final image underlies at least rough regularities.

2.3 Segmentation Refinement

As depicted in Fig. 1 the segmentation obtained from PBT needs to be smoothed to get an acceptable final result. In our approach this is done by means of an adversarially seeded fast marching approach: Two speed images are computed assigning detection times to every voxel relative to two sets of seed points. For that the current volume is interpreted as a graph where each voxel becomes a node with connectivity to its six spatially nearest neighbors. The resulting edges between voxels i and j are assigned edge weights

$$w_{ij} = e^{\|x_i - x_j\|_1} \quad (6)$$

where x_i and x_j denote intensity values of a median filtered version (index radii $r_x = 1$, $r_y = 1$, and $r_z = 1$) of the initial data. Starting with zero detection time for the seed points the voxels' detection times are determined by computing shortest paths following the graph edges from the set of seed voxels to the voxels of interest. One set of seed points is obtained by smoothing the current segmentation with a binary median filter with index radii $r_x = 3$, $r_y = 3$, and $r_z = 3$ from the Insight Segmentation and Registration Toolkit (ITK, www.itk.org). After morphological erosion this gives a first estimate of voxels that are likely to be inside of the object to segment. The second set of seed points is obtained by morphologically dilating the current segmentation and selecting those voxels that still remain part of the background, which gives an estimation of voxels outside of the object of interest that are not too close to the object boundaries. An adaptation to local image characteristics (edges and homogeneous regions) of the initial segmentation is then achieved through assigning each voxel to that region, i.e., foreground or background, for whoever's seeds it is detected first. Finally the resulting segmentation is again subject to binary median filtering with slightly smaller index radii $r_x = 2$, $r_y = 2$, and $r_z = 2$.

2.4 Processing Pipeline

The whole procedure used for segmentation of either the left or the right caudate nucleus within a given MR volume data set is described in Algorithm 1.

Algorithm 1: Left/right caudate nucleus segmentation

Input: MR volume

Output: Binary voxel classification

- 1.1 Brain extraction;
 - 1.2 Re-sampling of input image to resolution suitable for PBT classifier;
 - 1.3 Normalization;
 - 1.4 PBT binary voxel classification;
 - 1.5 Re-sampling of classification to original image resolution;
 - 1.6 Adversarial fast marching;
 - 1.7 Binary median filtering;
-

Correl	UNC Ped	UNC Eld	BWH PNL	Total
Left	0.4183	0.5532	0.5223	0.4980
Right	0.1763	0.1045	0.0456	0.1088
Average	0.2973	0.3289	0.2840	0.3034

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.

3 Material and Experimental Setting

For training and evaluation of the proposed method several volumetric T1-weighted MR brain scans of varying spatial resolution and size from multiple sources were used. The vast majority of data (29 scans) has been provided by the Psychiatry Neuroimaging Laboratory (PNL) at the Brigham and Women’s Hospital (BWH), Boston. They all are accompanied by expert annotations. The other 20 data sets arose from several studies carried out mainly at the University of North Carolina’s (UNC) Neuroimaging Laboratory (NIAL), Chapel Hill. The studies contain a pediatric study, a Parkinson’s Disease study, and a test/re-test study. For the latter no ground-truth was available. The training data for both the classifiers for the right and the left caudate nucleus were taken from the data sets provided by the BWH PNL (serial numbers 1–8). In this paper our method is first of all quantitatively evaluated on the BWH PNL scans 16–29, which are all considered to be routine scans, 5 of the pediatric scans, and 5 scans of patients older than 55 years. (See Tabs. 1-3) Additionally, our algorithm was tested on 10 datasets of the same young healthy person acquired within 60 days on 5 different scanners. (See Tab. 4) The coefficient of variation of the volumetric measurements is an indicator on how stable the method operates in a test/re-test situation including scanner variability. We refer to [8] and [9] for details on the used evaluation measures and scoring system.

The PBT voxel classifiers were built from approximately one million randomly selected training samples, i.e., voxels inside the head of the patients, uniformly distributed over all the input slices of the training scans. The scans were re-sampled to a voxel spacing of $1.0\text{mm} \times 1.0\text{mm} \times 2.0\text{mm}$ and size $256 \times 192 \times 128$, which was mainly due to hardware limitations that our implementation currently faces. Since 2-D contexts of voxels that are very close to each other are likely to be very similar only every second voxel in each of the three spatial dimensions was considered for random selection. The maximum number of features selected by AdaBoost in each tree node were increased level-wise beginning with 2 at the trees’ root node. The maximum depth of the trees learned was restricted to 10 and soft thresholding was turned off. The 2-D voxel context chosen for computing the 1,908 Haar-like features used per individual voxel sample was of size 61×61 centered at the voxel of interest.

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	0	1301	0	0	1825	10692000	0	1060	1158	0	1069734	3380920	316.1
Right	2121	2160	512	671	4030	104	962	2783	2409	0	1575	1324	84.1
Total						-					-	-	200.1

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.

It takes about 2–3 minutes to process one of the MRI volumes in a non-optimized C++ implementation of our segmentation method on a Fujitsu Siemens Computers notebook equipped with an Intel Pentium M 2.0 GHz processor and 2 GB of memory. Most of the computational cost is due to post-processing the initial segmentation result obtained by the PBT. With the same hardware as above building one classifier takes about 6 hours.

4 Results and Discussion

From the performance scores given in Tab. 1 it can be clearly recognized that our algorithm suffers from significant over fitting mainly due to scanner specific contrast characteristics. This may also eclipse potential over fitting to sub-group dependent anatomical characteristics: segmentation results are, with minor exceptions, only acceptable for those data sets that come from the same routine study as the training data. This can also be seen by examining individual segmentations obtained by our method on representative scans from each of the data sub-groups as in Fig. 2. As one of the results of this problem the performance of our method in a test/re-test situation is poor. (See Tab. 3) The reliability of our method when used for volumetric measurements can be estimated from figures in Tab. 2. Due to the very low segmentation performance on the pediatric and the elderly patient datasets associated correlation measures are not very expressive. The observable positive association between volume measurements for the left caudate nuclei in the routine data set is caused by, to some extent, regular under-segmentation provided by our algorithm. However, there is little association for the segmentations of the right caudate nuclei as our method seems to arbitrarily over- and under-segment the object of interest in equal manner.

5 Conclusion

In this paper we have presented a fully automatic approach to segmentation of the left and right caudate nuclei in 3-D MR images. The novelty of the proposed method lies both in composition and pipelining of different previously published methods as well as in the specific field of application. Experimental

results show that acceptable results can be obtained very fast and without any user-interaction if it has been possible to train the classifier on images that are acquired with the same MR scanner and associated settings as those images that are about to be segmented. The problem of over fitting to scanner and acquisition protocol dependent contrast characteristics currently present in the method could possibly be weakened by the integration of more reliable MR inter-scan intensity standardization approaches into the framework. Once the PBT shape detection step has been improved this way, more sophisticated approaches to refine the initial segmentation could be utilized. This may make the presented knowledge-based PBT-approach for fast fully automatic segmentation of the caudate nuclei a valuable tool in clinical practice.

References

1. Smith, S.M.: Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**(3) (2002) 143–155
2. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: *IEEE Intl. Conf. Comput. Vis.*, Beijing, China. (2005) 1589–1596
3. Tu, Z., Zhou, X., Comaniciu, D., Bogoni, L.: A learning based approach for 3D segmentation and colon detagging. In: *Europ. Conf. Comput. Vis.*, Graz, Austria. (2006) 436–448
4. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Stanford University, Department of Statistics, Technical Report (1998)
5. Viola, P., Jones, M.: Robust real-time object detection. In: *2nd International Workshop on Statistical Learning and Computational Theories of Vision*, Vancouver, Canada. (2001) 1–25
6. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In Michaelis, B., Krell, G., eds.: *DAGM-Symposium*. Volume 2781 of *Lect. Notes in Comp. Sci.*, Springer (2003) 297–304
7. Rex, D.E., Ma, J.Q., Toga, A.W.: The LONI pipeline processing environment. *NeuroImage* **19**(3) (2003) 1033–1048
8. Gerig, G., Jomier, M., Chakos, M.: Valmet: A new validation tool for assessing and improving 3D object segmentation. In: *Int. Conf. Med. Image Comput. and Comp.-Assist. Interv.*, Utrecht, The Netherlands. (2001) 516–523
9. Heimann, T., Styner, M., van Ginneken, B.: Workshop on 3d segmentation in the clinic: A grand challenge. mbi.dkfz-heidelberg.de/grand-challenge2007 (2007)