Validation of appearance-model based segmentation with patch-based refinement on medial temporal lobe structures

Shiyan Hu, Pierrick Coupé, Jens C. Pruessner, Louis D. Collins

Validation of appearance-model based segmentation with patch-based refinement on medial temporal lobe structures

Shiyan Hu, Pierrick Coupé, Jens C. Pruessner, and D. Louis Collins

McConnell Brain Imaging Centre, Montreal Neurological Institute,
McGill University, Montréal, Québec, Canada H3A 2B4

Abstract. This paper presents a new automatic segmentation scheme that combines active appearance (AAM) modeling and patch-based label fusion into a segmentation framework. AAM, which uses eigen-decomposition to analyze the statistical variation of image intensity and shape information over the population, is used to capture the global shape characteristics of the structure of interest with a generative model, while patch-based label fusion, which uses a non-local means method to compare the image local intensity properties, is applied to locally refine the segmentation results along the structure boundary area to improve the segmentation accuracy. The proposed segmentation scheme is used to segment human medial temporal lobe structures, which have low intensity contrast in MRI and complexity in shape. The experiments demonstrate that this new segmentation scheme is computationally efficient and robust. In a leave-one out validation with fifty-four normal young subjects, the method yields a mean Dice κ of 0.87 for hippocampus, 0.81 for amygdala, 0.73 for para-hippocampal complex cortex, and 0.73 for perirhinal cortex between manual and automatic labels.

1 Introduction

The medial temporal lobe, located in the inner side of each temporal lobe, is the most important part of the limbic system, and includes the hippocampus (HC), the amygdala (AG), and surrounding cortical areas (entorhinal (EC), perirhinal (PRC), and parahippocampal (PHC) cortices). These structures play important roles in learning, memory, and emotion. HC is the the most critical component of the medial temporal lobe memory system, and plays an essential role in memory and spatial navigation. AG is strongly involved in emotional and social processing, particularly fear and worry. EC is the main interface between HC and neocortex, and plays an important role in the formation and optimization of spatial memories. PRC is involved in both visual perception and memory, and PHC is involved in scene recognition and social context [1] [2]. Recently, medial temporal lobe structures have received considerable attention due to their importance in neurological diseases and disorders [3]. For example, many researchers [4] [5] showed that HC volume changes were an important marker of the early stage of Alzheimer's disease and temporal lobe epilepsy. Because of the importance of these structures in the neurodegeneration, there is significant interest in developing accurate, robust, and reliable segmentation techniques to

automatically extract those structures from magnetic resonance (MR) imaging for volume and shape analyses.

Manual segmentation is considered highly accurate, and treated as the gold standard so far. However, manual segmentation is time consuming, requires anatomical expertise, and may have important intra and inter-rater variability. To overcome the disadvantages of manual segmentation, many automatic segmentation techniques have been proposed. Most model-based segmentation techniques can be grouped into the following three categories: deformable models [6], appearance-based models [7] [8], and atlas-based techniques [9] [10].

To avoid the bias in the segmentation by using a single template, Heckemann et al. [11], Shattuck et al. [12] and Aljabar et al. [13] proposed multi-atlas based methods to improve the segmentation efficiency. Collins and Pruessner [14] incorporated the label fusion into the multi-atlas warping, and Wang et al. in [15] used the multi-atlas technique with error correction to yield the best published results for HC segmentation. Coupé et al. [16] used non-local means patch-based approach to weight the expert manual segmentation in a library of templates (see Table 1 for the segmentation method review). However, neither atlas-based or patch-based methods explicitly incorporate the structure's shape information into the segmentation.

Table 1. Segmentation method review

Author	Method summary	Result in term of κ		
		HC	AG	PHC
Fischl 2002 [9]	FreeSurfer	0.8	0.75∼0.78	
Heckemann 2006 [11]	Multi-atlas based	0.82	0.8	0.81*
Chupin 2007 [6]	Seeding+region growing	0.84	0.76∼0.8	
Powell 2008 [17]	Machine learning	0.85		
Lijn 2008 [18]	Multi-atlas+graph cuts	0.86		
Morey 2009 [19]	FST/FIRST	0.79	0.73	
Aljabar 2009 [13]	Multi-atlas+template selection	0.84	0.78	
Lotjonen 2010 [20]	Multi-atlas+intensity modeling	0.82∼0.88	0.77	
Collins 2010 [14]	Multi-atlas+label fusion	0.89	0.83	
Benavides 2010 [21]	FreeSurfer	0.78		
Sabuncu 2010 [22]	Label fusion	0.82∼0.87	0.8∼0.82	
Coupé 2011 [16]	Non-local means	0.88		
Patenaude 2011 [8]	Bayesian appearance	0.81	0.74	
Wang 2011 [15]	Multi-atlas+error correction	0.89∼0.91		
Bishop 2011 [23]	FMASH	0.8∼0.82		
Khan 2011 [24]	Multi-atlas+spatially-local selection	0.83∼0.85		

*parahippocampal + ambient gyri.

To integrate the shape constraint into the segmentation, we combine the appearance model and patch-based technique into a general segmentation frame-

work. The main contribution of this paper includes: 1) the proposed method combines the appearance model and non-local means patch-based methods where appearance models are used to capture the global shape variation, and the non-local means method is used to locally refine the segmentation, 2) the proposed method is applied to segment all medial temporal lobe structures. Compared to HC and AG, the other medial temporal lobe structures, like PRC and PHC, have much greater anatomical variability, and 3) nonlinear registration instead of linear registration is applied to both training and testing data to improve the alignment between subjects.

2 Method

2.1 Appearance-model based Global Segmentation

Appearance-model based global segmentation applies the eigen-decomposition technique on gray MR images and shape data to capture the statistical variations of the gray intensity and shape information of the training data. Based on the eigenvectors derived from the training data, using the notation from [25] the final shape and gray image can be given by,

$$\begin{aligned} \phi &= \bar{\phi} + \mathbf{P}_{\phi}\mathbf{Q}_{s}\bar{w}_{s}^{-1}\,\mathbf{c} \\ \mathbf{g}_{t1} &= \bar{\mathbf{g}}_{t1} + \mathbf{P}_{g,t1}\mathbf{Q}_{g,t1}\,\mathbf{c} \end{aligned} \tag{1}$$

The segmentation is achieved by minimizing the difference between the test image and the one synthesized from Eq. 1. The cost function in the least square measure can be written as:

$$\mathcal{E} = \sum_{j=1}^{N_p} (I_{t1,j} - g_{t1,j})^2 \tag{2}$$

where $I_{t1,j}$ and $g_{t1,j}$ are the intensity of the j-th voxel of the T1 test MR image and the synthesized T1 image respectively. N_p is the total number of voxels in each image. The detailed appearance-based segmentation method was described in [25]. In this study, all data including the test and training volumes, are non-linearly registered into an unbiased non-linear average template [26] to minimize the size, orientation and position differences between subjects.

2.2 Non-local Means Patch-based Segmentation

The non-local means patch-based segmentation uses the local intensity distance between patches to estimated the weighting of manual segmentations in the training data, and the local label fusion is based on the weighted average of manual segmentations. The method described here is the same as in [16]. In particularly, for each voxel x_i in the test image, the weighting of each patch $w(x_i, x_{s,j})$ is calculated by non-local means filter as:

$$w(x_i, x_{s,j}) = e^{\frac{-\|p(x_i)-p(x_{s,j})\|_2^2}{h^2}} \tag{3}$$

where $p(x_i)$ is the cubic patch centered at x_i, $p(x_{s,j})$ is the cubic patch centered at x_j in the $s-th$ training subject and $||.||_2$ is the normalized intensity distance between two patches. The final segmentation is the weighted labels of all labeled samples inside the search volume v_i of N training subjects.

$$v(x_i) = \frac{\sum_{s=1}^{N} \sum_{j \in v_i} w(x_i, x_{s,j}) y_{s,j}}{\sum_{s=1}^{N} \sum_{j \in v_i} w(x_i, x_{s,j})} \tag{4}$$

where $y_{s,j}$ is the manual segmentation at Voxel x_j of the $s-th$ training subject, and $w(x_i, x_{s,j})$ is the weighting of the patch.

2.3 Combine Appearance Modeling and Non-local Means Patch in Segmentation

Appearance-model based global segmentation is good at capturing the global shape variation but not sensitive to the local shape change; while non-local means patch-based segmentation is sensitive to the local geometry, but doesn't have global shape constraints since it only uses the local intensity property in the segmentation. To take advantage of the global shape constraints and local intensity information, we combine these two methods together to improve the segmentation accuracy. For this purpose, we first apply appearance-model based segmentation to the test subject and then use the result from the global segmentation as input of the patch-based segmentation to locally refine the segmentation along the structure boundary area. A summary of the proposed algorithm is as follows.

- Do appearance-model based global segmentation and obtain the segmented distance function ϕ.
- Define local refinement area Ref_{ring} which are the voxels inside the distance range $[d_1, d_2]$ of ϕ.
- For each voxel x_i inside Ref_{ring}, re-calculate the patch-similarity function of $\phi(x_i)$ using non-local means patch-based method described in Sec 2.2. Instead of using the manual labels in Eq.4, the signed Euclidean distance functions of the manual labellings are integrated into the equation, i.e.

$$\phi(x_i) = \frac{\sum_{s=1}^{N} \sum_{j \in v_i} w(x_i, x_{s,j}) \phi_{s,j}}{\sum_{s=1}^{N} \sum_{j \in v_i} w(x_i, x_{s,j})} \tag{5}$$

- The final segmentation is achieved by threshing ϕ.

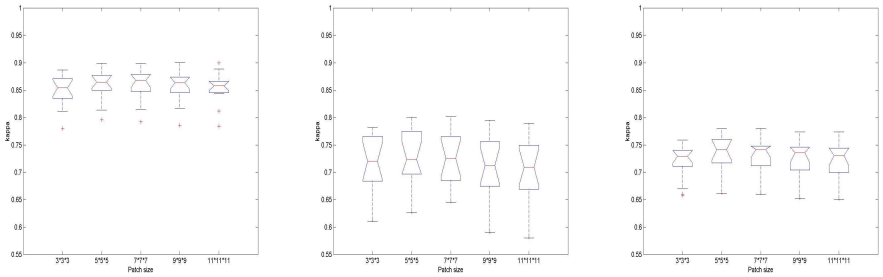
3 Illustrative Experiments and Results

The proposed segmentation algorithm is applied to segment human medial temporal lobe structures from real MR images. The validation uses the subset of the International Consortium for Brain Mapping (ICBM) database. The medial

temporal lobe structures (HC, AG, EC, PRC, PHC) on T1w MRI data for 54 subjects from the ICBM database (Philips 1.5T gyroscan, TE=10ms, TR=30ms, ang=30^o, $1mm^3$ isotropic voxels) were manually segmented using the protocol defined by Pruessner [27] [28]. The automatic segmentation results are compared with the manual labels. The similarity of these two labellings are measured by Dice kappa (κ) [29] [$\kappa = 2 * (V(M \cap A))/(V(M) + V(A))$].

3.1 Effects of Different Patch-size on Segmentation Performance

To study the impact of different patch size on segmentation accuracy, we segmented HC, para-hippocampal complex cortex (PCC=EC+PHC), and PRC using different patch sizes. The kappa results were presented in Fig. 1 which demonstrates that the best median kappa values were obtained with a patch size of $7 \times 7 \times 7$ for all structures. These results indicated that too small patch size might not be able to capture the local geometry, and too big patch size might fail to find the best matched patches in the training data.

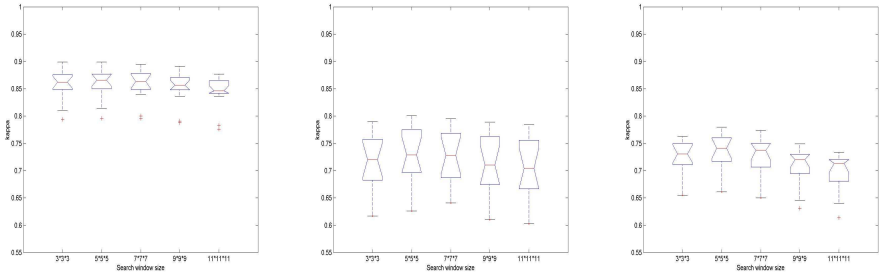


(a) HC (b) PCC (c) PRC

Fig. 1. Impact of patch size on segmentation performance. κ values of 14 test subjects under different patch sizes. (a) Segmentation for HC (b) Segmentation for PCC, and (c) Segmentation for PRC.

3.2 Effects of Different Search-size on Segmentation Performance

The impact of different search window size on segmentation accuracy was also analyzed by HC, PCC, and PRC. The κ values of 14 test subjects were shown in Fig. 2. From Fig. 2, we could see that the best median kappa values were obtained with a search window size of $5 \times 5 \times 5$ or $7 \times 7 \times 7$. Compared to the results in [16], the optimal search window size was a litter bit smaller. However, the data in [16] were aligned by linear registration and the data in this experiment were aligned by non-linear registration. Thus, this result suggested that the better alignment among subjects could reduce the search size.



(a) HC (b) PCC (c) PRC

Fig. 2. Impact of search window size on segmentation performance. κ values of 14 test subjects under different search window sizes. (a) Segmentation for HC (b) Segmentation for PCC, and (c) Segmentation for PRC.

3.3 Validation of Segmentation Accuracy on MTL Structures

The proposed method was applied to segment both left and right of HC, AG, PCC, PRC, and temporopolar cortex (TPC) from 54 MRI volumes using leave-one out method. The segmentation performance in terms of κ values based on the appearance-model based method with and without local correction was shown in Table 2. From these experiments, we can see: 1) for all structures, the individual κ value from the Appearance-based segmentation with local correction was always higher than that without local correction, and 2) the non-local means patch-based local correction increased the mean κ by 1.5% $\sim$ 3.5%, and those improvements were statistically significant ($p \leq 0.001$). The improvement of the patch-based refinement on segmentation accuracy could also be visually observed in Fig. 3 which showed segmentation examples of all structures for three test subjects. The experiments also showed that the proposed method was able to quickly segment a new subject (less than 30 seconds for appearance-based global segmentation, and 1.5 minutes for patch-based refinement) on a 1.5GHz Linux PC once data was aligned (6 minutes per subject for nonlinear registration).

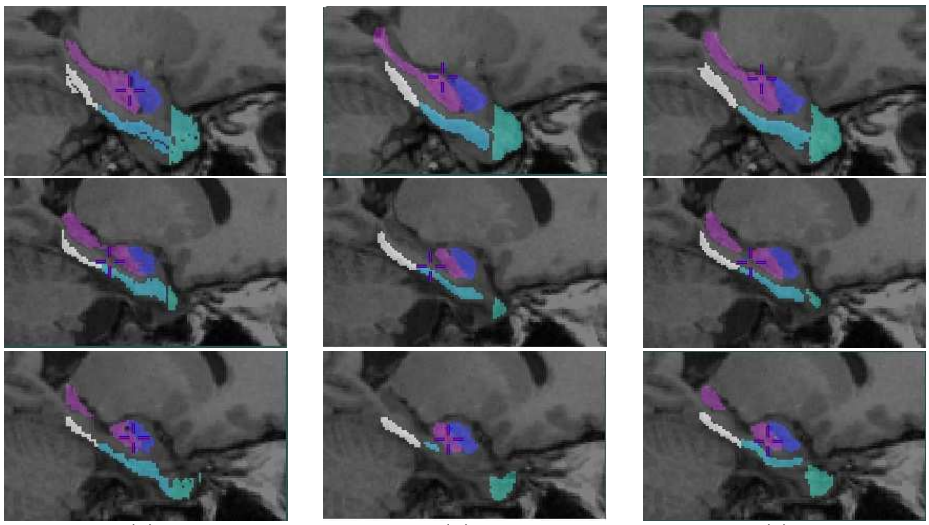
4 Discussion and Conclusions

In this paper, we present a novel segmentation algorithm which combines the appearance modeling and non-local means patch-based segmentation into a general segmentation framework, and applied it to the cortical structures of the medial temporal lobe. During segmentation, appearance modeling is used to capture the global shape variation, and non-local means patch-based method is used to improve the local fitting of the segmentation result. The experimental results have demonstrated feasibility, good performance, and robustness of this algorithm in 3D image segmentation.

Table 2. κ statistical result of segmentation

	Appearance-based segmentation without local correction		Appearance-based segmentation with local correction		Matched-pairs t-test (p-value)	
	Left	Right	Left	Right	Left	Right
HC	0.851 (0.028)	0.862 (0.02)	0.867 (0.025)	0.873 (0.019)	< 0.001	< 0.001
AG	0.80 (0.048)	0.792 (0.055)	0.812 (0.043)	0.803 (0.053)	< 0.001	< 0.001
PCC	0.711 (0.068)	0.697 (0.083)	0.735 (0.066)	0.714 (0.082)	< 0.001	< 0.001
PRC	0.696 (0.067)	0.707 (0.060)	0.730 (0.048)	0.739 (0.047)	< 0.001	< 0.001
TPC	0.684 (0.065)	0.675 (0.048)	0.712 (0.057)	0.703 (0.047)	< 0.001	< 0.001

Values in columns 2-5 are mean κ and the standard deviations are in parentheses.



(a) (b) (c)

Fig. 3. Visualization of segmentation results in 2D of three median test subjects: each row represents one test subject, and columns shows the segmentation results from (a) Manual labeling, (b) Appearance-based segmentation without local refinement, and (c) Appearance-based segmentation with local refinement. Note that the segmented labels of different structures are at the top of corresponding T1 test MR image, and presented by different colors: purple for HC, blue for AG, sky blue for PCC, white for PRC, and green for TPC.

In the experiments, we tested the proposed method using the ICBM data set with 54 young healthy subjects. The leave-one out validation experiments of 54 subjects in 3D volumes demonstrated the segmentation accuracy (the mean κ of 0.87 for HC, 0.81 for AG, 0.73 for EPC, 0.73 for PRC, and 0.71 for TPC). The direct comparison between our technique and others in the literature is difficult because of the different anatomic definitions of the structures, different types of input data, and different qualities of manual segmentations. However, taking these caveats into consideration, our results are among the best results of previous publications in HC and AG (Detail see Table. 1). For other medial temporal lobe structures, there is no published result available except for PHC [11]. Those promising initial results provide the impetus for future studies where the procedure will be applied to MR data from patients with temporal lobe epilepsy or Alzheimer's disease.

References

[1] Barense, M. D., Bussey, T. J., Lee, A. C., Rogers, T. T., Davies, R. R., Saksida, L. M., Murray, E. A., Graham, K. S.: Functional specialization in the human medial temporal lobe. J. Neurosci. 25, 10239-10246 (2005)

[2] Baxter, M. G.: Involvement of medial temporal lobe structures in memory and perception. Neuron 61, 667-677 (2009)

[3] Mori, E., Yoneda, Y., Yamashita, H., Hirono, N., Ikeda, M., Yamadori, A.: Medial temporal structures relate to memory impairment in Alzheimers disease: an MRI volumetric study. J. Neurol. Neurosurg. Psychiatry 63, 214-221 (1993)

[4] Wang, L., Miller, J. P., Gado, M. H., McKeel, D., Miller, M. I., Morris, J. C., Csernansky, J. G.: Hippocampal shape abnormalities in early AD: A replication study. Alzheimers & Dementia: The Journal of the Alzheimers Association 1(1), 52-53, (2005)

[5] Duzel, E., Schiltz, K., Solbach, T., Peschel, T., Baldeweg, T., Kaufmann, J., Szentkuti, A., Heinze, H. J.: Hippocampal atrophy in temporal lobe epilepsy is correlated with limbic systems atrophy. Journal of Neurology 253(3), 294-300(2005)

[6] Chupin, M., Mukuna-Bantumbakulu, A. R., Hasboun, D., Bardinet, E., Baillet, S., Kinkingnehun, S., Lemieux, L., Dubois, B., Gamero, L.: Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: method and validation on controls and patients with Alzheimers disease. NeuroImage 34, 996-1019 (2007)

[7] 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), 166-171 (2004)

[8] Patenaude, B., Smith, S. M., Kennedy, D. N., Jenkinson, M.: A bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage 56(3), 907-922 (2011)

[9] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A. M.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341-355 (2002)

[10] Collins, D. L., Holmes, C. J., Peters, T. M., Evans, A. C.: Automatic 3D model-based neuro-anatomical segmentation. Human Brain Mapping 3, 190-208 (1995)

[11] Heckemann, R. A., Hajinal, J. V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33(1), 115-126 (2006)
[12] Shattuck, D., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K., Poldrack, R., Bilder, R., Toga, A.: Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39(3), 1064-1080 (2008)
[13] Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46, 726-738 (2009)
[14] Collins, D. L., Pruessner, J. C.: Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting animal with a template library and label. NeuroImage 52, 1355-1366 (2010)
[15] Wang, H., Das, S., Suh, J. W., Altinay, M., Pluta, J., Craige, C., Avants, B., Yushkevich, P. A., The Alzheimer's Disease Neuroimaging Initiative: A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. NeuroImage 55, 968-985 (2011)
[16] Coupé, P., Manjon, J. V., Fonov, V., Pruessner, J., Robles, M., Collins, D. L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage 54(2), 940-954 (2011)
[17] Powell, S., Magnotta, V. A., Johnson, H., Jammalamadaka, V. K., Andreasen, N. C., Pierson, R.: Registration and machine learning based automated segmentation of subcortical and cerebellar brain structures. NeuroImage 39(1), 238-247 (2008)
[18] Lijn, V., Heijer, T., Breteler, M. M. B., Niessen, W. J.: Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. NeuroImage 43, 708-720 (2008)
[19] Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., II, H. R. W., Lewis, D. V., LaBar, K. S., Styner, M., McCarthy, G.: A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. NeuroImage 45, 855-866 (2009)
[20] Lotjonen, J. M., Wolz, R., Koikkalainen, J. R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., The Alzheimer's Disease Neuroimaging Initiative: Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage 49, 2352-2365 (2010)
[21] Benavides, G. S., Gomez-Ansonc, B., Sainzd, A., Vivesd, Y., Delfinod, M., Pea-Casanovaae, J.: Manual validation of freesurfers automated hippocampal segmentation in normal aging, mild cognitive impairment, and alzheimer disease subjects. Psychiatry research: NeuroImaging 181, 219-225 (2010)
[22] Sabuncu, M. R., Yeo, B. T. T., Leemput, K. V., Fischl B., Gollnad, P.: A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imag. 29(10), 1714-1729 (2010)
[23] Bishop, C. A., Jenkinson, M., Andersson J., Declerck, J., Merhof, D.: Novel Fast Marching for Automated Segmentation of the Hippocampus (FMASH): Method and validation on clinical data. NeuroImage 55, 1009-1019 (2011)
[24] Khan, A. R., Cherbuin, N., Wen, W., Anstey, K. J., Sachdev, P.: Optimal Weights for Local Multi-atlas Fusion using Supervised Learning and Dynamic Information (SuperDyn): Validation on Hippocampus Segmentation. NeuroImage 56, pg:126-139 (2011)
[25] Hu, S., Collins, D. L.: Joint level set shape modeling and appearance modeling for brain structure segmentation. NeuroImage 36, 672-683 (2007)

[26] Fonov, V. S., Evans, A. C., K. Botteron, C. R. A., McKinstry, R. C., Collins, D. L., BDCG: Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54(1), 313-327 (2011)
[27] Pruessner, J. C., Li, L. M., Serles, W., Pressner, M., Collins, D. L., Kabani, N., Lupien, S., Evans, A. C.: Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. Cerebral Cortex 10, 433-442 (2000)
[28] Pruessner, J. C., Kohler, S., Crane, J., andC. Lord, M. P., Byrne, A., Kabani, N., Collins, D. L., Evans, A. C.: Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: Considering the variability on the collateral sulcus. Cerebral Cortex 12,1342-1353 (2002)
[29] Dice, L.: Measures of the amount of ecological association between species. Ecology 26(3), 297-302 (1945)