

Improved Reproducibility in Dopamine D₂-Receptor Studies with Automatic Segmentation of Striatum from PET Images

Jussi Tohka, Esa Wallius, Jussi Hirvonen, Jarmo Hietala, Ulla Ruotsalainen

Abstract—Manual region-of-interest (ROI) definition is the conventional basis for calculating regional receptor binding potential values in neuroreceptor PET studies. In this study, we propose a fully automatic method to extract the left and right caudate and putamen from ¹¹C-raclopride PET studies. The automatic extraction of these brain structures is challenging due to a small size of the structures compared to the imaging resolution and image noise. The automation of the ROI extraction could remedy the problems relating to intra- and inter-observer variability introduced by the manual ROI drawing. We compared our method to the manual ROI drawing. The automatic ROI extraction of increased test-retest reproducibility of the calculated BP values compared to the manual ROI drawing. In addition, the method was tested on a Monte Carlo simulated PET image. In this experiment, the automatic method yielded more accurate BP values than those computed based on the anatomical ground-truth. Hence, our method for the automatic ROI extraction is reliable and it can simplify and speed up the analysis of large sets of PET scans e.g. in drug development studies.

I. INTRODUCTION

Positron emission tomography (PET) imaging enables the measurement of regional receptor occupancy in the brain. This methodology has been proven to be useful in guiding the dose-finding procedures of early drug development. Therefore, the extraction of specific regions of interest (ROIs) from parametric binding potential (BP) images is important for drug development. The datasets needed for drug development are usually large. This makes automatic extraction of ROIs necessary as traditional, manual MRI-based extraction of ROIs [1] is time-consuming and introduces intra- and inter-observer variabilities to the results. Due to these, a reliable automatic method can increase significantly reproducibility of ROI extraction and consequently improve accuracy of regional statistical analysis of PET neuroreceptor studies.

In the present study, we introduce a new completely automatic method for the segmentation of striatum into left and right caudate and putamen from BP images with D₂-receptor

This work has been supported by Academy of Finland (grant no. 104834), TEKES Drug 2000 Technology Program, and Tampere Graduate School of Information Sciences and Engineering (TISE).

J. Tohka, E. Wallius, and U. Ruotsalainen are with the Institute of Signal Processing, Tampere University of Technology, Tampere, Finland. Email: {jussi.tohka, esa.wallius, ulla.ruotsalainen}@utu.fi. J. Hirvonen is with the Turku PET Centre, Turku University Central Hospital, Finland. Email: jussi.hirvonen@utu.fi. J. Hietala is with the Department of Psychiatry, University of Turku, Finland. Email: jarmo.hietala@utu.fi

ligand ¹¹C-raclopride. In addition to image noise, there are certain specific challenges in the segmentation of striatum. The left and right striatum must be identified as separate brain structures. They typically feature highly similar BP values in the images rendering e.g. histogram based segmentation approaches infeasible. The division of striatum into the caudate and the putamen is not straightforward because these structures are (almost) connected in the brain. We combine several modern techniques for robust image segmentation in a novel way to overcome these problems.

II. METHODS

Our method for the segmentation of striatum proceeds in a top-down manner. The steps of the method are depicted in Fig. 1. In the first step, a PET brain image is segmented into the left and right hemispheres allowing the segmentation of striatum into the left and right components. The image representing the left (right) hemisphere is here referred as the left (right) *hemisphere image*, cf. Fig. 1 b) and c). The hemisphere segmentation is performed using a reliable automatic method for the mid-sagittal plane determination that is based on the extraction of the brain surface from a PET image [2]. In the second step, the left and right striatum are extracted from the corresponding hemisphere images using a deformable surface model [3] as described in Section II.A. In the final step, described in Section II.B, the extracted striatum (left or right) is divided into two volumes corresponding to caudate and putamen. For this, we apply a modification of Normalized Cuts (Ncuts) algorithm [4].

A. Extraction of striatum

The striatum is extracted from a hemisphere image applying a global-optimization based deformable surface mesh [3]. The basic idea behind deformable surfaces is to convert the surface extraction problem into an energy minimization problem. The energy of the deformable surface, whose minimum argument is the solution to the surface extraction problem, is a weighted sum of two energies. The external energy couples the surface with the salient image features and the internal energy is used to constrain the shape of the surface.

Simplex meshes are used for surface representation [5]. We denote by $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ the simplex mesh, where $\mathbf{w}_i \in \mathbb{R}^3$ are coordinates of its vertices. In a simplex mesh each vertex

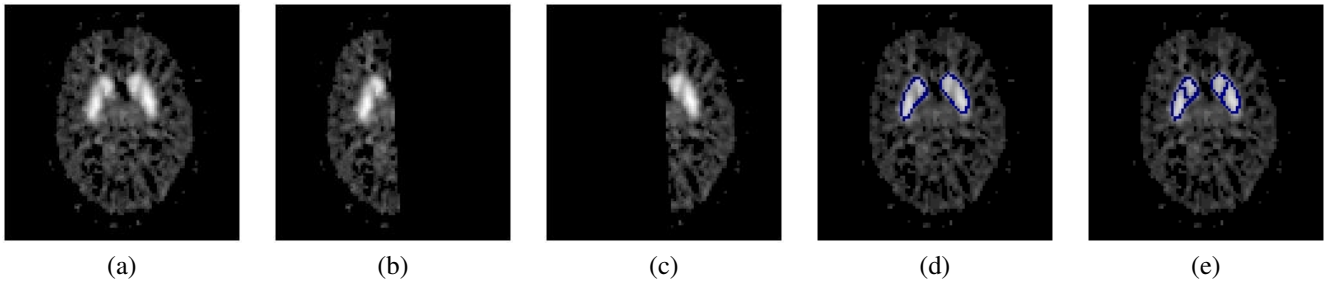


Fig. 1. The steps of the segmentation method. (a) The central transaxial cross-section of the original image. (b) and (c) The cross-sections of the images containing left (b) and right (c) striatum. These were defined by automatic mid-sagittal plane determination from the original image. (d) The surface cross-sections (in blue colour) of the left and right striatum extracted from the images depicted in (b) and (c) overlaid on the original image cross-section. (e) The left and right striatum in (d) divided into caudate and putamen by the N-cuts algorithm.

has exactly three neighbors, whose coordinates are denoted by \mathbf{w}_{i_j} . The energy of the mesh is

$$E(\mathbf{W}) = \lambda \sum_{i=1}^N [E_{int}(\mathbf{w}_i | \mathbf{w}_{i_1}, \mathbf{w}_{i_2}, \mathbf{w}_{i_3}) + (1 - \lambda) E_{ext}(\mathbf{w}_i)], \quad (1)$$

where $\lambda \in [0, 1]$ is the regularization parameter and E_{int} denotes the internal energy and E_{ext} denotes the external energy. The internal energy is defined as

$$E_{int}(\mathbf{w}_i | \mathbf{w}_{i_1}, \mathbf{w}_{i_2}, \mathbf{w}_{i_3}) = \frac{\|\mathbf{w}_i - \frac{1}{3} \sum_{j=1}^3 \mathbf{w}_{i_j}\|^2}{A(\mathbf{W})}, \quad (2)$$

where $A(\mathbf{W})$ is the area of the mesh. The normalization by the area is required for scale invariance of the internal energy. For external energy, we first construct an intensity limited image I_T from the original image I : If $I(\mathbf{x}) > T$ then $I_T(\mathbf{x}) = T$ and $I_T(\mathbf{x}) = I(\mathbf{x})$ otherwise. The threshold T is automatically determined for each image based on the expected volume of striatum¹. The external energy

$$E_{ext}(\mathbf{w}_i) = 1 - \frac{(I_T(\mathbf{w}_i)) \|\nabla I_T(\mathbf{w}_i)\|}{\max_{\mathbf{x}} I_T(\mathbf{x}) \|\nabla I_T(\mathbf{x})\|}, \quad (3)$$

where the edge image (∇I_T) is computed using the 3-D Sobel operator [7].

The energy function (1) is minimized with the global DSM-OS algorithm (dual surface minimization-outer surface) [3]. The initialization for the algorithm is generated automatically in essentially the same way as described in [2] for the brain surface extraction. In all our experiments, we set $\lambda = 0.3$.

B. Segmentation of striatum into caudate and putamen

Once the surface of striatum has been extracted, one can compute which voxels lie inside of the striatum. For these

¹If Vol is the volume of a single voxel in expressed in cm^3 , we set T to be the $\lfloor 10cm^3/Vol \rfloor$ th largest intensity value in the hemisphere image. The volume of the left or right striatum is approximately $5cm^3$ [6] and the constant $10cm^3$ is selected to take partial volume effect into account. The intensity limiting is required because the partial volume effect, noise, and other imaging artifacts cause intensities within the striatum to be highly inhomogeneous. The intensity limiting also prevents exceptionally high (outlying) intensity values within striatum from influencing the surface extraction result. In section IV we show that the way in which the value T is selected does not essentially influence our results.

voxels a weighted affinity graph $G = (V, E)$ is constructed. The nodes $\mathbf{v} \in V$ of the graph G represent voxels within the striatum. The arcs² $\mathbf{uv} \in E$ of the graph G connect two nodes if the corresponding voxels are 26-neighbours of each other. Note that the graph is connected because the deformable surface based segmentation results in connected volumes. A weight $w(\mathbf{u}, \mathbf{v}) > 0$ describing affinity of two neighboring voxels is assigned to every arc. A high weight means that the two voxels connected by the arc are likely to belong to the same structure. Our purpose is to divide set of nodes of G into two connected components corresponding to putamen and caudate. The graph G is divided into two parts, $P, C \subset V$ by minimizing the normalized cut criterion [4]

$$Ncut(P, C) = \frac{cut(P, C)}{assoc(P, V)} + \frac{cut(P, C)}{assoc(C, V)}, \quad (4)$$

where $cut(P, C) = \sum_{\mathbf{u} \in P, \mathbf{v} \in C} w(\mathbf{u}, \mathbf{v})$, $assoc(P, V) = \sum_{\mathbf{u} \in P, \mathbf{t} \in V} w(\mathbf{u}, \mathbf{t})$. The $Ncut$ criterion essentially minimizes the affinity between the sets P and C , or equivalently, maximizes the affinity within the sets.

Let $\mathbf{u}, \mathbf{v} \in V$ be two neighbouring striatal nodes and let $I(\mathbf{u}), I(\mathbf{v})$ be the corresponding image intensity values linearly normalized to span the range from 0 to 1. Then, the weight of the arc \mathbf{uv} is

$$w(\mathbf{u}, \mathbf{v}) = 2\sqrt{3} - \|\mathbf{u} - \mathbf{v}\| [2 - I(\mathbf{u}) - I(\mathbf{v})], \quad (5)$$

where $\|\mathbf{u} - \mathbf{v}\|$ is the distance between the two voxel coordinates. This selection of weights differs from edge-based criteria applied commonly for the image segmentation with Ncuts. In our case, both the caudate and putamen contain large intensity variations within the structures, which yield edge based criteria troublesome. Instead, we utilize the fact that striatum contains a thin 'plate' between the two structures, which features considerably lower intensity values than the structures of interest. The weights in Eq. (5) essentially state that the cut is likely to be in the area of low intensity values.

The exact minimization of the $Ncut$ criterion is a NP-complete problem. However, it is possible to efficiently approximate the correct solution by solving a certain generalized

²These are perhaps more commonly termed edges of the graph but to avoid confusion with image edges we have chosen to use an alternative term.

eigenvalue system (for details see [4]). This approximation provides us a qualitative membership value ($\in \mathbb{R}$) for each voxel. Thresholding qualitative membership values then gives the partitioning into caudate and putamen. Currently, the threshold is found based on brute-force: We compute the value of (4) for all admissible thresholds and select the one that minimizes (4). The approach is efficient since the number of voxels in the striatum is limited. In addition, the approach gives us the possibility to embed constraints in the partitioning. We try only those threshold values which respect the constraints. Particularly, we have found it useful to constrain the volumes of the Putamen and Caudate, so it does not happen that one of them would consist just, say, one voxel.

Still, we have to label the two parts of the striatum with the correct names. There is no way to predict which of the sets produced by the Ncuts-algorithm corresponds to putamen and which corresponds to caudate. To solve this problem and label the structures automatically, we utilize neuroanatomical knowledge; Putamen is nearer to the back of the head than caudate, and hence the anterior (i.e. y)-coordinate of its center of gravity should have lower value of the two structures.

III. EXPERIMENTS

The method was tested with a dynamic Monte Carlo simulated phantom image [8] built upon a high resolution anatomical phantom image [9]. The details of the generation of the dynamic image can be found in [10]. Based on the dynamic image, the parametric binding potential (BP) image was computed using the cerebellum as the reference region [11]. Based on the time-activity curves applied for the generation of the phantom, we computed the reference Binding Potential (BP) values for putamen and caudate. These were then compared to the values found by automatically segmenting the parametric BP image. For reference, the BP values based on the anatomical ground-truth were also computed.

The automatic method was compared to manual ROI drawing with eight pairs of raclopride images of healthy males acquired for test-retest purposes [1]. Each test-retest subject underwent two raclopride PET scans during the same day, at least 2 hours apart. The ROIs corresponding left and right caudate and left and right putamen were manually defined using an MR aided technique. These were used to measure the BP values for each of the structures.

We measured the reliability of the regional BP values using the intraclass correlation coefficient (ICC) ranging from -1 to 1 . ICC measures the relative homogeneity within groups (e.g. test-retest variation of BP values within subjects) in ratio to the total variation and it is derived using the one-way random effects analysis of variance model (ICC(1,1) in [12]). ICC of at least 0.7 is commonly interpreted as sufficient reliability [13] and then the mean-squared within-subject variation is much smaller than the mean-squared between-subjects variation. Poor reliability (ICC near -1) takes place when within-subject variation composes most of the total variation.

The reproducibility of regional BP values was measured by computing the normalized absolute differences (NADs) defined as

$$|BP_{scan1} - BP_{scan2}|/BP_{scan1}, \quad (6)$$

where BP_{scan1} (resp. BP_{scan2}) is the BP value computed from the test-image (retest-image). The NAD value is composed only of within-subject variation and it can be regarded as a measure of absolute signal change. The test-retest variation (i.e. NAD) of BP measurements with PET is typically $5 - 10\%$ and this is usually used as a confidence limit when measuring receptor occupancy (see [14]). We list the maximum and mean NAD (in percents) over the eight test-retest pairs of images. The results of the manual method differ from those presented in [1], since we omit PET-to-PET image coregistration between test-retest scans to ensure a fair comparison.

IV. RESULTS

A. Monte Carlo phantom

The true BP values for the Monte Carlo phantom were 2.67 for caudate and 2.77 for putamen (Left and right structures were simulated using the same time activity curves). By our automatic method we obtained values of 1.82 (left), 1.92 (right), and 1.87 (average) for caudate and 2.10 (left), 2.17 (right), and 2.13 (average) for putamen. These values seem low, but are higher than the values computed using the anatomical ground-truth (1.65 for caudate average and 1.84 for putamen average). Hence, low BP values were caused by artifacts in the PET imaging (e.g. partial volume effect), not unreliable and inaccurate segmentation. The segmentation result for the Monte Carlo simulated phantom is shown in Fig. 2.

With the phantom image, we studied how our intensity limiting procedure that was explained in the footnote in Sect II.A affects the results of our segmentation procedure. Particularly, we studied how changing the expected volume of striatum (EVS) parameter required for the intensity limiting affects the segmentation. The findings are plotted in Fig. 3. It can be seen from Fig. 3 that the change in EVS did not affect the BP values found for caudate. With putamen, the trend was that the BP values decreased when EVS was increased. The hemisphere images of the Monte Carlo phantom may be regarded as a test-retest pairs of images and the BP values changed similarly for both hemispheres. This is an indication that the selection of EVS has no essential influence to the results of the statistical analysis when subjects or subject groups are compared.

B. Test-retest analysis with human studies

The ICCs in eight pairs of test-retest scans showed that the reliability with the automatic method was slightly higher both in putamen average and in caudate average compared with manual ROI extraction (Table 1). The automatic method had higher ICCs in left putamen, right caudate and left caudate, but ICCs were similar in right putamen for the compared methods (Table 1). The ICCs seemed sufficiently reliable with both the automatic and manual segmentation method.

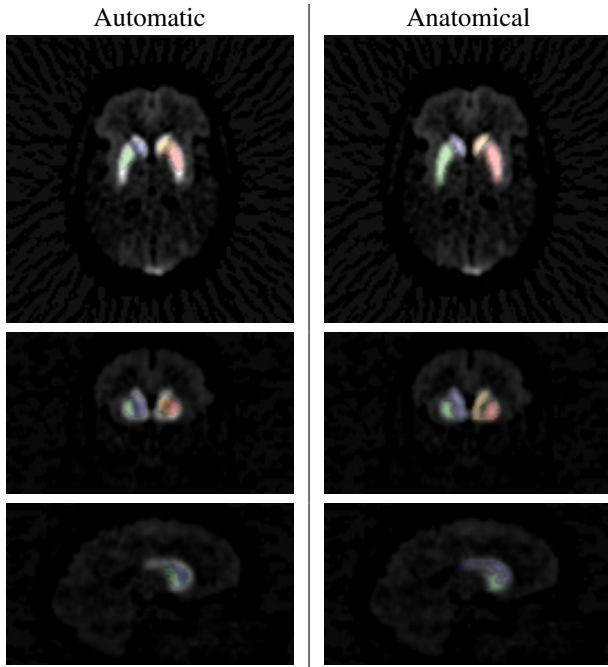


Fig. 2. Left: Example cross-sections of the automatic segmentation result of the Monte-Carlo phantom overlaid on the parametric BP image. Right: The anatomical ground-truth used in the generation of the Monte Carlo phantom compared to the parametric image. The influence of the partial volume effect shows clearly. From top axial, coronal, and sagittal image cross-sections are shown. In the transparent segmentation mask images, blue and yellow represent, respectively, the left and right caudate and green and red represent, respectively, left and right putamen.

The percentual NADs measure reproducibility and depict the absolute effect of test-retest variation to measured receptor occupancy. Mean NADs and max NADs were smaller with the automatic method than with the manual method in all striatal regions (Table 1). The superiority of automatic method was especially clear (mean NAD less than half of the manual method) with putamen average, left caudate and caudate average (Table 1). Concerning max NADs, clear superiority of the automatic method was seen in left caudate and right caudate. Compared with the typical test-retest variation (5 - 10 %), the mean NADs with the automatic segmentation method were in all regions less than 5 % whereas with the manual segmentation NADs exceeded 5 % in most regions (Table 1). The max NADs with the automatic segmentation were all less than 10 %, but with the manual segmentation they exceeded 10 % in some regions (Table 1). In addition, the mean BP values were fairly similar with the manual and automatic segmentation suggesting that the automatically found BP values were also close to the correct BP values. Comparison of the BP value differences between manually and automatically extracted ROIs for each image in the case of left caudate is shown in Fig. 4. In this case, the automatic segmentation yielded typically a bit higher BP values compared to the manual ROI drawing, and the differences were in most cases similar for the two test-retest pairs. An example of the segmentation result with one of the rest-retest image pairs is shown in Fig. 5.

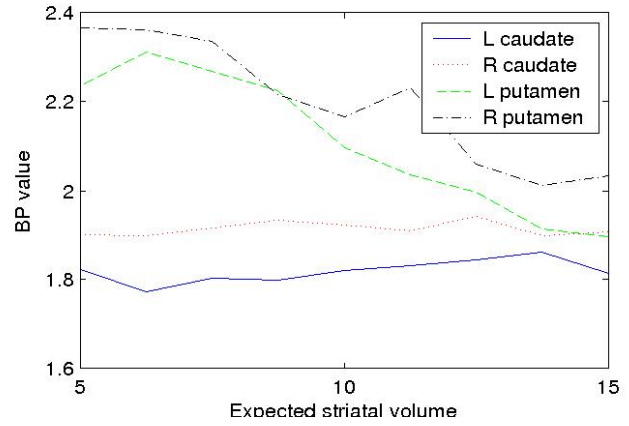


Fig. 3. The influence of changing the value of the expected volume of striatum to the BP values with the automatic segmentation of the Monte Carlo phantom.

		mean BP	ICC	mean NAD	max NAD
putamen					
left	manual	3.37	0.84	6.31%	12.0%
left	autom	3.38	0.97	4.21%	8.32%
right	manual	3.28	0.90	5.27%	8.63%
right	autom	3.39	0.89	3.81%	8.39%
average	manual	3.32	0.87	5.84%	9.43%
average	autom	3.38	0.93	2.71%	5.54%
caudate					
left	manual	2.71	0.91	5.46%	12.8%
left	autom	2.78	0.98	2.24%	5.16%
right	manual	2.67	0.86	6.44%	20.7%
right	autom	2.79	0.93	3.39%	7.93%
average	manual	2.69	0.94	4.31%	8.24%
average	autom	2.79	0.98	1.68%	5.29%

TABLE I
THE TEST-RETEST RELIABILITY (ICC) AND REPRODUCIBILITY (NAD) MEASURES OBTAINED BY MANUAL ROI DRAWING AND THE AUTOMATIC METHOD INTRODUCED IN THIS STUDY. NAD IS DEFINED AS $|BP_{scan1} - BP_{scan2}|/BP_{scan1}$. THE ROWS LABELED AVERAGE PROVIDE RESULTS WHEN BOTH LEFT AND RIGHT HEMISPHERES ARE TAKEN INTO ACCOUNT.

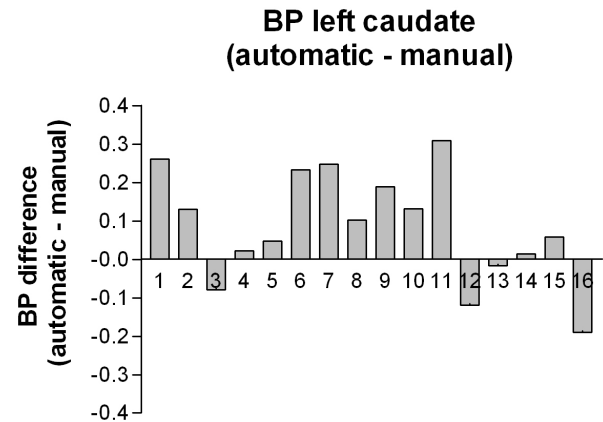


Fig. 4. Difference in the BP values for the left caudate between automatic and manual segmentation. The numbers used in labeling the bars refer to the image numbers in such way that consecutive images form a test-retest pair.

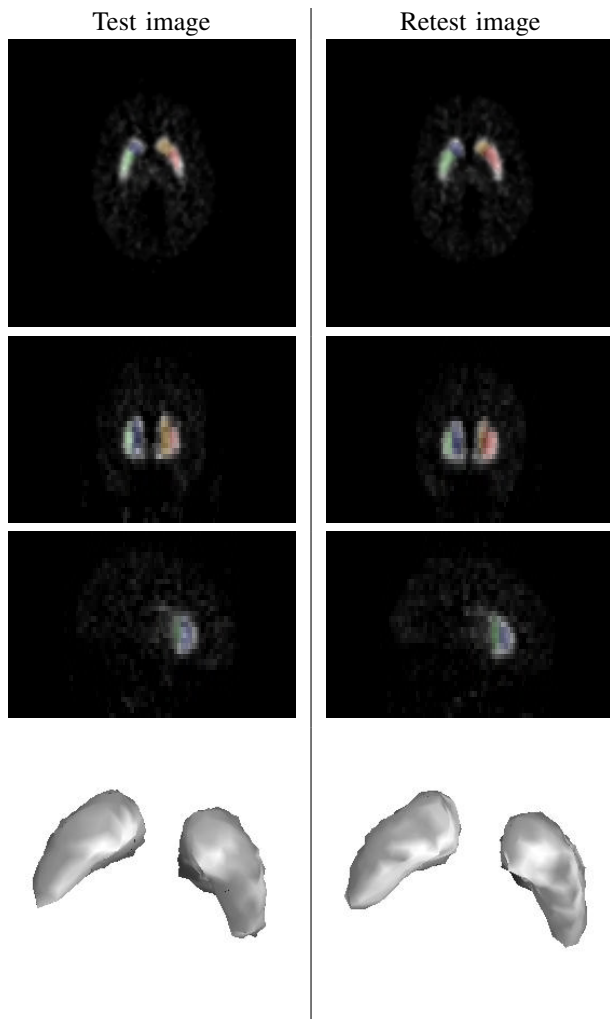


Fig. 5. An example of the automatic segmentation with one of the test-retest image pairs. From top, axial, coronal and sagittal cross-sections of the mask images are overlaid on the corresponding cross-sections of the BP images. The colours are as in Fig. 2. In the bottom row, 3-D renderings of the automatically extracted striatum surfaces are shown. The image pair for this example was chosen so that it represents an average result in the terms of NADs. We note that since the test-retest image pairs were not co-registered the image cross-sections shown above do not necessarily correspond to each other maximally well.

V. CONCLUSION

We have introduced a new method for the segmentation of the striatum from raclopride BP images. The striatum is among most interesting brain structures when studying dopamine D₂-receptor binding and hence fully automatic method for extraction of the putamen and caudate benefits the brain research considerably. The automatic method for the segmentation of striatum from raclopride BP images turned out to be superior in terms of the reproducibility of results when compared to traditional, manual ROI extraction. Reliability of automatic segmentation was at the least similar and in some subregions of striatum a bit higher than with manual ROI extraction. In our experiment with the Monte Carlo simulated phantom image, the BP values computed based on the automatic segmentation

were closer to the correct values than those computed based on the anatomical ground-truth.

Similar methods for automatic ROI extraction could be developed for other inner brain structures and radioligands. The introduced and alike methods are especially useful for the statistical analysis of preclinical and clinical trials for drug development.

REFERENCES

- [1] J. Hirvonen, S. Aalto, V. Lumme, K. Någren, J. Kajander, H. Vilkmann, N. Hagelberg, V. Oikonen, and J. Hietala, "Measurement of striatal and thalamic dopamine d-2 receptor binding with c-11-raclopride," *Nucl. Med. Comm.*, vol. 24, pp. 1207 – 1214, 2003.
- [2] J. Mykkänen, J. Tohka, J. Luoma, and U. Ruotsalainen, "Automatic extraction of brain surface and mid-sagittal plane from PET images applying deformable models," Department of Computer and Information Sciences, University of Tampere, Finland, Tech. Rep. A-2003-1, June 2003, available at <http://www.cs.uta.fi/reports/r2003.html>.
- [3] J. Tohka and J. Mykkänen, "Deformable mesh for automated surface extraction from noisy images," *International Journal on Image and Graphics*, vol. 4, no. 3, pp. 405– 432, 2004.
- [4] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans Patt Anal Mach Intell*, vol. 22, no. 8, pp. 888 – 905, 2000.
- [5] H. Delingette, "General object reconstruction based on simplex meshes," *International Journal of Computer Vision*, vol. 32, no. 2, pp. 111–142, 1999.
- [6] G. Firnau, E. Garnett, R. Chirakal, R. Sood, S. Nahmias, and G. Schrobilgen, "[¹⁸F]fluoro-L-dopa for the in vivo study of intracerebral dopamine," *Applied Radiation Isotopes*, vol. 37, no. 8, pp. 669 – 675, 1986.
- [7] S. Zucker and R. Hummel, "A three-dimensional edge operator," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, no. 3, pp. 324 – 331, May 1981.
- [8] A. Reilhac, C. Lartizien, N. Costes, S. Sans, C. Comtat, R. Gunn, and A. Evans, "PET-SORTEO: A Monte Carlo-based simulator with high count rate capabilities," *IEEE Transactions on Nuclear Science*, vol. 51, no. 1, pp. 46 – 52, 2004.
- [9] I. Zubal, C. Harrell, E. Smith, Z. Rattner, G. Gindi, and P. Hoffer, "Computerized three-dimensional segmented human anatomy," *Med. Phys.*, vol. 21, no. 2, pp. 299 – 302, 1994.
- [10] J. Tohka, A. Kivimäki, A. Reilhac, J. Mykkänen, and U. Ruotsalainen, "Assessment of brain surface extraction from PET images using Monte Carlo simulator," *IEEE Trans Nucl Sci*, vol. 51, no. 5, October 2004.
- [11] A. Lammertsma and S. Hume, "Simplified reference tissue model for PET receptor studies," *Neuroimage*, vol. 4, pp. 153 – 158, 1996.
- [12] P. Shrout and J. Fleiss, "Intraclass correlation: Uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420 – 428, 1979.
- [13] G. Hripesak and D. Heitjan, "Measuring agreement in medical informatics reliability studies," *Journal of Biomedical Informatics*, vol. 35, no. 99 – 110, 2002.
- [14] J. Passchier, A. Gee, A. Willemsen, W. Vaalburg, and A. van Waarde, "Measuring drug-related receptor occupancy with positron emission tomography," *Methods*, vol. 27, pp. 278 – 286, 2002.