ELSEVIER

Medical Image Analysis 4 (2000) 189–200

MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Segmentation and measurement of brain structures in MRI including confidence bounds ☆

Miguel Ángel González Ballester*, Andrew Zisserman, Michael Brady

Medical Vision Laboratory, Robotics Research Group, Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

Received 18 September 1998; received in revised form 15 February 1999; accepted 30 June 1999

Abstract

The advent of new and improved imaging devices has allowed an impressive increase in the accuracy and precision of MRI acquisitions. However, the volumetric nature of the image formation process implies an inherent uncertainty, known as the partial volume effect, which can be further affected by artifacts such as magnetic inhomogeneities and noise. These degradations seriously challenge the application to MRI of any segmentation method, especially on data sets where the size of the object or effect to be studied is small relative to the voxel size, as is the case in multiple sclerosis and schizophrenia. We develop an approach to this problem by estimating a set of bounds on the spatial location of each organ to be segmented. First, we describe a method for 3D segmentation from voxel data which combines statistical classification and geometry-driven segmentation; then we discuss how the partial volume effect is estimated and object measurements are obtained. A comprehensive validation study and a set of results on clinical applications are also described.

Keywords: Morphometry; Segmentation; Confidence intervals; Partial volume effect; Bias field correction

1. Introduction

Magnetic resonance imaging (MRI) has become one of the most important medical imaging modalities. Its excellent soft tissue contrast and the fact that it is virtually non-invasive mean that it is particularly good for studies involving organs such as the brain (Rinck, 1993). MRI machines have significantly increased in resolution, and acquisition protocols with voxel sizes of 1 mm^3 are widely in use. This not only means that more accurate measurements can be made; but it also opens up the possibility of applying MRI to new research areas where extremely high resolution is needed.

In most cases, in order to analyse MR images, a crucial first step is to segment the object of interest. The segmentation of three-dimensional structures is a broad subject in itself, and a fundamental problem in medical vision research. Many methods have been proposed and these can be broadly classified as: voxel-based (Röll et al., 1994), slice-based (Kass et al., 1987; Blake et al., 1993; Marais et al., 1996), and 3D surface methods (Delingette, 1994; McInerney and Terzopoulos, 1996; Székely et al., 1996). It is, however, noticeable that few methods provide a measure of confidence with which the result of the segmentation is obtained. This is a serious issue, not least because MRI acquisitions suffer from what is known as the *partial volume effect* (PVE). This derives from the fact that the intensity value assigned to a voxel is the average of the contributions of all the tissues present in the corresponding volume. Therefore, there exists an inherent uncertainty in any segmentation derived from an MR image that is obviously related to the image resolution. We argue that it is important to quantify this uncertainty and to establish confidence intervals on every measurement derived from an MR volume.

☆ A preliminary version of the article was presented at the Workshop on Biomedical Image Analysis held in Santa Barbara (USA) in June 1998.

*Corresponding author. Present address: Creare Toshiba Nasu, room no. 120, 1–1384–2 Wakakusa, Otawara City, Tochigi, Japan 324-0021. E-mail: miguel.gonzalez@glb.toshiba.co.jp.

E-mail address: magb@robots.ox.ac.uk (M.Á. González Ballester).

1361-8415/00/$ – see front matter
PII: S1361-8415(00)00013-X

In fact, some clinical applications rely fundamentally both on the accuracy of the imaging technique and the subsequent segmentation method. This is particularly the case when the size of the effect to be studied is small relative to the voxel size. Such is the case, for example, in multiple sclerosis (MS) and schizophrenia studies. MS lesions must be quantified with high precision in order to analyse accurately the temporal evolution of the disease or the effect of a certain treatment (Colchester et al., 1997). In the case of schizophrenia, there is a somewhat controversial literature surrounding the contention that the brains of schizophrenic patients are significantly more symmetric (loss of normal asymmetry) than those of normal individuals, and that the difference stems from a disruption to the normal process of lateral dominance during puberty (Crow, 1990; Delisi et al., 1995). The differences studied can be very subtle, which is why, to date, no technique has been able to verify or refute the hypothesis. It is clear that an estimate of the uncertainty arising from the imaging method and subsequent segmentation is critical for clinical studies of MS and schizophrenia.

Statistical and geometrical methods are the two main paradigms for segmentation. They have traditionally been regarded as opposites, since statistical classification is a local process, whereas geometrical segmentation deals with global shape. Some attempts have been made to reconcile the two kinds of technique, for example using *Markov Random Fields* (MRF) to incorporate global constraints to local classification methods (Kapur et al., 1998; Zhang et al., 1999), or *region competition* techniques that gather statistics of the region being segmented to guide edge search in a region-growing process (Brady et al., 1999; Zhu and Yuille, 1995).

In this paper, we develop a method for segmenting 3D data sets which combines characteristics both of statistical classification methods and geometry-driven segmentation. The approach taken bears some similarity with the work of Davatzikos and Bryan (1995), who apply a soft tissue segmentation followed by a ribbon surface deformation to capture the cortex, and Zeng et al. (1998), who extract grey matter based on tissue distribution and then apply a coupled level set surface propagation for cortex segmentation and measurement.

Sections 2 and 3 describe the statistical and geometrical methods that form the foundation of our technique. Next, we summarise our principal contributions, in the following order: a new shape model for modelling complex biological surfaces is devised in Section 4, a new feature for detecting boundaries between tissues in MRI is established (Section 5), and the segmentation framework is completed by incorporating a method for bounding the real surface of the object to be segmented using an *inner* and an *outer* surface, and this is described in Section 6. Finally, a set of validation and clinical application results is presented in Section 7, followed by discussion and future work.

2. Statistical classification and bias field correction

The first of the elements of our framework is based on the work of Wells et al. (1996) and Guillemaud and Brady (1997). Following their method, the range of image intensities corresponding to a certain tissue type is modelled as a Gaussian distribution with small variance around a mean intensity value. A straightforward classification procedure based on thresholding could be directly applied to the image; but real MR acquisitions are subject to a number of degrading factors, not least bias fields and partial volume effects (PVE). This method estimates and corrects for the bias field and produces a set of probability maps which will prove useful for the segmentation process. PVE will be discussed and treated in detail in Section 5.

The *expectation-maximisation* (E-M) algorithm interleaves the estimation of two coupled distributions. This is used in (Wells et al., 1996) to estimate simultaneously the (multiplicative) bias field and the probabilities of each voxel belonging to each tissue class. As a first step, the image is log-transformed, so the multiplicative field becomes an addition. Let $Y = \{Y_i\}$, $i = 1, \ldots, n$, be the (log-transformed) set of voxels in the MR volume. Each voxel has an additive contribution β_i from the bias field:

$$I_{\text{real}} = I_{\text{restored}} \times B \Rightarrow \ln I_{\text{real}} = \ln I_{\text{restored}} + \ln B = Y + \beta.$$

The distribution on log-intensities of voxels containing (only) tissue j is assumed to be Gaussian around a mean intensity value μ_j, with small variance φ_j. Therefore, the probability that a voxel i, which contains tissue j, and which is affected by a bias field contribution β_i, has log-intensity value Y_i is

$$p(Y_i \mid \Gamma_j, \beta_i) = G_{\varphi_j}(Y_i - \mu_j - \beta_i).$$

The usual independence assumption is made, so that the expression for the probability of an image, given the bias field, is:

$$p(Y \mid \beta) = \prod_i \left[\sum_{\Gamma_j} p(Y_i \mid \Gamma_j, \beta_i)\; p(\Gamma_j) \right].$$

Likewise, since the bias field varies slowly spatially, it is assumed that it can be modelled as an n-dimensional Gaussian prior probability density with zero mean:

$$p(\beta) = G_{\varphi_\beta}(\beta),$$

where $\beta = \{\beta_i\}$, $i = 1, \ldots, n$.

Using Bayes' rule, a expression for the posterior probability of the bias field given the observed intensity data is obtained:

$$p(\beta \mid Y) = p(Y \mid \beta)\, \frac{p(\beta)}{p(Y)},$$

where $p(Y)$ is considered an unimportant constant. A zero-gradient condition on the logarithm of this posterior probability is used to assess its maximum. Two coupled formulae may then be obtained, one estimating the probability of a voxel i belonging to tissue class j ($W_{i,j}$); the other computing the estimated bias field β. Both expressions are inter-dependent, and the E-M algorithm is used to interleave their estimation iteratively. Convergence is guaranteed, and, in practice, a good estimate is usually obtained after four or five iterations. See Fig. 1 for typical results.

The method developed by Wells et al. (1996) requires that every voxel in the image be assigned to one of the Gaussian distributions corresponding to each tissue class. Guillemaud and Brady (1997) argue that, in most cases, there are a number of voxels in the volume that cannot be included into any of the tissue classes explicitly modelled, due to their high variance. An additional class *other* with uniform probability distribution is introduced to model such voxels.

The algorithm requires the user to provide the number of tissues to be explicitly modelled, as well as their parameters μ_j and φ_j. We estimate these values interactively, by allowing the user to select a set of points in the tissue of interest and determining the statistics of the sample. This method has proven correct for our tests, but more refined techniques, like the one described in (Zhang et al., 1999), which introduces an update of the tissue parameters into the E-M framework, could be used.

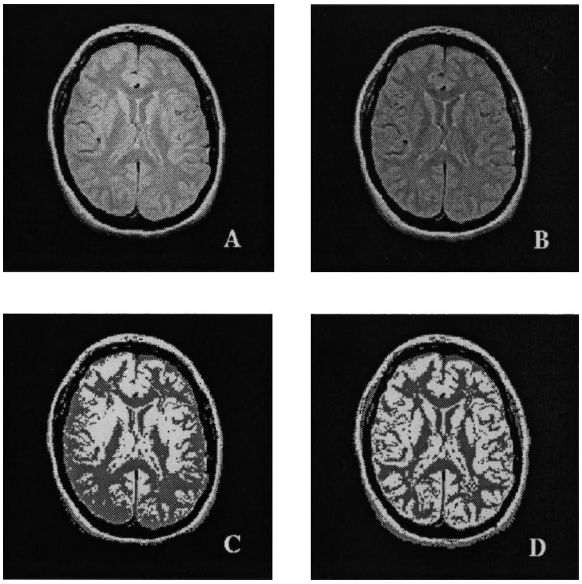

Fig. 1. Bias correction process: A) original image; B) image corrected for bias field; C) and D) segmentations corresponding to A) and B), respectively, obtained by assigning the tissue with most probability. Notice the improved performance of the segmentation after bias correction.

3. Geometry-driven segmentation

Active surfaces are the natural extension to 3D of active contours (Kass et al., 1987), and have become one of the most popular paradigms of 3D segmentation. These techniques search for the outer bounding surface of the object to be segmented using a dynamic process that updates the location and shape of a surface based on edge searches and shape constraints. A review of such methods can be found in (McInerney and Terzopoulos, 1996). Since they are capable of modelling a wide range of complex shapes, we have chosen simplex meshes (Delingette, 1994) as the basis for our active surface model.

The fundamental property of a simplex mesh is that all its nodes have the same number of connections to other nodes. Formally, a k-simplex mesh M of $\Re^3$ is defined as a pair $\{V(M), N(M)\}$, where $V(M)$ is a set of vertices and $N(M)$ is a connectivity function between these vertices. A k-simplex mesh has $(k+1)$-connectivity, i.e. each vertex is connected to exactly $(k+1)$ other vertices. A set of complementarity conditions ensures the mathematical correctness of the construction of the mesh, by not allowing loops and guaranteeing the existence of a path connecting any two nodes (Delingette, 1994). In this work we use 2-simplex meshes.

The topology of a simplex mesh is defined by its connectivity function $N(M)$. A set of operations is defined in order to provide tools to alter the mesh topology. These operations are based on the addition and removal of nodes and on changes to the connectivity between nodes.

The mesh is initialised, and then exposed to a set of forces which make it lock on to the target data. Both internal (shape) and external (fit to data) forces are introduced in the model, which is subjected to Newtonian dynamics.

Internal forces determine the response of a physically-based model to external constraints. Rather than minimising a global elastic energy, the internal force is expressed in terms of the local parameters of each vertex, so shape control is conceived as a local process of a vertex relative to its neighbours. Following this model, different types of constraints can be implemented, such as continuity of the normal, surface orientation, or mean curvature.

External forces are used to fit the mesh to three-dimensional data (in our case, the brain surface or some organ of interest contained in the MR volume). The method used in the implementation of the simplex mesh is based on distances from each mesh vertex to the closest data point, and follows the *iteratively closest point* approach (Besl and McKay, 1992). The term *data point* refers to some feature derived from the image data. A detailed discussion of the feature used in our framework is given in Section 5.

The original formulation of the simplex mesh provides additional topological control by introducing two extensions to the normal behaviour of the mesh. First, a

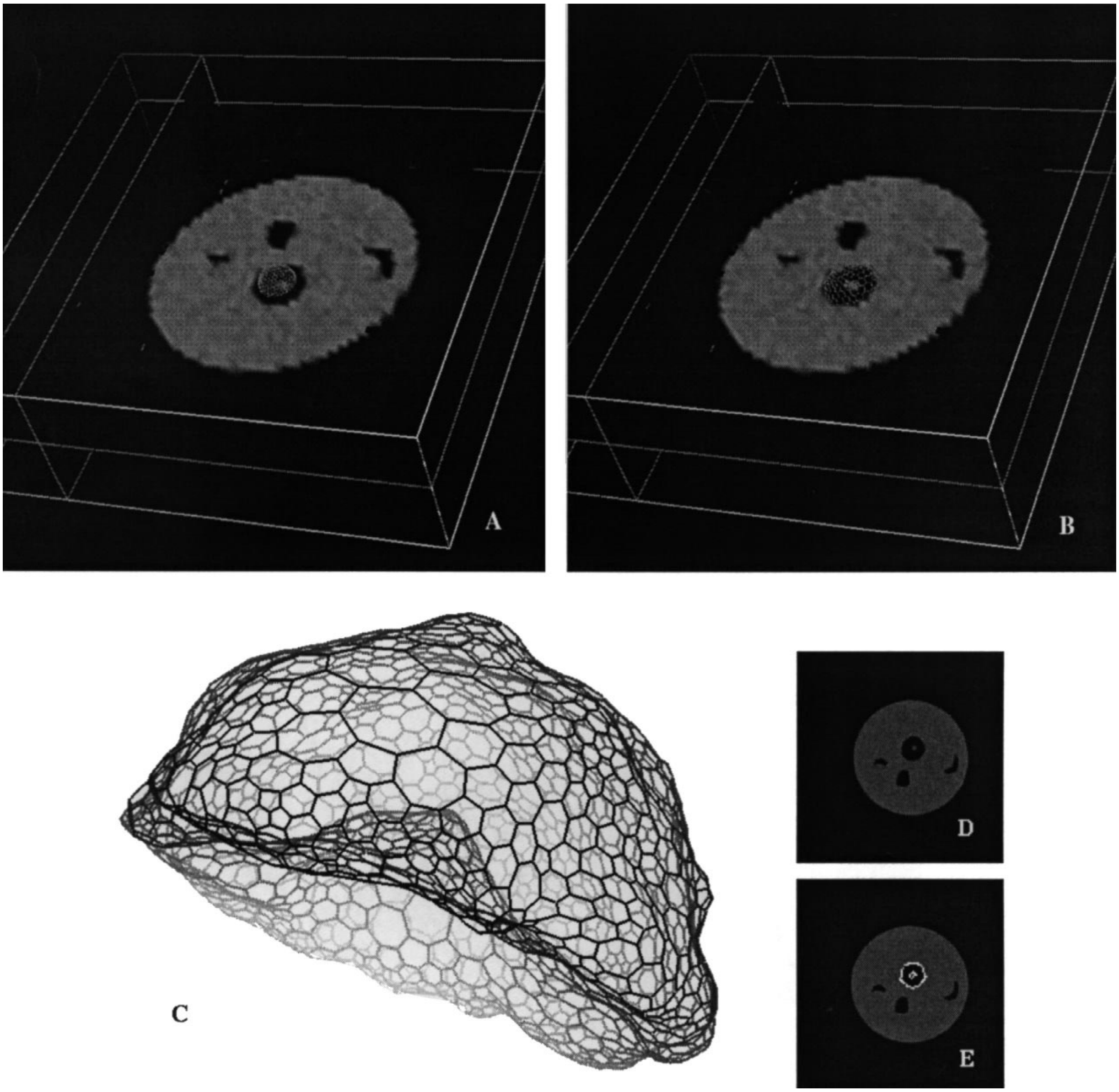

Fig. 2. Simplex mesh fitting process: A) Original mesh (sphere) superimposed on a 2D slice of the MR volume; B) the mesh is locking on to the boundary; C) refined mesh; D) one slice of the target MRI data, showing a cross-section of the object to be segmented at top (c.f. Section 7.2 for details), and E) intersection of the resulting mesh with the slice.

procedure is implemented that adapts the mesh by moving vertices towards areas of high curvature, where more information is needed to represent the data. As a complement to this procedure, a method exists whereby the mesh is refined, i.e. the number of vertices is increased, in these areas, so increasing the flexibility of the shape model, thus allowing a closer fit to the target data (Delingette, 1994). See Fig. 2 for an example of the fitting process.

4. Modelling complex biological shape

The simplex mesh is a powerful tool for three-dimensional segmentation due to its topological flexibility; but much of this flexibility derives from the simplicity of the shape model, consisting of a set of connected 3D points. The representation of biological shape, and in particular the enormously complex shape of the brain surface, requires a more sophisticated representation technique. In addition, the applications of interest demand high accuracy in the location of the boundary of an object, and this is at odds with the simplicity of the simplex mesh.

Our approach is to construct a G^1-continuous surface, interpolating the positions and normals of the nodes of the simplex mesh. A triangulation is first derived from the mesh by adding a node at the centroid of each of its polygons. This node's position is updated to the nearest data point, which is computed by means of a local search, in the same way that nodes are updated during the fitting process. Then, a set of triangular Gregory-Bézier (tGB) patches (Schmitt et al., 1991) are interpolated to the triangulation. TGB patches are expressed, in barycentric coordinates, as follows (Fig. 3):

$$\begin{aligned}\mathrm{GB}(u, v, w) &= u^3 P_0 + v^3 P_1 + w^3 P_2 \\ &+ 12u^2 vwP_{211} + 12uv^2 wP_{121} + 12uvw^2 P_{112} \\ &+ 3u^2 v(1-w)P_{01} + 3uv^2(1-w)P_{02} \\ &+ 3v^2(1-u)wP_{11} + 3(1-u)vw^2 P_{12} \\ &+ 3u(1-v)w^2 P_{21} + 3u^2(1-v)wP_{22},\end{aligned}$$

where $0 \leqslant u,\ v,\ w \leqslant 1$, $u + v + w = 1$ and

$$\begin{aligned}P_{211} &= \frac{wP^{v}_{211} + vP^{w}_{211}}{w+v}, \\ P_{121} &= \frac{uP^{w}_{121} + wP^{u}_{121}}{u+w}, \\ P_{112} &= \frac{vP^{u}_{112} + uP^{v}_{112}}{v+u}.\end{aligned}$$

An important property of tGB patches is that the expressions for the first derivatives with respect to u, v and w do not share any inner control point. In practice, this means that G^1 continuity can be guaranteed simply by constraining the connection along the boundary between two adjacent patches, avoiding the tedious process of considering continuity at corners. G^1 continuity across patches is ensured by enforcing a coplanarity constraint between the two radial first derivative vectors

$$\frac{\partial \mathrm{GB}_R(u_R, v_R, w_R)}{\partial \theta_{u_R}}\Big|_{u_R=0} \quad \text{and} \quad \frac{\partial \mathrm{GB}_L(u_L, v_L, w_L)}{\partial \theta_{u_L}}\Big|_{u_L=0}$$

(where θ_{u_L} and θ_{u_R} take values in $\{u,v,w\}$, depending on the orientation of the patch), and the first order derivative vector $\Gamma^{(1)}(v)$ of the common boundary (see Fig. 4):

$$\begin{aligned}&\alpha(v)\frac{\partial \mathrm{GB}_R(u_R, v_R, w_R)}{\partial \theta_{u_R}}\Big|_{u_R=0} \\ &\quad + \beta(v)\frac{\partial \mathrm{GB}_L(u_L, v_L, w_L)}{\partial \theta_{u_L}}\Big|_{u_L=0} \\ &\quad + \gamma(v)\Gamma^{(1)}(v) = 0, \quad 0 \leqslant v \leqslant 1.\end{aligned}$$

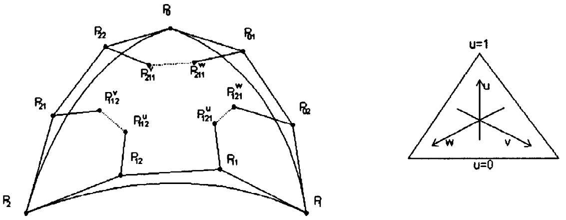

Fig. 3. Triangular Gregory-Bézier (tGB) patch, defined by 15 control points. A G^1-continuous mesh of tGB patches is used to interpolate the nodes of the simplex mesh and its corresponding normals.

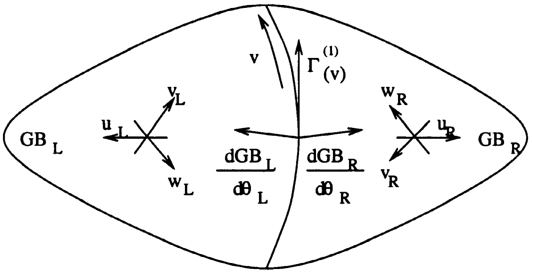

Fig. 4. Enforcing G^1 continuity.

There are two degrees of freedom per edge, corresponding to two of the neighbouring inner control points. One is used to ensure continuity, while the other one is free to move as desired. In our case, its position is chosen to minimise fluctuations in the surface, in a similar vein to (Schmitt et al., 1991). It is however conceivable to use these extra degrees of freedom on the surface to improve the fit to the data. An investigation of this possibility is currently underway.

5. Detection of PVE voxels

In the previous sections, we have described the fitting process of our shape model, but no explicit reference has been made regarding the nature of the target data points. A careful consideration of the elements present in the imaging process leads us to the conclusion that partial volume effects (PVE) are the best choice for a feature for segmentation and establishment of confidence bounds. In fact, we are interested in detecting boundaries between tissues, but these are not accurately localised in MR images, due to the discretisation of the images into voxels. PVE voxels contain more than one tissue, thus indicating that boundaries between tissues are present in them.

In our framework, PVE voxels are detected by analysing the output of the EM-based statistical method, which consists of: an estimation of the bias field corrupting the MR (3D) image, a corrected version of the image after removing the bias field, and a set of probability maps for each tissue class. The usual approach is to obtain a segmentation of the image into tissue classes by assigning to each voxel the tissue class with maximum probability (see Fig. 1), i.e.,

$$t_i = \operatorname{argmax}_j W_{i,j},$$

where $W_{i,j}$ is the probability of voxel i contains tissue j.

We contend that a more careful treatment of the probability maps can provide additional useful information. In particular, for certain voxels the value of the maximum-probability t_i is not very large, indicating that the voxel does not accord well with any of the tissue classes (including the uniform class *other*). This is par-

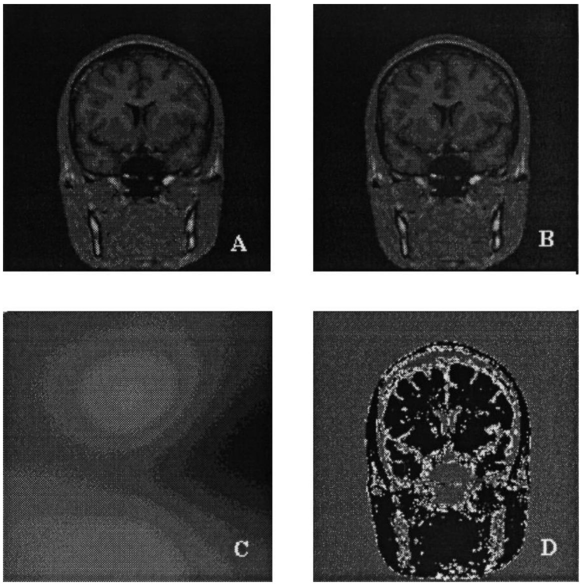

Fig. 5. A) Original slice; B) corrected slice; C) estimated bias field; D) segmentation using a tissue model for white matter and grey matter, plus a class with a uniform probability distribution for CSF, air, and other tissues. The voxels deemed to be places where the PVE is significant are coloured white.

ticularly the case for PVE voxels. That is, a low value in the maximum of the probability maps is often a good indicator of boundariness. Thus, we detect PVE voxels by setting a threshold thrs on the minimum value allowed for this maximum probability (typically, a value in the range of 90–95% probability is used). This is demonstrated in Fig. 5. Thus, tissue classes are assigned to each voxel, in a preliminary segmentation map, using the following criterion:

$$T_i = \begin{cases} t_i & \text{if} \quad t_i > \text{thrs}, \\ \text{PVE} & \text{otherwise.} \end{cases}$$

6. Estimation of confidence bounds

The overall scheme of our segmentation framework is as follows. First, the image is processed using the EM algorithm, which corrects for the bias field and provides an indication of the location of voxels which are good candidates to be PVE voxels. This classification may be somehow noisy, but in general these voxels can typically be seen as a band of variable width delineating the interface between the tissues (see Fig. 5). We interpret this width, corresponding to the transition from *pure* voxels of tissue *A* and *pure* voxels of tissue *B*, as the confidence interval that we seek for our segmentation.

The geometric constraints introduced by the simplex mesh take care of spurious noise in the classification and enforce continuity in the segmentation. Also, the fact of obtaining a parametric surface as output of the process is particularly appealing for further processing and shape description. Thus, the mesh is used to obtain two surfaces representing the *inner* and *outer* estimates on the location of the true surface of the object. The exact location of these surfaces depends on the width of the PVE band described above.

In particular, during each update step of the mesh, local searches in the direction of the normal of each node of the mesh are performed. When these searches hit PVE voxels, they produce upper and lower estimates by computing the intersection of the profile with the closest and farthest PVE voxel boundary. By employing this process on all the vertices of the mesh, the two bounding surfaces are built (see Fig. 6).

7. Results

Since MR imaging takes place in vivo, it is difficult to define ground truth with which to compare the results obtained from a segmentation method on clinical data. We follow several approaches to counter this problem. First, we simulate MR scans using a recursive subdivision procedure that we describe below. This allows us to apply our segmentation and measurement method to objects of known geometry. Next, our method is tested on an MR acquisition of a phantom, developed for MS, with known volume. Finally, the measurement precision on clinical MRI is tested by scanning twice, in rapid succession, a patient whose head is oriented differently in the MRI machine (about 20° separation) and comparing the results (which we assume should be the same). A further example demonstrates the flexibility of the shape model using simplex mesh and tGB patches to segment the cortex from a clinical MR data set.

The measurements shown below are upper and lower bounds on the volume of the segmented object. Several methods are compared: the basic simplex mesh; the simplex mesh after applying the refinement process (c.f. Section 3) to lock closer to the data; the simplex mesh with tGB patches; and voxel-counting measurements obtained using an advanced thresholding tool (Colchester et al., 1996; Bello, 1998).

Volume computations on the simplex mesh are obtained by computing the centroid of each polygon and joining it to the neighbouring nodes, so creating a triangulation $T = \{p_1^i, p_2^i, p_3^i\}_{i=1,\dots,n}$. Then, the volume is determined by the following expression, which is a simplification of Gauss' divergence theorem (Boas, 1983) when applied to triangulated surfaces:

$$V = \frac{1}{6} \sum_{i=1}^{n} \langle p_1^i, p_2^i, p_3^i \rangle,$$

where $p_j^i \in \Re^3$ and $\langle\rangle$ denotes the scalar triple product operator. It is worth noting that computing the volume of a simplex mesh, based on a triangulation of the surface,

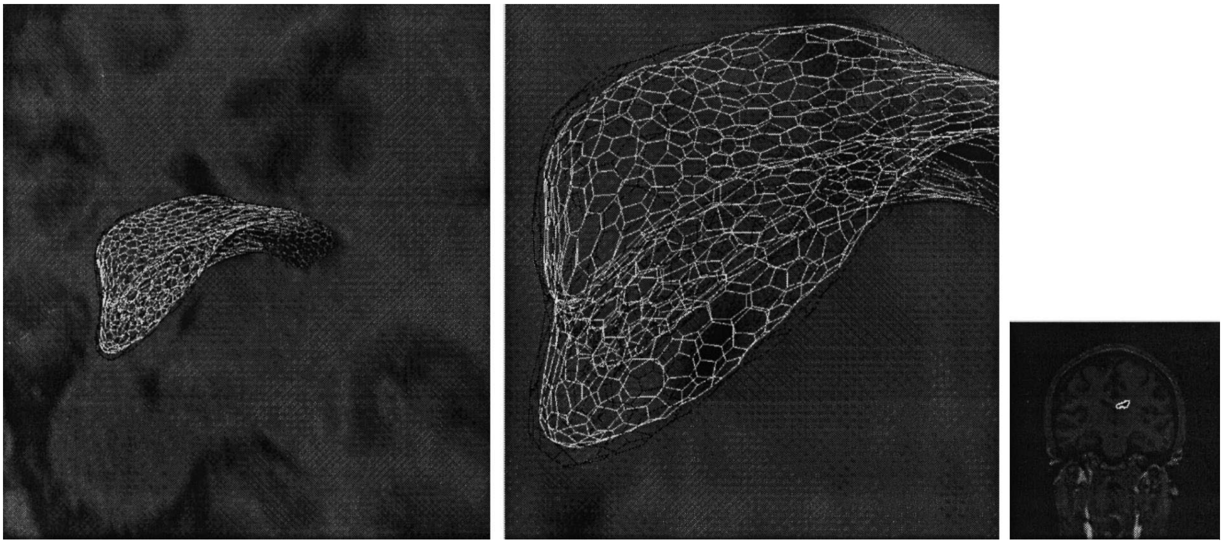

Fig. 6. Outer (black) and inner (white) meshes segmenting the lateral ventricle of a patient. Left: mesh superimposed on a coronal slice of the data set, shown on the right; middle: closeup showing that the lower bound is contained inside the upper bound; right: intersection of both meshes with a slice. The separation between upper and lower bounds is quite small, so it is only reflected by a slight thickening of the white line in the picture on the right.

poses a problem when it comes to establishing an upper bound on the volume, since ondulations of the surface can ‘overflow’ the triangulated surface. We circumvent this problem by displacing the position of the centre of each triangle to the nearest voxel boundary and updating the position of the nodes of the mesh accordingly. Fig. 7 shows a 2D representation of this process. Grey circles represent nodes of the triangulation, and black circles are the data points found for the upper and lower bounds, which correspond to boundaries between voxels of different tissue or furthest boundaries of PVE voxels. A one pixel wide PVE band is shown in the 2D scheme. The real surface of the object is represented by the curve *r*, while the upper and lower meshes are represented by *u* and *l*, respectively. The flat nature of the triangulation produces an underestimation using the upper mesh. To solve this problem, the position of the mid-point *m* of *u* in Fig. 7 (centroid of the triangle in 3D) is updated to the nearest data point, *m*′, and the vertices of the upper mesh are recomputed.

In the case of tGB patches, the volume is determined using the divergence theorem in its full form and the expression for the first fundamental form of a surface, found in classical differential geometry (Boehm and Prautzsch, 1994). Generally, the volume enclosed by a parametric surface $S(u, v)$ is

$$V=\int_{\Sigma} S(u,v)\cdot n(u,v)\,\mathrm{d}\sigma,$$

where Σ is the surface enclosing the object and $n(u, v)$ is the surface normal at the point (u, v). The surface element $d\sigma$ is computed using the usual Jacobian expression from differential geometry:

$$\mathrm{d}\sigma=\sqrt{EG-F^2}\,\mathrm{d}u\,\mathrm{d}v,$$

where

$$E=\left(\frac{\partial S}{\partial u}\right)^2,\quad G=\left(\frac{\partial S}{\partial v}\right)^2\quad\text{and}\quad F=\frac{\partial S}{\partial u}\cdot\frac{\partial S}{\partial v}.$$

Therefore, the expression for the volume enclosed by the surface $S(u,v)$ is

$$V=\int\!\!\int S(u,v)\ \cdot n(u,v)\,\sqrt{EG-F^2}\,\mathrm{d}u\,\mathrm{d}v.$$

For the results reported for the voxel-based thresholding tool, upper and lower estimates for the segmentation are obtained by setting two different thresholds. The results shown below were visually validated by an expert in the use of the tool as good segmentations of the object. Volume measurements are obtained by simple voxel counting.

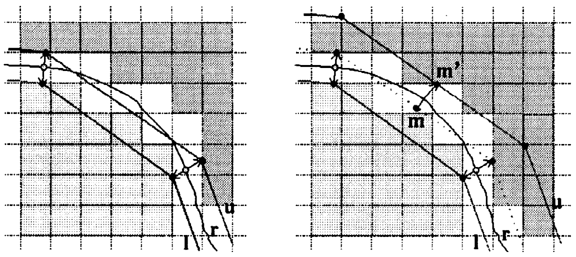

Fig. 7. 2D illustration of the construction of the upper estimate on surface location using the simplex mesh. In order to solve problems caused by ondulations of the boundary, the mid-point *m* is updated to the nearest data point, *m*′, and a parallel surface is then built (see text for details).

7.1. Validation with simulated images

We first simulate MR acquisitions of objects of known geometry, for which an analytic formula is available (e.g. an ellipsoid), by means of a *recursive subdivision* procedure. At each voxel location, it is determined whether the 8 corners of the voxel are inside the simulated object by evaluating its formula. If this is the case, the voxel is assigned intensity value I_{in}, whereas value I_{out} is given to voxels whose 8 corners are all outside the object. The remaining voxels correspond to boundaries and their intensity values should simulate the partial volume effect. In order to determine the proportion α of the voxel that is inside the object, a recursive subdivision procedure is initiated by dividing the voxel by its centre into 8 smaller cubes. Testing for inclusion continues in the manner described above until all boxes are assigned a value or a recursion limit is reached. Then, intensity value

$$I = \alpha I_{in} + (1 - \alpha) I_{out}$$

is assigned to the voxel. The recursion limit is chosen so that the contribution to the final value of α in the limit is smaller than a certain small value ϵ.

This process does not simulate noise or bias fields, but it effectively generates PVE. The primary interest of this validation technique is its flexibility, since different protocols can be simulated by acting on the intensity values to simulate image modalities (T_1, T_2, proton density, etc.), and different voxel sizes and inter-slice gaps can be used to simulate different spatial resolutions. Also, the fact that virtually any shape can be used for the simulated object provides sufficient flexibility to study the effect of size (relative to voxel size), concavities, texture and global shape on the tested segmentation technique.

In our experiments, we simulated the several acquisition protocols in use at our laboratory. Volume dimensions, voxel sizes, and inter-slice gaps are detailed in Table 1. For each of these protocols, several objects are used, ranging from spheres and ellipsoids to smooth objects with concavities or local textures simulating brain gyration.

Volume measurements are obtained by applying the simplex mesh, simplex mesh with tGB patches, and voxel counting after thresholding. The real volume of the object is computed using its analytic formula. Results are reported next for one of the simulated objects (Fig. 8). Volume measurements for the different protocols using the different tested methods are shown in Table 2 and Fig. 9.

Table 1
Simulated acquisition protocols. Protocols 1, 2 and 3 correspond to real clinical practices applied on data sets available in our data pool, whereas protocols 4 and 5 were included to estimate the effect of voxel size on volume measurements

Protocol	Dimensions	Voxel size (mm)	Gap (mm)
1. SUNY	256×256×20	0.9375×0.9375×5	2
2. Oxford	240×240×30	0.9375×0.9375×5	0
3. High res.	256×256×140	1×1×1	0
4. Med. res.	256×256×70	1×1×2	0
5. Low res.	256×256×35	1×1×4	0

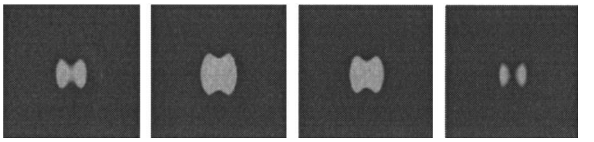

Fig. 8. Slices 8, 11, 14 and 16 of simulation of the object for which results are reported, using protocol SUNY. Notice the considerable blurring due to PVE, especially at slices close to the top and bottom of the object.

Table 2
Volume measurements obtained for the test object. The middle row shows the real volume of the object[a]

	SUNY	Oxford	Low res.	Med. res.	High res.
Sm upper	220999	221591	218653	214162	212171
Voxel upper	199410	200566	197520	190968	188200
tGB upper	191130	193318	191575	191220	190156
Real volume	178447	178447	178447	178447	178447
tGB lower	153555	156952	160265	168950	172774
Voxel lower	151668	152473	155200	162312	166008
Sm lower	146130	149576	152505	160609	164326

[a] The measurements are presented in mm^3, sm = simplex mesh.

The first conclusion that can be drawn from these results is that the confidence bounds are correct in the sense that they bound the real volume. Also, the width of the bounds decreases as the resolution of the protocol increases. A more careful analysis of this is underway, in order to build a model for the expected uncertainty as a function of the different parameters described above. The use of a smooth mesh of tGB patches improves significantly the performance of the simplex mesh. The tGB results are also better (except for the highest resolution protocol, in this example) than those obtained by voxel count based on thresholding. As we will see, these results hold consistently throughout our experiments.

7.2. Validation with MRI phantoms

A phantom consisting of a group of shapes made from paraffin wax and embedded in an agarose gel is used for the second experiment. By measuring the density of the wax, the true volume can be derived from their weight to

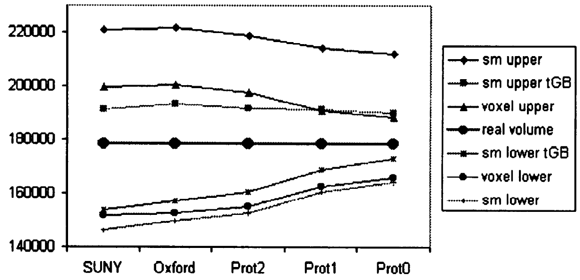

Fig. 9. Volume measurements for the test object. The horizontal axis reflects the different simulated protocols, from lowest to highest resolution. The vertical axis shows the volume.

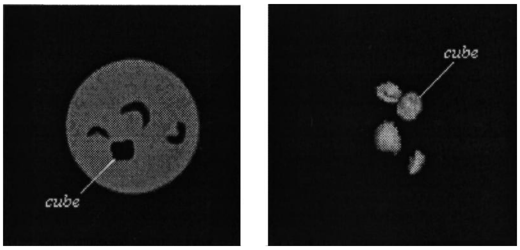

Fig. 10. One slice through and a rendering of the physical phantom. Object *cube*, for which results are reported below, is indicated.

within a confidence interval of 2% (Röll et al., 1994). The phantoms were developed to simulate the size and shape of MS lesions, which usually are very small relative to the resolution of the MR acquisition. Fig. 10 shows a slice through an MRI (1 mm^3 voxels, no gap) and a rendering of the shapes embedded in it.

We present the (typical) results for the phantom object *cube*, shown in Fig. 10. Simplex mesh segmentation, tGB fitting, and simplex mesh refinement volume measures are obtained as described above. Voxel-based volumes are derived from intensity thresholding and posterior voxel count. Tables 3 and 4 and Figs. 11 and 12 show these results.

The results show that the best volume bound estimates are obtained by using a mesh of tGB patches interpolating the simplex mesh. This gives better volume estimates than those obtained using voxel-based methods, while also providing a continuous surface segmenting the object, which can be used for further shape description and processing. It is worth noting that, although the refinement method of the simplex mesh improves the volume measurement significantly, the use of a continuous surface results in a much better estimate. Similar results were obtained when applying the method to the other shapes in the MRI phantom.

Table 3
Volume measurements for object cube of the physical phantom L = lower bound; U = upper bound

	Volume (mm^3)
L simplex mesh	776
L refined sm	797
L voxel-based	821
L sm + tGB	823
Real volume	1028
U sm + tGB	1114
U voxel-based	1152
U refined sm	1228
U simplex mesh	1238

Table 4
Width of the confidence interval relative to the volume of the object $\left(\frac{V_{\text{upper}} - V_{\text{lower}}}{V_{\text{real}}} \times 100\right)$

	Width (%)
Simplex mesh	44.9
Refined sm	41.9
Voxel-based	32.2
Sm + tGB	28.3

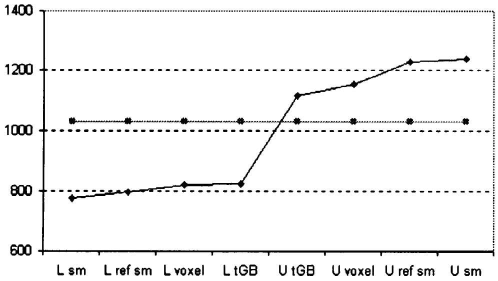

Fig. 11. Volume measurements, in mm^3, for cube (sm = simplex mesh, U = upper, L = lower). The real volume is plotted as a straight line.

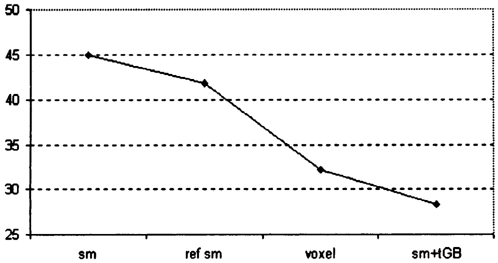

Fig. 12. Width of the confidence interval as a percentage of the real volume. From left to right, simplex mesh, refined simplex mesh, voxel measurements by thresholding, and tGB patches fit over simplex mesh.

7.3. Validation with clinical data

Next, we test the method on in vivo clinical data. We scanned a volunteer twice, in quick succession, the second time with his head rotated through about $20-30°$ (Fig. 13). The data sets are T_1-weighted and each consists of 124 slices of 256×256 voxels of size $0.781251 \times 0.78125 \times 1.7$ mm^3 (TE = 9000 ms, TR = 24 000 ms). The left lateral ventricle of the patient is segmented and results are shown below. No preliminary registration step was necessary.

The data set is first bias-corrected, assuming only one tissue encompassing white matter and grey matter, plus a

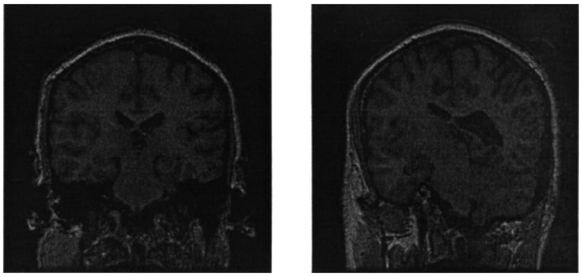

Fig. 13. Slice 56 of the MRI acquisitions used for the study: normal (left) and rotated (right). Note the significant rotation of the right image.

Table 5
Measured volume (in mm^3) of the left ventricle (initial and rotated positions). U = upper bound estimate; L = lower bound estimate. Simplex mesh measurements were obtained after using the refinement process

	Straight	Rotated
U simplex mesh	12192	11100
U voxel-base	10090	9770
U sm + tGB	9454	9574
L sm + tGB	7835	7998
L simplex mesh	7373	7483
L voxel-based	7314	7162

Table 6
Width of confidence interval, in mm^3 $(U - L)$

	Straight	Rotated
Simplex mesh	4819	3617
Voxel-based	2776	2608
Sm + tGB	1619	1576

uniform class modelling the rest of tissues plus CSF and air. Probability maps for the different tissues are generated, and a pre-segmentation step labels voxels with a probability smaller than 95% of belonging to one of the tissues as PVE voxels (see Fig. 5). Simplex meshes are fitted to the data using the information derived from the maps to guide it and the refinement process is used to obtain better fit to the target data. Two meshes are fitted to obtain an inner and an outer bound of the location of the surface. The numbers of vertices for the fitted refined meshes are 1538 (771 polygons), and 1558 for the rotated set (781 polygons). A set of tGB patches for each mesh is then built. For comparison, voxel-based segmentations are performed and validated by an expert. Volume measurements are shown in Tables 5 and 6.

Fig. 14 shows the upper and lower bounds on the volume of the ventricle, and Fig. 15 shows the width of the confidence interval. There are two main points to note about Fig. 15. First, the confidence interval is significantly smaller for the tGB model, and second, the interval is almost invariant to the patient head rotation.

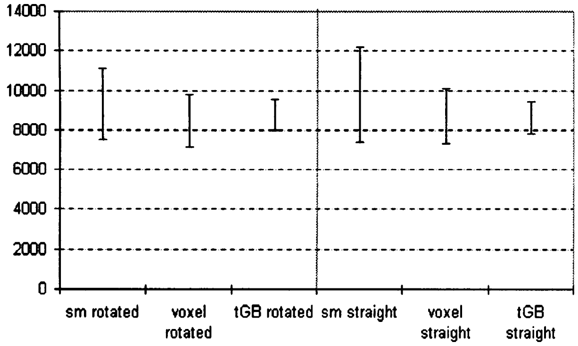

Fig. 14. Volume estimates for the left ventricle (mm^3), with the conditions indicated in the tables above, and for the original and rotated configurations. The vertical bars show the difference between the upper and lower volume estimates.

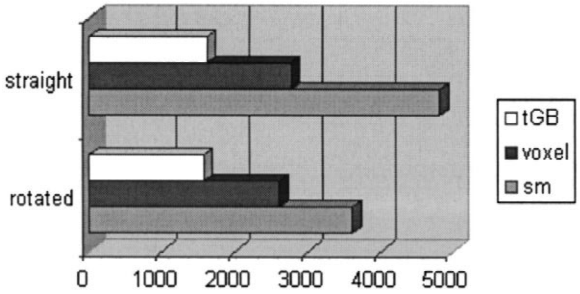

Fig. 15. Width of the confidence interval in the original and rotated condition. The best results are obtained by using a mesh of tGB patches interpolating the simplex mesh. Note that the results for the tGB patches are almost invariant to patient head rotation.

7.4. Modelling the cortical surface

Finally, we investigate the flexibility of our shape model by segmenting the cortical surface of a patient from a clinical MR data set (T_1-weighted, 124 slices, voxel size 0.9375×0.9375×1.2 mm^3). The cortex is known for its highly convoluted shape, which makes it a very demanding surface for shape modelling. The volume is first bias corrected and a prior segmentation with a tissue model encompassing both grey matter and white matter, and another model with uniform probability density function to cater for other tissues, is created (c.f. Fig. 5). An ellipsoidal simplex mesh with low number of nodes is first located roughly near the brain, and the fitting process is started, driven by the probability maps computed during the bias correction step. Also, the refinement process of the mesh is enabled so more nodes are added in areas where more detail is required to fit closely to the target data. Fig. 16 illustrates the process and shows visual results.

8. Discussion and future work

We have described a method for segmentation and measurement of structures embedded in volumetric data. The method itself is novel in the sense of combining statistical classification and geometry-driven segmentation. However, the most important point in the technique is the establishment of confidence intervals in every measurement, by bounding PVE. This is a crucial issue in morphometry which has too often been ignored. We described above several techniques for segmentation and compared their relative performance both in synthetic and clinical data, concluding that the smooth surface model based on the combination of the simplex mesh and triangular Gregory-Bézier patches provides the best results, i.e. gives narrower confidence bounds while still ensuring that the real surface is contained between them. This is of

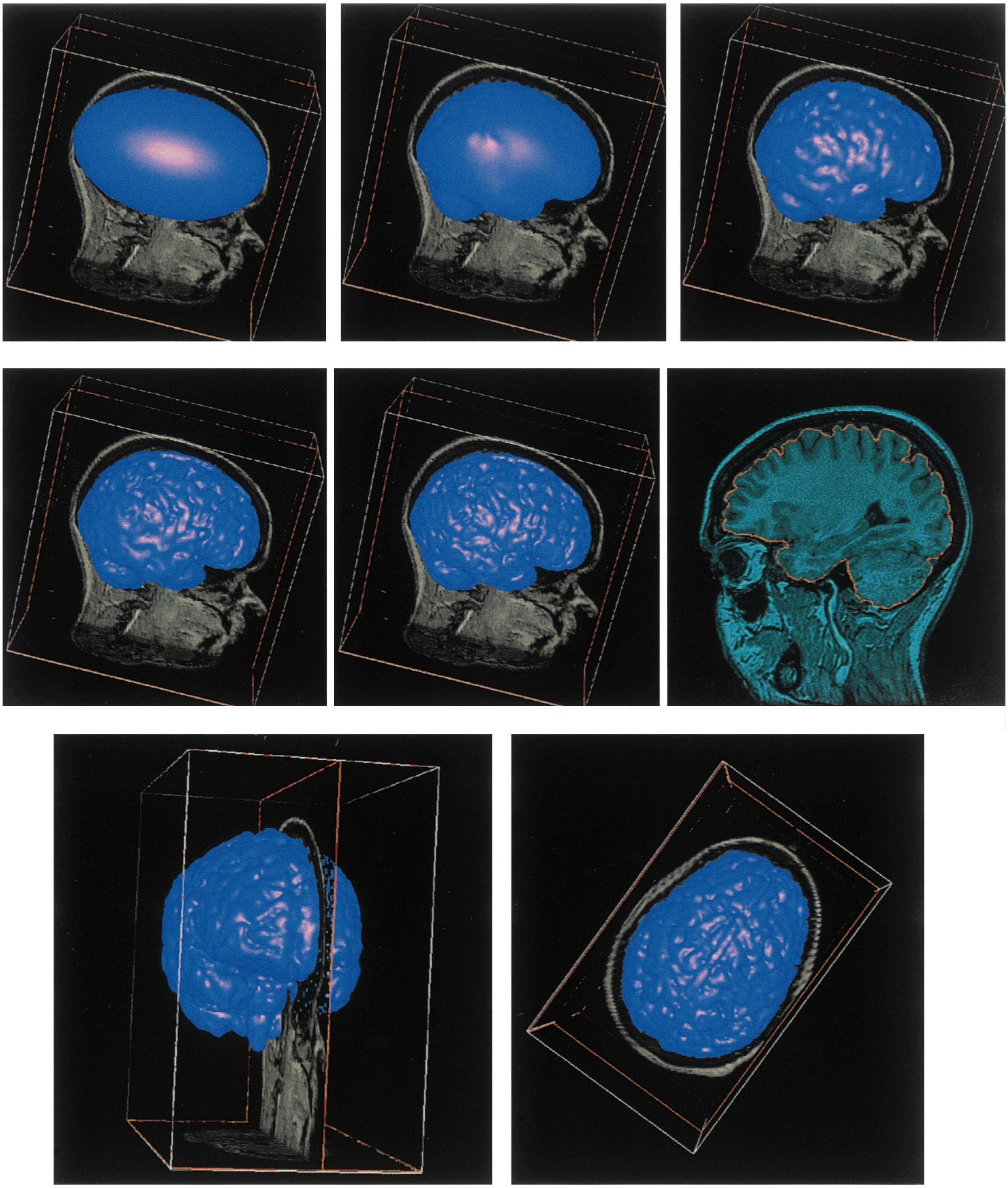

Fig. 16. Mesh fitting process, starting from a mesh with the shape of an ellipsoid and low number of nodes. Refinement is applied to the mesh in order to add nodes in the areas where more detail is required. Different stages on the refinement-fitting process are shown in the first two rows, together with the intersection of the model with one of the slices of the MRI. Enlarged front and top views of the final model are shown in the last row.

vital importance for our clinical partners studying schizophrenia and multiple sclerosis, and an in-depth study is underway to test whether our method can provide the sufficient resolution to validate hypothesis that remain unproved due to the lack of resolution of current morphometric methods.

Regarding the computational cost of this technique, by far the most expensive part is the EM algorithm. When applied to data sets of size $256 \times 256 \times 124$, it may take up to 1.5 h. However, this task can be a batch process and requires no user intervention, other than the prior estimation of tissue parameters. The other components of the method are acceptably fast even on a relatively slow machine (SGI INDY R4400, 200 MHz, 64 Mb RAM). In particular, for the ventricle example reported in Section 7.3, mesh fitting takes approximately 4 min, including refinement of the mesh, and computation of tGB parameters is performed in less than 4 s.

Our approach to bounding the real surface of the object to be segmented has so far been based on establishing conservative upper and lower bounds. The logical next step in our work will be to study the PVE area between the bounds in order to establish a probability distribution function (pdf) on the location of the surface. This will not only allow for further narrowing of the width of the confidence intervals to a certain probability, but will also incorporate the statistical apparatus required for robust morphometry.

The shape model, which is a G^1-continuous parametric surface, provides good flexibility for further processing in order to derive shape descriptors. In particular, little or few descriptors have been established for studying brain asymmetry, a key issue in schizophrenia studies. The direction of our further work encompasses this problem in a more general framework, namely the global, anatomy-driven, parameterisation of the brain. This will allow for comparison of homologous regions in both hemispheres, in the sense of relative position to established cortical and inner landmarks, emulating the process followed by clinicians when analysing a brain, but also providing the quantitative nature required for morphometric purposes.

Acknowledgements

This work was supported by the EC-funded BIOMORPH project 95-0845, a collaboration between the Universities of Kent and Oxford (UK), ETH Zurich (Switzerland), INRIA Sophia Antipolis (France) and KU Leuven (Belgium). The authors would like to thank Dr. Tim Crow from the Department of Psychiatry of the University of Oxford, for driving our research towards clinically useful areas. The phantom images used in Section 7.2 were kindly provided by Dr. Fernando Bello of the University of Kent. Also, we are very grateful to Dr. Hervé Delingette (Epidaure Project, INRIA, Sophia-Antipolis, France) for greatly useful comments and sharing his simplex mesh code, and Dr. Régis Guillemaud (CEA, Grenoble, France) for discussion and also for the use of his brain as a test object in Section 7.3.

References

Bello, F., 1998. Rescue: An integrated software package for the analysis and visualisation of 3d data sets in a clinical research environment. Technical Report TR98-4, NeuroMedIA Group, University of Kent at Canterbury.

Besl, P.J., McKay, N.D., 1992. A method for registration of 3d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (2), 239–256.

Blake, A., Curwen, R., Zisserman, A., 1993. A framework for spatiotemporal control in the tracking of visual contours. Intenational Journal of Computer Vision 11 (2), 127–145.

Boas, M.L., 1983. Mathematical Methods in the Physical Sciences, 2nd edition. John Wiley.

Boehm, W., Prautzsch, H., 1994. Geometric Concepts for Geometric Design. A.K. Peters.

Brady, M., Li, F., Xie, Z. 1999. Texture segmentation from non-parametric statistical analysis of wavelet local energy. IEEE PAMI, to appear.

Colchester, A.C.F., Gerig, G., Crow, T., Ayache, N., Vandermeulen, D., 1997. Development and validation of techniques for brain morphometry (biomorph). In: Proceedings of CVRMED'97.

Colchester, A.C.F., Zhao, J., Holton-Tainter, K.S., Henri, C.J., Maitland, N., Roberts, P.T.E., Harris, C.G., Evans, R.J., 1996. Development and preliminary evaluation of vislan, a surgical planning and guidance system using intra-operative video imaging. Medical Image Analysis 1 (1), 73–90.

Crow, T.J., 1990. Temporal lobe asymmetries as the key etiology of schizophrenia. Schizophrenia Bulletin 16, 433–443.

Davatzikos, C., Bryan, R.N., 1995. Using a deformable surface model to obtain a shape representation of the cortex. In: Proceedings of the International Symposium on Computer Vision, pp. 212–217.

Delingette, H., 1994. Simplex meshes: a general representation for 3d shape reconstruction. Technical Report 2214, INRIA.

Delisi, L.E., Tew, W., Xie, S., Hoff, S.L., Sakuma, M., Kusher, M., Lee, G., Shedlack, K., Smith, A.M., Grimson, R., 1995. A prospective follow-up study of brain morphology and cognition in first-episode schizophrenic patients. Preliminary Findings in Biological Psychiatry 38, 349–360.

Guillemaud, R., Brady, M., 1997. Estimating the bias field of mr images. IEEE Transactions on Medical Imaging 16 (3), 238.

Kapur, T., Grimson, W.E.L., Wells III, W.M., Kikinis, R., 1998. Enhanced spatial priors for segmentation of magnetic resonance imagery. In: Proceedings of MICCAI'98.

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: Active contour models. In: Proceedings of ICCV'87. pp. 259–268.

Marais, P., Guillemaud, R., Sakuma, M., Zisserman, A., Brady, M., 1996. Visualising cerebral asymmetry. In: Hoehne, K.H. Kikinis, R. (Eds.), Proceedings of VBC'96. Vol. 1131 of Lecture Notes in Computer Science. Springer, pp. 411-416.

McInerney, T., Terzopoulos, D., 1996. Deformable models in medical image analysis: A survey. Medical Image Analysis 1 (2), 91–108.

Rinck, P.A. (Ed.), 1993. Magnetic Resonance Imaging. Blackwell Scientific.

Röll, S.A., Colchester, A.C.F., Summers, P.E., Griffin, L.D., 1994. Intensity-based object extraction from 3d medical images including a correction for partial volume errors. In: Proceedings of BMVC'94. pp. 195–204.

Schmitt, F., Chen, X., Du, W.H., 1991. Geometric modelling from range image data. In: Proceedings of Eurographics'91. Elsevier, pp. 317–328.

Székely, G., Kelemen, A., Brechbüler, C., Gerig, G., 1996. Segmentation of 2d and 3d objects from mri volume data using constrained elastic deformations of flexible fourier contour and surface models. Medical Image Analysis 1 (1), 19–34.

Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A., 1996. Adaptive segmentation of mri data. IEEE Transactions on Medical Imaging 15 (4), 429–442.

Zeng, X., Staib, L.H., Schultz, R.T., Duncan, J.S., 1998. Volumetric layer segmentation using coupled surfaces propagation. In: Proceedings of CVPR'98. pp. 708–715.

Zhang, Y., Smith, S., Brady, M., 1999. Segmentation of brain MR images using Markov random field. In: Proceedings of MIUA'99. Oxford, UK.

Zhu, S.C., Yuille, A.L., 1995. Region competition and its analysis: A unified theory for image segmentation. Technical Report 95-07. Harvard Robotics Lab., Harvard University.