

# Measuring Size and Shape of the Hippocampus in MR Images Using a Deformable Shape Model

Dinggang Shen<sup>1</sup>, Scott Moffat<sup>2</sup>, Susan M. Resnick<sup>2</sup>, and Christos Davatzikos<sup>1</sup>

<sup>1</sup>Center for Biomedical Image Computing, Department of Radiology  
Johns Hopkins University School of Medicine

<sup>2</sup>Laboratory of Personality and Cognition, National Institute on Aging

## Abstract

A method for segmentation and quantification of the shape and size of the hippocampus is proposed, based on an automated image analysis algorithm. The algorithm uses a deformable shape model to locate the hippocampus in magnetic resonance images and to determine a geometric representation of its boundary. The deformable model combines three types of information. First, it employs information about the geometric properties of the hippocampal boundary, from a local and relatively finer scale to a more global and relatively coarser scale. Second, the model includes a statistical characterization of normal shape variation across individuals, serving as prior knowledge to the algorithm. Third, the algorithm utilizes a number of manually defined boundary points, which can help guide the model deformation to the appropriate boundaries, wherever these boundaries are weak or not clearly defined in MR images. Excellent agreement is demonstrated between the algorithm and manual segmentations by well-trained raters, with a correlation coefficient equal to 0.97 and algorithm/rater differences statistically equivalent to inter-rater differences for manual definitions.

## Introduction

The hippocampus is a critical structure of the human limbic system, which plays an essential role in learning and memory processing. In particular, researchers have shown that abnormalities in the volume and architecture of the hippocampus are associated with a number of neurological and psychiatric illnesses, including Alzheimer's disease, epilepsy, and schizophrenia [Jack 1994a, Jack 1994b, Dean 1996]. Since the early 1990's, Magnetic Resonance Imaging (MRI) has been used to produce accurate hippocampal volume measurements, by separating hippocampal structures not only from the surrounding white matter (WM), but also from contiguous areas of gray matter (GM).

The hippocampus is a small GM structure that is adjacent to other GM structures (e.g. amygdala, parahippocampal gyrus). This feature of the hippocampus means that in Magnetic Resonance (MR) brain images, the hippocampus has relatively low contrast and no distinguishable boundaries along significant portions of its surface. Moreover, the hippocampus has a high surface to volume ratio, and the hippocampal

surface voxels are most difficult to define. These difficulties complicate the accurate automatic segmentation of the hippocampus. To date, three different approaches have been proposed for quantification of hippocampal volumes: manual segmentation, fully automatic, and semi-automatic methods.

In manual segmentation, an expert rater identifies and labels the hippocampus on each slice of a volumetric MR brain image in order to obtain a 3-D reconstruction of the hippocampus. This method requires extensive human interaction and considerable training, and both intra-rater reproducibility and inter-rater reliability may be difficult to achieve.

A fully automatic method is ideal, but the segmentation of the hippocampi with conventional methods, such as edge tracking, thresholding, or region growing, is not reliable, due to the small size, low contrast, and apparent discontinuity of the edges of the hippocampus. One attempt proposed for automatic segmentation involves warping of an atlas to the individual MR image [Webb 1999]. However, when the internal object is relatively small and its shape is highly variable, as is the case for the hippocampus, this approach may not generate an accurate result, due to its sensitivity to the imperfections involved with the registration and warping steps. Alternatively, a deformable model [Kelemen 1999] has been proposed for the segmentation of the hippocampus. This method uses shape and boundary profile information to guide the segmentation procedure of hippocampus.

Semi-automatic methods may provide a more realistic approach for hippocampal segmentation because they combine the automatic techniques with *a priori* operator knowledge of the hippocampal location, anatomical boundaries and shape [Ashton 1997, Hogan 1999, Hogan 2000, Ghanei 1998, Schnabel 1999]. All of these methods use the powerful and flexible capability of deformable models [McInerney 1996, Staib 1992]. For example, in [Ashton 1997], a semi-automated feature extraction algorithm is used for the extraction and measurement of the hippocampus from volumetric MR brain images, including both deformable model and region growing techniques, as well as *a priori* operator knowledge of the hippocampal location and shape.

Notable is the method presented in [Csernansky 1998, Haller 1997, Miller 1997, Joshi 1997], where a high-dimensional fluid transformation is used to warp a template of the hippocampus and the surrounding anatomy to an individual MR image, constrained by a number of user-defined points that lie on the hippocampal boundary. Although this method has been shown to achieve remarkable accuracy in segmenting the hippocampus, it can be computationally very demanding. Most importantly, it is based on matching image intensities and not shape characteristics; as a result the definition of the structure can potentially flow to other regions that are not necessarily anatomical counterparts, but merely have similar image intensities. Finally, this approach assumes that the template has the same signal characteristics as the images to be segmented, which makes its use difficult across modalities and for aging or diseased populations, in which MR signal characteristics can vary substantially.

We have developed a deformable shape modeling framework, for the segmentation and reconstruction of anatomical shapes, and for determining morphology-based correspondence across individuals from tomographic images [Shen 2000, Shen 2001]. By the term *correspondence* here we mean a map that associates each point in a template of the hippocampal anatomy with a point in an individual’s image. Establishing accurate correspondences is relatively less important in segmentation and size estimation of the hippocampus. However, it is very important in quantifying the shape characteristics of the structure. Our framework is based on an Adaptive-Focus Deformable Model (AFDM). In AFDM, for a given structure or set of structures, a shape model is first constructed to represent a typical shape of these structures. This shape model includes two kinds of information: information about the geometry of the structures [Ip 1998] and information about the statistical variation of these structures within a given population [Chen 1999, Cootes 1995, Duta 1998, Wang 2000]. This anatomical model represents the prior knowledge that helps identify the structure in an individual image. In particular, in order to segment a given image, the deformable shape model is placed close to the structure of interest and is subsequently allowed to deform according to features extracted from the images, seeking objects that have similar geometry, but also objects that fall within the expected range of shape variation. In this paper, we use AFDM to segment the hippocampus from the volumetric MR brain images, with the help of the manual landmarks. This approach reduces the required user interaction significantly, relative to manual drawing of the whole structure, thereby saving time and improving accuracy and reproducibility. The landmarks are used for initializing the deformable model and for guiding its deformation process.

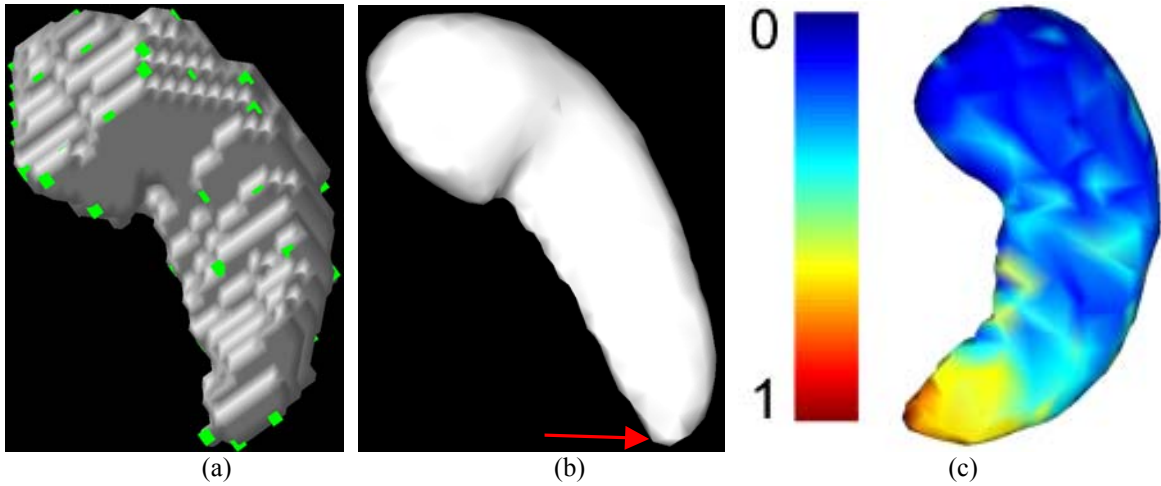
Compared to other published semi-automated methods for hippocampal segmentation, our approach has several advantages. First, the attribute vectors, in conjunction with the deformation mechanism of the model, warranty that the model not only deforms to nearby edges, as is customary in most deformable surface models, but also that it determines point correspondences based on geometric similarity at different scales. Second, there is no strict requirement on the placement of landmarks, which are used to guide the model deformation to the appropriate hippocampal boundaries. The landmarks are placed on the boundaries of the left and right hippocampi, but they do not need to define anatomical homologies (correspondence). Third, like other methods, our model includes a statistical characterization of normal shape variation across individuals, serving as prior knowledge to the algorithm. But, the shape of the deformable model is not fully constrained by statistical information of normal hippocampi, which enables our model to segment the hippocampus that is not seen in the training set, i.e. hippocampus undergoing large shape deformations due to moderate Alzheimer’s disease. In particular, the shape of the deformable model is initially constrained more by statistical information of normal hippocampi, and is relatively less constrained later on.

We present results based on images from the neuroimaging study of the Baltimore Longitudinal Study of Aging [Resnick 2000], where MR images are collected over a 9-year period from elderly subjects aged 55-85. In order to validate our methodology, we compare the volumes and overlap errors of our segmentation algorithm with

differences between manually defined anatomical measurements by two raters that were trained extensively to achieve high intra- and inter-rater agreement. The results indicate that the performance of our algorithm is equivalent to manual segmentation, in that the algorithm/rater agreement is comparable to the inter-rater agreement. Moreover, it significantly reduces human rater time, including extensive time for training, and allows for shape analysis in addition to volume measurements.

## Deformable hippocampal surface model

In the following sections, we first describe the procedures by which we generate a 3D hippocampal surface model. Then, we introduce the concept of *attribute vectors*, which reflect shape characteristics of the hippocampus from a local and fine scale to global and coarser scale. We also describe the placement of landmarks, whose goal is to assist the automated surface model warping procedure to avoid being trapped by spurious edges or skipping through weak or non-existing edges. Finally, we design an energy function that incorporates attribute vector similarity between the model and the individual, as well as conformation to the manually defined landmarks. This function is designed to be at a minimum when the hippocampal surface model is warped into perfect alignment with the subject's hippocampus, thus providing a shape representation of the subject's hippocampus.



**Figure 1.** Generating the hippocampal surface model from the manually segmented hippocampi. (a) A 3D rendering of an expert's manual segmentation of the left hippocampus, (b) its surface representation, (c) the degree of similarity between the attribute vector of the end point (indicated by the red arrow in (b)) and the attribute vectors of all other hippocampal points. The similarity degrees are normalized between 0 and 1, with red and blue colors respectively denoting the similarity degrees 1 and 0, as shown in the color bar. Colors for other similarity degrees are provided in the color bar. This figure demonstrates that the attribute vectors uniquely characterize certain parts of the hippocampal boundary, based on their geometric structure, and therefore help establish anatomically meaningful correspondences based on attribute vector similarity.

Generation of a hippocampal surface model. Our hippocampal model is a surface model including both the left and the right hippocampi. This model is represented by a triangular mesh with a total of 906 vertices and 1804 triangles and was generated from a single manually labeled image. In Fig. 1, we display one view of the left hippocampus (Fig. 1a) and the corresponding surface model (Fig. 1b). The green points in Fig. 1a are the landmarks, which will be described below.

Affine-invariant attribute vector on each model point. To capture the geometry of anatomical structures in a hierarchical fashion, we introduced the concept of an attribute vector that is attached to each point of the hippocampal surface model and reflects the geometric structure of the model from a global scale to a local scale. Local scale attributes reflect differential geometric characteristics of the underlying structure, such as curvature, while global scale attributes capture spatial relationships among distant model points. It is not hard to see that the attribute vector corresponding to, say, a high-curvature region is completely different from attribute vectors of flat segments of the surface. More importantly, even vertices of similar curvatures might have very different attribute vectors, depending on the number of neighborhood layers, or equivalently the number of components of the attribute vector. Each attribute is defined as the volume of a tetrahedron, formed by a model point  $V_i$  and three other points in its certain neighborhood layer; the distance from  $V_i$  to these other points determines whether the corresponding attribute is local or global. For each vertex  $V_i$  in the model, the first *neighborhood layer* includes its immediate neighbors. The second neighborhood layer includes the immediate neighbors of the immediate neighbors, and so forth. The neighborhood layers are constructed so that no vertex is repeated twice in the neighborhood of another vertex. For more detailed meanings of attribute vectors, please refer to [Shen 2000, Shen 2001].

For each model point  $V_i$ , the volumes that are calculated from different neighborhood layers are stacked into an attribute vector  $F(V_i)$ , which can be made affine-invariant as  $\hat{F}(V_i)$  by normalizing it over the whole model [Shen 2000, Shen 2001]. The attribute vectors are an important aspect of our hippocampal model, since they provide a means for finding correspondences across individuals by examining the similarity of the underlying attribute vectors. For example, a point along the anterior-most end of the hippocampus, indicated by the red arrow in Fig. 1b, has very different attribute vector from any other point along the hippocampal boundary. Fig. 1c demonstrates this feature. The similarity of attribute vectors is normalized between 0 and 1, and it is shown color-coded in Fig 1c, with high similarity as red. Other similarity degrees are color-coded according to the color bar in Fig 1c.

Manual hippocampal measurement and placement of landmarks. Manual segmentation of the hippocampus was performed by an experienced rater. Measurements of hippocampal volumes were performed according to criteria similar to [Pruessner 2000] and in consultation with neuroanatomic atlases [Duvernoy 1998, Mai 1997]. Briefly,

hippocampal volumes included the hippocampus proper (cornu ammonis regions), the dentate gyrus, the alveus, the subiculum and the fimbria. The parahippocampal gyrus, entorhinal cortex and fornix were not included. The hippocampus extended anteriorly to the amygdala. The hippocampal-amygdalar boundary was identified with reference to the temporal horn of the ventricles and, where possible, the presence of the alveus. Posteriorly, the hippocampal tail extended to the grey matter within the trigone of the lateral ventricle, ventral to the splenium of the corpus callosum. All manual hippocampal volume measurements and landmark placements were performed using “Display” neuroimaging software (McConnell Brain Imaging Center, Montreal Neurological Institute and Hospital, Montreal Quebec, H3A 2B4).

We have previously applied AFDM to a variety of brain and non-brain anatomical structures, in a largely or completely automated way. However, since the hippocampus is a small structure, with surrounding edges and boundaries that may be confusing to an automated algorithm, the placement of landmarks that guide the deformation of the model to a subject’s image is particularly important for accurate segmentation. The anatomically meaningful landmarks definitely add precision to the method. But it implies a relatively high human burden. The goal of our approach here is to reduce human burden in placing landmarks. In our approach, the number of the landmarks is flexible, and the correspondences among the landmarks are not required. In particular, these landmarks are placed on the boundaries of the left and right hippocampi, but they do not need to define anatomical homologies (correspondence). The general procedure for the placement of landmarks was to place 4-5 points on the outer edges of every fourth slice through both the coronal and sagittal sections of the hippocampus. Some care was taken to ensure placement of several landmarks at the hippocampal-amygdalar boundary and along the region of the subiculum-parahippocampal gyrus. This resulted in the placement of approximately 200 landmarks. Since the hippocampal boundary is very pixilated and uneven, more landmarks usually produce a better segmentation result. It has been proved in our experiments next. Approximately 20 minutes are required for a skilled user to place the landmarks in both the left and the right hippocampi. The green points in Fig. 1a show the positions of the landmarks. These landmarks are incorporated into the definition of the energy function of our hippocampal model, and therefore they guide the model’s deformation to the boundaries of the subject’s hippocampi.

*Energy function of hippocampal surface model.* The deformation of the shape model of the hippocampus is posed, as customary in the deformable model literature, as an optimization problem. The energy function to be optimized reflects the tendency of the deformable model to adapt to the individual shape, while maintaining the *a priori* known shape characteristics of the hippocampus. Three terms compose this energy function, as described next. The first term favors shapes that have attribute vectors similar to the model. The second term demands that the deformable model adhere to edges extracted from the MR images. Finally, the third term demands that the reconstructed hippocampus pass through the landmarks that are defined manually. The

solution is found iteratively, and it is a trade-off between the three aforementioned requirements. More specifically, the energy function being optimized is defined as:

$$E = \sum_{i=1}^N (E_i^{\text{model}} + \alpha_i E_i^{\text{data}}) + \sum_{j=1}^L \beta_j E_j^{\text{landmark}}, \quad (1)$$

where  $N$  is the number of the vertices in the hippocampal model and  $L$  is the number of the landmarks placed on the boundaries of the hippocampi under investigation.  $\alpha_i$  and  $\beta_j$  are the parameters to control the relative importance of each energy term.

The model energy term  $E_i^{\text{model}}$  is defined to allow our hippocampal model to determine correspondences, in addition to segmenting the hippocampal structures. In particular, the model energy term  $E_i^{\text{model}}$  is defined as the difference between the attribute vectors of the model and its deformed configuration on the vertex  $V_i$ , and it is given by

$$E_i^{\text{model}} = \|\hat{F}^{\text{Def}}(V_i) - \hat{F}^{\text{Mdl}}(V_i)\|^2,$$

where  $\hat{F}^{\text{Def}}(V_i)$  and  $\hat{F}^{\text{Mdl}}(V_i)$  are, respectively, the normalized attribute vectors of the deformed model configuration and the model at the point  $V_i$ . The data energy term  $E_i^{\text{data}}$  is designed to move the deformable model towards an object boundary. Since our deformation mechanism deforms a surface segment around each model point  $V_i$  at a time, we design a data energy term  $E_i^{\text{data}}$  that reflects the fit of the whole surface segment, rather than a single model point, with image-derived edges [Shen 2001].

$E_j^{\text{landmark}}$  denotes an energy term of the  $j$ -th landmark  $P_j$ . It is designed as a spring energy that is defined to be proportional to the distance between the landmark  $P_j$  and a selected model point  $U_j$ :

$$E_j^{\text{landmark}} = \|P_j - U_j\|^2.$$

Spring energy  $E_j^{\text{landmark}}$  acts to pull the hippocampal model toward  $P_j$ . The further away the model is from  $P_j$ , the greater the spring energy  $E_j^{\text{landmark}}$ . The model point  $U_j$  is selected by finding the closest point in the hippocampal model to the landmark  $P_j$  using a heuristic search. Here, the selected model point  $U_j$  belongs to the set of the model points, i.e.  $U_j \in \{V_i, 0 \leq i < N\}$ .

Local deformation mechanism on the surface segments of the hippocampal model. Once a deformable shape model has been constructed and its initial configuration has been defined, the process of object detection becomes that of minimizing the energy function in (1). In the literature, many different approaches for minimizing similar energy functions have been reported, i.e. finite difference methods, finite element methods, dynamic programming, greedy algorithms, and neural networks [McInerney, 1996]. All of these approaches

have trade-offs between the quality of the solution and the computational complexity. In this paper, we will use the greedy algorithm [Williams 1992] as an optimization technique to minimize the energy function in equation (1), which has been found to be one of the most reliable, fast, and flexible local strategies. It is customary in deformable model algorithms to move each model point to the position in its neighboring region that makes the energy function minimal. This approach is often problematic, as it can result in unrealistic “pulling” or “pushing” of the model on individual points. This problem is even greater under the presence of spurious or missing edges. To overcome this problem, AFDM deforms surface segments instead of individual points, which also greatly helps avoid local minima. The size of the surface segment decreases progressively. Thus, AFDM captures progressively finer details of the hippocampal shape in a hierarchical fashion.

*Shape statistical information of the hippocampi.* The deformation process described above can potentially warp the hippocampal model to a configuration that does not look like a hippocampus. This can happen, for example, if noisy or adjacent edges attract the model, which tends to seek edges. In order to avoid this undesirable situation, we use a statistical prior that constrains the deformation of the model according to a range of normal shape variation [Cootes 1995]. This range is determined from a *training set*, i.e. a set in which the hippocampus is manually traced. Based on the hand-labeled images, AFDM reconstructs the shapes of the hippocampi in the training samples and determines correspondences. These correspondences are then used to determine a multivariate Gaussian distribution describing the expected range of shape variation. With a robust training algorithm, more training samples usually lead to better shape model. But in the reality, it’s impossible to encompass all types of training samples. In this way, we hierarchically include statistical shape information in the deformation procedure, i.e. constraining the shape of the deformable model strictly in the initial stages and relaxing it in the final stages.

In the deformation process, if the deformable hippocampal model falls outside the range of normal shape variation, it is projected back into the 95% confidence interval of the underlying Gaussian distribution. Since this projection takes place after the iterations of the numerical optimization procedure, it insures that the resulting segmented hippocampi “look like hippocampi”.

## **Detailed implementation of AFDM**

Segmentation of the hippocampus is carried out in two steps: the first step is manual, and the second step is fully automatic. In the first step, a user manually draws landmarks on the boundaries of the left and right hippocampi of an individual MR image, following the procedure described above. In the second step, the deformable model is applied to images that have been pre-classified into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) [Goldszal 1998]. The second step consists of three major parts: 1) generation of a good GM edge map by enlarging the size of the volumetric image, 2) automatic initialization of the hippocampal model based



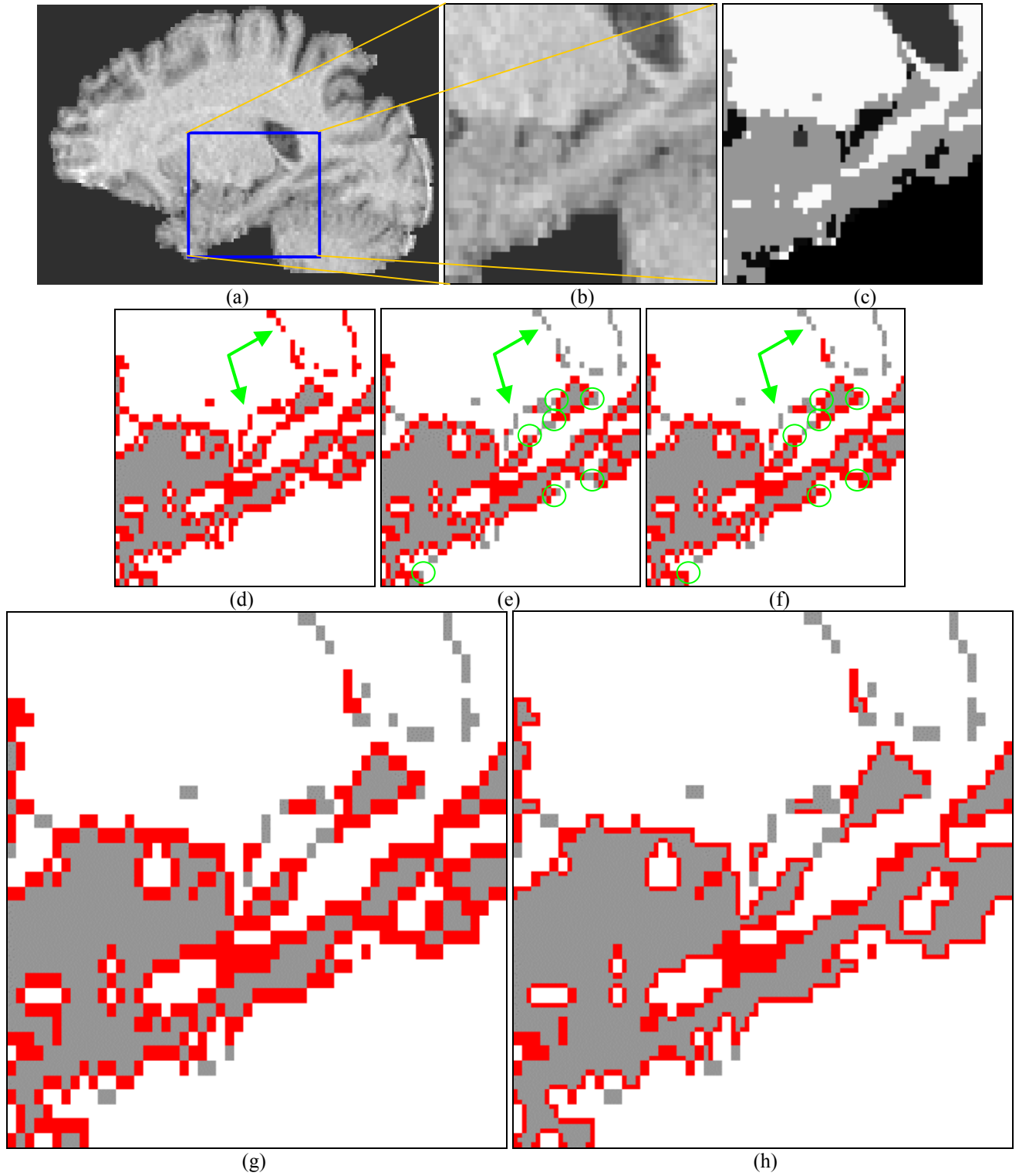
on the landmarks, and 3) deformation of the hippocampal model under the forces from both the GM edge map and the landmarks. These three steps are described in detail in the next section.

Generation of an edge map of GM. AFDM is designed to operate on data that has previously been classified into three tissue types, namely GM, WM, and CSF. Tissue segmentation was completed by a robust tissue segmentation method in [Goldszal 1998], which has been tested on many brain images in our lab, with high success. This method is good enough to be used as a tissue segmentation tool for our approach. Since the hippocampus is almost entirely a GM structure, we calculate the edge map of GM and expect the hippocampal model to finally converge to the desired GM boundaries, driven by forces seeking to minimize the second term in equation (1). The procedure for extracting the GM boundaries involves *three* steps. *First*, for each GM voxel, we check its six neighboring voxels to determine whether it is a GM edge point. In this application, if one of its neighboring voxels is not GM, then this GM voxel is determined as a GM edge point (see Fig. 2d). *Second*, we remove the confusing GM edges in the narrow GM regions that are usually located between CSF and WM (see green arrows in Figs. 2(d-f)). We complete this by directly looking for the GM edge points for which most neighbors are also GM edges. We also use another step to undo the GM edges that are removed from the sharp parts of the large GM regions. This difference is depicted inside of the green circles respectively in Figs. 2e and 2f. *Third*, we enlarge the GM edge map in order to obtain a better and more detailed representation of the GM edges around the hippocampus. Since the hippocampus is an extremely thin structure, GM edges in the hippocampal region typically end up being overly thick relative to the full thickness of the structure (see Fig. 2g). Therefore, a detailed representation of the shape of the hippocampus cannot be obtained, unless a better extraction of edge maps is obtained. To avoid this problem, we increase the resolution of the GM edge map, by upsampling the image to approximately .5mm in each dimension (see Fig. 2g). As a result, the edge maps are thinner and allow for a better shape reconstruction. Fig. 2h shows the edge map of the upsampled GM image. Although upsampling is not an issue for structures that are much larger than the voxel resolution, it is an important issue for the hippocampus since pixelation effects along the periphery of the hippocampus can introduce substantial errors in volume measurements.

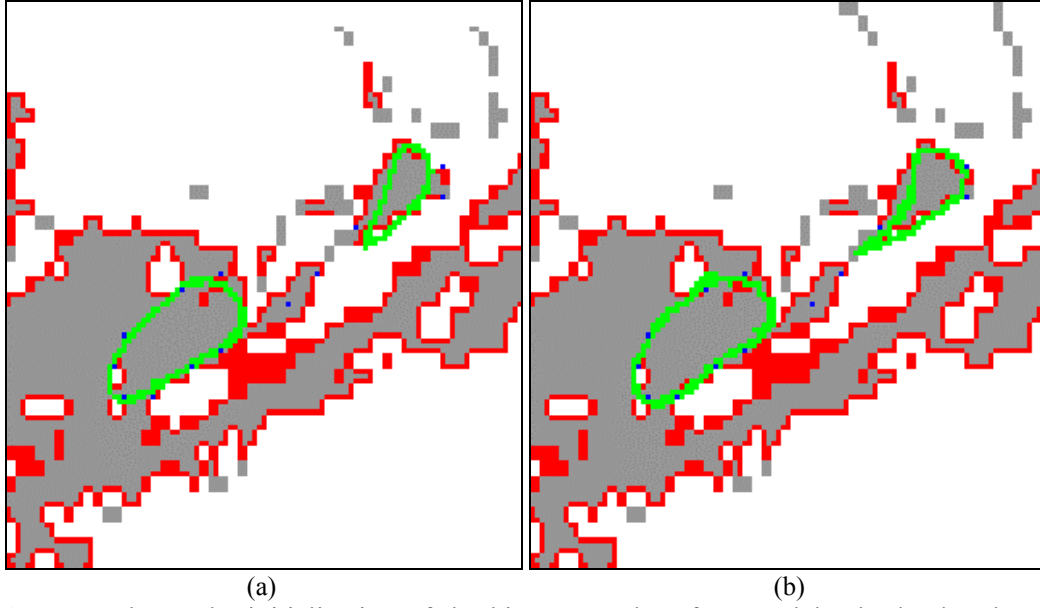
Initialization of the hippocampal surface model. For the algorithm to be effective in finding the hippocampal boundaries, the deformable shape model must be initially placed close to the boundary of interest. This initialization process is guided by the manually placed landmarks. In particular, we perform a series of rigid transformations (translation and rotation) on our surface model and then calculate the total distance of the landmarks to each transformed model. The hippocampal model is finally initialized to a pose that has the minimal total distance. Fig. 3 displays an initialization example, where the green contour denotes the position of the initialized model in this slice and the blue points denote the landmarks in the same slice. The GM edge map that is displayed as red in this figure is the same as the one shown in Fig. 2h.

Deformation procedure of the hippocampal surface model. Starting from the initialization described above, the hippocampal model then automatically deforms to the boundaries of the desired hippocampi under the forces from both the landmarks and the GM edge map. Our model is hierarchically implemented, by using relatively smoother deformations in the beginning, and by gradually capturing the finer details of the hippocampal shape via more localized deformations. This implementation helps preserve the global shape of the model, while allowing the model to adapt to the local shape of the hippocampus. This increases the algorithm's robustness to spurious and missing edges. At each iteration of the deformation procedure, the hippocampal surface model deforms as described next.

- First, for each landmark, a heuristic search method is used to find the nearest model point on the surface model. In the initial deformation stages, one model point could be selected as a nearest point to a few of the landmarks. That is, multiple landmarks produce the spring forces on the same model point. In this way, the spring energies or spring forces will be summed and applied on this model point.
- Second, for each model point, if it has spring energy, the surface segment around this model point will be deformed only under the forces from its spring energy. For the model point where no spring forces are performed, its surface segments will be deformed only under the forces from GM edge map, if the deformation of this surface segment decreases a certain amount of energy. Otherwise, the position of this model point just follows the deformations of the model points around it. In other words, manually placed landmarks supercede edge maps extracted from the images.
- Third, statistical information is used to refine the tentative hippocampal shapes obtained after each iteration of the deformation mechanism. If the deformable hippocampal model happens to fall outside the range of normal shape variation, it is projected back into the 95% confidence interval of the underlying Gaussian distribution. Since our current set of the training hippocampal samples is very small, the range of normal shape variation estimated from the samples is only considered as approximate, and therefore used only as a partial constraint to refine the shape of the deformable model. That is, statistical information is used to constrain the shape of the deformable model strictly in the initial deformation stages, and relatively less later on.



**Figure 2.** Calculating a good GM edge map from the tissue labeled image. (a) A slice of the original gray-level MR image, (b) a cropped and zoomed image from (a), (c) tissue segmentation result, with the white as WM, the grey as GM, and the black as CSF and background, (d) GM and its edge points (red), (e) a GM edge map with the edges on narrow GM removed, (f) a corrected GM edge map from (e), (g) the enlarged version of (f), and (h) a final GM edge map used to derive our hippocampal surface model.



**Figure 3.** An example on the initialization of the hippocampal surface model. The landmarks are shown as blue, and the hippocampal surface model is displayed as green. (a) The initialization of the model in a slice, (b) an intermediate state of the model in the deformation procedure.

## Results

In the following, we provide a set of experiments to evaluate the performance of our algorithm in segmenting the hippocampi, by calculating the volume and the overlap errors between our segmentation and the manually defined gold standard. The results show that the performance of our algorithm is statistically equivalent to manual segmentation. All subjects used in this study were scanned with a GE Signa 1.5 T scanner (GE Medical Systems, Waukesha, WI) employing a T1-weighted Spoiled GRASS (SPGR) pulse sequence. The parameters of the SPGR sequence were TR = 35 ms, TE = 5 ms, flip angle =  $45^\circ$ , image size =  $256 \times 256 \times 124$  voxels, and voxel size =  $0.9375 \times 0.9375 \times 1.5 \text{ mm}^3$ .

*Performance of AFDM in hippocampus segmentation.* We evaluate our algorithm’s performance in the segmentation of the hippocampi from the volumetric MR brain images by calculating the volume error and the overlap error between AFDM segmentations and manually defined gold standard (hippocampi that are carefully identified by experts). As shown in Table 1, we found a mean volume error of 2.9% and a mean overlay error of 11.5% for our algorithm in comparison to the expert rater. By comparison, two expert raters, using manual hippocampal definitions with a slightly different hippocampal definition excluding the subiculum, achieved a volume error of 2.6% and an overlap error of 13.8%. In addition, we evaluated the performance of our algorithm as a function of the number of landmarks used. We selected three cases, for which we used all of about 200 landmarks, then 50% of these landmarks, and finally only 25% of these landmarks. We selected the subset of the landmarks (i.e. 25% and 50%) by simply removing the landmarks from the original set of the landmarks (i.e. 100%).

A qualitative evaluation of our algorithm is shown in Fig. 4, which demonstrates a good agreement between the AFDM result and the manual tracing. This example is selected at random. In all subfigures, the green contours denote the boundaries of the algorithm-segmented hippocampi in representative slices. In order to view the overall overlay of our segmentation with the manual segmentation, in Fig. 5 we provide a 3D rendering of these two superimposed segmentations. The algorithm's segmentation is displayed as red, and the manual segmentation is displayed as green. Yellow is the overlay between these two. Fig. 5a shows the initialization of the hippocampal model, with the manual segmentation of the hippocampus. Fig. 5b gives the final deformed position and shape of the hippocampal model, overlaid on the manual segmentation of the same hippocampus. In Fig. 5b, the segmentation result of the algorithm shows a good agreement with the manual segmentation.

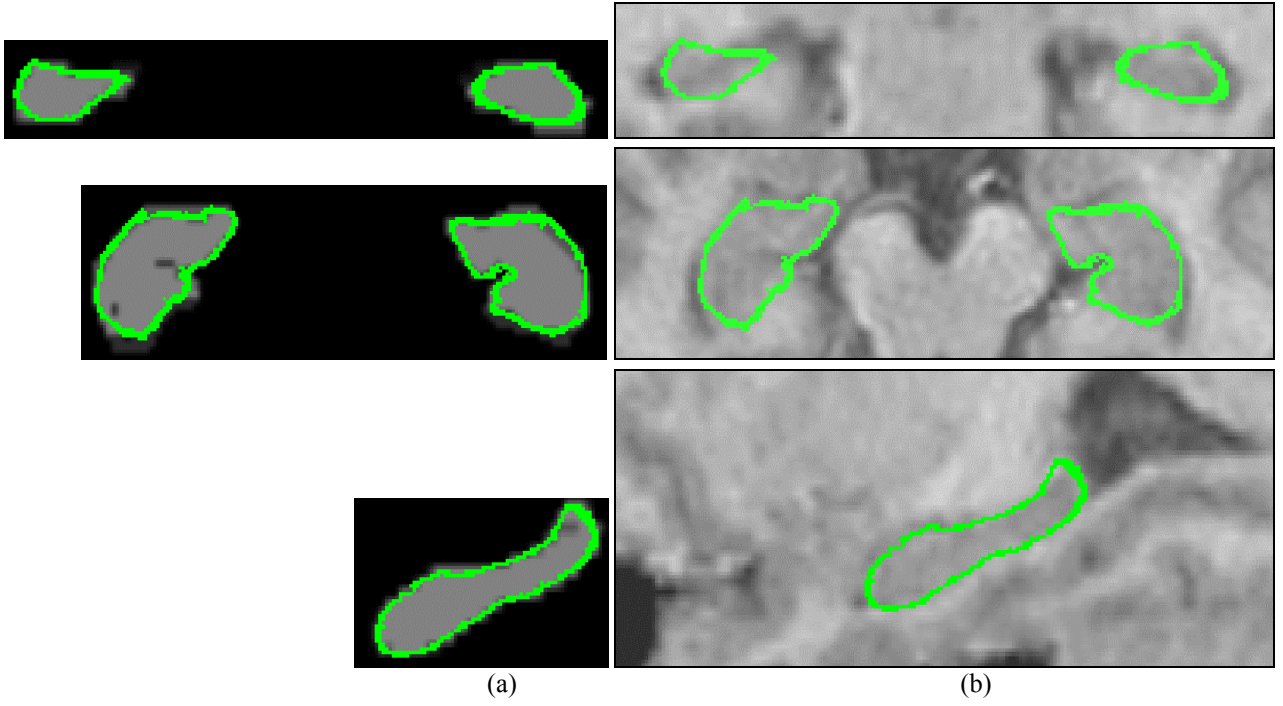
Fig. 6a displays the volumes in  $\text{mm}^3$  of the left and right hippocampi (combined) for the 10 subjects, with the horizontal axis plotting the volume of the manually defined gold standard and the vertical axis plotting the volume of the hippocampus segmented by AFDM. The correlation coefficient between the proposed method and the manually defined gold standard is 0.97. The almost symmetric distribution of the data points around the diagonal line illustrates that AFDM does not over-segment or under-segment the hippocampi when all landmarks are used. Fig. 6d displays the volume and the overlap errors of our algorithm in the segmentation of the hippocampi for each of the 10 subjects, individually. The actual error percentages of our algorithm in Fig. 6d are also listed in Table 1, with the mean volume error 2.9% and the mean overlap error 11.5%.

Figs. 6b and 6e display the analogous results to those in Figs. 6a and 6d, using 50% of the landmarks. From Table 1, we find that the mean volume error increased to 4.1% when only half of the landmarks were used, compared to 2.9% when using 100% of the landmarks, while the mean overlap error remained about the same. Fig. 6c and 6f shows the results when only 25% of the landmarks were used. Similarly, the mean volume error increased, while the mean overlap error changed only slightly. However, Fig. 6c clearly demonstrates a bias toward an over-estimation of the hippocampus volume using our method following a reduction in the number of landmarks to 25% of the original, despite the high correlation of 0.98. This bias is due primarily to pixelation effects, because our geometric model is a smooth surface. Hence, unless adequately constrained by landmarks, the algorithm tends to smooth pixelation effects and overestimate hippocampal volume by bridging between adjacent GM voxels and thus including some WM voxels. This bias can be easily removed numerically if we were only to include voxels classified as GM in the volumetric calculations. However, our definition of the hippocampus includes some WM regions (i.e. alveus, fimbria), and therefore masking with the GM classification is not possible, unless a slightly different ROI is defined that includes only hippocampal GM.

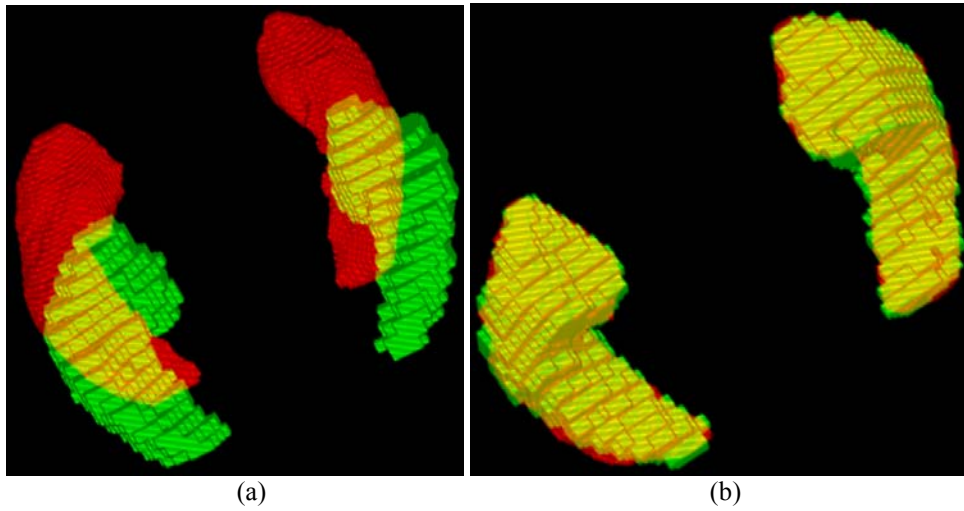
The volume and overlap errors of the two well-trained raters are respectively displayed as two curves in Fig. 6g and also listed in Table 1. The volume errors are ranging from 0.1% to 7.7%, with the mean 2.6%. The overlap

errors range from 11.9% to 16.0%, with the mean 13.8%. Notice here, the maximal overlap error of our algorithm is only 14.0%, regardless of whether we use 100%, 50%, or 25% of the original landmarks.

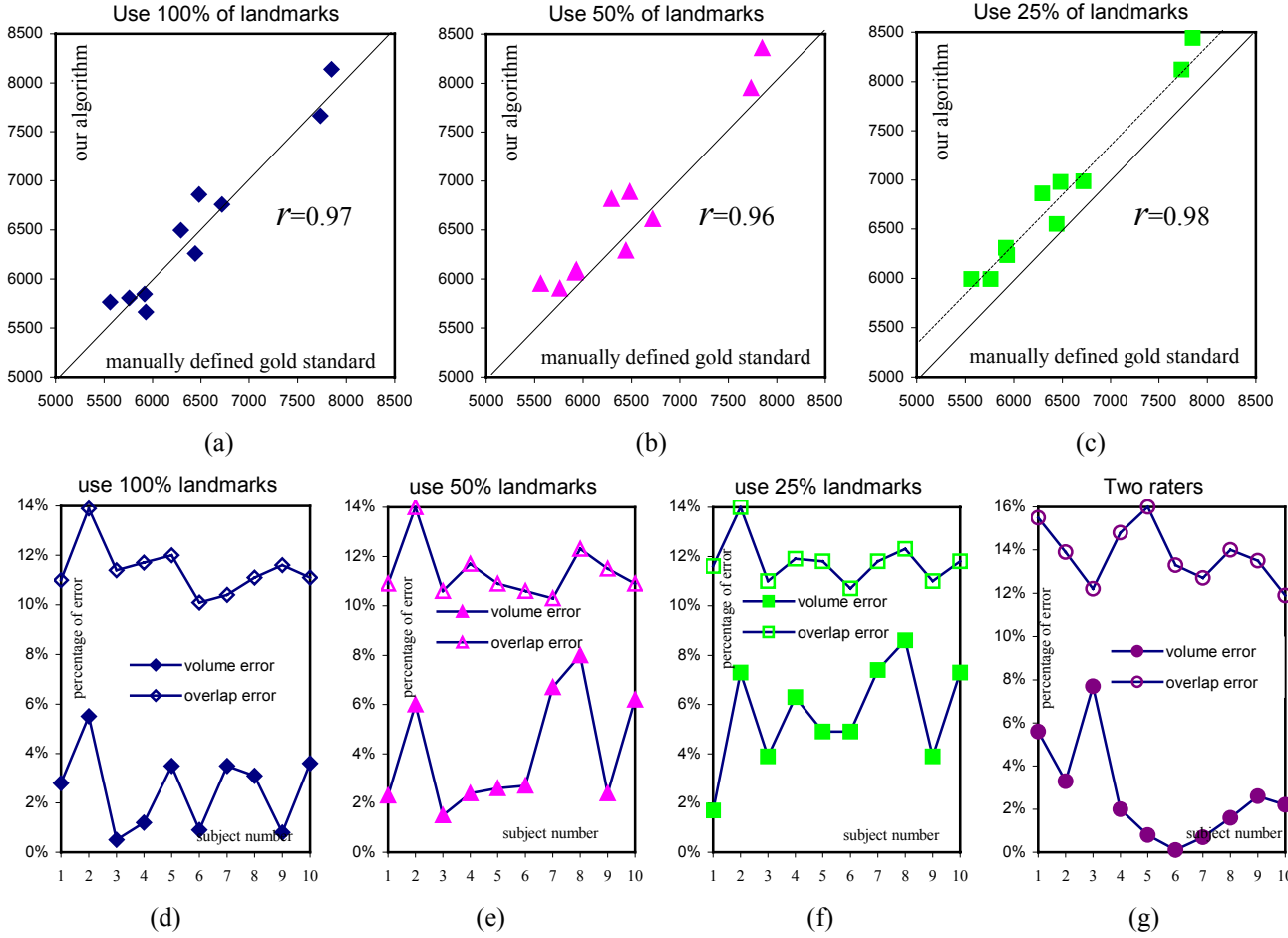
In order to test the reproducibility of our algorithm, we designed two separate experiments. In the first experiment, we randomly selected a subject, and compared the segmentation results of the same subject respectively from the different initializations of the hippocampal model. By using the manual landmarks for the selected subject, our algorithm automatically estimated a good initialization for the hippocampal model, and then segmented the hippocampi from this subject. The volume error and the overlap error between our segmentation and expert's segmentation are respectively 0.5% and 11.4%. For evaluating the robustness of our algorithm, we shifted the hippocampal model from the automatically estimated initialization, and then deformed the shifted model to the hippocampi of the subject. Each shifting results in a separate hippocampus segmentation on the same selected subject. By calculating the volume error and the overlap error for each segmentation, we can observe the robustness of our algorithm. Table 2 gives the volume errors and the overlap errors, with the amounts of shifts respectively along  $x$ ,  $y$ , and  $z$  axes. For example, when the automatically initialized model was shifted at the  $z$  axis for -5 voxels, the resulted segmentation has volume error 0.95% and overlap error 11.2%. From the mean volume errors and the mean overlap errors in the last row of Table 2, it's easy to conclude that our algorithm has high reproducibility in segmenting hippocampi. In the second experiment, we test the reproducibility of our model to the placement of the landmarks. In this experiment, we select the first 5 subjects from the 10 subjects that have been used in Table 1. One rater placed the landmarks again for each of these 5 selected subjects, which are called the second placements of the landmarks. The definition for the second placements of the landmarks is a little different from the first placements, and for each subject the total landmarks in the second placement are 15% less than that in the first placement. Table 3 compares the results from these two different placements of the landmarks. As to the results of the first placements, they are copied from Table 1 for comparison. According to the mean volume errors and the mean overlap errors, we can conclude that the reproducibility of our algorithm is satisfactory.



**Figure 4.** An example of hippocampal segmentation using the AFDM algorithm. Final segmentation (green) is overlaid on (a) the manual segmentation result, and (b) the original images. The slice numbers of the images in (a) and (b) are the same.



**Figure 5.** A 3D rendering of the segmentation results in Fig. 4. (a) The model (red) with the carefully manually segmented hippocampi (green), (b) Overlay of the final deformed model with the manually segmented hippocampi.



**Figure 6.** Quantitative evaluation of the proposed algorithm in segmenting the hippocampus. (a-c) The volume of the hippocampus, (d-f) the volume error and the overlap error of the hippocampus segmented by our algorithm and by an expert. (a, d) Using all the landmarks, (b, e) 50% of the landmarks, and (c, f) 25% of the landmarks. (g) The volume error and the overlap error of the two raters. Parameter  $r$  in (a-c) is the correlation coefficient between the proposed method and the manually defined gold standard.

**Table 1** Quantitative evaluation of the performance of our algorithm. Volume and overlap errors are computed between our segmentation and the real hippocampus, and also between two well-trained raters.

Volume error:

Subjects	1	2	3	4	5	6	7	8	9	10	Mean	Corr. Coeff.
100% landmarks	2.8%	5.5%	0.5%	1.2%	3.5%	0.9%	3.5%	3.1%	0.8%	3.6%	<u>2.9%</u>	<u>0.97</u>
50% landmarks	2.3%	6.0%	1.5%	2.4%	2.6%	2.7%	6.7%	8.0%	2.4%	6.2%	<u>4.1%</u>	<u>0.96</u>
25% landmarks	1.7%	7.3%	3.9%	6.3%	4.9%	4.9%	7.4%	8.6%	3.9%	7.3%	<u>5.6%</u>	<u>0.98</u>
Two raters	5.6%	3.3%	7.7%	2.0%	0.8%	0.1%	0.7%	1.6%	2.6%	2.2%	<u>2.6%</u>	<u>0.98</u>

Overlap error:

Subjects	1	2	3	4	5	6	7	8	9	10	Mean
100% landmarks	11%	13.9%	11.4%	11.7%	12.0%	10.1%	10.4%	11.1%	11.6%	11.1%	<u>11.5%</u>
50% landmarks	10.9%	14.0%	10.6%	11.7%	10.9%	10.6%	10.3%	12.3%	11.5%	10.9%	<u>11.4%</u>
25% landmarks	11.6%	14.0%	11.0%	11.9%	11.8%	10.7%	11.8%	12.3%	11.0%	11.8%	<u>11.8%</u>
Two raters	15.5%	13.9%	12.2%	14.8%	16.0%	13.3%	12.7%	14.0%	13.5%	11.9%	<u>13.8%</u>



**Table 2** Reproducibility of our algorithm to the initializations. The automatically initialized model was shifted respectively for different voxels along  $x$ ,  $y$ , and  $z$  axes. The resulted segmentations were evaluated by volume errors and overlap errors. From the mean volume errors and the mean overlap errors in the last row of the Table, we can see the high reproducibility of our algorithm.

	Along $x$ dimension		Along $y$ dimension		Along $z$ dimension	
Number of Voxels Shifted	Volume error	Overlap error	Volume error	Overlap error	Volume error	Overlap error
-5	2.03%	12.4%	0.16%	11.5%	0.95%	11.2%
-4	2.01%	11.5%	1.11%	11.9%	0.75%	11.2%
-3	0.45%	11.4%	0.14%	11.5%	0.88%	11.3%
-2	1.08%	11.3%	0.88%	12.1%	0.78%	11.4%
-1	0.57%	11.4%	0.93%	11.3%	0.36%	11.3%
<u>0</u>	<u>0.50%</u>	<u>11.4%</u>	<u>0.50%</u>	<u>11.4%</u>	<u>0.50%</u>	<u>11.4%</u>
1	0.33%	11.4%	0.53%	11.2%	0.94%	11.4%
2	1.14%	11.3%	0.32%	11.9%	0.26%	12.1%
3	1.07%	11.5%	0.08%	11.5%	0.47%	11.3%
4	0.19%	11.8%	0.08%	11.2%	1.57%	11.4%
5	0.59%	11.2%	1.39%	12.3%	1.38%	11.6%
Mean	0.90%	11.4%	0.55%	11.5%	0.80%	11.3%

**Table 3** Reproducibility of our algorithm to the different placements of the landmarks. For the first 5 subjects in Table 1, the segmentation results from the two different placements of landmarks are compared. According to the volume and overlap errors between our segmentations and the real hippocampi, the reproducibility of our algorithm is satisfactory. The second placements lead to a little bit worse results, since the definition of the second placements of the landmarks is a little different from the first placement, and the total landmarks of each subject in the second placements is about 15% less than that in the first placements.

Volume error:

Subjects	1	2	3	4	5	Mean
The first placement	2.8%	5.5%	0.5%	1.2%	3.5%	<u>2.7%</u>
The second placement	0.87%	0.82%	6.66%	0.38%	6.2%	<u>2.98%</u>

Overlap error:

Subjects	1	2	3	4	5	Mean
The first placement	11%	13.9%	11.4%	11.7%	12.0%	<u>12.0%</u>
The second placement	13.0%	13.9%	11.2%	11.2%	11.9%	<u>12.2%</u>

## Conclusion and future work

In this paper, we described a hippocampal surface model and its validation for segmentation and size estimation of the hippocampi in volumetric MR brain images. This model integrates geometric, statistical, and user-defined information, and segments the hippocampus by deforming a deformable shape model to an individual MR

image. After the manual placement of a number of landmarks, our segmentation procedure is fully automatic; the landmarks simply must be located along the hippocampal surface with some representation along the difficult-to-define boundaries, but need not define correspondences.

From our validation experiments, we found that the mean volume and overlap error of our segmentation with manual segmentation is within the limits of the reliability and accuracy achieved for manual definitions in our laboratory and in the literature generally. Difference between our automatic and manual segmentations was similar to the difference between manual segmentations of two well-trained raters using a slightly different definition of the hippocampus. Therefore, the proposed algorithm can dramatically reduce processing time and increase reproducibility, without compromising accuracy.

The main characteristics of our algorithm, which make its performance robust, are summarized next:

- (1). Each point along the hippocampal model has an attribute vector, which reflects the geometric structure of the model from a local to a global scale. The attribute vectors are a key component of our system, for two reasons. First, they help establish anatomical correspondences between the model and an individual hippocampus, to the extent that such correspondences are reflected by the geometry of the structure. Fig. 1c vividly demonstrates an example of the unique identification of a part of the hippocampus via its attribute vector. Second, they help train the model in the stage in which the normal range of shape variation is determined from a training set, without the need for manual definition of homologies, which would be practically impossible in 3D.
- (2). The statistical prior helps the deformable model stay within the range of shapes that “look like hippocampi”, and thus avoids following incorrect or noisy edges.

The main shortcoming of our approach is the need for manual definition of a number of points along the hippocampal boundary. The very high accuracy required in longitudinal studies necessitates the use of manually placed points. Moreover, our current definition of the hippocampus includes some white matter and runs along boundaries that are not visible in MRI. Therefore, since there is no edge information in those regions, the manual placement of boundary points is necessary.

Some extensions of our methodology are possible. Currently, only the left and right hippocampi are included in our hippocampal model. In order to clearly differentiate the boundaries of the hippocampi in the MR volumetric images, other surrounding structures will be included into the hippocampal model. In this way, information about the relative positions of these structures will help our model in localizing hippocampal boundaries. Moreover, image-derived forces, which guide the deformable model’s deformation, are currently based only on edge maps. Edge maps can be noisy and can contain a lot of nearby extraneous information that can adversely affect the model’s deformation. We will investigate the use of other information in addition to edges. Finally, since our algorithm determines point correspondences, in addition to segmenting the hippocampus, it can be

used in conjunction with our previously published RAVENS method for regional volumetric analysis [Davatzikos 2001, Davatzikos 1998, Goldszal 1998, Davatzikos 2001a].

## Acknowledgements

This work was supported in part by NIH contract NIH-AG-93-07 and NIH grant R01-AG14971.

## References

- [Ashton, 1997] Edward A. Ashton, Kevin J. Parker, Michel J. Berg, Chang Wen Chen. "Novel volumetric feature extraction technique with applications to MR images". *IEEE Transactions on Medical Imaging*, 16(4): 365-371, Aug. 1997.
- [Chen, 1999] M. Chen, T. Kanade, D. Pomerleau, J. Schneider. "3-D deformable registration of medical images using a statistical atlas". *MICCAI*, pages 621-630, Sept. 1999.
- [Cootes, 1995] T.F. Cootes, D. Cooper, C.J. Taylor and J. Graham. "Active shape models-their training and application". *Computer Vision and Image Understanding*, 61(1):38-59, Jan. 1995.
- [Csernansky, 1998] J.G. Csernansky, S. Joshi, L. Wang, J.W. Haller, M. Gado, J.P. Miller, U. Grenander, M.I. Miller. "Hippocampal morphometry in schizophrenia by high dimensional brain mapping". *Proc. Natl. Acad. Sci. USA*, 95 (19): 11406-11411 SEP 15 1998.
- [Davatzikos, 1998] C. Davatzikos. "Mapping of Image Data to Stereotaxic Spaces: Applications to Brain Mapping". *Human Brain Mapping*, 6:334-338, 1998.
- [Davatzikos, 2001] C. Davatzikos, A. Genc, D. Xu, and S.M. Resnick. "Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy", submitted.
- [Davatzikos 2001a] C. Davatzikos. "Measuring biological shape using geometry-based shape transformations". *Image and Vision Computing*, 19:63-74, 2001.
- [Dean, 1996] D. Dean, P. Buckley, F. Bookstein, J. Kamath, D. Kwon, L. Friedman and C. Lys. "Three-dimensional MR-based morphometric comparison of schizophrenic and normal cerebral ventricles". *Vis. in Biom. Computing, Lecture Notes in Comp. Sc.*, p. 363-372, 1996.
- [Duta, 1998] N. Duta, M. Sonka. "Segmentation and interpretation of MR brain images: An improved active shape model". *IEEE Trans. on Medical Imaging*, 17: (6) 1049-1062, Dec. 1998.
- [Duvernoy, 1998] H.M. Duvernoy. *The human hippocampus: Functional anatomy, vascularization and serial sections with MRI*. New York, NY: Springer, 1998.
- [Ghanei, 1998] A. Ghanei, H. Soltanian-Zadeh, J.P. Windham. "A 3D deformable surface model for segmentation of objects from volumetric data in medical images", *Computers in Biology and Medicine*, 28(3):239-253, May 1998.

- [Goldszal, 1998] A.F. Goldszal, C. Davatzikos, D.L. Pham, M.X.H. Yan, R.N. Bryan, S.M. Resnick. "An image-processing system for qualitative and quantitative volumetric analysis of brain images". *Journal of Computer Assisted Tomography*, 22 (5): 827-837, Sept.-Oct. 1998.
- [Haller, 1997] J.W. Haller, A. Banerjee, G.E. Christensen, M. Gado, S. Joshi, M.I. Miller, Y. Sheline, M.W. Vannier, J.G. Csernansky. "Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas". *Radiology*, 202 (2): 504-510, FEB 1997.
- [Hogan, 1999] R.E. Hogan, K.E. Mark, L. Wang, S. Joshi, M.I. Miller, M.E. Bertrand, R.D. Bucholz. "MR imaging deformation-based segmentation of the hippocampus in patients with mesial temporal sclerosis and temporal lobe epilepsy". *Epilepsia*, 40:192-192, Suppl. 7 1999.
- [Hogan, 2000] R.E. Hogan, K.E. Mark, I. Choudhuri, L. Wang, S. Joshi, M.I. Miller, R.D. Bucholz. "Magnetic resonance imaging deformation-based segmentation of the hippocampus in patients with mesial temporal sclerosis and temporal lobe epilepsy". *Journal Of Digital Imaging*, 13 (2): 217-218, Suppl. 1 MAY 2000.
- [Ip, 1998] H.H.S. Ip, D. Shen. "An affine-invariant active contour model (AI-snake) for model-based segmentation". *Image and Vision Computing* 16(2):135-146, 1998.
- [Jack, 1994a] C.R. Jack. "MRI-based hippocampal volume measurements in epilepsy". *Epilepsia*, 35: S21-S29, Suppl. 6 1994.
- [Jack, 1994b] C.R. Jack, M.D. Bentley, C.K. Twomey, *et al.* "MR imaging-based volume measurements of the hippocampal-formation and anterior temporal-lobe-validation studies". *Radiology*, 176(1): 205-209, July 1990.
- [Joshi, 1997] S.C. Joshi, M.I. Miller, U. Grenander. "On the geometry and shape of brain submanifolds". *Intl. J. of Pattern Recognition and Artificial Intelligence*, 11:1317-1343, 1997.
- [Kelemen, 1999] A. Kelemen, G. Szekely, and G. Gerig. "Elastic model-based segmentation of 3-D neuroradiological data sets". *IEEE Trans. on Medical Imaging*, 18(10): 828-839, October 1999.
- [Mai, 1997] J.K. Mai, J. Assheuer, G. Paxino. *Atlas of the human brain*. New York, NY: Academic Press, 1997.
- [McInerney, 1996] T. McInerney and D. Terzopoulos. "Deformable models in medical image analysis: a survey". *Medical Image Analysis*, 1(2): 91-108, 1996.
- [Miller, 1997] M. Miller, A. Banerjee, G. Christensen, S. Joshi, N. Khaneja, U. Grenander, L. Matejic. "Statistical methods in computational anatomy". *Statistical Methods in Medical Research*, 6: 267-299, 1997.
- [Pruessner, 2000] J.C. Pruessner, L.M. Li, W. Series, M. Pruessner, D.L. Collins, N. Kabani, S. Lupien, A.C. Evans. "Volumetry of hippocampus and amygdala with high-resolution MRI and three dimensional analysis software: Minimizing the discrepancies between laboratories". *Cerebral Cortex*: 10:433-442, 2000.
- [Resnick, 2000] S.M. Resnick, A. Goldszal, C. Davatzikos, S. Golski, M.A. Kraut, E.J. Metter, R.N. Bryan, A.B. Zonderman. "One-year age changes in MRI brain volumes in older adults". *Cerebral Cortex*, 10:464-472, 2000.
- [Schnabel, 1999] J.A. Schnabel, L. Lemieux, U.C. Wiesmann, S. R. Arridge. "Measurement of hippocampal volume changes in serial MRI scans". In K. Hanson, editor. *Proc. SPIE Medical Imaging 1999: Image Processing*, vol. 3661-II, pp. 1367-1376, 20-26 February 1999, San Diego, CA. SPIE, 1999.
- [Shen, 2000] D. Shen, C. Davatzikos. "An adaptive-focus deformable model using statistical and geometric information". *IEEE Trans. on PAMI*. 22(8):906-913, August 2000.

- [Shen, 2001] D. Shen, E. Herskovits, and C. Davatzikos. “An adaptive-focus statistical shape model for segmentation and shape modeling of 3D brain structures”. *IEEE Trans. on Medical Imaging*. 20(3), March 2001.
- [Staib 1992] L.H. Staib and J.S. Duncan. “Boundary finding with parametrically deformable models”. *IEEE Trans. on PAMI*, 14(11):1061-1075, 1992.
- [Wang, 2000] Y. Wang and L.H. Staib. “Boundary finding with prior shape and smoothness models”. *IEEE Trans. on PAMI*, 22(7):738-743, 2000.
- [Webb, 1999] J. Webb, A. Guimond, P. Eldridge, D. Chadwick, J. Meunier, J.P. Thirion, N. Roberts. “Automatic detection of hippocampal atrophy on magnetic resonance images”. *Magnetic Resonance Imaging*. 17(8):1149-1161, Oct. 1999.
- [Williams, 1992] D.J. Williams, M. Shah. “A fast algorithm for active contours and curvature estimation”. *Computer Vision, Graphics, Image Processing*, 55:14-26, 1992.