Fast and Robust 3-D MRI Brain Structure Segmentation

Michael Wels[1,4], Yefeng Zheng[2], Gustavo Carneiro[3,⋆], Martin Huber[4], Joachim Hornegger[1], and Dorin Comaniciu[2]

[1] Chair of Pattern Recognition, Department of Computer Science, University Erlangen-Nuremberg, Germany, `michael.wels@informatik.uni-erlangen.de`
[2] Integrated Data Systems, Siemens Corporate Research, Princeton, NJ, USA
[3] Institute for Systems and Robotics, Electrical and Computer Engineering Department, Technical University of Lisbon, Portugal
[4] Siemens CT SE5 SCR2, Erlangen, Germany

Abstract. We present a novel method for the automatic detection and segmentation of (sub-)cortical gray matter structures in 3-D magnetic resonance images of the human brain. Essentially, the method is a top-down segmentation approach based on the recently introduced concept of Marginal Space Learning (MSL). We show that MSL naturally decomposes the parameter space of anatomy shapes along decreasing levels of geometrical abstraction into subspaces of increasing dimensionality by exploiting parameter invariance. At each level of abstraction, i.e., in each subspace, we build strong discriminative models from annotated training data, and use these models to narrow the range of possible solutions until a final shape can be inferred. Contextual information is introduced into the system by representing candidate shape parameters with high-dimensional vectors of 3-D generalized Haar features and steerable features derived from the observed volume intensities. Our system allows us to detect and segment 8 (sub-)cortical gray matter structures in T1-weighted 3-D MR brain scans from a variety of different scanners in on average 13.9 sec., which is faster than most of the approaches in the literature. In order to ensure comparability of the achieved results and to validate robustness, we evaluate our method on two publicly available gold standard databases consisting of several T1-weighted 3-D brain MR scans from different scanners and sites. The proposed method achieves an accuracy better than most state-of-the-art approaches using standardized distance and overlap metrics.

1 Introduction

Currently, many scientific questions in neurology, like the revelation of mechanisms affecting generative or degenerative processes in brain development, require quantitative volumetric analysis of (sub-)cortical gray matter structures
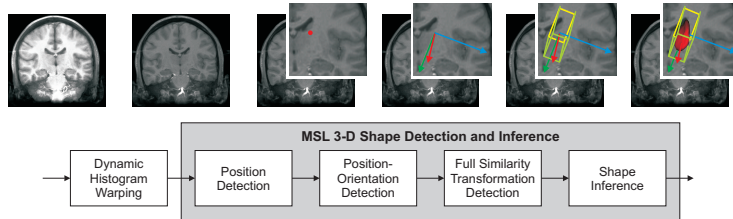
Fig. 1. The processing pipeline of the proposed 3-D shape detection and inference method. Each image (detection and delineation of the left caudate) schematically represents the input and/or output of individual processing steps. Please view in color.

in large populations of patients and healthy controls. For instance, atrophy in the presence of Alzheimer's disease considerably affects morphology of the hippocampus. In addition, 3-D segmentation of various deep gray matter structures facilitates image-based surgical planning, therapy monitoring, and the generation of patient-specific geometrical models from imaging data for further processing. As a result of unclear boundaries, shape complexity, and different anatomical definitions, precise manual delineation is usually time consuming and user dependent. Moreover, typical artifacts present in MR imaging (Rician noise, partial volume effects, and intra-/inter-scan intensity non-uniformities) challenge the consistency of manual delineations. Therefore, a system for the automatic detection and segmentation of (sub-)cortical gray matter structures not only has the potential to increase segmentation consistency, but also has the capability of facilitating large-scale neuromorphological studies.

We propose a fully automatic method for the detection and delineation of the following eight (sub-)cortical gray matter structures: the left and right caudate nucleus, hippocampus, globus pallidus, and putamen. Our method consists of two major steps: 1) we standardize the observed MR intensities by non-rigidly aligning their histogram to a template histogram by means of Dynamic Histogram Warping (DHW) [1]; and 2) for each (sub-)cortical structure of interest we detect and infer its position, orientation, scale, and shape in an extended Marginal Space Learning (MSL) framework [2], which explicitly integrates shape inference into the overall MSL formulation. The overall system block diagram is depicted in Fig. 1.

In contrast to other methods [3–5] where a partly manually initialized nine parameter registration is part of the approaches we do not require the input volumes to be spatially normalized. In some cases [3, 4], the feature pools used for discriminative model generation are enriched with features explicitly encoding normalized location. In accordance with this observation, the approaches are only evaluated on spatially normalized data sets from one type of MR scanner that are not publicly available [4, 5]. Nevertheless, Morra et al. [3] report state-of-the-art results on data sets that have not been subject to spatial normalization. Apart from that, the mentioned methods make use of machine learning in a similar manner as we do, but follow a bottom up approach ascending from the lowest level of abstraction, i.e., the level of individual voxels, to the level of complete anatomical entities.

Alignment of a probabilistic atlas by means of an affine registration also plays an important role in further approaches [6, 7]. While sometimes [6] quantitative evaluation is only carried out on simulated data, the method of Akselrod-Ballin et al. [7] is trained and evaluated on only one publicly available dataset that has been subject to a specific preprocessing including intensity standardization. By generating observation or discriminative models based on intensity values without explicitly allowing for inter-scan intensity variations [3–5, 7] the resulting models are at the risk of being over-adapted to specific contrast-characteristics of the data at hand.

2 Methods

2.1 Combined 3-D Shape Detection and Shape Inference

For combined 3-D rigid anatomy detection and shape inference we use a method based on the concept of Marginal Space Learning (MSL) [2]. We estimate the structure of interest's center $\boldsymbol{c} = (c_1, c_2, c_3) \in \mathbb{R}^3$, orientation $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in [-\pi, \pi] \times [-\pi/2, \pi/2] \times [-\pi, \pi]$ represented as Euler angles in $z - x - z$ convention, scale $\boldsymbol{s} = (s_1, s_2, s_3) \in \{ \boldsymbol{s} \in \mathbb{R}^3 \, | \, s_i > 0, i = 1, 2, 3 \}$, and shape $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^{3 \times n}$. The latter consists of canonically sampled 3-D points on the surface of an object to be segmented. Note that θ is relative to $\boldsymbol{c}$, $\boldsymbol{s}$ is relative to $\boldsymbol{c}$ and $\boldsymbol{\theta}$, and $\boldsymbol{X}$ is relative to $\boldsymbol{c}$, $\boldsymbol{\theta}$, and $\boldsymbol{s}$. Let $\mathcal{V} = \{ 1, 2, \ldots, N \}$, $N \in \mathbb{N}$, be a set of indices to image voxels, $\boldsymbol{Y} = (y_v)_{v \in \mathcal{V}}$, $y_v \in \{ -1, 1 \}$, a binary segmentation of the image voxels into object and non-object voxels, and f be a function with $\boldsymbol{Y} = f(\boldsymbol{I}, \boldsymbol{\Theta})$ that provides a binary segmentation of volume $\boldsymbol{I}$ using segmentation parameters $\boldsymbol{\Theta} = (\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{X})$. Let $\boldsymbol{Z} = (\boldsymbol{z}_{\boldsymbol{\Theta}})$ be a family of high-dimensional feature vectors extracted from a given input volume $\boldsymbol{I} = (i_v)_{v \in \mathcal{V}}$ and associated with different discretized configurations of $\boldsymbol{\Theta}$. In our context $\boldsymbol{Z}$ includes voxel-wise context encoding 3-D generalized Haar features [8] to characterize possible object centers and steerable features [2] that are capable of representing hypothetical orientations and optionally scaling relative to a given object center or shape surface point. These features were chosen for our method because of their fast computation and effective representation [2].

We search for the optimal parameter vector

$$\boldsymbol{\Theta}^* = \arg \max_{\boldsymbol{\Theta}} p(y = 1 | \boldsymbol{\Theta}, \boldsymbol{I}, \boldsymbol{M}^{(\boldsymbol{\Theta})}) = \arg \max_{\boldsymbol{\Theta}} p(y = 1 | \boldsymbol{Z}, \boldsymbol{M}^{(\boldsymbol{\Theta})}) \qquad (1)$$

maximizing the posterior probability of the presence, i.e., $y = 1$, of a sought anatomy given the discriminative model $\boldsymbol{M}^{(\boldsymbol{\Theta})}$ and the features $\boldsymbol{Z}$ extracted from the input volume $\boldsymbol{I}$ using a certain set of values for the parameters $\boldsymbol{\Theta}$.

Let $\pi^{(\boldsymbol{c})}(\boldsymbol{Z})$, $\pi^{(\boldsymbol{c}, \boldsymbol{\theta})}(\boldsymbol{Z})$, $\pi^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s})}(\boldsymbol{Z})$, $\pi^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{X})}(\boldsymbol{Z})$ denote the vectors of components of $\boldsymbol{Z}$ associated with individual groups of elements $(\boldsymbol{c})$, $(\boldsymbol{c}, \boldsymbol{\theta})$, $(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s})$, and $(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{X})$ of the parameter vector $\boldsymbol{\Theta}$. The MSL method avoids exhaustively searching the high-dimensional parameter space spanned by all the possible $\boldsymbol{\Theta}$ by exploiting the fact that ideally for any discriminative model for center detection with parameters $\boldsymbol{M}^{(\boldsymbol{c})}$ working on a restricted amount of possible features

$$\boldsymbol{c}^* = \arg \max_{\boldsymbol{c}} p(y = 1 | \pi^{(\boldsymbol{c})}(\boldsymbol{Z}), \boldsymbol{M}^{(\boldsymbol{c})}) \qquad (2)$$

holds, as the object center $\boldsymbol{c}$ is invariant under relative reorientation, relative rescaling, and relative shape positioning. Similarly, we have

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} p(y = 1 | \pi^{(\boldsymbol{c}^*, \boldsymbol{\theta})}(\boldsymbol{Z}), \boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta})}) \tag{3}$$

for combined position-orientation detection with model parameters $\boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta})}$ where only features $\pi^{(\boldsymbol{c}^*, \boldsymbol{\theta})}(\boldsymbol{Z})$ with $\boldsymbol{c} = \boldsymbol{c}^*$ are considered. This is due to the fact that position and orientation are invariant under relative rescaling and relative shape positioning. Analogous considerations yield

$$\boldsymbol{s}^* = \arg \max_{\boldsymbol{s}} p(y = 1 | \pi^{(\boldsymbol{c}^*, \boldsymbol{\theta}^*, \boldsymbol{s})}(\boldsymbol{Z}), \boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s})}) \tag{4}$$

for the object's scaling, and

$$\boldsymbol{X}^* = \arg \max_{\boldsymbol{X}} p(y = 1 | \pi^{(\boldsymbol{c}^*, \boldsymbol{\theta}^*, \boldsymbol{s}^*, \boldsymbol{X})}(\boldsymbol{Z}), \boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{x})}, \boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{X})}) \tag{5}$$

for the object's shape where $\boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{x})}$ are the parameters of a local shape model with respect to individual surface points $\boldsymbol{x}$ and parameters $\boldsymbol{M}^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{X})}$ represent a global shape model. Equations (2)–(5) naturally establish a chain of discriminative models exploiting search space parameter invariance for combined 3-D shape detection and shape inference. It allows us to apply different discriminative models descending along geometrical abstraction as, in our framework, the object center $\boldsymbol{c}$ alone is the most abstract and the complete set of parameters $\boldsymbol{\Theta}$ is the least abstract shape representation. Therefore, MSL establishes a hierarchical decomposition of the search space along decreasing levels of geometrical abstraction with increasing dimensionality of the considered parameter subspace.

2.2 3-D Shape Detection: Similarity Transformation Estimation

Let $\mathcal{Z}$ be the set of annotated image volumes in their transformed feature representation as mentioned above. We will refer to $\mathcal{Z}$ as the training data. In order to find the first parts of the optimal parameter vector $\boldsymbol{\Theta}^*$ describing a nine parameter similarity transformation, i.e., $\boldsymbol{c}^*$, $\boldsymbol{\theta}^*$, and $\boldsymbol{s}^*$, we have to learn discriminative models $p(y = 1 | \pi^{(\boldsymbol{c}^*)}(\boldsymbol{Z}))$, $p(y = 1 | \pi^{(\boldsymbol{c}^*, \boldsymbol{\theta})}(\boldsymbol{Z}))$, and $p(y = 1 | \pi^{(\boldsymbol{c}^*, \boldsymbol{\theta}^*, \boldsymbol{s})}(\boldsymbol{Z}))$. Following the concept of MSL [2] we generate a set of positive and negative training examples $\mathcal{C} = \{ (\pi^{(\boldsymbol{c})}(\boldsymbol{Z}), y) \,|\, \boldsymbol{Z} \in \mathcal{Z} \}$ to train a probabilistic boosting tree (PBT) model [9] for position detection. The feature vectors $\pi^{(\boldsymbol{c})}(\boldsymbol{Z})$ consist of 3-D generalized Haar features [8] encoding voxel context of candidate object centers based on observed intensity values. Decreasing the level of geometric abstraction we analogously train a PBT model for combined position-orientation detection based on an extended set of training examples $\mathcal{P} = \{ (\pi^{(\boldsymbol{c}, \boldsymbol{\theta})}(\boldsymbol{Z}), y) \,|\, \boldsymbol{Z} \in \mathcal{Z} \}$ where $\pi^{(\boldsymbol{c}, \boldsymbol{\theta})}(\boldsymbol{Z})$, associated with $(\boldsymbol{c}, \boldsymbol{\theta})$ and an image volume, is made of steerable features [2]. They allow varying orientation and scaling to be encoded in terms of aligned and scaled intensity sampling patterns. In accordance with this scheme, steerable features are also used to finally train a PBT for full nine parameter similarity transformation detection based on $\mathcal{S} = \{ (\pi^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s})}(\boldsymbol{Z}), y) \,|\, \boldsymbol{Z} \in \mathcal{Z} \}$ where $\pi^{(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s})}(\boldsymbol{Z})$ is derived from $(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{s})$ and the associated image volume.

Table 1. Average segmentation accuracy (left and right structures grouped together) for IBSR 18 of models trained from mutually exclusive training and test data. Except for the Dice coefficient $(2 \cdot TP/(2 \cdot TP + FP + FN)$ where TP, FP, and FN denote the number of true positive, false positive, and false negative voxels, respectively) see [13] on details on the used accuracy metrics.

Structure	Overlap Err. [%]	Dice Coeff. [%]	Volume Diff. [%]	Abs. Dist. [mm]	RMS Dist. [mm]	Max. Dist. [mm]
Caudate nucleus	32.42 ± 6.14	80.49 ± 4.51	9.57 ± 8.45	0.67 ± 0.17	1.10 ± 0.23	7.76 ± 1.82
Hippocampus	41.96 ± 4.69	73.34 ± 3.73	21.14 ± 17.29	0.91 ± 0.15	1.33 ± 0.21	6.34 ± 1.63
Globus pallidus	39.72 ± 7.05	74.97 ± 5.88	20.97 ± 12.38	0.79 ± 0.24	1.24 ± 0.37	5.53 ± 1.63
Putamen	29.82 ± 5.20	82.37 ± 3.65	13.76 ± 7.59	0.72 ± 0.20	1.15 ± 0.28	6.60 ± 1.85

2.3 3-D Shape Inference under Global Shape Constraints

For the final object shape we further decompose

$$\pi^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{X})}(\boldsymbol{Z}) = \left(\pi^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)}(\boldsymbol{Z})\right)_{i=1,\ldots,n}$$

where $\pi^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)}(\boldsymbol{Z})$ are the features associated with an image volume and individual relatively aligned candidate points $(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)$ for the surface of the object of interest. In order to apply discriminative modeling we assume the $\boldsymbol{x}_i$ and correspondingly $\pi^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)}(\boldsymbol{Z})$ to be independently and identically distributed (i.i.d.) and approximate

$$\boldsymbol{X}^* = \arg\max_{\boldsymbol{X}} p(y = 1 | \pi^{(\boldsymbol{c}^*,\boldsymbol{\theta}^*,\boldsymbol{s}^*,\boldsymbol{X})}(\boldsymbol{Z}), \boldsymbol{M}^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x})}, \boldsymbol{M}^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{X})})$$

$$\approx \arg\max_{\boldsymbol{X}} \left[\prod_{i=1}^{n} p(y_i = 1 | \pi^{(\boldsymbol{c}^*,\boldsymbol{\theta}^*,\boldsymbol{s}^*,\boldsymbol{x}_i)}(\boldsymbol{Z}), \boldsymbol{M}^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x})}) \right] p(\boldsymbol{X} | \boldsymbol{c}^*, \boldsymbol{\theta}^*, \boldsymbol{s}^*, \boldsymbol{M}^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{X})}) \quad (6)$$

in an iterative manner. The term $p(y_i = 1 | \pi^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)}(\boldsymbol{Z}))$ describes the probability that the relatively aligned point $(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)$ is part of the shape to be inferred, i.e., lies on its surface, and $p(\boldsymbol{X} | \boldsymbol{c}^*, \boldsymbol{\theta}^*, \boldsymbol{s}^*, \boldsymbol{M}^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{X})})$ is a global shape model [10]. We estimate $p(y = 1 | \pi^{(\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i)}(\boldsymbol{Z}))$ with a PBT model [9] using steerable features [2] trained on $\mathcal{X} = \{ (\pi_{\boldsymbol{c},\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{x}_i}(\boldsymbol{Z}), y) \,|\, i = 1,\ldots,n; \boldsymbol{Z} \in \mathcal{Z} \}$. An iterative approach for (6) is suitable as, in practice, $\boldsymbol{X} \in \mathbb{R}^{3 \times n}$ only varies around the mean shape positioned relatively to the $(\boldsymbol{c}^*, \boldsymbol{\theta}^*, \boldsymbol{s}^*)$ detected before at time $t = 0$ and the previous most likely anatomy shape in each iteration $t = 1,\ldots,T$.

3 Material and Experimental Setting

For training and quantitative evaluation of our system there were four sets of T1-weighted MRI scans available. The first one is a subset of the "Designed Database of MR Brain Images of Healthy Volunteers"[5] [11] (DDHV) containing 20 scans. The associated ground-truth annotations were manually recovered

---

[5] The database was collected and made available by the CASILab at the University of North Carolina, Chapel Hill. The images were distributed by the MIDAS Data Server at Kitware, Inc. (insight-journal.org/midas). The authors would like to thank Martin Styner, Clement Vachet, and Paul Pandea for helping to preprocess parts of the data.

from automatically generated segmentations [12] of the structures of interest. The second collection of 18 MRI scans was provided by the Center of Morphometric Analysis at the Massachusetts General Hospital and is publicly available on the Internet Brain Segmentation Repository[6] (IBSR 18). The scans are accompanied by detailed ground-truth annotations including the (sub-)cortical structures of interest in this paper.[7] A subset[8] of the data provided by the NIH MRI Study of Normal Brain Development[9] consisting of 10 pediatric data sets states another collection (NIH) of annotated MR scans used for model generation. They have been manually annotated by the authors for training purposes. Additionally, we use data provided by the ongoing "3-D Segmentation in the Clinic: A Grand Challenge" competition[10] [13] for training and evaluation of the proposed method. The collection consists of several volumetric T1-weighted MR brain scans of varying spatial resolution and size from multiple sources (MICCAI'07 training/testing). 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. The other 20 data sets arose from a pediatric study, a Parkinson's Disease study, and a test/re-test study carried out at the University of North Carolina's (UNC) Neuroimaging Laboratory (NIAL), Chapel Hill. A predefined evaluation protocol is carried out fully automatically after uploading the testing fraction of the data to the Cause'07 file server. We refer to Heimann et al. [13] for details on the used evaluation measures and scoring system.

All the images were re-oriented to a uniform orientation ("RAI"; right-to-left, anterior-to-posterior, inferior-to-superior) and resampled to isotropic voxel spacing ($1.0 \times 1.0 \times 1.0$ mm^3) for processing. For increasing the amount of training data we exploited natural brain symmetry and therefore doubled the size of any training data set used for model generation by mirroring all the data sets with respect to the mid-sagittal plane. Throughout all our experiments we ensured that training and testing data are mutually exclusive: We trained on DDHV,

---

[6] www.cma.mgh.harvard.edu/ibsr

[7] We corrected the ground-truth annotations for the left and the right caudate in the IBSR 18 data set to better meet the protocol applied by the "3-D Segmentation in the Clinic: A Grand Challenge" competition where the caudate is grouped with the nucleus accumbens in the delineations [13, 14].

[8] The following 10 data sets were used: defaced_native_100{2,3,7}_V{1,2}_t1w_r2, defaced_native_100{1,4,8}_V2_t1w_r2, and defaced_native_1005_V2_t1w_r2.

[9] The NIH MRI Study of Normal Brain Development is a multi-site, longitudinal study of typically developing children, from ages newborn through young adulthood, conducted by the Brain Development Cooperative Group and supported by the NICHD, the NIDA, the NIMH, and the NINDS (Contract #s N01-HD02-3343, N01-MH9-0002, and N01-NS-9-2314, -2315, -2316, -2317, -2319 and -2320). A listing of the participating sites and a complete listing of the study investigators can be found at www.bic.mni.mcgill.ca/nihpd/info/participating_centers.html. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH.

[10] www.cause07.org

Table 2. Average left/right caudate segmentation accuracy for the MICCAI'07 testing data set. The complete results can be found at www.cause07.org ("Segmentation Team"). As of 03/10/2009 our method ranks number 2 in the overall ranking list.

Cases	Overlap Err. [%]	Score	Volume Diff. [%]	Score	Abs. Dist. [mm]	Score	RMS Dist. [mm]	Score	Max. Dist. [mm]	Score	Total Score
Average UNC Ped	27.31	82.82	7.24	86.70	0.65	76.08	1.29	76.99	10.73	68.44	78.21
Average UNC Eld	34.48	78.31	10.73	81.18	0.73	72.82	1.31	76.59	11.57	65.96	74.97
Average BWH PNL	31.72	80.05	-16.54	70.20	0.65	76.08	1.26	77.56	11.19	67.09	74.20
Average All	31.38	80.27	-5.91	75.92	0.66	75.40	1.27	77.24	11.17	67.14	75.19

NIH, and IBSR 18 1-9 to evaluate on IBSR 18 10-18 and on DDHV, NIH, and IBSR 18 10-18 to evaluate on IBSR 18 1-9. We trained on DDHV, NIH, IBSR 18, and MICCAI'07 training to evaluate on MICCAI'07 testing.

As pointed out by Heimann et al. [13] there are differences in the annotation protocols used for annotating the caudate nuclei in data sets originating from the BWH and the UNC. In the former the "tail" of the caudate is continued much further dorsally. We therefore decided to detect it as a separate structure that can be attached to the caudate if required. We did not try to automatically determine the annotation protocol used from the imaging data itself as this may lead to over-fitted systems.

As our real discriminative models are not ideal as assumed for theoretical considerations we keep the top 100 candidates after position detection and the top 25 candidates after position-orientation detection for further processing steps in order to make the full similarity transformation detection more robust. For shape inference we use $T = 3$ iterations.

In an optimized and parallelized C++ implementation of our segmentation method it takes on average 13.9 sec. to detect and segment 8 (sub-)cortical structures in an MRI volume on a Fujitsu Siemens notebook equipped with an Intel Core 2 Duo CPU (2.20 GHz) and 3 GB of memory. Intensity standardization takes 1–2 sec. Our method is therefore faster than other state-of-the-art approaches whose timing is 50 sec. for 8 structures [5], 60 sec. for 1 structure [3], and 8 min. for 8 structures [4].

4 Experimental Results

As can be seen from Table 1 in terms of the Dice coefficient our method achieves better results (80%,73%,75%,82%) for the segmentation of the caudate nuclei, hippocampi, globi pallidi, and putamina on the same IBSR 18 data set than the methods of Akselrod-Ballin et al. [7] (80%, 69%, 74%, 79%) and Gouttard et al. [12] (76%,67%,71%,78%) except for the caudate nuclei in comparison to the method of Akselrod-Ballin et al. [7], where we reach a comparable accuracy. It also reaches a higher score for the caudate nuclei and putamina on IBSR 18 than the method of Bazin and Pham [14] (78%,81%), which does not address segmentation of the hippocampi and globi pallidi.

The overall average score in Table 2 shows that for segmenting the caudate nuclei our method performs better than the methods of Morra et al. [3] (73.38), Bazin and Pham [14] (64.73) and Tu et al. [4] (59.71). All the mentioned methods were evaluated on the same MICCAI'07 testing data set.

5 Conclusions

In this paper we integrated shape inference into the overall MSL methodology from the theoretical point of view. We showed that MSL decomposes the parameter space of anatomy shapes along decreasing levels of geometrical abstraction into subspaces of increasing dimensionality and applied MSL to the difficult problem of (sub-)cortical gray matter structure detection and shape inference. In an evaluation on publicly available gold standard databases our method works equally fast, robust, and accurate at a state-of-the-art level.

References

1. Cox, I.J., Hingorani, S.L.: Dynamic histogram warping of image pairs for constant image brightness. In: Proc. ICIP, Washington, D.C., USA. (1995) 366–369
2. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. IEEE T. Med. Imag. **27**(11) (2008) 1668–1681
3. Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Toga, A.W., Thompson, P.M.: Automatic subcortical segmentation using a contextual model. In: Proc. MICCAI, New York, NY, USA. (2008) 194–201
4. Tu, Z., Narr, K.L., Dollár, P., Dinov, I., Thompson, P.M., Toga, A.W.: Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE T. Med. Imag. **27**(4) (2008) 495–508
5. Corso, J.J., Tu, Z., Yuille, A., Toga, A.: Segmentation of sub-cortical structures by the graph-shifts algorithm. In: Proc. IPMI, Kerkrade, the Netherlands. (2007) 183–197
6. Scherrer, B., Forbes, F., Garbay, C., Dojat, M.: Fully Bayesian joint model for MR brain scan and structure segmentation. In: Proc. MICCAI, New York, NY, USA. (2008) 1066–1074
7. Akselrod-Ballin, A., Galun, M., Gomori, J.M., Brandt, A., Basri, R.: Prior knowledge driven multiscale segmentation of brain MRI. In: Proc. MICCAI, Brisbane, Australia. (2007) 118–126
8. Tu, Z., Zhou, X.S., Barbu, A., Bogoni, L., Comaniciu, D.: Probabilistic 3D polyp detection in CT images: The role of sample alignment. In: Proc. CVPR, New York, USA. (2006) 1544–1551
9. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: Proc. ICCV, Beijing, China. (2005) 1589–1596
10. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Comp. Vis. Image Understand. **61**(1) (1995) 38–59
11. Bullitt, E., Zeng, D., Gerig, G., Aylward, S., Joshi, S., Smith, J.K., Lin, W., Ewend, M.G.: Vessel tortuosity and brain tumor malignancy: A blinded study. Acad. Radiol. **12**(10) (2005) 1232–1240
12. Gouttard, S., Styner, M., Joshi, S., Smith, R.G., Cody, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. In: Proc. SPIE Med. Imag., San Diego, CA, USA. (2007) 65122J–1–11
13. Heimann, T., Styner, M., van Ginneken, B.: Workshop on 3D segmentation in the clinic: A grand challenge. In: T. Heimann, M. Styner, B. van Ginneken (Eds.): 3D Segmentation in the Clinic: A Grand Challenge (2007) 7–15.
14. Bazin, P.L., Pham, D.L.: Homeomorphic brain image segmentation with topological and statistical atlases. Med. Image Anal. **12**(5) (2008) 616–625