Automatic Segmentation of the Caudate Nuclei using Active Appearance Models

K.O. Babalola, V. Petrovic, T.F. Cootes, C.J. Taylor, C.J. Twining, T.G. Williams, and A Mills

Division of Imaging Science and Biomedical Engineering, The University of Manchester, Manchester, UK
kola.babalola@manchester.ac.uk

Abstract. We describe the application of an active appearance model (AAM) based method to segmentation of the caudate nuclei. A "composite" 3D profile AAM was constructed from the surfaces of 15 subcortical structures using a training set of 50 subjects, and individual AAMs of the left and right caudate constructed from 227 subjects.
Segmentation starts with affine registration to initialise the composite model within the image, then a search using the composite model. This provides a reliable but coarse segmentation, used to initialise search with the individual caudate models. The results are further refined by reclassifying voxels within a one-voxel neighbourhood of the surface.
Application to 24 subjects resulted in 2 failed searches (L+R of same subject) and a mean Tanimoto overlap of 73.1% (failed searches excluded). The overall official score (failed searches included) was 71. The coefficient of variation when applied to 10 independently acquired datasets of the same subject was 3.5% (1.9% for left and 5.1% for right).

1 Introduction

We approach the task of segmenting the caudate nuclei provided for the 2007 MICCAI segmentation challenge using an AAM based method. The results returned by the AAM are post-processed to reclassify voxels within a one voxel neighbourhood of the surface. There are many other approaches to segmenting structures from 3D volumetric images, including deformable surface meshes, medial representations, segmentation by registration, and shape constrained level set approaches, amongst others. The AAM approach is based on the original work of Cootes *et al.*[1]. It requires the construction of an appearance model that captures the variation of shape and gray level across a training set. Shape variation is learnt by placing homologous points within each image in the training set, and variation in gray level by sampling image intensity within or around the model points. Matching is performed by a steepest descent type search in which the sum of squares of residual differences between model and image is driven toward a minimum. It can be very efficient, because it pre-computes the Jacobian describing the average change in residuals with respect to changes in

T. Heimann, M. Styner, B. van Ginneken (Eds.): 3D Segmentation in The Clinic:
A Grand Challenge, pp. 57-64, 2007.

model parameters on a training set, allowing parameter updates to be computed quickly [2].

Although AAMs are well known and their use in 3D has been described e.g. [3], a number of choices have to be made in tailoring an AAM approach to a specific task. These include the method of establishing correspondence, whether to model the entire volume within the structures of interest or limit this to sampling across the surface(s) of the structure(s) of interest, and how to update the model parameters during search.

Here we use profile AAMs, which involve placing points on the surfaces of the structures of interest and sampling the intensity along vectors normal to the surface. We establish correspondence amongst the surfaces of interest by groupwise registration of binary images. We are interested in the caudate nuclei in particular, however, we build a "composite" model of 15 subcortical structures including the caudate nuclei, and use the results of the search by this model to initialise single structure models of the left and right caudate. We have found this approach gives slightly better overall results, but more importantly, prevents the search from falling into local minima or diverging.

2 Method

2.1 Computing Correspondence

We constructed statistical models of shape and appearance using images in which a number of different structures have been given voxel labels. Thus for each image in the training set we are able to extract the set of voxels belonging to each individual structure. To construct the shape models, we require sets of points which define correspondences across the images.

We use a volumetric approach, in which we find the deformations of space which map the 3D binary images of a structure in each subject to one another. In particular, we use the "groupwise" approach of Cootes *et al.*[4], which seeks to find the deformations of space which allow us to construct the most compact model of a set of images. The output of this stage is a set of control points at corresponding points in the volume of each image.

2.2 Model Construction

By applying the above method to binary images describing examples of a particular structure, we can estimate the deformation fields mapping each image into a mean reference image. To estimate the mean surface we warp each binary image into the mean reference frame, compute the mean image and threshold it at 0.5 to generate a mean surface. We place points densely over the mean surface by triangulating the zero level set of its signed distance transform, and project the points to the native space of each training image using our estimates of the deformation fields. For each training image this gives an approximation of the surface of the structure with a given number of points which are on homologous parts

of the surface of that structure in all the other training images. When building "composite" models of two or more structures, the above steps are carried out independently for each structure and the resulting points concatenated.

2.3 Active Appearance Models

An Active Appearance Model is a statistical model of both the shape of a structure and its appearance, together with an algorithm for matching it to an image. The model is capable of synthesising an image of the object of interest, and the residual differences between the synthesised image and the target image are used to drive the search.

Although the original approach [2] was to model the intensities across an entire region containing the model, it is equally possible to only model the intensities in sparse regions, such as those around the surface of the object. A natural approach in 3D is to model the intensities along profiles normal to the boundary at each model point.

Thus if we have n_p model points, and at each point we sample the intensities along a profile of length n_s, centred on the point, we obtain a combined vector of intensities, $\mathbf{g}$, of dimension $n_g = n_s n_p$.

To construct a statistical model of such textures, we sample each training image in the same way, normalise the resulting vectors (so the mean of their elements is zero and their length is one unit), then apply PCA. As is done with the original appearance models, we combine the shape and texture together into a single model with the following form.

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \; ; \; \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape, $\bar{\mathbf{g}}$ the mean texture in a mean shaped patch, $\mathbf{Q}_s$,$\mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set, and $\mathbf{c}$ is a vector of parameters controlling both shape and texture.

2.4 Searching an Unseen Image

AAM searching is carried out in a coarse to fine manner on images at different levels in an image pyramid to prevent it from diverging or falling into local minima [2] (our searches were at levels 2, 1 and 0 - image resolution, and used the updating algorithm described by Cootes [5]). However, as it is a local search it needs to be initialised near the structure of interest. This can be done interactively, but an automated alternative is to register the image to be segmented with the image in the training set that determined the pose of the model, in this case the first image in the training set. This registration does not need to be accurate - rigid body translation is sufficient, although for the purposes of this task we carried out affine registration on skull stripped versions of the images using FLIRT and BET from the FSL library [6].

A preliminary search is carried out in the registered space initialised using the mean of the composite model. This provides a segmentation of all the structures in the model. A large model trained on too few images will not have sufficient

degrees of freedom to match to the general population accurately due to the model flexibility being constrained, but is less likely to fall into local minima or diverge. The points of the caudate are extracted and used to initialise the individual caudate models in the native space of the image being segmented, which provides a more accurate segmentation.

2.5 Boundary reclassification

The preliminary search using the AAM yields only an approximate estimate of the true caudate boundary constrained by the variation of shape contained within the model constructed from a limited number of examples. So a boundary reclassification step is performed in order to capture the true boundary of each example. Due to the limited resolution and size of the caudate region reclassification is performed on a limited scale - one voxel either side of the boundary returned by the AAM.

Boundary reclassification is based on voxel intensity. Explicit intensity distribution models are constructed from the 33 training images in the MICCAI WS data. Initially, the training set is normalised within a bounding box 5 voxels wider than each true caudate label in all directions to the same intensity mean and variance. Intensity distributions of the caudate and an immediate background obtained by a morphological dilation of the caudate label using a 3x3x3 structuring element $p(g|caudate)$ and $p(g|background)$ are evaluated for each example. The intensity models for the two regions are defined as their mean across the training set.

When a new example is encountered, intensities within the bounding box defined using the initial model fit are normalised and a narrow consideration region is defined around the initial boundary. The boundary is extended outwards through morphological dilation with a 3x3x3 structuring element and inwards through erosion with a 6-neighbourhood element. Each voxel within the eroded and dilated region is then reassigned to caudate or background depending on the likelihood of their corrected intensity $g + \delta$, being the caudate or the background according to the training set distribution models. The correction term δ is the difference of the mean of the trained caudate intensity model and mean caudate intensity in the AAM fit of the current image.

3 Results

We built the single structure profile AAMs of the caudate using a dataset of 227 MR brains in which a variety of subcortical structures had been manually labelled. Fifty of these were used in building the composite model. The data was provided courtesy of the Centre for Morphometric Analysis, Boston and included control subjects as well as subjects with Alzheimer's Disease, Schizophrenia, Attention Deficit Hyperactivity Disorder, and prenatal drug exposure. The age range spanned at least 4.5yrs to 83yrs.

To build the composite model the labels of 15 subcortical structures were used – the brain stem, the left and right pairs of the amygdala, caudate, hippocampi,

Correl	UNC Ped	UNC Eld	BWH PNL	Total
Left	0.7525	0.8273	-0.2634	0.4388
Right	0.9116	0.7775	-0.4098	0.4264
Average	0.8320	0.8024	-0.3366	0.4326

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	3401	3370	3407	3588	3533	3471	3406	3443	3476	3465	3456	66	1.9
Right	3683	3527	3523	3850	3198	3649	3385	3687	3532	3662	3570	182	5.1
Total	-										-	-	3.5

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.

lateral ventricles, thalamus, pallidum and putamen. 50 subjects were randomly selected from the 227 to constitute the training set for this model. Individual models of the left and right caudate were built using all 227 subjects. The models were applied to skull stripped versions of the test images and the model results post-processed as described in sections 2.4 and 2.5. The composite model had 4030 points including 861 for the left caudate and 871 for the right. The AAMs used 12 voxel long profiles, 6 voxels on each side of the surface boundary.

Binary versions of the segmentation output were submitted to the workshop organisers for analysis using the agreed volumetric and distance measures defined in [7]. The qualitative results are shown in Figure 1 and Tables 1–4 contain the quantitative results.

4 Discussion and Conclusions

The qualitative results in Figure 1 show good agreement with the gold standard and are typical of the results returned by our method. The method was unsuccessful for one subject (left and right of BWH PNL 26) because the affine registration failed and hence the model search was initialised in an incorrect part of the brain. Unfortunately, this failure skewed the results of the BWH PNL group and to a lesser extent the overall results, most noticeably in the correlations presented in Table 1. Table 4 shows summary data when the unsuccessful result was left out (compare with summary data in Table 3). A similar recalculation of the correlation coefficient for the BWH dataset gave a value of 0.76 (compared to -0.34 with the failures included).

The breakdown of volumetric and distance based measures in Table 1 show two trends worth noting. Firstly the different volumetric measures are not consistent across the groups. For all groups the overlap is generally between 70% and 80%, however the percentage volume difference of the UNC Eld group is signifi-

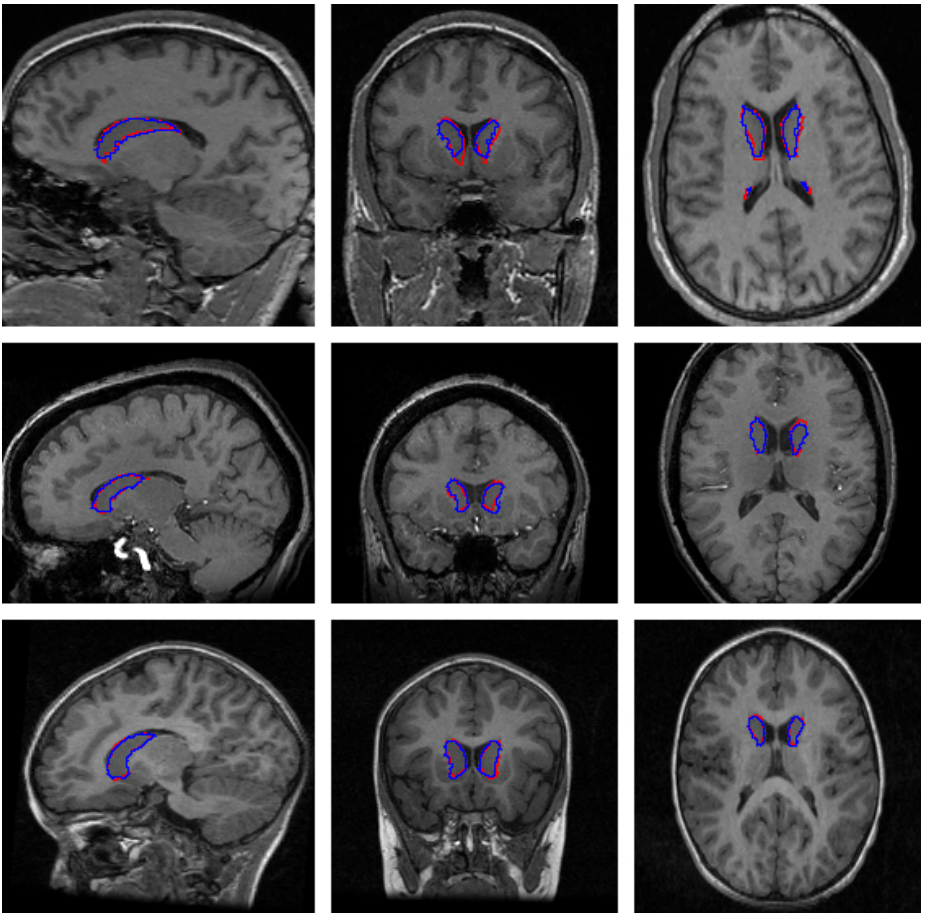

Fig. 1. From left to right, a sagittal, coronal and transversal slice from a a subject in the adults BWH group (top), one in the elderly UNC group (middle) and one in the pediatric UNC group (bottom). The outline of the reference standard segmentation is in red, the outline of the segmentation of the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

cantly greater than those of the other groups. We cannot provide an explanation for this, but it would be worth investigating.

Secondly, the distance measures for the PNL BWH group are signifcantly lower than that of the other groups (this is not only because of the failed segmentation). However, this trend is consistent across all the distance measures and we attribute it to the manner in which the tails of this group were segmented. We noticed from the training data that the tails were significantly elongated, and we believe our model failed to capture this feature.

The method we have presented offers the advantages that it is a general method which can be applied to a wide variety of segmentation problems in the medical imaging field and elsewhere. The compactness of the AAM representation allows the inclusion of a large number of subjects - a limitation that some

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	26.5	84	4.6	92	0.6	80	1.1	81	9.1	73	82
UNC Ped 14	21.0	87	-0.7	98	0.4	86	0.7	87	5.1	86	88
UNC Ped 15	20.5	87	7.0	88	0.4	86	0.7	88	3.8	89	88
UNC Ped 19	24.1	85	-3.7	92	0.5	82	0.8	86	4.0	88	87
UNC Ped 30	18.3	88	1.5	96	0.3	88	0.6	88	3.7	90	90
UNC Eld 01	28.2	82	27.4	52	0.5	82	0.9	84	5.7	83	77
UNC Eld 12	28.6	82	17.3	70	0.5	82	0.9	84	5.6	84	80
UNC Eld 13	22.0	86	9.2	84	0.4	86	0.7	88	3.2	90	87
UNC Eld 20	23.5	85	24.8	56	0.4	85	0.7	88	3.9	88	80
UNC Eld 26	27.6	83	28.5	50	0.4	84	0.8	86	3.9	88	78
BWH PNL 16	31.3	80	-12.7	78	0.8	70	2.4	56	23.5	31	63
BWH PNL 17	25.6	84	-12.7	78	0.8	72	2.8	50	28.9	15	60
BWH PNL 18	28.6	82	-20.2	64	0.7	76	1.5	74	12.6	63	72
BWH PNL 19	30.3	81	-6.5	88	0.8	72	2.2	60	24.0	30	66
BWH PNL 20	29.0	82	-2.5	96	0.7	74	2.3	58	27.0	20	66
BWH PNL 21	31.4	80	-14.8	74	0.9	66	2.6	54	25.9	24	60
BWH PNL 22	32.2	80	-21.0	64	0.9	66	2.6	54	24.0	30	58
BWH PNL 23	25.4	84	1.7	97	0.5	82	1.0	82	9.7	72	84
BWH PNL 24	22.9	86	-1.8	97	0.4	85	0.9	84	7.1	80	86
BWH PNL 25	28.7	82	-4.4	92	0.8	70	2.4	58	23.2	32	67
BWH PNL 26	100.0	0	280.5	0	61.5	0	62.4	0	85.7	0	0
BWH PNL 27	29.3	82	-13.2	77	1.0	62	3.0	46	25.2	26	59
BWH PNL 28	28.4	82	-10.0	82	0.9	66	2.9	48	25.0	26	61
BWH PNL 29	36.1	78	-1.7	86	0.7	73	1.4	76	9.0	74	77
Average All	30.0	80	11.5	77	3.2	74	4.1	69	16.6	58	71
Average UNC Ped	22.1	86	1.7	93	0.4	84	0.8	86	5.1	85	87
Average UNC Eld	26.0	84	21.4	62	0.4	84	0.8	86	4.5	87	80
Average BWH PNL	34.2	76	11.5	77	5.1	67	6.5	57	25.1	37	63

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.

other 3D methods struggle with. The total time needed to process one image is under 6 minutes on a 3Ghz pentium 4 machine with 1GB RAM. This time can be reduced significantly by eliminating the registration step which takes about 3.5 minutes, or simply using a higher spec system.

Our method is robust as demonstrated by the fact that the only unsuccessful case was associated with a failure in the registration stage. This robustness extends to images with different levels of noise and acquisitions from different scanners as demonstrated by the repeatability experiments (Table 2). The overall value obtained for repeatability was 3.5%, comparable with that of manual segmentations (3.1%). The value for the left was lower (1.9% vs 3.1%), and that for the right was higher (5.9%) mainly due to one case of under-segmentation (UNC 17 right). However, the reliance on registration for model initialisation is not desirable and we are investigating other approaches to automatic initialisation. Two possibilities are using multiple initialisations and selecting that which

	Overlap Err		Volume Diff		Abs. Dist.		RMS Dist.		Max. Dist.		Total Score
Average All	26.9	83	-0.2	80	0.6	77	1.6	72	13.6	60	75
Average BWH PNL	29.2	82	-9.2	83	0.8	72	2.2	62	20.4	40	68

Table 4. Averages of Table 3 excluding results for the failed subject.

gives the lowest residual, or alternatively, building a model of the brain surface which could be used to determine the pose of the brain in the image to be segmented.

It is difficult to comment on the accuracy of the method. It achieved 73% overlap on average when the failed subject is excluded. An analysis of the supplied annotated data showed that 65-70% of voxels labelled as caudate are within 1 voxel of the boundary. Thus small errors in locating the boundary can lead to large changes in the overlap measure. The use of the boundary reclassification was an attempt to address this, and we found it increased the overlap measure by between 5% and 10% in the training data.

In conclusion, we have constructed AAMs of the caudate and other subcortical structures in the brain using groupwise registration to establish correspondence. We have demonstrated the application of the AAMs to the problem of automatically segmenting the caudate nuclei and our results show that the method is about as repeatable as that of a human expert, and that the segmentations produced are in good agreement with the gold standard.

Acknowledgements
We would like to thank David Kennedy of the CMA Boston for images, and members of the Integrated Brain Image Modelling (IBIM) consortium for their contributions to helping us devise our segmentation process. Also thanks to Dave Love and Brian Blower of our IT department for keeping our HPC clusters up and running.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D., Graham, J.: ASMs - their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(6) (2001) 681–685
3. Klemencic, J., Pluim, J., Viergever, M., Schnack, H., Valencic, V.: Non-rigid registration based active appearance models for 3D medical image segmentation. Journal of Imaging Science and Technology **48**(2) (2004) 166–171
4. Cootes, T., Twining, C., V.Petrović, R.Schestowitz, Taylor, C.: Groupwise construction of appearance models using piece-wise affine deformations. In: 16th British Machine Vision Conference. Volume 2. (2005) 879–888
5. Cootes, T., Taylor, C.: An algorithm for tuning an active appearance model to new data. In: Proc. British Machine Vision Conference. Volume 3. (2006) 919–928
6. Smith *et al.*: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage **23(S1)** (2004) 208–219
7. Gerig, G., Jomier, M., Chakos, M.: Valmet: A new validation tool for assessing and improving 3D object segmentation. In: Proceedings of the 4th MICCAI LCNS Vol. 2208. (2001) 516–523