

# Robust Surface-Based Multi-template Automated Algorithm to Segment Healthy and Pathological Hippocampi

Hosung Kim<sup>1</sup>, Tommaso Mansi<sup>2</sup>, Neda Bernasconi<sup>1</sup>, and Andrea Bernasconi<sup>1</sup>

<sup>1</sup> Department of Neurology and McConnell Brain Imaging Center, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada

<sup>2</sup> Siemens Corporate Research, Image Analytics and Informatics, Princeton, NJ, USA  
{khs001, neda, andrea}@bic.mni.mcgill.ca,  
{tommaso.mansi}@siemens.com

**Abstract.** The most frequent drug-resistant epilepsy is temporal lobe epilepsy (TLE) related to hippocampal atrophy. In addition, TLE is associated with atypical hippocampal morphologies. Automatic hippocampal segmentations have generally provided unsatisfactory results in this condition. We propose a novel segmentation method (SurfMulti) to statistically estimate locoregional texture and shape using a surface-based approach that guarantees shape-inherent point-wise correspondences. To account for inter-subject variability, including shape variants, we used a multi-template library derived from a large database of controls and patients. SurfMulti outperformed state-of-the-art volume-based single- and multi-template approaches, with performances comparable to controls (Dice index: 86.1 vs. 87.5%). Furthermore, the sensitivity of SurfMulti to detect atrophy was similar to that of manual volumetry. Given that the presence of hippocampal atrophy in TLE predicts a favorable seizure outcome after surgery, the proposed automated algorithm assures to be a robust surrogate tool in the presurgical evaluation for the time-demanding manual procedure.

**Keywords:** multi-template, segmentation, surface-based texture, SPHARM-PDM, template library, hippocampus, MRI, epilepsy.

## 1 Introduction

The most frequent form of drug-resistant epilepsy is temporal lobe epilepsy (TLE) related to hippocampal sclerosis, which generally appears as atrophy on MRI. Detecting hippocampal atrophy is clinically relevant as it allows lateralizing the side of seizure focus and defining the surgical target. The removal of this pathology is the only effective treatment, offering seizure freedom in 80% of cases [1].

Manual labeling is considered the gold standard to measure hippocampal volume, as it is accurate, reproducible, and sensitive. High time requirement, rater-bias, and the demand to study large cohorts of healthy and diseased populations, however, have motivated the automation of hippocampal volumetry.

Most segmentation methods employ deformable model- [2], appearance- [3], or atlas-based approaches [4-6]. The use of deformable parametric surfaces based on locoregional features has improved performance in healthy controls [2]. However, this approach estimates features on vertices arranged without anatomical correspondences, which may ultimately produce inaccurate segmentation in case of topological discrepancies due to pathology.

In patients with TLE, automatic algorithms have generally yielded less satisfactory results than in healthy subjects, with Dice indices ranging from 0.63 to 0.77 and 0.75 to 0.89, respectively [3, 7]. The reduced performance in TLE may stem from factors other than atrophy, as previous studies achieved results similar to controls in patients with Alzheimer’s disease, a neurodegenerative disorder associated with marked hippocampal atrophy [7]. Indeed, 40% of TLE patients show atypical morphologies of the hippocampus and its surroundings, which are referred to as malrotation [8]. In the presence of such abnormal anatomical variant, automatic approaches may fall into local minima when using a template or prior-knowledge based on healthy subjects. Multi-template approaches offer an appropriate framework to account for structural variability by selecting from a database a subset of atlases that best describe anatomical characteristics of the target structure. To date, these techniques have been volume-based and applied to segment healthy hippocampi [4-5]. However, they rely on nonlinear registration of image-intensity that may also fail in case of atypical morphology.

We propose a novel hippocampal segmentation method based on a multi-template approach that statistically models parametric surfaces and locoregional texture features. We applied this method to a large database of healthy controls and patients with TLE.

## 2 Methods

Our approach TLE consists of a template library construction stage and a segmentation stage, as illustrated in Fig. 1. Each stage is detailed in the following sections.

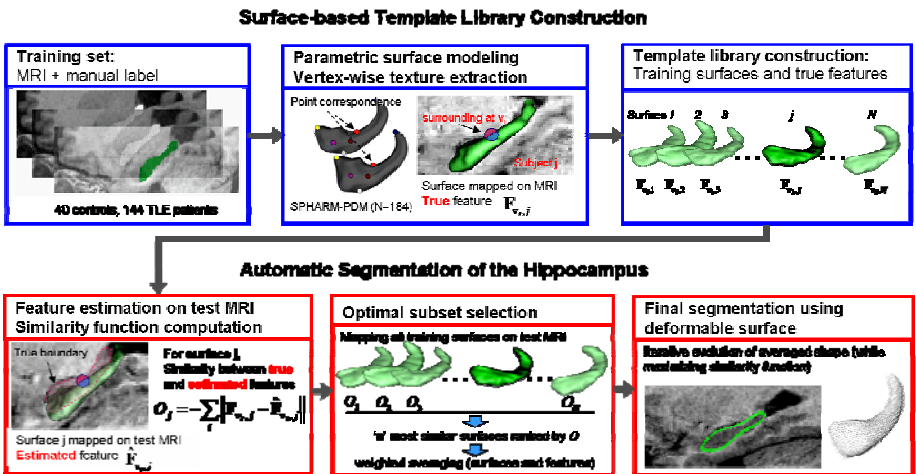


Fig. 1. Flowchart of the proposed segmentation technique

## 2.1 Template Library Construction

### 2.1.1 Surface Extraction from Manual Labels

Labels of the hippocampus obtained from manual segmentation were converted into surface meshes using the spherical harmonics parameterization and point distribution models (SPHARM-PDM), an area-preserving, distortion-minimizing spherical mapping that ensures shape-inherent point-wise correspondence through an icosahedron subdivision of the spherical harmonic parameterization [9].

### 2.1.2 Regional Texture Models

Each surface was mapped to its corresponding image to estimate regional textures. For each hippocampal vertex, a spherical neighborhood with various radii (3mm, 5mm, 7mm) was defined. The “inner region” (IR) and “outer region” (OR) of these local neighborhoods were determined with respect to the surface boundary. The following texture features were then computed at each vertex  $\mathbf{v}_i$ .

- i. *Normalized intensity* (NI) to capture regional tissue homogeneity. Let  $\mu_{IR, i} / \mu_{OR, i}$  be the mean of intensities within IR or OR at  $\mathbf{v}_i$  and  $SD_{IR, i} / SD_{OR, i}$  be the standard deviation. We defined  $NI_{IR, i} = \mu_{IR, i} / SD_{IR, i}$  and  $NI_{OR, i} = \mu_{OR, i} / SD_{OR, i}$ .
- ii. *Relative intensity* (RI) to assess the contrast between IR and OR voxels. RI was defined as  $RI_i = 2 \times (\mu_{OR, i} - \mu_{IR, i}) / (\mu_{OR, i} + \mu_{IR, i})$ .
- iii. *Gabor energy* (GE) to capture image texture through a multi-channel filtering strategy [10]. Mimicking human visual perception, this feature portrays the complexity, directionality and repetition of the intensity distribution. Let  $x, y, z$  be the spatial coordinates and  $\mathbf{R}_\theta$  a 3x3 rotation matrix whose 3D Euler angle  $\theta$  defines the orientation of the normal to the parallel stripes of a Gabor function. The Gabor filter [10] was defined by:

$$g_{\theta, \lambda, \sigma, \gamma, \varphi}(x, y, z) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2 + \gamma^2 z'^2}{2\sigma^2}\right) \exp\left(-2\pi \frac{x'}{\lambda} + \varphi\right), \quad (x', y', z') = \mathbf{R}_\theta(x, y, z) \quad (1)$$

$\gamma$  is the aspect constant,  $\varphi$  is the phase offset and the ratio  $\sigma/\lambda$  describes the bandwidth, namely the filter size  $b$  according to:

$$b = \log_2 \frac{\pi\sigma / \lambda + \sqrt{0.5 \ln 2}}{\pi\sigma / \lambda - \sqrt{0.5 \ln 2}}, \quad \sigma / \lambda = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2} \frac{2^b + 1}{2^b - 1}} \quad (2)$$

At voxel  $I(x, y, z)$ , the Gabor energy was given by  $GE_{\theta, \lambda, \sigma, \gamma, \varphi}(x, y, z) = \|g_{\theta, \lambda, \sigma, \gamma, \varphi}(x, y, z) * I(x, y, z)\|$ , where  $*$  is the convolution operator. In this work, we calculated the Gabor energy of the immediate surroundings only along the surface normal  $\theta$ . We fixed  $\varphi$  as 0 (no offset) and  $\gamma$  as 1 (same amount of information horizontally and vertically). Multiscale texture analysis was performed by varying the bandwidth  $b = \{0.5, 1, 2\}$ .

- iv. *Intensity gradient* (IG): to capture edge information. Gradients along  $x, y, z$  direction were computed and interpolated on the vertices.

### 2.1.3 Regional Shape Models

The following features constrained the deformable model evolution within the range of the anatomical variability in the template.

- i. *Distance between adjacent vertices*: to prevent irregular vertex topology.
- ii. *Gaussian curvature*: to constrain local convexity/concavity.
- iii. *Local orientation*: manual tracing in regions where anatomical boundaries are not visible often rely on arbitrary oblique lines as geometric landmarks (e.g., the inferomedial border separating CA1 from the subiculum) [11]. To model this feature, we projected the surface normals to the xy, yz and zx planes and computed their angles with respect to their orthogonal axis x, y and z-axis.

For each subject, texture and shape features were normalized across vertices using a z-transform and combined at each vertex  $\mathbf{v}_i$  into a vector  $\mathbf{F}_v = [\mathbf{F}_{texture\ v} \ \mathbf{F}_{shape\ v}]$ .

## 2.2 Automatic Segmentation of the Hippocampus

### 2.2.1 Automatic Selection of an Optimal Shape and Feature Template

Let  $\mathbf{S}_j = [\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_i \dots, \mathbf{v}_L]$  be a SPHARM-PDM surface of the template library, which is initially mapped on its own MR image. Let  $\mathbf{F}_{v_i, j}$  be the *true features*, i.e. the set of features computed at  $\mathbf{v}_i$ . Given a test image, we mapped the templates  $\mathbf{S}_1, \mathbf{S}_2 \dots \mathbf{S}_j \dots \mathbf{S}_N$ , to the test image and computed a set of *estimated features*  $\hat{\mathbf{F}}_{v_i, j}$ . For each mapping, we computed a similarity measure as:

$$O_j = -\sum_i \frac{\|\mathbf{F}_{v_i, j} - \hat{\mathbf{F}}_{v_i, j}\|}{\sqrt{\frac{1}{N} \sum_k (\mathbf{F}_{v_i, k} - \bar{\mathbf{F}}_{v_i})^2}}, \quad \bar{\mathbf{F}}_{v_i} = \frac{1}{N} \sum_{k=1}^N \mathbf{F}_{v_i, k} \tag{3}$$

Equation (3) represents a normalized similarity between the j-th *true features* and *estimated features*. Thus, it enables the selection of the closest template surface. Two approaches were investigated:

- a) Traditional approach: Given a test image, we selected the n most similar templates (surfaces and their corresponding features) [4-5]. We then computed the average surface from these ‘n’ surfaces ( $\mathbf{S}_{opt\ 1}, \mathbf{S}_{opt\ 2}, \dots, \mathbf{S}_{opt\ n}$ ) to generate an initial shape for the later segmentation process. Best results were obtained experimentally with n=10 templates.
- b) Advanced adaptive approach: Recently, Coupe et al. [12] proposed a weighted averaging of the volume labels in the selected subset according to their similarity measure. Here, we adapted this technique by optimally weighting surfaces and vertex-wise features separately. Let  $\mathbf{w}_S$  and  $\mathbf{w}_F$  be  $n \times 1$  weight vectors for optimal surfaces and features, respectively. We defined the new average surface as:

$$\bar{\mathbf{S}} = \sum_{j=1}^n w_{S,j} \mathbf{S}_{opt j}; \quad \sum w_{S,j} = 1 \quad (4)$$

Similarly, we defined the weighted mean and SD of features at vertex  $\mathbf{v}_i$  by:

$$\bar{\mathbf{F}}_{\mathbf{v}_i} = \sum_{j=1}^n w_{F,j} \mathbf{F}_{\mathbf{v}_i, opt j}; \quad \sum w_{F,j} = 1; \quad \sigma_{F \mathbf{v}_i} = \sqrt{\sum_{j=1}^n w_{F,j} (\mathbf{F}_{\mathbf{v}_i, j} - \bar{\mathbf{F}}_{\mathbf{v}_i})^2} \quad (5)$$

We then redefined the similarity function based on the n-top ranked subset as:

$$O_{subset} = -\sum_i \frac{\|\bar{\mathbf{F}}_{\mathbf{v}_i} - \hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}}\|}{\sigma_F} \quad (6)$$

$\hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}}$  is the *estimated* feature-set computed on the surface  $\bar{s}$  mapped on the test image. Finally, both weights were determined by maximizing the similarity between the current template-subset and the test image.

$$\mathbf{w} = [\mathbf{w}_S \ \mathbf{w}_F] = \arg \max_{\mathbf{w}} O_{subset} \quad (7)$$

We initially set all the components of  $\mathbf{w}$  as 1/n. We then iteratively perturbed every  $w_j$  by  $\pm\delta \cdot 1/n$  and updated it if the similarity function (7) increased. The step-size parameter  $\delta$  was initialized to 1 and decreased at each iteration by 0.1.

### 2.2.2 Automatic Segmentation: Evolution and Objective Function

The final segmentation was obtained as follows. We first linearly mapped the template computed in the previous section to the test image. Then, we locally deformed the surface at each vertex along the surface normal based on a multi-level b-spline interpolation. Analogous to Eq. (6), the cost function was defined at iteration k as:

$$O_k = -\sum_i \frac{\|\bar{\mathbf{F}}_{\mathbf{v}_i} - \hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}_k}\|}{\sigma_F} \quad (8)$$

$\bar{s}_k$  is a deformed surface at iteration k and  $\hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}_k}$  is its *estimated* feature vector. We maximized  $O_k$  using the gradient descent approach.

## 3 Experiments and Results

### 3.1 Experiments

Our training-set included 40 healthy controls (18 males; mean age  $33 \pm 12$  yrs) and 144 TLE patients (61 males; mean age  $36 \pm 11$  yrs). TLE diagnosis and lateralization of the seizure focus into left TLE (LTLE; n=73) and right TLE (RTLE; n=71) were determined by a comprehensive evaluation including video-EEG recordings and MRI evaluation. Our ethics committee approved the study. MR images were acquired on a 1.5T scanner, with 1mm-isotropic voxels. All images underwent automated correction for intensity non-uniformity, intensity standardization, and were linearly registered

into a stereotaxic space [13]. The hippocampus was manually segmented by an expert in all subjects. Using z-score normalization based on the distribution of healthy controls, we identified 91 (63%) patients with hippocampal atrophy (i.e.  $z < -2$ ) ipsilateral to the seizure focus.

### 3.1.1 Evaluation of Template Selection Approaches

We evaluated the proposed surface-based multi-template methods (SurfMulti) through a leave-one out strategy. For each test data, we selected optimal subsets and built the initial shapes and feature models using both the traditional approach and the adaptive approach. The segmentations resulting from each initialization were compared using Dice index and paired t-test.

### 3.1.2 Comparison with Volume-Based Single- and Multi-template Approaches

We obtained hippocampal segmentation using FreeSurfer, a volume-based single-template approach [6], and a multi-template approach [4-5] (that we named VolMulti). For VolMulti, we used ANIMAL as the nonlinear registration method [13] and chose the optimal subset of 11 as suggested by [4].

We evaluated the performance of each automated segmentation algorithm against the manual label using the Dice index. We compared Dice indices between automated methods in controls and each patient group (i.e. hippocampi ipsilateral / contralateral to the seizure focus) using Student's t-tests.

In a separate analysis, we assessed the sensitivity of each algorithm to detect atrophy in TLE relative to controls by computing Cohen's  $d = (\text{mean volume controls} - \text{mean volume TLE}) / \text{pooled SD}$ , indicating the effect size of a 'between group' difference. We computed the significance of the observed effect using t-tests.

Significances of all statistical tests were adjusted for multiple comparisons using Bonferroni-correction.

## 3.2 Results

The adaptive template selection approach outperformed the traditional strategy (all groups:  $p < 0.02$ , Table 1). We therefore took as reference the adaptive approach.

In all groups, performance of SurfMulti was superior to the two volume-based approaches (vs. FreeSurfer:  $p < 10e-15$ ; vs. VolMulti:  $p < 0.0004$ ). Moreover, our algorithm performed equally well in TLE patients and controls ( $p > 0.1$ ), whereas volume-based approaches segmented poorly hippocampi in patients (FreeSurfer:  $p < 10e-6$ , VolMulti:  $p = 0.02$ ).

Group-wise comparisons identified hippocampal atrophy ipsilateral to the seizure focus in TLE patients irrespective of the method, i.e. manual or automated ( $p < 0.05$ , Table 2). However, the effect size of atrophy detected using SurfMulti was closest to manual labeling (Cohen's  $d$ : Manual=1.71,  $t=7.6$ ; SurfMulti=1.60,  $t=7.0$ ; VolMulti=1.38,  $t=6.1$ ; FreeSurfer=0.91,  $t=3.9$ ).

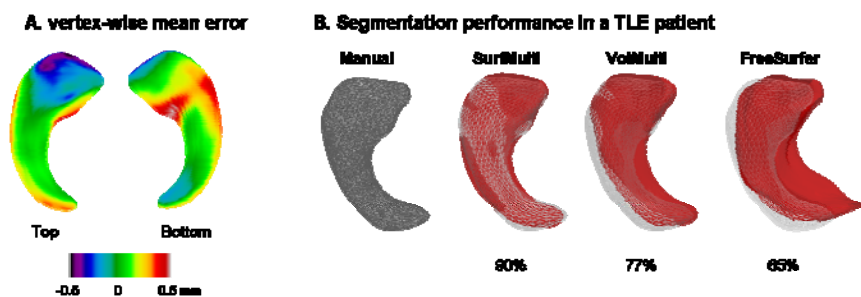
For each test data, we computed vertex-wise displacements between the SPHARM-PDM surfaces of ground truth and SurfMulti segmentations. Fig. 2-A illustrates the mean and SD of the normal displacement across subjects ( $n=184$ ). This surface-based analysis showed that SurfMulti yielded results with sub-voxel error (absolute mean error  $< 0.51$ mm; mean SD=0.9mm).

**Table 1.** Segmentation accuracy as defined by Dice similarity index (% mean $\pm$ SD)

Group	FreeSurfer	VolMulti	SurfMulti	
			Equal weight	Adaptive weight
<i>Controls</i>	72.1 $\pm$ 2.5	84.5 $\pm$ 4.5	86.4 $\pm$ 3.2	87.5 $\pm$ 2.6
<i>Ipsilateral</i>	65.8 $\pm$ 4.5	81.2 $\pm$ 5.7	84.5 $\pm$ 3.3	86.1 $\pm$ 2.9
<i>Contralateral</i>	70.0 $\pm$ 3.7	82.7 $\pm$ 5.5	86.6 $\pm$ 3.5	87.7 $\pm$ 2.7

**Table 2.** Hippocampal volumetry. Values represent z-scores with respect to controls (mean $\pm$ SD) and the effect size of atrophy as Cohen's d (within brackets; 0.2 indicates a small effect, 0.5 a medium effect, and >0.8 is a large effect).

	Manual	FreeSurfer	VolMulti	SurfMulti
<i>Ipsilateral</i>	-2.41 $\pm$ 1.85 (1.71)	-1.35 $\pm$ 2.06 (0.91)	-1.49 $\pm$ 1.40 (1.38)	-1.78 $\pm$ 1.45 (1.60)
<i>Contralateral</i>	-0.65 $\pm$ 1.40 (0.56)	-0.05 $\pm$ 1.91 (0.03)	-0.22 $\pm$ 1.29 (0.22)	-0.57 $\pm$ 1.41 (0.47)

**Fig. 2.** Vertex-wise comparison between manual (ground-truth) and automated segmentation. (A) Mean segmentation error of SurfMulti across all subjects ( $n=184$ ) mapped on the template. (B) Representative example of a TLE patient showing the overlap between manual labeling (wireframe) of the left hippocampus and the three automatic segmentation methods (in red). The Dice index is indicated below each method.

## 4 Discussion and Conclusion

We propose a novel hippocampal segmentation algorithm that integrates deformable parametric surfaces and multiple templates in a unified framework. Our technique achieved a level of accuracy in TLE patients virtually identical to healthy controls, with a Dice index of 86.1% for the atrophic hippocampus ipsilateral to the seizure focus. Such performance has not yet been paralleled in epilepsy. Vertex-wise surface-based shape mapping showed that SurfMulti with adaptive weight had an excellent overlap with manual labels, with sub-millimetric precision. Furthermore, we achieved the same sensitivity than manual volumetry in detecting atrophy ipsilateral to the focus. Given that the presence of hippocampal atrophy in TLE predicts a favorable

seizure outcome after surgery, the proposed technique assures to be a robust surrogate tool for the time-demanding manual procedure in the presurgical evaluation.

To segment brain structures, multi-template algorithms use either mutual information [4] or entropy [5] as a measure of intensity similarity between the target and template. In these methods, if intensity distributions of the two images are different, segmentation may fail. Our features, on the contrary, capture intrinsic image characteristics by computing higher order semantic features from a given image, i.e. image homogeneity, contrast, gradient and texture. Moreover, features are weighted according to both mean and SD of the optimal subset.

Although the volume-based multi-atlas method performed globally better than the single-template approach, its agreement with manual segmentation was significantly lower in patients than controls. On the other hand, our method outperformed both state-of-the-art algorithms in both controls and patients and sensitivity in detecting atrophy. This likely results from the integration of: *i*) surface-based shape-inherent point-wise correspondences guaranteed by SPHARM-PDM; *ii*) vertex-wise sampling scheme with respect to the surface boundary, allowing for a better characterization of locoregional texture and shape of structures neighboring the hippocampus; and *iii*) multi-template library derived from a large cohort of healthy controls and patients, that accounts for inter-subject variability, in particular shape variants.

## References

1. Schramm, J., Clusmann, H.: The surgery of epilepsy. *Neurosurgery* 62 S2, 463–481; discussion 481 (2008)
2. Pitiot, A., Delingette, H., Thompson, P.M., Ayache, N.: Expert knowledge-guided segmentation system for brain mri. *NeuroImage* 23, S85–S96 (2004)
3. Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C.: The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49(3), 2457–2466 (2010)
4. 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 fusion. *Neuroimage* 52(4), 1355–1366 (2010)
5. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images. *NeuroImage* 46, 726–738 (2009)
6. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der, K.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(3), 341–355 (2002)
7. Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinnet, E., Duncan, J.S., Garnero, L., Lemieux, L.: Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints. *NeuroImage* (2009)
8. Bernasconi, N., Kinay, D., Andermann, F., Antel, S., Bernasconi, A.: Analysis of shape and positioning of the hippocampal formation: An mri study in patients with partial epilepsy and healthy controls. *Brain* 128(Pt 10), 2442–2452 (2005)
9. Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Gerig, G.: Statistical shape analysis of brain structures using spharm-pdm. In: *MICCAI 2006 Opensource Workshop* (2006)

10. Grigorescu, S.E., Petkov, N., Kruizinga, P.: Comparison of texture features based on gabor filters. *IEEE Trans. Image Process.* 11(10), 1160–1167 (2002)
11. Konrad, C., Ukas, T., Nebel, C., Arolt, V., Toga, A.W., Narr, K.L.: Defining the human hippocampus in cerebral magnetic resonance images—an overview of current segmentation protocols. *Neuroimage* 47(4), 1185–1195 (2009)
12. Coupé, P., Manjón, 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)
13. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. *J. Comput. Assist. Tomogr.* 18(2), 192–205 (1994)