Caudate Segmentation using Deformable M-reps

Joshua H. Levy, Kevin Gorczowski, Xiaoxiao Liu, Stephen M. Pizer, and Martin Styner

Department of Computer Science, University of North Carolina at Chapel Hill
levy@cs.unc.edu

Abstract. We use object scale and then atom scale Bayesian optimization of m-reps to automatically segment the caudate nucleus in brain MRI images. Our shape priors are learned after alignment of m-reps fit to 15 manual segmentations of caudates. The alignment is to the m-rep means at the respective scale levels. Our appearance likelihood is learned from regional intensity quantile functions from images that have been aligned, corrected for inhomogeneity, and intensity normalized. We begin the segmentation of a target image by the image preprocessing steps described above followed by an initialization of the mean m-rep model using the image alignment transformation. The segmentation then proceeds by optimizing the posterior probability over a shape space encompassing eigenmodes of full m-rep shape variation capturing 87% of the total variance in the training population of left caudates (83% right). The segmentation concludes by successively optimizing the posterior probability of the residual changes in the medial atoms making up the m-rep.
Since only a weak variant of a part of our segmentation method was inadvertently applied in the Grand Challenge, its results are not representative of the performance of our method. Qualitative results of the actual method are reported within; quantitative results will be reported separately.

1 Introduction

A popular approach for segmenting 3D objects from a medical image in a largely automatic way has been to deform a geometric model into the target image. Such methods optimize an objective function in which a major term measures the match of the model to the target image. Another term can reflect knowledge about the anatomy.

A Bayesian point of view allows one to reflect both of these pieces of knowledge. One seeks the most probable object model $\mathbf{m}$ given the image $\mathbf{I}$ over objects with the specified geometric properties i.e. one optimizes $p(\ \mathbf{m} \mid \mathbf{I}\ )$. By Bayes' theorem,

$$\arg\max_{\mathbf{m}} p(\ \mathbf{m} \mid \mathbf{I}\ ) = \arg\max_{\mathbf{m}} \log p(\ \mathbf{m}\ ) + \log p(\ \mathbf{I} \mid \mathbf{m}\ ) \tag{1}$$

The log prior term, $\log p(\ \mathbf{m}\ )$, reflects what is known about anatomic geometry and how it can vary. The log likelihood term, $\log p(\ \mathbf{I} \mid \mathbf{m}\ )$, reflects what image intensity patterns are consistent with the anatomy and how they can vary.

Because they are especially effective for training probability densities, we use the m-rep [1, 2] as the object representation and discrete regional intensity quantile functions (DRIQFs) [3] as the image intensity representation. We use principal component analysis (PCA) on DRIQFs as the means of estimating the likelihood distributions. We use principal geodesic analysis (PGA) [4], i.e., PCA on the tangent space to the Fréchet mean of the training m-reps (or their medial atom residuals at the atom scale) as the means of estimating the prior distributions. The PGA also yields a limited-dimensional space of credible objects within which the optimal object is sought.

In Sec. 2 we describe the training procedure we used to learn probability distributions on m-reps and DRIQFs for caudates. We then describe our method for segmenting the target data set. Section 3 contains a discussion of those segmentation results.

2 Methods

2.1 Image Preprocessing

The brain MRI images are preprocessed according to a scheme that we have also used for unbiased atlas production and for atlas based segmentation [5]. This preprocessing pipeline consists of four steps. Each image is rigidly aligned to an atlas. The aligned images undergo an automatic tissue segmentation [6, 7] and an intensity inhomogeneity correction of that removes gradual variations in the image intensities mainly due RF coil imperfections. All training images are then normalized into the same intensity range via a spline based histogram transfer function that matches the mean intensities of the tissue classes of WM, GM and CSF, as well as the overall range of the image intensities. The final preprocessing step is skull stripping using a procedure based on a mask generated from the binary tissue segmentation obtained in step 2.

2.2 Training m-reps and the prior distribution on them

The m-rep for a caudate nucleus is mathematically defined [1, 2] as a 2-manifold with boundary of medial atoms (Fig. 1, Left), where for all interior points of the manifold with boundary the atom consists of a hub location, the directions of two spokes emanating from the hub to the implied boundary of the object, and the common length of the two spokes. On the boundary (edge) of each manifold the atom has a third crest spoke bisecting the two spokes and having an additional length parameter that controls the sharpness of the associated crest on the medially implied object boundary. Our m-rep representation for a caudate samples this 2-manifold with boundary into a rectilinear grid of 3×7 atoms (Fig. 1).

Training both the log prior and the log likelihood begins with binary images of manual segmentations of the left (respectively right) caudate from the brain aligned training images. M-reps are fit to each of these binary training images using a variant of our segmentation program [8].

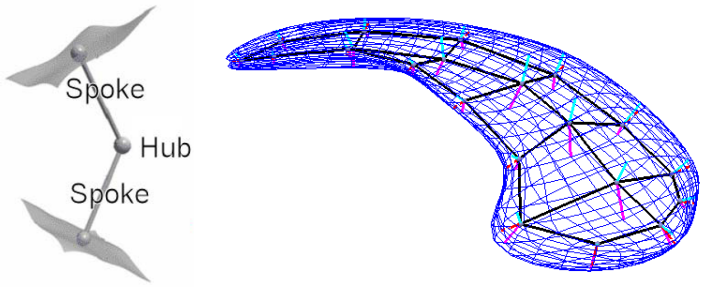

Fig. 1. A medial atom (Left). A discrete m-rep, and its implied boundary, for a caudate nucleus (Right).

The resulting m-reps are aligned to their mean and a new mean left (resp. right) caudate is computed. PGA yielded shape spaces made from the 8 dominant components, which captured 87% and 83% of the variance for the left and right caudates. The coefficients of these 8 components in any candidate m-rep within the object scale stage of the segmentation, together with the associated principal variances, yields the log prior penalty term in the objective function, weighted to have unit variance.

To train the segmentation at the atom scale, the residue from the projection of each training case into the object scale shape space is aligned according to the orientation and position of a local 3-5 atom patch in the projection. PGA on the aligned residues yields 5 dominant components, which captures 90%of the variance for the left caudate and 90% for the right caudate. The coefficients of these components in any candidate m-rep within the atom scale stage of the segmentation, together with the associated principal variances yields the log prior penalty term in the objective function for optimizing that atom, weighted to have unit variance.

2.3 Training DRIQFs

The local image match function that we use as our log likelihood function is based on Broadhurst's discrete regional intensity quantile functions (DRIQF) [3, 9]. We use the training m-reps to subdivide each MRI image into local (overlapping) regions. In particular, we use the m-rep spoke ends and directions to define 58 interior and 58 exterior regions for the left (resp. right) caudate. Image intensity histograms are recorded for each region of each training image.

The local image match function measures statistical distance between the observed histogram in a region of the target image and the population of histograms observed for the corresponding region of the training set. Broadhurst has shown that although the histogram space requires a non-Pythagorean metric, each histogram can be uniquely mapped onto a quantile function through the inverse of its cumulative density function and that linear statistics are valid

on the resulting quantile space. PCA on the training cases yields a mean and eigenmodes of variation for the quantile functions. The log likelihood term in the objective function for segmentation is a Mahalanobis distance in this PCA space, weighted to have unit variance.

2.4 Segmentation

The aforementioned training was done separately for the left and right caudates, and segmentation proceeded on each caudate using the appropriate probability distributions. The segmentation is performed via an iterative application of conjugate gradient optimization. First the template model (the side-specific mean computed in Sec. 2.2) is initialized into the target image via a 7 parameter similarity transform that optimizes the log likelihood of the target object according to the probability distribution of DRIQFS learned in Sec. 2.3. Then the model is deformed in the PGA shape space, first at the whole object scale and then at the local atom scale, so that the initialization transform applied to the deformed model optimizes the objective function (1). We then iterate, first computing a new initialization transform that optimally brings the deformed model into the target image, and then computing the optimal deformation for this improved initialization. Our experience has been that two iterations of this method yield high quality segmentations and can prevent certain local minima problems.

3 Results

The segmentations we submitted to the workshop were produced in a pipeline that deviated from the method outlined in Sec. 2 in ways which we feel invalidate our results. Because of serious data preparation errors these segmentations can be explained as *garbage in, garbage out*.

Several problems with our log likelihood training were due to confusion over units of *cm* vs. *mm*. The sampling radius of each DRIQF was too narrow by a factor of 10 voxels. This led to an undersampling of exterior intensities during DRIQF training. Furthermore, the union of the regions did not cover the entire caudate, i.e. there were large gaps which were never sampled. Consequently the DRIQFs poorly described the image intensity relative to the object.

The object-scale log prior distribution was trained from models that had been aligned at the whole-brain scale, but not aligned at the caudate scale. As a result of this, the PGA statistics were dominated by orientation components and also allowed transformations that produce atypical objects.

Our segmentation consisted of only a variation of a subset of the procedure described in Sec. 2.4. Our segmentation consisted of only a single round of object scale PGA deformation, initialized by the identity transform. We did not iterate over initialization and segmentation, and we did not use the image data to drive initialization. Because the atom stage had not been adequately debugged prior to the workshop deadline, the log-prior was trained only at the object scale. Thus the segmentations we submitted had only been deformed at the object

scale; typically this produces segmentations that are in many places a few voxels off.

Fixing all of these mistakes makes such strong qualitative improvements that our submitted results (Figs. 3-6) should be ignored. Fig. 2 shows typical examples of the segmentations produced by our correct method in comparison to our submitted segmentations from the incorrect version of our method. It is our intention to work with the workshop organizers to apply their quantitative analysis to the results of the correct version of our method.

3.1 Training sample needs

Any probabilistic training requires a large sample size in order to robustly estimate the underlying distributions. This applies to our methods of applying PGA on m-reps and m-rep residuals, and applying PCA on DRIQFs. It is our opinion that the set of 15 training images provided by the workshop was too small to adequately describe the variation in caudate shape across patients. We expect that the quality of our target image segmentations would greatly benefit from 20 additional training cases.

The success of our method, and indeed any method based on probabilistic training, in segmenting a novel image depends on that image being typical with respect to the training population. The physical differences between the adult population on which we trained, and the pediatric target population are such that the performance of our method *as trained on adults* applied to the pediatric images should be poor; and indeed they were. It is our belief that with a pediatric training population we could learn probability distributions that would allow our method to segment pediatric targets.

3.2 Benefits expected from multiobject segmentation

We believe that our method would be significantly improved if, instead of focusing on segmenting a single object, it simultaneously segmented that object and others surrounding it. This multiobject segmentation should use geometric probability distributions not only on each of the objects, but also on their interrelations. A method to do such multiobject segmentation and the associated learning of probability distributions is well along in its development in our laboratory [10]. For the caudate nucleus in particular, a multiobject segmentation together with the putamen, nucleus incumbens, and lateral ventricle should provide superior results. In particular, those of our results in which the segmented caudate leaks into the nucleus incumbens should be improved.

3.3 Run-time performance

Our training is performed off-line, so we exclude it from this discussion. Preprocessing of an image typically requires on 15 to 20 minutes of computation using a single modern Linux workstation. Most of this time is spent on the registration

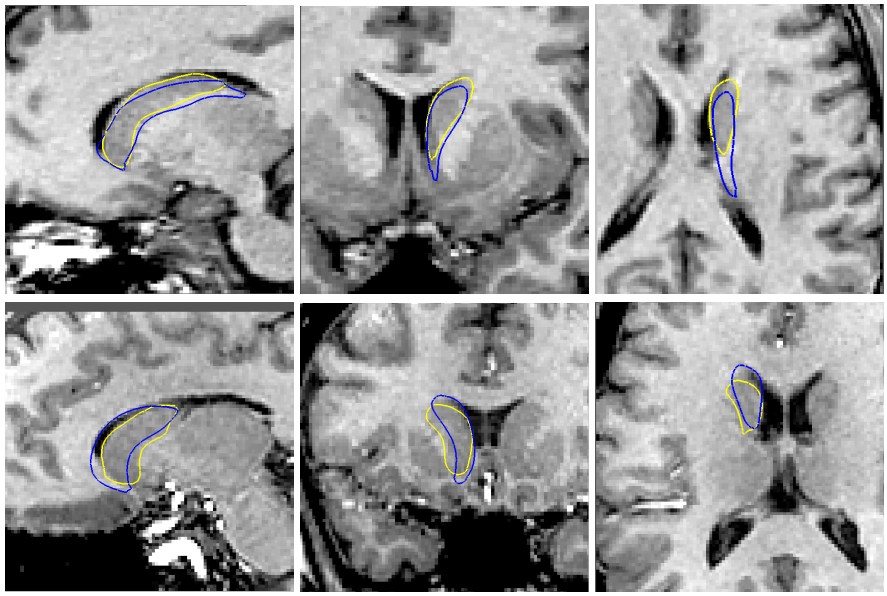

Fig. 2. From left to right, a sagittal, coronal and transversal slice from a a subject in the adults BWH group (top) and one in the elderly UNC group (bottom). The outline of the segmentation produced by *correctly* applying our method is shown in yellow. The submitted result from the incorrect version of our method is shown in blue.

and tissue segmentation steps. The object scale initialization and deformations combined typically run in under 5 minutes. Atom scale segmentation runs on average in 20 minutes. Overall our research implementation of our method can fully segment an image in 45 minutes.

References

1. Pizer, S., Fletcher, T., Fridman, Y., Fritsch, D., Gash, A., Glotzer, J., Joshi, S., Thall, A., Tracton, G., Yushkevich, P., Chaney, E.: Deformable m-reps for 3d medical image segmentation. International Journal of Computer Vision - Special UNC-MIDAG issue **55**(2) (2003) 85–106
2. Pizer, S.M., Han, Q., Joshi, S., Fletcher, P.T., Yushkevich, P.A., Thall, A.: Synthesis, deformation, and statistics of 3d objects via m-reps. In Siddiqi, K., Pizer, S., eds.: Medial Representations: Mathematics, Algorithms, and Applications. Kluwer (2007)
3. Broadhurst, R., Stough, J., Pizer, S., Chaney, E.: A statistical appearance model based on intensity quantiles histograms. ISBI (2006)
4. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. Medical Imaging, IEEE Transactions on **23** (2004) 995–1005

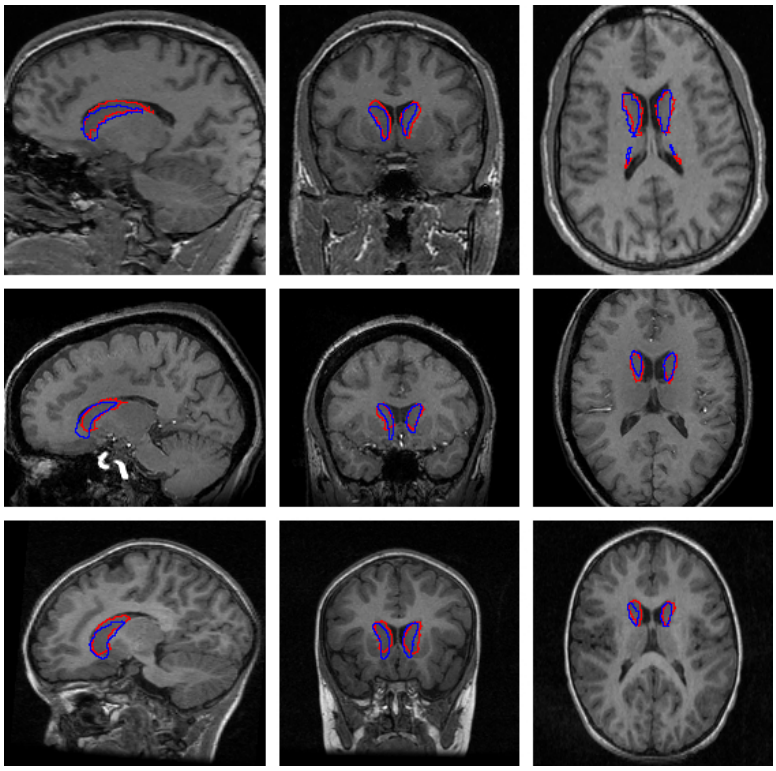

Fig. 3. From left to right, a sagittal, coronal and transversal slice from a a subject in the adults BWH group (top), one in the elderly UNC group (middle) and one in the pediatric UNC group (bottom). The outline of the reference standard segmentation is in red, the outline of the segmentation of the method described in this paper is in blue.

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	70.6	56	9.0	84	2.5	6	3.4	38	13.1	62	49
UNC Ped 14	46.5	70	10.6	81	1.0	61	1.4	75	5.3	84	74
UNC Ped 15	49.1	70	-5.3	90	1.2	56	1.5	72	5.9	82	74
UNC Ped 19	91.9	42	45.5	36	13.7	0	14.8	0	32.9	4	16
UNC Ped 30	97.2	38	32.6	42	7.7	0	8.8	0	19.7	42	24
UNC Eld 01	56.0	64	4.9	82	1.4	48	1.9	66	7.3	78	68
UNC Eld 12	52.4	67	2.1	96	1.2	54	1.7	70	6.4	81	74
UNC Eld 13	38.3	76	3.1	94	0.8	68	1.3	78	5.3	84	80
UNC Eld 20	43.3	72	-4.8	88	1.0	64	1.3	76	6.7	80	76
UNC Eld 26	50.2	68	16.9	70	1.1	60	1.5	73	5.8	83	71
BWH PNL 16	48.7	70	-14.9	74	1.3	51	2.6	53	23.1	32	56
BWH PNL 17	47.8	70	-15.6	73	1.4	49	2.9	48	27.1	20	52
BWH PNL 18	69.5	56	-23.9	58	2.2	18	2.9	48	11.1	68	50
BWH PNL 19	55.5	65	-30.9	46	1.6	42	3.0	46	25.7	24	45
BWH PNL 20	48.6	70	-13.6	76	1.3	54	2.6	54	26.3	22	55
BWH PNL 21	62.1	61	-26.6	53	1.8	33	3.0	47	24.4	28	44
BWH PNL 22	46.3	71	-32.0	44	1.5	45	3.6	36	28.9	15	42
BWH PNL 23	45.9	71	6.4	88	1.0	62	1.4	74	9.5	72	74
BWH PNL 24	46.2	71	-18.9	67	1.0	64	1.4	76	7.6	78	71
BWH PNL 25	44.4	72	-13.7	76	1.2	54	2.6	53	23.4	32	58
BWH PNL 26	50.6	68	-1.8	86	1.2	56	1.8	68	13.5	60	68
BWH PNL 27	43.8	72	-22.4	61	1.5	46	3.4	39	26.1	23	48
BWH PNL 28	46.9	70	-8.4	85	1.4	47	3.2	42	25.4	25	54
BWH PNL 29	56.0	64	-6.5	88	1.3	54	1.7	70	7.8	77	70
Average All	54.5	66	-4.5	73	2.1	45	3.1	54	16.2	52	58
Average UNC Ped	71.0	55	18.5	67	5.2	25	6.0	37	15.4	55	48
Average UNC Eld	48.0	70	4.4	86	1.1	59	1.5	72	6.3	82	74
Average BWH PNL	50.9	68	-15.9	70	1.4	48	2.6	54	20.0	41	56

Fig. 4. 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	0.7317	0.1468	0.4778	0.4521
Right	-0.1184	0.6791	0.3710	0.3106
Average	0.3067	0.4129	0.4244	0.3813

Fig. 5. 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	3177	3330	2724	2415	1888	2838	2605	2642	2011	2020	2565	489	19.1
Right	3256	2869	3272	3476	3284	3315	3717	2661	3778	2943	3257	356	10.9
Total	-										-	-	15.0

Fig. 6. The volumetric measurements of the 10 datasets 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.

5. Gouttard, S., Styner, M., Joshi, S., Smith, R., Hazlett, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. Proceedings of SPIE **6512** (2007) 65122J
6. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of thebrain. Medical Imaging, IEEE Transactions on **18**(10) (1999) 897–908
7. Prastawa, M., Gilmore, J., Lin, W., Gerig, G.: Automatic Segmentation of MR Images of the Developing Newborn Brain. Med Image Anal **9**(5) (2005) 457–66
8. Han, Q., Merck, D., Levy, J., Villarruel, C., Damon, J., Chaney, E., Pizer, S.: Geometrically Proper Models in Statistical Training. In: Proceedings of Information Processing in Medical Imaging. (2007) 751–762 10.1007/978-3-540-73273-0_62.
9. Stough, J., Broadhurst, R., Pizer, S., Chaney, E.: Regional Appearance in Deformable Model Segmentation. In: Proceedings of Information Processing in Medical Imaging. (2007) 532–543 10.1007/978-3-540-73273-0_44.
10. Jeong, J.Y., Pizer, S.M., Ray, S.: Statistics on anatomic objects reflecting inter-object relations. In: MICCAI Conference: Workshop on Mathematical Foundations of Computational Anatomy. (2006) 136–145