

# Appearance-Based Segmentation of Medial Temporal Lobe Structures

S. Duchesne,<sup>1</sup> J. C. Pruessner, and D. L. Collins

McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Canada H3A 2B4

Received October 8, 2001

**A new paradigm for the characterization of structure appearance is proposed, based on a combination of gray-level MRI intensity data and a shape descriptor derived from a priori principal components analysis of 3D deformation vector fields. Generated without external intervention, it extends into 3D more classic, 2D manual landmark-based shape models. Application of this novel concept led to a method for the segmentation of medial temporal lobe structures from brain magnetic resonance images. The strategy employed for segmentation aims at synthesizing, using the appearance model, a deformation field that maps a new volume onto a reference target. Any information defined on the reference can then be propagated back on the new volume instance, thereby achieving segmentation. The proposed method was tested on a data set of 80 normal subjects and compared against manual segmentation as well as automated segmentation results from ANIMAL, a nonlinear registration and segmentation technique. Experimental results demonstrated the robustness and flexibility of the new method. Segmentation accuracy, measured by overlap statistics, is marginally lower (< 2%) than ANIMAL, while processing time is six times faster. Finally, the applicability of this concept toward shape deformation analysis is presented.** © 2002 Elsevier Science (USA)

## 1. INTRODUCTION

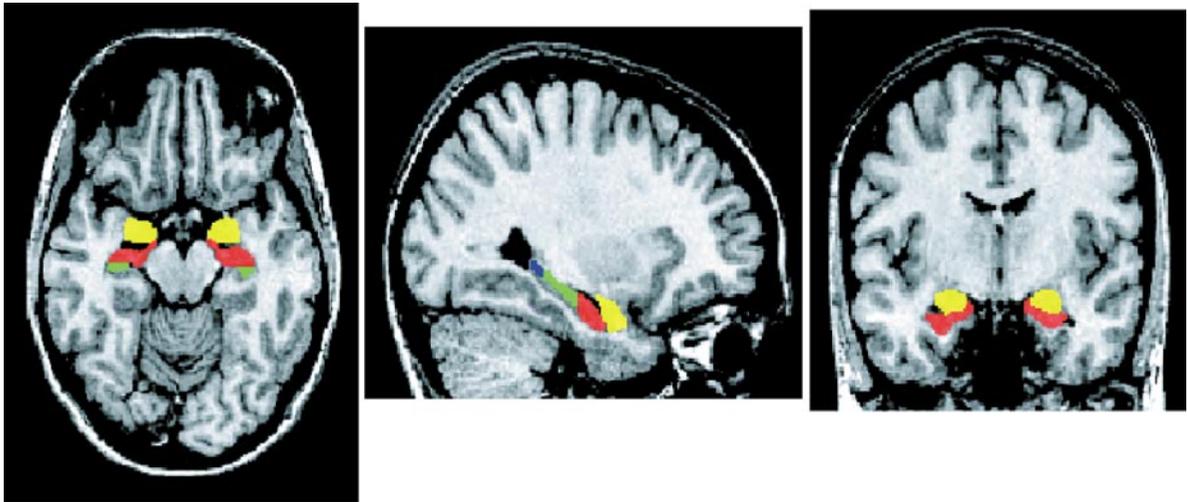
Brain structures like the hippocampus (HC) and the amygdala (AG), located in the medial temporal lobe (MTL), have received a lot of attention due to their importance in neurological diseases and disorders. For example, clinical consensus is now clear on the implication of the HC and the AG in temporal lobe epilepsy, and study of these structures (and others in the MTL) cannot but help in understanding other forms of epileptic activity. Current approaches in the study of MTL structures are heavily reliant on *in vivo* imaging techniques such as magnetic resonance imaging (MRI). Its

high contrast and spatial resolution allow morphological studies of MTL structures by extracting their positions, shapes, volumes, and other properties from the information contained in the images.

The work presented in this article is concerned with the segmentation of 3D brain structures, such as those found in the MTL, from MR images. Manual segmentation, as shown in Fig. 1, is considered highly accurate, but subject to intra/inter-observer variability, especially in the absence of a thorough segmentation protocol. It is also obvious for researchers that this is a time-consuming endeavor. Available automatic segmentation techniques include ANIMAL (Automatic Nonlinear Image Matching and Anatomical Labeling), a registration and segmentation tool based on image intensity features, and developed at the Montréal Neurological Institute (MNI) by Collins *et al.* (1995). While the segmentation accuracy of ANIMAL is comparable to that of expert manual segmentation, there is no statistical information embedded in the process, making each application of the segmentation task a completely new one from the system's point of view. Further, one must use an iterative multiscale approach leading to a segmentation time of around 2 h for the MTL region, which is not practical for many tasks and studies. Since the inception of ANIMAL, simple models have been developed elsewhere that could be integrated in some fashion to enhance the capabilities of an intensity-based technique without "memory" or *prior knowledge* of the structure of interest. For example, shape descriptors such as the point distribution models (PDMs) of Cootes *et al.* (1993) are now being used extensively in a number of applications, including segmentation in medical images (Kelemen *et al.*, 1999). There exist a few hybrid systems that combine intensity and shape data into a description of the *appearance* of an object, one of which has been successfully applied to medical imaging segmentation tasks (Cootes *et al.*, 1998).

The original motivation of the work described in this article was to develop a new segmentation method, embedding the paradigm of statistically relevant a priori information, that could achieve or surpass the accuracy of ANIMAL while possibly reducing computa-

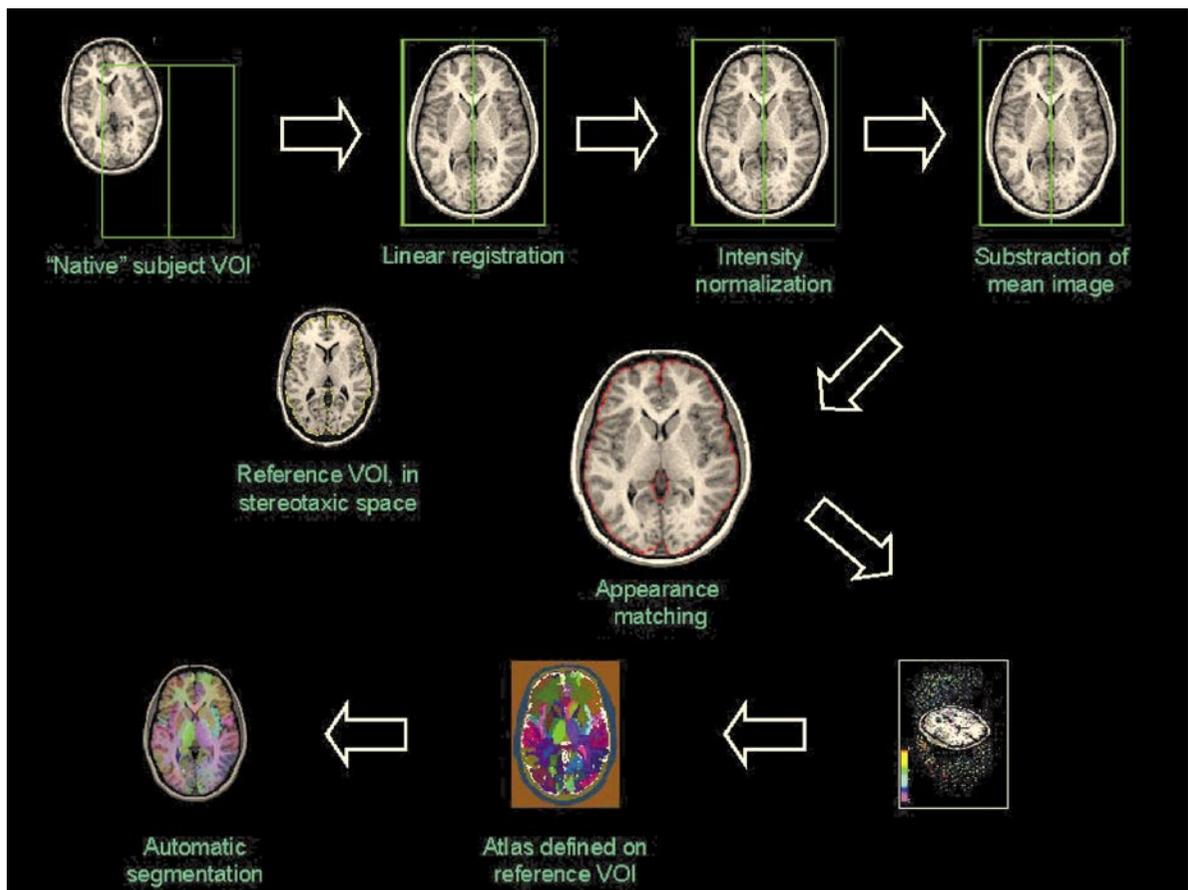
<sup>1</sup> To whom correspondence should be addressed. Fax: (514) 398-2975. E-mail: {duchesne, jens, louis}@bic.mni.mcgill.ca.



**FIG. 1.** Example of  $T_1$  MR image with manually labeled HC (blue, green, red) and AG (yellow). Image courtesy N. Bernasconi, BIC, MNI.

tional costs (Fig. 11). The reader is first invited to read an overview of applicable segmentation strategies forming the necessary background for this work. Draw-

ing on these models, an appearance-based method incorporating analysis of dense deformation fields is introduced and rigorously developed. The main



**FIG. 2.** Processing pipeline. All MRI data are processed through the pipeline shown. After preprocessing to correct for intensity nonuniformity, the data are linearly registered into stereotaxic space and resampled onto a 1-mm isotropic grid. Nonlinear transformation to stereotaxic space is used to produce 3D dense deformation fields for PCA analysis and construction of the WDM.

contribution of this work is the extension of 2D appearance-based techniques into three dimensions by including nonlinear registration vector fields into a 3D Warp Distribution Model. The image matching and segmentation strategy to be employed by the new technique is presented in the same section. The performance of the appearance-based method is characterized through a series of experiments. The accuracy of the segmentation method is evaluated against manual segmentation results from an expert neuroanatomist, and also compared with results from ANIMAL using the same input data. A discussion of experimental results and properties of the technique is followed by some general observations on this current work and ideas for future research.

## 2. BACKGROUND

Segmentation of the HC/AG on MR images has been approached in a number of ways. While a complete review of segmentation techniques is beyond the scope of this article, this section endeavors to describe those that have been applied to the MTL.

### 2.1. Manual Segmentation

Manual segmentation techniques involve contour delineation of MTL structures by one or more trained neuroanatomists. While expert human intervention remains the most accurate segmentation technique, serious drawbacks undermine its usefulness in a number of situations.

The main difficulty resides in the subjective interpretation of anatomic variations (Hogan *et al.*, 2000). In the case of HC and AG, for example, differences in border definitions among research groups—and for that matter, among investigators of the same group—have hindered the comparison of results (Pruessner *et al.*, 2000). Commonality in the definition of segmentation protocols is needed if results are to be compared.

Second, research groups use different software packages to trace the targeted structure. Most employ 2D visualization tools for brain images, without the possibility of adjusting resolution or image contrast. A common 2D error is interslice misregistration, which leads to nonsmooth 3D surfaces (Hogan *et al.*, 2000). On the other hand, scalable 3D imaging is available in some centers, allowing for precise display and enlargement of regions of interest in coronal, sagittal, and transverse orientations (Pruessner *et al.*, 2001).

Finally, this process is obviously lengthy, as the specialist must delineate the structures on a slice-by-slice basis. Taking the HC as an example, its longest axis of 4–5 cm generates on the order of 40–50 slices with isotropic 1-mm<sup>3</sup> voxels. Hogan *et al.* (2000) report a total segmentation time of 2 h per HC. This result has been corroborated locally by one of the authors (J.C.P.) for his own research.

These drawbacks motivated Pruessner *et al.* (2001) to conduct a manual segmentation study which has been retained in this work as the gold standard for HC and AG segmentation. Following a thorough literature review, Pruessner *et al.* developed a protocol for contour tracing using high-resolution, uniform, and standardized  $T_1$ -weighted MR images with 1-mm isotropic voxels. He used a 3D analysis software called DISPLAY, developed at the Montreal Neurological Institute, which allows simultaneous display of coronal, transverse, and sagittal images for visualization and segmentation. The results indicate that with 3D visualization, the irregularly shaped HC and AG can be reliably and precisely segmented. The reported intraclass interrater overlap coefficients are in the range  $\kappa = 0.83$ – $0.94$ , which indicates very good accord between raters (see Section 3 for a mathematical description of the overlap coefficient  $\kappa$ ). Regarding the relative volumes of the targeted structures, comparison of left and right HC and AG volumes in normal controls revealed a significantly larger right HC volume and no differences in the AG volume.

### 2.2. Automated Segmentation Techniques

To discuss some of the many segmentation techniques that have been developed for purposes related to brain imaging, we introduce a simple classification scheme based on two design features:

- *Segmentation paradigm*: The prevalent paradigm is one of forward segmentation, where a predefined template is made to match the new volume to achieve segmentation. In most cases experts are required to initialize the segmentation process by choosing landmarks. The matching process can be done in the way of an initial contour which will be propagated through some form of elastic matching onto the image until a proximity criterion is maximized. An example of this technique is the high-dimensional brain mapping technique using fluid transformations, proposed by Christensen *et al.* (1997). Other techniques, such as the active shape model (ASM) of Cootes and Taylor *et al.* (2000), relies on the placement of landmarks on the image to derive a model that is then globally deformed to match some intensity features of the new image. Backward segmentation, on the other end, can be thought as a reversion of the original paradigm. In those cases the new image is registered through various means onto a reference volume; segmentation information that had been previously obtained for this volume can be propagated back onto the new subject, using the inverse transformation matrix. This paradigm can be observed in Thirion's *demons* (1996) and Collins and co-workers' (1995) ANIMAL algorithms.

- *Generic or model-based*: Modeling is used to constrain the solution space while making the convergence process more efficient. At the same time, the frame-

work should be flexible enough to allow instant retraining on a new area of research, without expensive modifications or user intervention. The ASM of Cootes and Taylor (2000) is a model-based approach but with limited flexibility; it requires extensive user intervention for retraining. Nonmodeled approaches, such as *demons* and ANIMAL, while avoiding most of the problems associated with landmarks, lack any form of predictability arising from statistical analysis of priors. This versatility can be seen as a strength or a weakness.

While many other segmentation techniques exist, it is outside the scope of this article offer a comprehensive review of this field. The reader is thus referred to Pham *et al.* (2000) as well as to Clarke *et al.* (1995). The following review concentrates on a small number of selected techniques, with emphasis being placed on ANIMAL and the active shape model of Cootes and Taylor due to their relevance in the design of the appearance-based segmentation technique being proposed in this article.

### 2.3. ANIMAL

This approach, introduced by Collins *et al.* (1995), depends on a general, iterative, hierarchical nonlinear registration procedure and a 3D digital model of human brain anatomy that contains both volumetric intensity-based data and a geometric atlas. The key element in ANIMAL is its inversion of the common segmentation paradigm (Collins *et al.*, 1995), which we labeled backward segmentation earlier. Instead of matching geometric contours from an idealized atlas directly onto the MRI data, segmentation is achieved by identifying the nonlinear spatial transformation that best maps corresponding intensity-based features between the new MRI data and a model image/volume. When completed, atlas contours defined a priori on the model image are mapped by applying the inverse transformation, thereby effectively segmenting and labeling individual structures in the new data set.

The method comprises three steps (Collins *et al.*, 1995). First, geometrically invariant image features are extracted, for example, blurred image intensity and image gradient magnitude. Next, a linear registration is performed automatically by estimating the best affine transformation matrix that maximizes the correlation of the invariant features from both target and source volumes. The final step is a nonlinear registration process, where local deformations are applied to the source volume to improve the alignment of specific regions. The deformations are estimated with a hierarchical multiscale, multiresolution strategy. At each step in scale space, the goal of the optimization procedure is to identify the nonlinear transformation that maximizes the similarity between features derived from the voxels in the two data sets. The concatenation

of linear and nonlinear transform yields a global transformation, able to warp the source volume in its native space directly onto the target volume, usually defined in a stereotaxic space (but not necessarily, since any volume can be a target volume).

Segmentation is achieved by applying the inverse global transformation on any model defined on the target image. Using manually segmented structure boundaries for comparison, measures of volumetric difference and volumetric overlap were less than 10% and better than 85% (Collins *et al.*, 1995), respectively.

### 2.4. High-Dimensional Brain Mapping

The work of Christensen *et al.* (1997) focused on the development of a segmentation method based on a local, fluid registration technique. Clinical work on the HC for epilepsy by Haller *et al.* (1996) and on the HC for schizophrenia by Csernansky *et al.* (1998, 2000) are quite demonstrative of the applicability of such a method. The entire automated process, including landmark identification and coarse and fine (fluid) transformations, takes approximately 2 h on a massively parallel supercomputer (Haller *et al.*, 1996).

### 2.5. Demons

Webb *et al.* (1999) used the Thirion (1996) method to identify significant atrophy in the left or right hippocampus of patients with clinical evidence of temporal lobe epilepsy (TLE). It is an intensity-based segmentation technique based on a registration method using the so-called *demons* algorithm. For small deformations the approach is similar to an optical flow technique (Thirion, 1996).

### 2.6. Medial Models

A new technique for segmentation based on medial models is emerging in the literature and has been presented recently by Joshi *et al.* (2001) and Pizer *et al.* (2001). Styner *et al.* (2001) have used those models to segment the HC. The paradigm is based on a shape representation by spherical harmonics and a coarse-scale sampled medial description. While HC segmentation is presented as an application of this technique, no overlap statistics from implementation data were given.

### 2.7. Active Shape Model (ASM)

The landmark-based active shape model approach suggested by Cootes *et al.* (1995), based on the point distribution model (PDM) (Cootes *et al.*, 1993), formed the basis for many other landmark-based variants that were used in the segmentation of structures from MR images. Most notably for MTL structures, the idea proposed by Kelemen *et al.* (1999) closely followed the seminal work of Cootes *et al.* on ASMs, but was based

on a hierarchical parametric object description rather than a PDM. The segmentation system included both the building of statistical models and the automatic segmentation of new image data sets via a restricted elastic deformation of shape models. The technique has been applied to automatically segment left and right HC, thalamus, putamen, and globus pallidus from volumetric MR scans taken from schizophrenia studies. No overlap statistics were given.

The aim of the PDM is to build a model of the shape of the structure of interest that describes both typical shape and typical variability using previous examples from a training set (Cootes *et al.*, 1995). In the PDM, shapes are represented by a set of points or landmarks that are manually placed on each image. The labeling of the points is important. Each labeled point represents a particular part of the object or its boundary. The method works by modeling how different labeled points tend to move together as the shape varies. If the labeling is incorrect, the method will fail to capture shape variability reliably. If a point is not in the correct position on each shape, the model will be unable to precisely represent the position of that point. The model includes terms describing the noise caused by errors in point location. Points are manually identified during the training phase. Landmarks for a given shape do not move about independently; their positions are partially correlated.

### 2.8. Active Appearance Model (AAM)

In the ASM, matching of the resulting PDM model occurs as control points are moved in normals to the boundary, trying to match intensity profiles with those of the image. The allowed modes of variation for the control points are those resulting from landmark positional modeling. Intensity, in this model, is used as a constraint, rather than a property of the structure of interest. The AAM was introduced by Cootes *et al.*, (1998) in part to further their original method and make more use of the intensity information in the images. Their goal in developing the AAM was to match a full, photorealistic model directly on a new image instance by minimizing the difference between the image under interpretation and one synthesized by the model.

Knowledge about the shape is incorporated into the AAM via a PDM. The intensity under the landmarks in the PDM is sampled and used to generate a gray-level model. The two models—gray-level and PDM—are concatenated, and a supermodel is created from principal components analysis (PCA) or Karhunen–Loeve expansion of the covariance matrix. The resulting principal components that explain the most variation in the supermodel are then selected. Those principal components can be considered as eigenmodes of appearance variation, embedding shape and intensity variability.

It is possible afterward to generate a new image instance each time the shape is deformed along those principal modes of variation. It is this synthesized image that is matched to the original image. Shapes are thus modified to reduce a cost function, in the least-squares sense, between the synthesized and original images, rather than moving the shape points on an imaginary normal to the boundary. The process becomes more robust since it matches known gray-level values within contiguous neighborhoods rather than possibly erroneous samples along a profile that may be offset by other factors, such as misregistration or errors in landmark placements.

In a publication (Cootes *et al.*, 1998), and especially a technical report (Cootes and Taylor, 2000), a direct optimization approach is presented, which leads to a rapid, accurate, and robust matching algorithm. In their proposed method, they do not attempt to solve a general optimization when a model is tried to fit to a new image. Instead the fact that the optimization problem is similar each time is exploited by performing training off-line. This allows the finding of directions of rapid convergence, even though the search space has very high dimensionality (Cootes and Taylor, 2000).

No applications of this method toward MTL segmentation has been found in the literature to date.

## 3. METHODS

The technique presented in this article could be considered at first glance as a variation of the active appearance model. The AAM was selected as a basis since it takes into account not only the shape of the structure of interest (SOI), but also its gray-level appearance. However, central to the AAM is the shape-descriptive PDM which, while being sufficient for a number of applications, cannot be used in the context of this work. In effect, PDMs have to make use of neuroanatomical expertise each time a new SOI needs to be segmented, a situation that is to be avoided. In this resides the novelty of the proposed approach: a nonsupervised, completely 3D means of characterizing shape.

In lieu of the PDM, it is proposed that a warp distribution model (WDM), which forms the main contribution of this work, be used. Based on a statistical analysis of dense 3D deformation fields, this model is able to effectively characterize shape, while at the same time making use of all voxels within an image and thus avoid subsampling. The following sections introduce the mathematical foundation of the WDM and its inclusion in the more generalized frame of the appearance-based (AB) method. Rueckert *et al.* (2001) has proposed a way to create a landmark model of the brain that follows many of the same lines as proposed here, but has not proposed an application in segmentation.

The reader will note that some of this work was first presented by the authors as preliminary works in

progress (Duchesne and Collins, 2001; Duchesne *et al.*, 2001).

### 3.1. Warp Appearance Model

To simplify the following derivations, the same notation as that of Cootes *et al.* (1998) is used.

The choice of shape representation in the WDM centers on 3D deformation fields from ANIMAL (Collins *et al.*, 1995). These deformation fields consist of volumes of vectors in 3D mapping source image voxels to target image voxels of similar intensity in the nearest-neighbor sense, and which can be considered equivalent (in terms of shape deformation information) to manual landmarks, albeit of each volume element. To simplify computations, the 3D deformation vector fields or *warps* are decomposed into volumes of orthogonal deformation components  $x$ ,  $y$ ,  $z$ .

Each warp in the deformation training set can be represented by a single point in a  $3n$  dimensional space ( $3n$ -D), where  $n$  is the number of voxels in each warp volume, and for each voxel there exist three orthogonal deformation magnitude components. Thus a set of  $N$  example warps gives a cloud of  $N$  points in this  $3n$  dimensional space. It is assumed that these points lie within some region of the space, which is called the “allowable warp domain” (AWD, similar to the ASD of Cootes *et al.* (1995)), and that the points give an indication of the shape and size of this region. Every point within this  $3n$ -D domain gives a warp whose shape is broadly similar to that of those in the original training set. Thus by moving about within the AWD, new warp instances can be generated in a systematic way.

The following mathematical reasoning is aimed at constructing independent linear models for each orthogonal warp component magnitude  $x$ ,  $y$ ,  $z$ . As such, it is given in its general form.

Given a set of  $N$  warp component magnitude examples, the mean warp component  $\bar{\mathbf{s}}$  is calculated using

$$\bar{\mathbf{s}} = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i, \quad (1)$$

where  $\mathbf{s}_i$  is the column vector of warp component for a given warp field instance  $i$ , of dimension  $n$ . The principal axes of an ellipsoid fitted to the data can be calculated by applying PCA to the data. Each principal component axis yields a direction of a mode of variation, a way in which the warp components tend to move together as the warp field varies. It can also be considered as a new orthonormal basis, optimally constructed for description of the space being studied.

For each warp field instance in the training set, the deviation from the mean warp component  $d\mathbf{s}_i$  is calculated, where

$$d\mathbf{s}_i = \mathbf{s}_i - \bar{\mathbf{s}}. \quad (2)$$

To find the basis for this space, the  $n \times n$  covariance matrix,  $\mathbf{S}$ , has to be calculated using

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^N d\mathbf{s}_i d\mathbf{s}_i^T. \quad (3)$$

The principal axes of the hyperellipsoid, giving the modes of variation of the warp component, are described by  $\mathbf{p}_k$  ( $k = 1, 2, \dots, n$ ), the unit eigenvectors of  $\mathbf{S}$  such that:

$$\mathbf{S} \mathbf{p}_k = \lambda_k \mathbf{p}_k \quad (4)$$

(where  $\lambda_k$  is the  $k$ th eigenvalue of  $\mathbf{S}$ ,  $\lambda_k \geq \lambda_{k+1}$ ).

The assumption is made that the shape of this cloud in a high-dimensional space is approximately ellipsoidal (Cootes *et al.*, 1995). Finding the center and the major axes of the ellipsoid yields a way to move around in the cloud, and hence to approximate new example warp components permitted by the model.

It can be shown that the eigenvectors of the covariance matrix corresponding to the largest eigenvalues describe the longest axes of the ellipsoid, and thus the most significant modes of variation in the variables used to derive the covariance matrix (Cootes *et al.*, 1995). Most of the variation can usually be explained

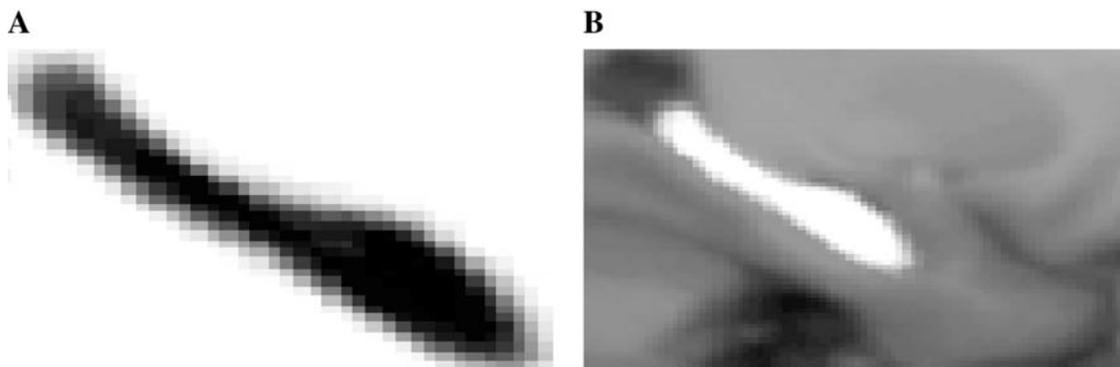
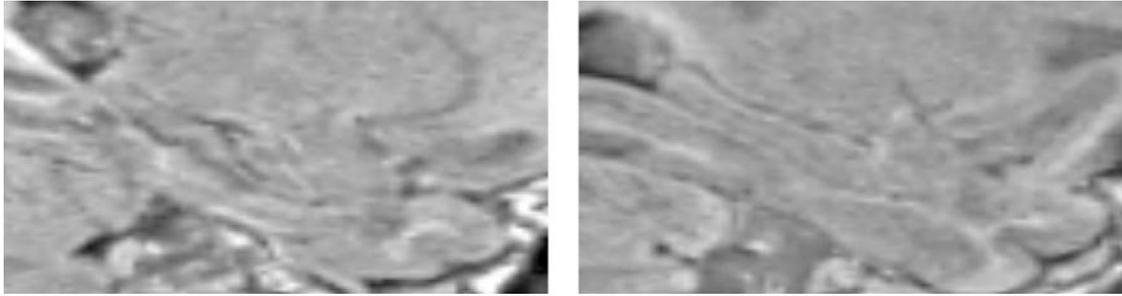


FIG. 3. (A) Medial slice through HC atlas. (B) HC atlas superimposed on average VOI.



**FIG. 4.** Subject 336: Sagittal views through the medial axis of the left and right VOI.

by a small number of modes,  $t$ , where  $t \ll n$ . This means that the  $n$ -dimensional ellipsoid is approximated by a  $t$ -dimensional ellipsoid, where  $t$  is chosen so that the original ellipsoid has a relatively small width along axes with indexes  $t + 1$  and greater (Cootes *et al.*, 1995). One method for calculating  $t$  is to choose the smallest number of modes such that the sum of their variances explains a sufficiently large proportion of  $\lambda_T$ , the total variance of all the variables, where

$$\lambda_T = \sum_{k=1}^{3n} \lambda_k \quad (5)$$

To know the variance of each eigenvector, or how each principal direction contributes in the description of the total variance of the system, the ratio of relative importance of the eigenvalue  $\lambda_k$  associated with the eigenvector  $k$  is used,

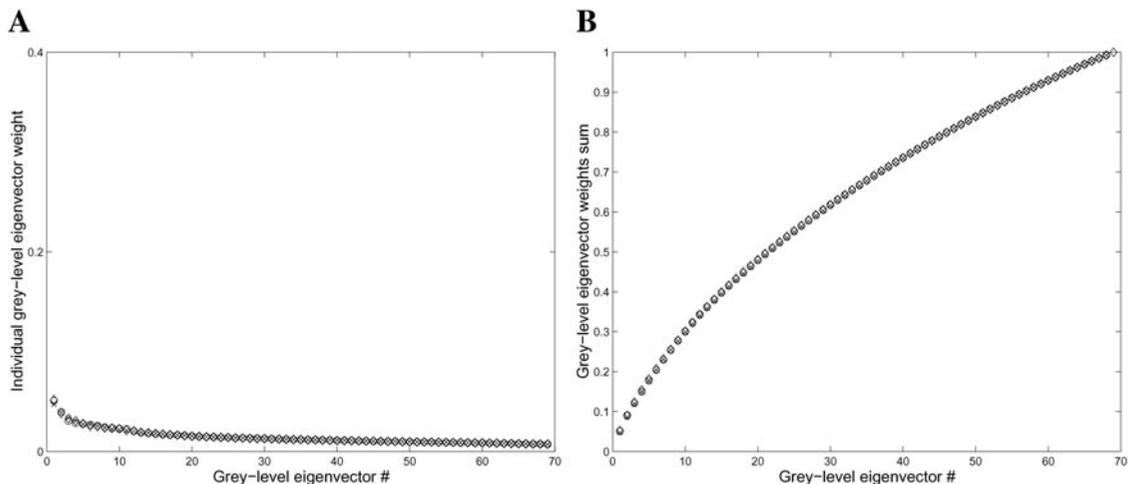
$$r_k = \frac{\lambda_k}{\sum_{j=1}^p \lambda_j}, \quad (6)$$

where  $r_k$  is the percentage relative importance for eigenvalue  $\lambda_k$ , over the sum of all  $\lambda$ , and  $p$  is the total number of eigenvectors. Incidentally,  $p = N - 1$  for the PCA implementation being used in this work, adapted from Beyrouti (1999) and similar to that described in Appendix A of Cootes and Taylor's (2000) technical report. This algorithm reduces the dimensionality of the matrices to be estimated and hence ensures that substantial changes in VOI size can be easily accommodated.

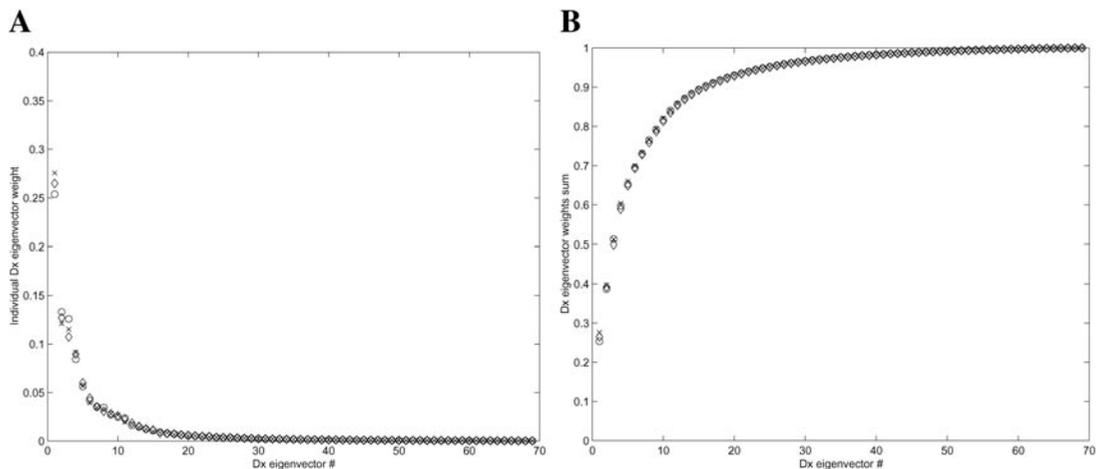
Setting a percentage  $f\%$ , one wants the smallest  $t$  eigenvectors such that the following condition is satisfied:

$$r_1 + r_2 + \dots + r_t > \frac{f}{100}. \quad (7)$$

Any point  $s$  in the reduced ellipsoidal space can be reached by taking the mean warp component and a weighted sum of the deviations obtained from the first  $t$  modes,



**FIG. 5.** Comparison of gray-level eigenvectors. (A) Individual eigenvector weights. (B) Eigenvector weights sum by training set. Results are identical for all three training sets, as is immediately apparent from the overlap of symbols.



**FIG. 6.** Comparison of  $x$ -model eigenvectors. (A) Individual eigenvector weights. (B) Eigenvector weights sum by training set. Agreement between training sets is very good.

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{P}\mathbf{B}_s, \quad (8)$$

where  $\mathbf{P} = (\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_t)$  is the matrix of the first  $t$  eigenvectors, and  $\mathbf{B}_s = (b_1 b_2 \dots b_t)^T$  is a vector of weights. The above equations allow us to generate new examples of the warp component by varying the parameters ( $b_k$ ) within suitable limits (e.g., three standard deviations around the mean), so the new warp component will be similar to those in the training set.

Thus in this fashion the linear warp variation models are constructed for each orthogonal direction:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_x \mathbf{B}_x, \quad (9)$$

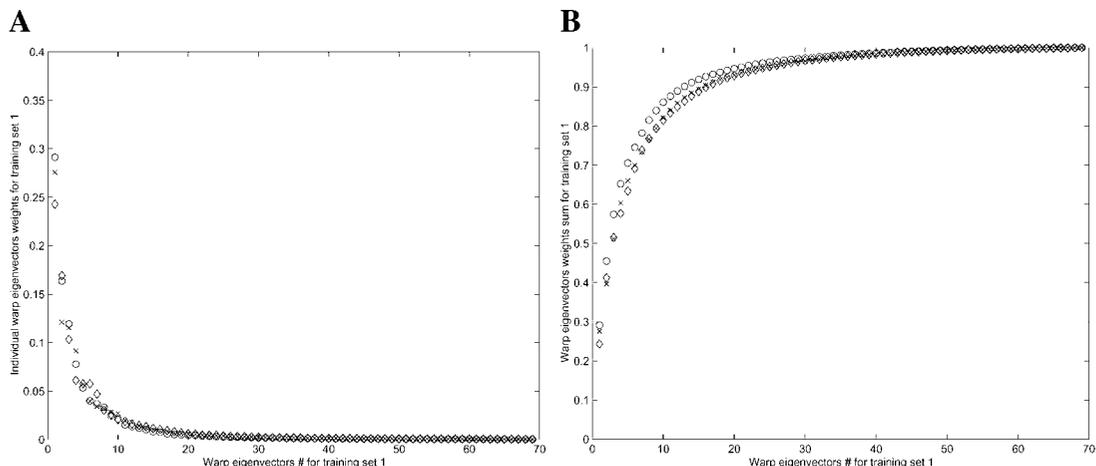
$$\mathbf{y} = \bar{\mathbf{y}} + \mathbf{P}_y \mathbf{B}_y, \quad (10)$$

$$\mathbf{z} = \bar{\mathbf{z}} + \mathbf{P}_z \mathbf{B}_z, \quad (11)$$

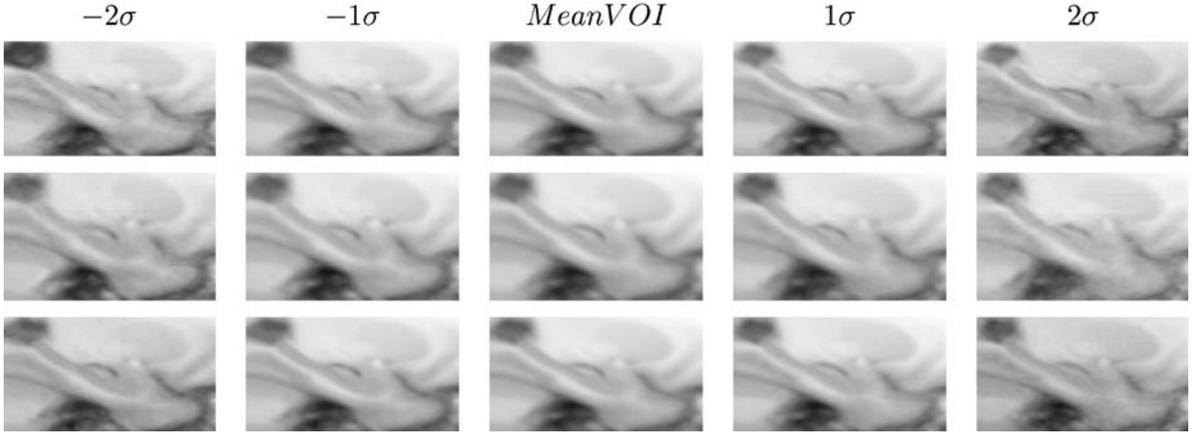
These linear models allow any new warp instance  $\mathbf{w}(\mathbf{x}, \mathbf{y}, \mathbf{z})$  to be approximated by the mean warp  $\bar{\mathbf{w}}$  and a weighted sum of the deviations obtained from the first  $t$  modes for each model, composed of  $\mathbf{P}_w$ , the set of orthogonal modes of warp variations, and  $\mathbf{B}_w$ , the set of warp parameters. Each mode of variation is linearly independent from others within the same  $x, y, z$  model, while nonlinearities may remain. Additionally, the possibility of covariance between directions is not excluded, but does not bear on the outcome of the technique being developed.

Mathematical reasoning similar to that above allows for the modeling of gray-level intensity. A linear model is constructed by applying PCA to the normalized gray-level data of  $N$  example volumes in the training set:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{B}_g, \quad (12)$$



**FIG. 7.** Comparison of warp eigenvectors for Training Set 1. Left: Individual eigenvector weights for all three deformation axes  $x, y, z$ . Right: Eigenvector weights sum for all three deformation axes. Agreement between deformation directions is very good.



**FIG. 8.** Rows 1 through 3 show variations around the mean VOI from  $-2$  to  $+2\sigma$  for PCs 1, 2, and 3, respectively. Some variations are readily observable, for example, the HC tail/horn of lateral ventricle boundary in PC1.

where  $\bar{\mathbf{g}}$  is the mean normalized gray-level vector,  $\mathbf{P}_g$  is a set of orthogonal modes of variation, and  $B_g$  is a set of gray-level parameters.

One will note that pose correction, as mentioned in Cootes and Taylor (2000), is unnecessary because all training set images are brought into a common reference space using rigid transformations. Similarly, global intensity parameters are accounted for in the normalization process performed before training. Both preprocessing steps are discussed in more detail in Section 3.3.

There may be correlations between the gray-level and warp variations that one wants to capture, and hence gray-level and warp parameters  $\mathbf{B}_g$  and  $\mathbf{B}_{x,y,z}$  are concatenated in a common matrix  $\mathbf{B}$ :

$$\mathbf{B} = \begin{pmatrix} \mathbf{W}_g \mathbf{B}_g \\ \mathbf{B}_x \\ \mathbf{B}_y \\ \mathbf{B}_z \end{pmatrix}, \quad (13)$$

where  $\mathbf{W}_g$  is a diagonal matrix of weights accounting for differences in dimensions between gray level (intensity) and warp variations (distances). The weights

were based on the ratio  $R$  of standard deviation variation in each model:

$$\mathbf{W} = R\mathbf{I} \quad (14)$$

$$\mathbf{R} = \sqrt{\frac{(\bar{\sigma}_x^2 + \bar{\sigma}_y^2 + \bar{\sigma}_z^2)^{1/2}}{\bar{\sigma}_{\text{gray}}}}$$

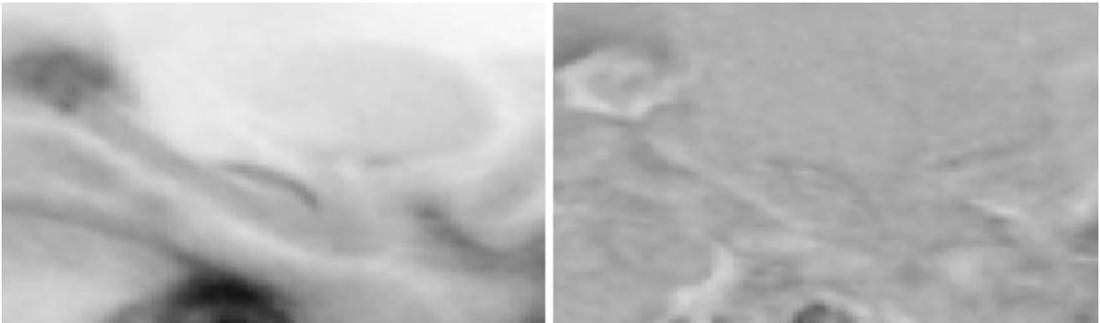
where  $\mathbf{I}$  is the identity matrix, and  $\sigma_{\text{gray},x,y,z}$  the standard deviations of the gray and warp models described above. PCA of the concatenated matrix  $\mathbf{B}$  (Eq. 13) yields a superset of parameters describing the complete appearance model;

$$\mathbf{B} = \mathbf{Q}\mathbf{C}, \quad (15)$$

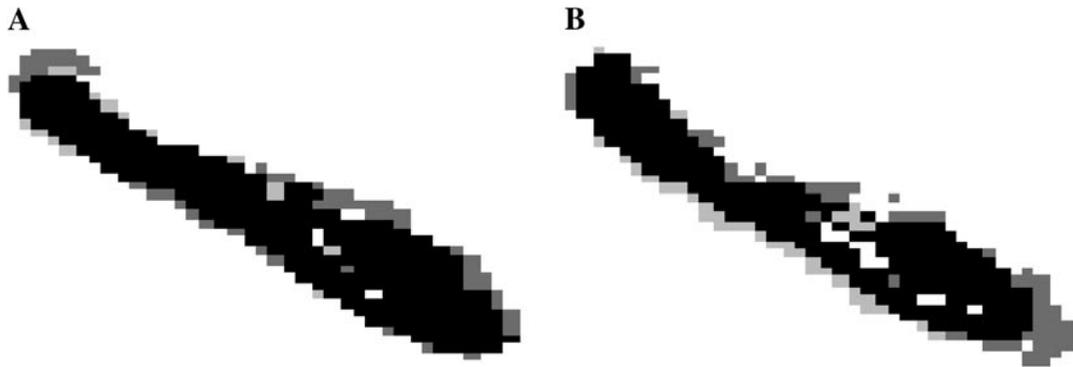
where  $\mathbf{Q}$  are appearance eigenvectors and  $\mathbf{C}$  is a vector of parameters controlling both the warp and the gray levels of the model.

### 3.2. AB Model Segmentation

Having built a model of the appearance of a training set of images and their associated deformation fields,



**FIG. 9.** Synthesized image (left) of the left VOI of subject 321 generated by the AB model and difference image (right) between the real input and the synthesized image. This difference image is in effect the residual that cannot be matched by the AB model.

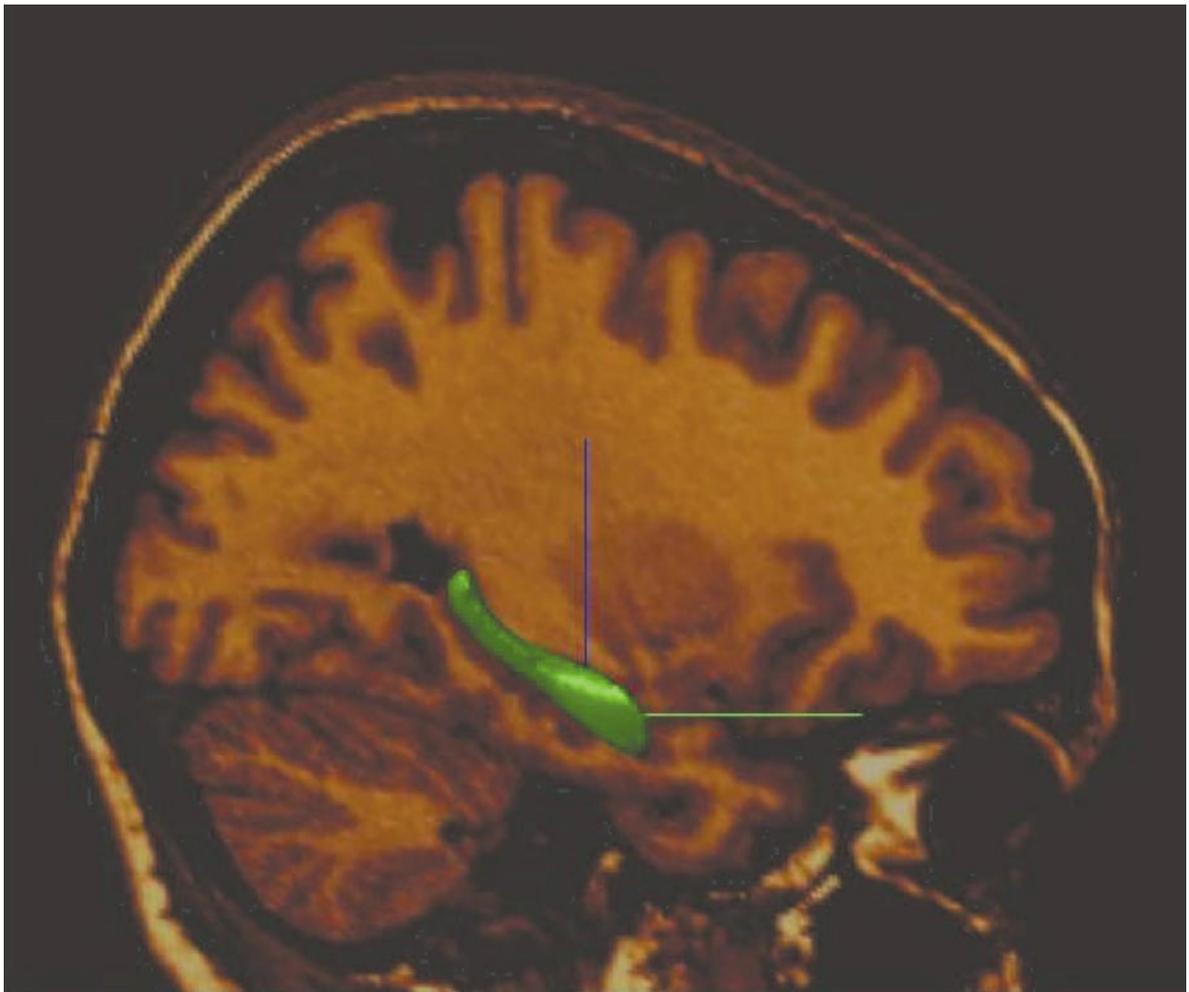


**FIG. 10.** Best and worst AB segmentation examples from Training Set I: Left: Subject 336—right hippocampus ( $\kappa = 75.8$ ). Right: Subject 347—right hippocampus ( $\kappa = 63.7$ ). Segmentation using the AB model in light gray, superimposed on manually segmented left hippocampus in dark gray. Structure overlap is displayed in black. Note that in both cases white matter folds were properly segmented.

the problem remains the same as before, albeit closer to a solution: How can one segment structures using such a model? The AB segmentation algorithm being proposed here has been inspired by the AAM segmen-

tation method presented by Cootes *et al.* (1998; Cootes and Taylor, 2000).

The AB model seeks to minimize the difference between a new image and one synthesized by the appear-



**FIG. 11.** Three-dimensional segmented left hippocampus (green) from subject 341 shown superimposed on a false color  $T_1$  MR image.

ance model. Given a set of model parameters, a hypothesis for the warp and texture of a model instance can be generated. The difference between the synthesized texture and that of the image is computed; the goal is to minimize the magnitude of this difference. Like Cootes *et al.* (1998), we observed that the pattern in the difference vector is related to the error in the model parameters. This means that the system can learn how to solve the problem off-line.

The first step in this approach consists of building a linear relationship between variations in appearance parameters and gray-level synthesized images. Defining  $\delta\mathbf{V}$  as the difference vector between  $\mathbf{V}_i$ , the vector of gray-level values in the image, and  $\mathbf{V}_m$ , the vector of gray-level values from the synthesized image,

$$\delta\mathbf{V} = \mathbf{V}_i - \mathbf{V}_m \quad (16)$$

a linear model was constructed between  $\delta\mathbf{V}$  and the error in the model parameters  $\delta\mathbf{C}$ ,

$$\delta\mathbf{C} = \mathbf{A}\delta\mathbf{V}, \quad (17)$$

where  $\mathbf{A}$  is the matrix of linear coefficients found for each voxel. To derive such coefficients, a multivariate linear regression is run on a sample of known model displacements. To generate such displacement a fraction of the standard deviation of superparameters found in  $\mathbf{B}$  (Eq. 15) is calculated for each element of  $\mathbf{V}_m$  and the resulting error is taken as  $\mathbf{E} = \text{norm}(\mathbf{V})$ . A more thorough discussion of this approach can be found in Cootes *et al.* (1998).

The second step in segmenting with the appearance model is to use an iterative algorithm to generate new estimates of the synthesized image  $\mathbf{V}_m$  that gradually approximated the new image. Varying model parameters  $\mathbf{C}$  along each vector of  $\mathbf{A}$ , the algorithm finds the closest match in the least-squares sense by minimizing the magnitude of the difference vector,  $\Delta = |\delta\mathbf{V}|^2$ .

By varying the parameters  $\mathbf{C}$  for each eigenvector, using the linear coefficients of matrix  $\mathbf{A}$ , new synthesized image instances can be created and thus a matching error (in the least-squares sense) can be calculated. Plotting the error as a function of increment, where the range of incrementation is also expressed as a fraction of standard deviation, the closest matching point is found by finding the minimum of a second-order fit to the error matching function.

After convergence, the solution explicitly contains warp variation parameters (Eq. 15), which can be expressed back into  $x, y, z$  components of the warp field and concatenated into ANIMAL vector format. Segmentation of the VOI is then possible using any structure model defined on the ANIMAL reference volume. It is achieved by applying the inverse of the deforma-

tion field to structures defined in the standard volume and then mapping those onto the subject.

Many such structure models exist already at the MNI compatible with the ANIMAL format. Hence, they render possible the automated, nonsupervised segmentation not only within the MTL, but also of different structures of the brain by simply retraining the AB model on another VOI.

### 3.3. MRI Preprocessing

A number of processing steps are required before training volumes are incorporated into the models, and before any new volume can be segmented. Preprocessing steps include

- image intensity nonuniformity correction (Sled *et al.*, 1998);
- linear registration (ANIMAL in linear mode (Collins *et al.*, 1994)) and resampling into stereotaxic space;
- delineation of the VOI (see Section 4.4);
- intensity normalization with respect to a common volume of reference (see Section 4.4)

Of course, for the construction of the warp models, all subjects in the training set had to be nonlinearly registered (ANIMAL in nonlinear mode (Collins and Evans (1997)) with respect to a common volume of reference (see Section 4.4).

These steps have been combined in a processing pipeline presented schematically in Fig. 2 which represents the steps undertaken to segment a new volume instance.

## 4. EXPERIMENTS

Three experiments have been designed to assess, in a quantitative manner, the accuracy and robustness of the AB segmentation method under a variety of conditions, for which the results are presented here. Depending on the experiment, different training set/validation set pairs were used, as follows:

- Training Set 1: first 70 subjects of group of 80
- Training Set 2: last 70 subjects of group of 80
- Training Set 3: first and last 35 subjects of group of 80

Validation sets (for segmentation) were implicitly defined as the remaining 10 subjects of the group of 80 that were not part of the training set.

### 4.1. AB Segmentation

The first experiment aimed at segmenting the structures of interest (SOIs), the hippocampus and the amygdala, from validation subjects. In this experiment, Training Set 1 was used to train the model,

which was then used to segment the left and right VOIs of subjects in Validation Set 1. The threshold (Eq. 7) for eigenvector retention in the model training was set at 97.5%.  $\kappa$  statistics was used to assess the overlap with the manual structures taken from the Pruessner *et al.* (2001) study.

#### 4.2. ANIMAL Segmentation

In the interest of thorough validation, the AB segmentation method was compared with another automated technique. The ANIMAL segmentation method described earlier (see Section 2.3) was a logical choice, being the primary technique used at the MNI and also the parent of the current method. Experiment 2 consisted of the segmentation of the same validation subjects as those of Experiment 1.

#### 4.3. AB Robustness Analysis

Experiment 3 was designed and performed to quantitatively assess the robustness of the AB segmentation method. The aim was to determine if, for different input training sets, the validation results would be consistent, all other conditions being the same. The degree of variance was not expected to change appreciably since there is a commonality of at least 50 subjects for each of the training sets, and hence eigenvectors from different training sets should not be appreciably different. Experiment 3 consisted of two additional model trainings and segmentations, and comparison of these results with those of Experiment 1. A global measure of accuracy for AB segmentation can be found by averaging results of Experiments 1 and 3.

#### 4.4. Choice of Reference and Determination of the VOI

In the standard application of ANIMAL, the target is an MRI volume from a single subject where all of the voxels within the volume have been anatomically labeled by a neuroanatomist to form an atlas (e.g., Collins *et al.*, 1998). In the paradigm described here, the target is a voxel-by-voxel intensity average of 152 MRI volumes of young normal normal subjects (86 male, 66 female, age  $24.6 \pm 4.8$ ). These data were acquired as part of the International Consortium for Brain Mapping (ICBM) project, a Human Brain Mapping-funded research project with the goal of building a probabilistic atlas of human neuroanatomy (Mazziotta *et al.*, 1995). Each volume was automatically registered and resampled in stereotaxic space using ANIMAL in linear mode (Collins *et al.*, 1994). The resulting MRI intensity average was termed ICBM 152 Symmetrical Average.

Delineation of the volume of interest for both hemispheres was done with the help of an expert anatomist. The extent of the VOI was selected to ensure that the

MTL would be included in the VOI regardless of individual anatomical variations. The resulting VOI for this work consisted of  $80 \times 70 \times 52$  voxels.

#### 4.5. Evaluation and Validation

To assess the accuracy and robustness of the AB technique in a quantitative manner, a gold standard with which to compare results was necessary. The one retained for this thesis was manual segmentation. The manual labels that are used for comparison were defined on the same subjects in an earlier study by Pruessner *et al.* (2001).

Consequently the atlas to be used for segmentation—with ANIMAL and AB methods—had to be defined using the same border definitions as those of Pruessner *et al.* As no such atlas existed, it had to be constructed using the same manual labels.

The HC and AG atlases were created by:

- mapping the manual labels (originally defined in stereotaxic space) in the space of the reference target (this was done using ANIMAL in nonlinear mode);
- averaging the structures in the nonlinear, reference target space;
- normalizing the averaged voxel distribution; and
- selecting a threshold (25%) for the extent of the averaged data (this threshold ensured that the atlas was limited to the best possible definition of HC and AG, according to local expert opinion (J.C.P.)).

The atlas was created using 70 of the 80 Pruessner labels for normal subjects from the ICBM database. The following figures show a medial slice through the left HC atlas (Fig. 3) and its superposition on the average VOI for 70 normal subjects (Fig. 3, inverted gray scale).

To compare the methods quantitatively, a similarity measure, first proposed by Dice (1945), was selected. As shown by Zijdenbos *et al.* (1994), this measure is a variant of the standard chance-corrected kappa ( $\kappa$ ) coefficient originally developed by Cohen (1960). This measure is the same as  $\kappa$  when the background is infinitely large,

$$\kappa_{\infty} = \frac{2a}{2a + b + c}, \quad (18)$$

where  $a$  is the number of voxels in the intersection of both labelings,  $b$  is the number of voxels only labeled automatically, and  $c$ , those only labeled manually.

A  $\kappa$  of 0.7 is usually considered to represent good agreement between labelings. In comparison, a single hippocampus, compared with itself after a displacement of only 1 mm in each direction ( $x$ ,  $y$ ,  $z$ ), results in  $\kappa = 0.80$ . The stringency of this test should be kept in

**TABLE 1**

AB Segmentation  $\kappa$  Statistic Average Results for Three Validation Sets of 10 Subjects Each

Validation set	Left		Right	
	$\kappa$ HC	$\kappa$ AG	$\kappa$ HC	$\kappa$ AG
1	65.0	62.3	69.6	60.6
2	65.4	61.3	68.7	62.5
3	69.9	65.6	72.3	67.0

mind in the discussion on the experimental results to follow.

## 5. RESULTS

### 5.1. Appearance-Based Model Training Results

Before proceeding with experiment results, it seemed appropriate to include other results from the training phase of the AB model. The method described in Section 3 was used to build a model of medial temporal lobe appearance. Figure 4 presents sagittal views through the medial axis of the volume of interest, after linear alignment, which form the input data into the gray-level PCA analysis. Note that the mean VOI has been subtracted prior to PCA; hence, those are difference images. Figure 9 presents the image synthesized by the model for the left VOI of validation subject 321, as well as the difference between the input and the synthesized images. One can thus see the residual that cannot be matched by the model.

For each model, all  $n - 1$  eigenvectors were found, but not all were retained for training and segmentation (see Eq. 6). It is interesting to have a look at the distribution and coherence of the eigenvector results and the weight associated with each using the ratio  $r_k$  (Eq. 6). Figures 5 and 6 display this information for the gray-level and  $x$  models. Graphs for the  $y$ ,  $z$  and concatenated appearance vectors (from Eq. 15) display similar patterns and thus are not shown here; the significance of this result is discussed in Section 6. Two graphs are presented for each model; the first shows the relative weight per eigenvector, across the three training sets. The second graph displays the cumulative weight of all eigenvectors. From this latter graph one can easily see how many eigenvectors will be retained for each model, given a user-driven desired percentage fraction (Eq. 7) of variations to be explained by the models.

The same data are now presented differently. For each training set, weights of all three deformation eigenvectors ( $x$ ,  $y$ ,  $z$ ) are presented on the same graph (Fig. 7). It is interesting to compare training sets in turn, looking for significant differences between deformation directions. Again, as the deviations were minimal, data

for training sets 2 and 3 are not presented here, the significance of which is also addressed in Section 6.

Equally interesting is the image of the eigenvectors themselves, or more aptly the principal directions. The reader will find in Fig. 8 images of the first three principal components from the gray-level VOI model of the left-side data, as well as variations around the mean for each PC. All images displayed are sagittal views through the medial axis of the VOI.

### 5.2. AB Segmentation Results

The AB model was used to segment the hippocampus from 10 subjects for validation and comparison purposes. Manual labels for those subjects were available (Pruessner *et al.*, 2001). Average  $\kappa$  statistic estimates of the overlap between AB segmented structures and manual labels are presented in Table 1. Processing time (the time required to compute the deformation field for a preprocessed, input volume), excluding off-line training, was approximately 20 min per side per subject for the VOI dimensions mentioned above. Experiments were run on a Pentium III 550-MHz PC, using MATLAB.

Figure 10 presents renditions of the best and worse  $\kappa$  subjects for right-side segmentations using Training Set 1. AB segmentation results are shown in light gray, superimposed on the manually segmented right hippocampi in dark gray. Structure overlap is displayed in black. Figure 11 presents the left HC for subject 347 segmented in three dimensions and superimposed on the individual's  $T_1$  MR image.

### 5.3. ANIMAL Segmentation Results

For Experiment 2, ANIMAL software was used to segment the hippocampus from the same validation set as Experiment 1. Average  $\kappa$  statistic estimates of the overlap between ANIMAL segmented structures and manual labels are presented in Table 2. Processing time (the time required to compute the deformation field for a preprocessed, input volume) has been approximated at 120 min per side per subject for the same VOIs. Experiments were run on a Pentium III 550-MHz PC using compiled C code.

**TABLE 2**

ANIMAL Segmentation  $\kappa$  Statistic Average Results for Same Validation Sets as Those Used for AB Segmentation

Validation set	Left		Right	
	$\kappa$ HC	$\kappa$ AG	$\kappa$ HC	$\kappa$ AG
1	67.1	64.5	71.5	62.8
2	67.5	64.0	70.0	64.9
3	73.0	67.7	73.9	68.0

TABLE 3

Segmentation Results, AB versus ANIMAL, Training set 1, Validation set 1, Left/Right

Method	Structure	Left			Right		$\bar{\kappa}$
		Lowest $\kappa$	Highest $\kappa$	$\kappa$	Lowest $\kappa$	Highest $\kappa$	
AB	HC	57.7	71.3	65.0	63.7	75.8	69.6
AB	AG	54.6	71.8	62.3	52.5	71.5	60.6
ANIMAL	HC	61.4	71.5	67.1	67.7	75.6	71.5
ANIMAL	AG	56.6	74.7	64.5	56.4	72.7	62.6

#### 5.4. AB Robustness Analysis Results

The mean  $\kappa$  statistics results of Experiment 3—segmentation of an additional two validation sets—are shown in Table 1 for AB segmentation and Table 2 for ANIMAL segmentation.

## 6. DISCUSSION

This section presents a discussion of the relevance of an appearance-based segmentation technique, further comments on experimental results, and finally some thoughts on the potential for shape analysis.

### 6.1. On the Relevance of AB Segmentation

The principle of an appearance-based model incorporating analysis of 3D deformation fields for segmentation has been demonstrated in the preceding sections of this article. While this technique presupposes the existence of an accurate nonlinear registration method for the creation of a deformation model, a method that in turn could be used for segmentation, it adds novelty rather than redundancy to the image segmentation community. Primarily, it extends the actual 2D nature of appearance matching into three dimensions. Second, it allows for the creation of an appearance model of the VOI which a technique like ANIMAL could not perform. This modeling may improve the robustness of the segmentation process in the presence of pathologies. Finally, other arguments include computational speed (AB being faster than ANIMAL) and the promise of shape analysis, which is discussed later in this section.

### 6.2. Concerning AB Model Training Results

For each model, before performing PCA, the mean element (mean gray-level image, mean  $x$ ,  $y$ ,  $z$  warp) is subtracted from each element. Analysis is therefore performed on the difference vectors, and the principal directions explain the variations around the mean. Consequently a large number of eigenvectors need to be kept in each model to reach the desired percentage of explained variation ( $f$ ).

The results displayed in Fig. 5 signify that (1) linear registration of the VOI worked well, as no principal

component seems to be related to positional variation; and (2) the amount of morphological variation from the mean is small, since the first 10 eigenvectors account for <20% of total variation in the system.

Interestingly, but not surprisingly, all three deformation directions are statistically equivalent; that is, no deformation direction seems favored over another. This forms another proof that (1) linear registration worked correctly, isotropically; and (2) within the VOI, differential data do not present any preferential direction of deformation. For normal subjects the morphological significance of this last result is not entirely surprising.

One can see in Fig. 7 that the resulting weights attached to each eigenvector (using the relative importance of its eigenvalue) are mostly equivalent, across training sets. This is in large part due to the commonality/overlap within the training sets, since 50 structures are common to the three training sets of Experiments 1 and 3. It was therefore not expected that the gray and warp eigenvectors be radically different from one training set to another.

Images of the principal eigenvectors demonstrate that the modes of variation follow high gradient lines (e.g., HC borders) and zones of low intensity (e.g., ventricles), which intuitively seems correct.

Improvements in model training philosophy could be implemented following the results of Baker and Matthews (2001) and Jones and Poggio (1998).

### 6.3. Comparing AB and ANIMAL Segmentation Results

Segmentation speed is approximately six times faster than ANIMAL or expert manual segmentation. Its accuracy, measured by statistics of overlap with respect to manual segmentation, is marginally below that of ANIMAL as shown in Table 3.

There are a few observations to make regarding the data presented in Table 3. First, one notes that the right HC and AG fare better for both techniques. Interestingly, the method is completely independent from the notion of side. On closer examination of the data, one finds that the atlas that was being used was actually better defined on the right side than the left

TABLE 4

Robustness Results: Average  $\kappa$  Statistics over Three Training Sets for Each Structure

Method	Structure	$\bar{\kappa}$	$\bar{\kappa}$	Final $\kappa$
		Left	Right	
AB	HC	66.7	70.2	68.5
AB	AG	63.1	63.4	63.3
ANIMAL	HC	69.2	71.8	70.5
ANIMAL	AG	65.4	65.2	65.3

side. The averaged data for the right structures constituting the atlas were more “compact,” yielding evidence of smaller positional variability. Hence, the resulting atlas had clearer boundaries, and since the variation in the contours was smaller, the resulting segmentation was better, driving  $\kappa$  statistics higher. The fact that this holds true for ANIMAL as well as AB segmentation is an indication that the proposed method, AB, is not side-discriminatory, as it should be.

The same observation holds true for HC results versus AG results. The AG is systematically less well-defined in the manually segmented structures, and hence the atlas is also of a poorer quality. The resulting segmentation is thus affected by the quality of the atlas, and hence the  $\kappa$  statistics. Equally, the fact that this holds true for ANIMAL as well as AG segmentation is an indication that the proposed method, AB, is not structure-discriminatory.

#### 6.4. Assessing AB Robustness

Results from this experiment, shown in Table 4, indicate that there are no significant differences between training sets, on the one hand, and validation sets, on the other. Incidentally, it gives an expanded range for statistical analysis of AB segmentation accuracy if one considers all 60 structures at once and for which the observations listed previously hold true.

While the values for AB segmentation presented in Table 4 are found, on the whole, marginally lower than those for ANIMAL, one must take into account the fact that the hippocampus has a large surface-to-volume ratio, so that even small errors of surface agreement adversely affect  $\kappa$ . It should also be recalled that the overlap statistic of a 1-mm isotropically displaced HC (one voxel off in each direction) against itself is  $\kappa = 0.80$ . From analysis of the images, both AB and ANIMAL segmentation techniques generally oversegmented the structures, a condition brought about by the poor definition of the atlas being used around its boundaries. Hence, results using a nonoptimized atlas indicate that the AB method measures favorably against established nonsupervised techniques and human observers, while being significantly faster.

The reader should also note that overlap statistics are hard to compare in an absolute way with results

from another technique, when the same data, models, and atlases are not used. It would have been easy to adjust the threshold parameters of the model to increase the  $\kappa$  statistic figures at the expense of anatomical trueness, but the point would have been lost. The central element of these experiments was to compare this novel approach with ANIMAL, itself an already established technique with a track record of accuracy. The key conclusion to retain is thus the ability of this technique to segment within a 2% (HC and AG) difference of ANIMAL. It is probable that another statistical test could be used that would be less stringent or sensitive than  $\kappa$ , while yielding more comprehensive information about the quality of the segmentation.

#### 6.5. On the Potential for Morphometry or Shape Analysis

Hippocampal shape is now at the forefront of brain imaging research. For example, analysis of shape deformations is being used for early diagnosis of dementia of the Alzheimer type (DAT) (Csernansky *et al.*, 1998). A means of quantifying 3D shape deformations, and thus performing shape analysis, is embedded in the segmentation technique proposed in this thesis.

The basis of this idea originates from a closer look at the allowable warp domain. The distribution of the eigencoordinates that arise from the projection of subjects in the spaces defined by the training set can be used as a probability function which serves to characterize the type of population being projected.

It is conceivable that, coupled with other methods in the proper statistical framework, results from such a deformation analysis could increase both specificity and sensitivity of diagnosis for a number of neurological disorders and possibly aid in early detection, as well, of course, as serving to characterize “normal” populations. Other techniques such as volumetric analysis and/or voxel-based morphometry could be correlated with the results from this approach in the form of an image processing *protocol* as an aid to diagnosis.

A second benefit from this analysis is not only to derive a simple factor of “distance” and relate it to the presence or not of pathologies, but also to correlate diagnostic information to a particular principal component. This idea is closer in spirit to the work of Csernansky *et al.* (1998). Trained eyes are able to give anatomical interpretations of at least the first few PCs. When this is not feasible, however, factor analysis can be used. Knowing which PC can be statistically associated with a disease state yields important information for the clinicians. For example, it has the potential to detect variations in the shape of an organ when there are no discernable volume changes or where volumetric data cannot determine the location of atrophy.

## 7. CONCLUSION

The original motivation of the work was to develop a new segmentation method, embedding statistically relevant *a priori* information, that could achieve the same accuracy as ANIMAL while possibly reducing computational cost. This goal has been met by the appearance based method, which incorporates analysis of dense 3D deformation fields from ANIMAL into the framework of appearance-based segmentation. This article has presented the theoretical basis for such a method, as well as a series of experiments and results that took this concept past the proof of principle stage, into the domain of applicability. While more work remains to increase its efficiency in terms of accuracy, the method is ready to be used for segmentation tasks in a number of areas of interest to researchers. The gains in speed, compared with available MNI segmentation techniques, render this approach immediately attractive for a number of prospective and retrospective studies on the MTL and other regions. The potential for morphometry should also be explored in detail to extract all possible diagnosis information from the data collected.

## ACKNOWLEDGMENTS

This study was supported by Fonds pour les Chercheurs et l'Aide à la Recherche (Government of Québec) and ICBM.

## REFERENCES

- Baker, S., and Matthews, I. 2001. Equivalence and efficiency of image alignment algorithms. In *Proceedings of the 2001 Conference on Computer Vision and Pattern Recognition*, pp. 1090–1097.
- Beyrouti, F. 1999. *An Improved Appearance-Based Approach to Image Retrieval and Classification*. Master's thesis, McGill University.
- Christensen, G., Joshi, S., and Miller, M. 1997. Volumetric transformation of brain anatomy. *IEEE Trans. Med. Imaging* **16**: 864–877.
- Clarke, L., Velthuizen, R., Camacho, M., Heine, J., Vaidyanathan, M., Hall, L., Thatcher, R., and Silbiger, M. 1995. MRI segmentation: Methods and applications. *Magn. Reson. Imaging* **13**: 343–368.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**: 37–46, 1960.
- Collins, D., and Evans, A. 1997. ANIMAL: Validation and applications of non-linear registration based segmentation. *Int. J. Pattern Recog. Artificial Intelligence* **11**: 1271–1294.
- Collins, D., Holmes, C., Peters, T., and Evans, A., 1995. Automatic 3D model-based neuroanatomical segmentation. *Hum. Brain Mapping* **3**: 190–208.
- Collins, D., Neelin, P., Peters, T., and Evans, A. 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach Space. *J. Comput. Assist. Tomogr.* **18**: 192–205.
- Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., and Evans, A. 1998. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging* **17**: 463–468.
- Cootes, T., and Taylor, C. 2000. *Statistical Models of Appearance for Computer Vision*. Technical report, University of Manchester.
- Cootes, T., Edwards, G., and Taylor, C. 1998. Active appearance models. In *Proceedings of the European Conference on Computer Vision*, pp. 484–498. ECCV Springer-Verlag, Berlin/New York.
- Cootes, T., Taylor, C., Cooper, D., and Graham, J. 1995. Active shape models, their training and application. *Comput. Vision Image Understanding* **61**: 38–59.
- Cootes, T., Taylor, C., Lanitis, A., Cooper, D., and Graham, J. 1993. Building and using flexible models incorporating grey-level information. In *Proceedings of the International Conference on Computer Vision*, pp. 242–246. IEEE Computer Soc. Press.
- Csernansky, J., Joshi, S., Wang, L., Haller, J., Gado, M., Miller, J., Grenander, U., and Miller, M. 1998. Hippocampal morphometry in schizophrenia by high dimensional brain mapping. *Proc. Natl. Acad. Sci. USA* **95**: 11406–11411.
- Csernansky, J., Wang, L., Joshi, S., Miller, J., Gado, M., Kido, D., McKeel, D., Morris, J., and Miller, M. 2000. Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus. *Neurology* **55**: 1636–1643.
- Dice, L. 1945. Measures of the amount of ecologic association between species. *Ecology*.
- Duchesne, S., and Collins, D. 2001. Analysis of 3D deformation fields for appearance-based segmentation. In *Proceedings of the Conference on Medical Image Computing and Computer-Assisted Intervention*.
- Duchesne, S., Pruessner, J., and Collins, D. 2001. Appearance-based modelling and segmentation of the hippocampus from MR images. In *Proceedings of the 23rd Annual Conference of the IEEE Engineering in Medicine and Biology Society*.
- Haller, J., Christensen, G., Joshi, S., Newcomer, J., Miller, M., Csernansky, J., and Vannier, M. 1996. Hippocampal MR imaging morphometry by means of general pattern matching. *Radiology* **199**: 787–791.
- Hogan, R., Mark, K., Wang, L., Joshi, S., Miller, M., and Bucholz, R. 2000. Mesial temporal sclerosis and temporal lobe epilepsy: MR imaging deformation-based segmentation of the hippocampus in five patients. *Radiology*.
- Jones, M., and Poggio, T. 1998. Multidimensional morphable models. In *Proceedings of the 1998 International Conference on Computer Vision*, pp. 683–688.
- Joshi, S., Pizer, S., Fletcher, P., Thall, A., and Tracton, G. 2001. Multi-scale 3D deformable model segmentation based on medial description. In *Proceedings of the Information Processing in Medical Imaging Conference*, pp. 64–77. IPMI, Springer-Verlag, Berlin/New York.
- Kelemen, A., Szekely, G., and Gerig, G. 1999. Elastic model-based segmentation of 3D neuroradiological data sets. *IEEE Trans. Med. Imaging*.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., and Lancaster, J. 1995. A probabilistic atlas of the human brain: Theory and rationale for its development. *NeuroImage* **2**: 89–101.
- Pham, D., Xu, C., and Prince, J. 2000. Current methods in medical image segmentation. *Annu. Rev. Biomed. Eng.*, **2**: 315–337.
- Pizer, S., Fletcher, P., Fridman, Y., Fritsch, D., Gash, A., Glotzer, J., Joshi, S., Thall, A., Tracton, G., Yushkevich, P., and Chaney, E. 2001. Deformable M-Reps for 3D medical image segmentation. *Med. Image Anal.*
- Pruessner, J., Collins, D., Pruessner, M., and Evans, A. 2001. Age and gender predict volume decline in the anterior and posterior hippocampus in early adulthood. *J. Neurosci.* **21**: 194–200.
- Pruessner, J., Li, L., Serles, W., Pruessner, M., Collins, D., Kabani, N., Lupien, S., and Evans, A. 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cereb. Cortex* **10**: 433–442.

- Rueckert, D., Frangi, A., and Schnabel, J. 2001. Automatic construction of 3D statistical deformation models using non-rigid registration. In *Proceedings of the 2001 Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 77–84.
- Sled, J., Zijdenbos, A., and Evans, A. 1998. A nonparametric method of automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imaging* **17**: 87–97.
- Styner, M., and Gerig, G. 2001. Medial models incorporating object variability for 3D shape analysis. In *Proceedings of the Information Processing in Medical Imaging Conference*, pp. 502–516. IPMI, Springer-Verlag, Berlin/New York.
- Thirion, J. 1996. Non-rigid matching using demons. In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition*, pp. 245–251. CVPR.
- Webb, J., Guimond, A., Elderidge, P., Chadwick, D., Meunier, J., Thirion, J.-P., and Roberts, N. 1999. Automatic detection of hippocampal atrophy on magnetic resonance images. *Magn. Reson. Imaging* **17**: 1149–1161.
- Zijdenbos, A., Dawant, B., Margolin, R., and Palmer, A. 1994. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Trans. Med. Imaging* **13**: 716–724.