

## A novel quantitative cross-validation of different cortical surface reconstruction algorithms using MRI phantom

Jun Ki Lee,<sup>a</sup> Jong-Min Lee,<sup>a,\*</sup> June Sic Kim,<sup>b</sup> In Young Kim,<sup>a</sup>  
Alan C. Evans,<sup>c</sup> and Sun I. Kim<sup>a</sup>

<sup>a</sup>Department of Biomedical Engineering, Hanyang University, 17 Haengdang-dong Sungdong-gu P.O. Box 55, Seoul 133-791, Republic of Korea

<sup>b</sup>Department of Neurosurgery, Seoul National University College of Medicine, Seoul, Republic of Korea

<sup>c</sup>McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada

Received 31 May 2005; revised 22 October 2005; accepted 23 December 2005  
Available online 24 February 2006

Cortical surface reconstruction is important for functional brain mapping and morphometric analysis of the brain cortex. Several methods have been developed for the faithful reconstruction of surface models which represent the true cortical surface in both geometry and topology. However, there has been no explicit comparison study among those methods because each method has its own procedures, file formats, coordinate systems, and use of the reconstructed surface. There has also been no explicit evaluation method except visual inspection to validate the whole-cortical surface models quantitatively. In this study, we presented a novel phantom-based validation method of the cortical surface reconstruction algorithm and quantitatively cross-validated the three most prominent cortical surface reconstruction algorithms which are used in Freesurfer, BrainVISA, and CLASP, respectively. The validation included geometrical accuracy and mesh characteristics such as Euler number, fractal dimension (FD), total surface area, and local density of points. CLASP showed the best geometric/topologic accuracy and mesh characteristics such as FD and total surface area compared to Freesurfer and BrainVISA. In the validation of local density of points, Freesurfer and BrainVISA showed more even distribution of points on the cortical surface compared to CLASP.

© 2006 Elsevier Inc. All rights reserved.

*Keywords:* Cortical surface reconstruction algorithm; Quantitative validation; MR phantom; Freesurfer; BrainVISA; CLASP

### Introduction

The surface of the human cerebral cortex is known as a highly folded sheet with 60–70% of its surface area buried within folds (Van Essen and Drury, 1997; Zilles et al., 1988). The relationship

between the gross cortical folding pattern and the cytoarchitectonic and functional organization of the underlying cortex is a subject of much current interest and debate (Essen et al., 2001; Thompson et al., 2003). Reconstructed cortices enable the visualization and the study of the sulcal and gyral patterns of an individual subject (Carman et al., 1995; Essen et al., 2001; Fischl et al., 1999a,b; Thompson et al., 2001) and allow morphometric measurements such as volume (Kim et al., 2000), surface area (Chung et al., 2003; Magnotta et al., 1999), thickness (Krugel et al., 2003; MacDonald et al., 2000; Thompson et al., 2001; Yezzi and Prince, 2003), and sulcal depth (Manceaux-Demiau et al., 1998). Correspondences between cortical reconstructions from different subjects can also be used for image registration (Cachier et al., 2001), digital atlas labeling (Jaume et al., 2002; Sandor and Leahy, 1997), and population-based probabilistic atlas generation (Thompson et al., 1997). In addition, cortical reconstruction is important for functional brain mapping (Dale and Sereno, 1993), surgical planning (Grimson et al., 1998), and cortical unfolding or flattening (Carman et al., 1995; Essen et al., 2001; Fischl et al., 1999a,b; Tosun and Prince, 2001).

To fulfil the needs described above, it is important to faithfully represent true cortical surface in terms of geometry and topology. The geometrically accurate cortical reconstruction must generate a representation of the cortex consistent with the true geometry of the brain cortex, completing with multiple lobes, gyral folds, and narrow sulci. This task is difficult because of artifacts such as image noise, partial volume effects, and intensity inhomogeneities (Dale et al., 1999; Han et al., 2001; MacDonald et al., 2000). Especially in tightly folded sulci, the exact boundaries of the cortex are hard to detect because opposing sulcal banks are closer than the magnetic resonance imaging (MRI) resolution. This causes inaccuracies in surface extraction and subsequent morphometric measures such as cortical thickness. The topology of the reconstructed surface is also an important consideration. Since the cerebral cortex has the topology of a two-dimensional sheet, the surface representation of the cortex should have the same topology

\* Corresponding author. Fax: +82 2 2296 5943.

E-mail address: [ljm@hanyang.ac.kr](mailto:ljm@hanyang.ac.kr) (J.-M. Lee).

Available online on ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)).

without holes and bridges, preventing the surface from being accurately flattened or inflated. Mesh characteristics such as point distribution on the cortical surface are also important factors which determine good representation of brain cortex. Properly distributed meshes on the cortical surface will give a good description of shape of the brain cortex, which can then contribute to visualization, shape analysis, and surface-based post-processing such as surface registration.

The computational methods to obtain a polygon mesh that represents the cortex can largely be grouped according to the type of representation upon which they operate. On one hand, there are segmentation methods, which are a kind of a bottom–up approach, operating directly on segmented anatomical MRI volumes (Dale et al., 1999; Mangin et al., 1995). This approach is usually geometrically accurate but typically generates a number of topological errors, called handles. Thus, additional processes that fix the topology errors are needed. Alternatively, there are iterative morphing methods, which use a top–down approach, operating on polygon mesh representations (Kim et al., 2005; MacDonald et al., 2000). In this approach, the resulting representation of the cortex is topologically correct, but some inaccuracies in surface extraction may occur due to overly strong deformation constraints. Furthermore, the extensive morphing of the initial standard model is computationally expensive, a diminishing issue as computational power increases.

Of the well-established methods for the cortical surface reconstruction, we have evaluated three algorithms, Freesurfer, BrainVISA, and CLASP. They have been chosen for evaluation according to some rules of accessibility to a tool, number of users, possibilities of file format, and coordinate transformations. Dale et al. introduced an automatic method that is implemented in the freely available Freesurfer (Dale et al., 1999) which is a mixed approach taking both the segmentation and the iterative morphing method. Mangin et al. (1995) also presented a method implemented in the freely available BrainVISA. This enables the shape analysis of the sulcal pattern as well as morphological measure of the cortex. Kim et al. (2005) have developed a method called CLASP (Constrained Laplacian Anatomical Segmentation using Proximities), an enhanced version of the iterative morphing method first developed by MacDonald et al. (2000). Although many studies about brain cortex have been based on the cortical surfaces created by these tools (Fischl et al., 1999a,b; Thompson et al., 1997), there has been until now, as far as we know, no quantitative comparison of the performances of these reconstruction methods. Such a detailed evaluation is difficult since each method has its own procedure, file format, coordinate system, and specific purpose for the surface reconstruction (Fischl et al., 1999a; Mangin et al., 1995; Kim et al., 2005). In this study, we performed a quantitative cross-validation of the three cortical surface reconstruction tools using an MRI simulator generating a realistic MRI incorporating the calculation of noise and partial volume effects (Kwan et al., 1996; Collins et al., 1998). The evaluation strategy presented here using simulated MR phantom provides “gold standard” with which to access the performance of cortical surface reconstruction algorithms. The validation study evaluated the topology, geometrical accuracy, and mesh characteristics of the cortical surfaces extracted by each algorithm. We focused most effort on the geometrical accuracy of surface extraction which is essential for the accurate measurement of morphological variables such as cortical thickness and cortical surface area.

## Methods

### Computational method used in each tool

Each algorithm has its own procedure to create cortical surface. Generally, the processing sequence in each tool includes pre-processing steps such as spatial normalization, intensity inhomogeneity correction, skull stripping, and tissue classification. For the comparison study of three different algorithms, we followed the complete series of automated procedures for extracting the cortical surface with their own pre-processing. Table 1 showed the whole procedure of each tool. Since each method was explained in detail in its original publication (Dale et al., 1999; Kim et al., 2005; Mangin et al., 1995), we briefly summarize them below:

- (i) BrainVISA (Mangin et al., 1995) first provides a binary mask of each hemisphere cortex with spherical homotopy. This sequence includes bias correction, brain mask segmentation, hemisphere mask segmentation, and detection of the gray/white interface. A standard facet tracking algorithm is used to compute a first spherical mesh made up of facets from the cortex mask. Then, the center of each facet is connected to the center of the neighboring facets in order to yield a spherical mesh of triangles. This algorithm preserving the initial topology relies on a look-up table of configurations like in the standard marching cube algorithm. Finally, a decimation including smoothing is performed to discard stair artefacts related to the underlying discretization. The embedded smoothing operation iteratively moves the nodes towards their neighborhood gravity center, which may be related to some usual surface evolution processes. This mesh construction includes some smoothing operations that may remove some interesting anatomical information.

Table 1  
The summary of the procedure performed on each tool to create cortical surface

Process	Subprocess	Freesurfer	BrainVISA	CLASP
Pre-processing	Inhomogeneity correction	o	o	o
	Skull stripping	o	o	o
	Segmentation (white)	o	o	o
	Segmentation (gray)	x	o	o
Create white surface	Cut planes	o	o	x
	Tessellation	o	o	x
	Smooth	o	x	x
	Inflate	o	x	x
	Decimation	x	x	x
	Deformable surface	o	x	o
Edit segmentation	Manually edit defects	o	x	x
Fix topology	Automatic correction	o	o	x
Create gray surface	Tessellation	x	o	x
	Decimation	x	o	x
	Deformable surface	o	x	o

Note. An ‘o’ indicates that a procedure is performed and an ‘x’ that it is not.

- (ii) Freesurfer (Dale et al., 1999; Fischl et al., 1999a,b) first corrects for intensity variations due to magnetic field inhomogeneities and creates a normalized intensity image from a high resolution, T1-weighted, anatomical 3-D MRI data set. Next, extracerebral voxels are removed using a skull stripping procedure. The intensity-normalized skull-stripped image is then operated on by a segmentation procedure based on the geometric structure of the gray matter (GM)/white matter (WM) interface. Cutting planes are then computed to separate the cerebral hemispheres and disconnect subcortical structures from the cortical component. This generates a preliminary segmentation which is partitioned using a connected components algorithm. Any interior holes in the components representing WM are filled, resulting in a single filled volume for each cortical hemisphere. Finally, the resulting volume is covered with a triangular tessellation and deformed to produce an accurate and smooth representation of the GM/WM interface as well as the pial, i.e. GM the cerebrospinal fluid (CSF) boundary, surface. This surface departs from a simple spherical topology due to subcortical GM as well as various midbrain structures. These topological defects are removed through a manual editing procedure that results in a surface with both accurate geometry and spherical topology.
- (iii) CLASP (Kim et al., 2005) first removes intensity non-uniformity in the raw MR images (Sled et al., 1998). The MR volume is re-sampled into stereotaxic space (Talairach and Tournoux, 1988; Collins et al., 1994) and a 3-D stereotaxic brain mask used to remove extracerebral voxels. The discrete classification employs Intensity-Normalized Stereotaxic Environment for Classification of Tissues (INSECT) (Zijdenbos et al., 1996, 1998). CLASP extracts the inter-cortical surface by deforming a sphere polygon model to the boundary between GM and WM. The number of mesh triangles is hierarchically refined from 320 to 80,920. Then, the pial surface is expanded from the white surface to the boundary between GM and CSF along a Laplacian map which smoothly increases potential surfaces between WM and CSF. A CSF fraction image is skeletonized to determine the outer boundary of the cortex in buried sulci (Ma and Wan, 2001).

Since default parameters used for the reconstruction generally provide the best results and give the objective performance which is not dependent on the subject, the values of the parameters in each method were default settings. You can see the reconstructed cortical surfaces through three tools respectively (Fig. 1).

#### Cross-validation of reconstruction algorithms

The quantitative evaluation of the three algorithms was performed in several ways. First, the geometric accuracies were validated using both volume-based and surface-based comparisons detailed below. In surface-based validation, an MRI simulator (Collins et al., 1998; Kwan et al., 1996) was used to create a “ground truth”. Second, mesh characteristics including topological accuracy (Worsley, 1995), fractal dimension (Sarraille and Myers, 1994), surface area, and point distribution were calculated for the surfaces obtained by each method. T1-weighted MR images with  $1.0 \text{ mm} \times 1.0 \text{ mm} \times 1.0 \text{ mm}$  resolution and  $181 \times 217 \times 181$  voxel dimensions were selected randomly from the data sets of the International Consortium for Brain Mapping (ICBM) (Mazziotta et al., 1995). We used 30 cases (18 M/12 F; age average 28.6 years, range 18–42 years) for the volume-based validation and 4 cases (2 M/2 F, age average 24.5 years, range 20–30 years) for the topology measure, surface-based validation, and mesh characteristic studies. Freesurfer was excluded in volume-based validation because the classified GM volume could not be defined in the reconstruction process. Thus, in volume-based validation only, BrainVISA and CLASP algorithms are accessed. In each brain, the white (i.e. GM/WM boundary) and pial (i.e. GM/CSF boundary) surface was extracted by each of the three methods. Then, we validated the cortical surfaces as follows.

#### Validation of geometrical accuracy

**Volume-based validation.** In volume-based validation, we compared the GM volume images obtained by tissue classification during the pre-processing steps, with those obtained by labeling all GM voxels between the pial and white surfaces of brain cortex. For the latter case, the pial and white surfaces were first extracted using

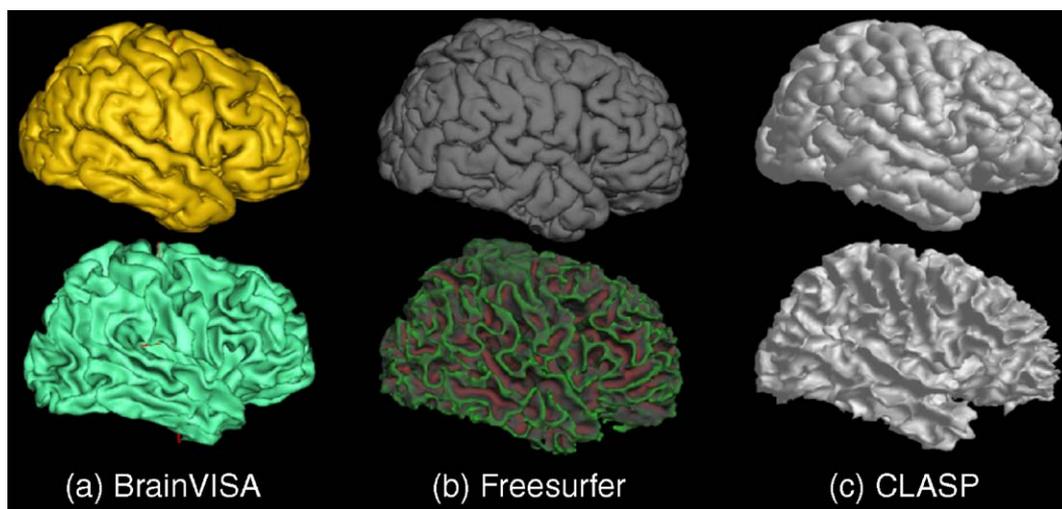


Fig. 1. Cortical surfaces reconstructed by three different methods. Cortical surfaces reconstructed by (a) BrainVISA (Mangin et al., 1995; Gordon and Udupa, 1989), (b) Freesurfer (Dale et al., 1999; Fischl et al., 1999a,b), and (c) CLASP (Kim et al., 2005) respectively. Each surface created from the same T1-weighted image randomly selected from ICBM data sets.

the surface extraction algorithm. Next, a GM map was created by filling all voxels between these two surfaces (Hearn and Baker, 1997). This process was repeated for all tools being compared. Freesurfer was excluded in this validation because it does not use an explicit gray matter segmentation as an intermediate in its computation (Dale et al., 1999; Fischl et al., 1999a,b). Note that a separating masking step was employed to remove non-cortical gray voxels from the tissue-classified GM map since we were not interested in deep GM structures for this study. The two GM maps from tissue classification and from intra-surface filling were then overlapped and compared. In Fig. 2, you can see a result of maps of geometric errors between the classified image and the surface-masked image generated by both tools.

In the overlapped volume, we calculated three statistical values: (1) percentage of matched GM voxels to total voxels between the classified image and the surface-masked image (True-Positive, TP), (2) percentage of background voxels by the surface masking but classified as GM voxels (False-Negative, FN), and (3) percentage of voxels classified as GM by the surface-based method but classified as background (False-Positive, FP). Equations are below:

$$TP = \sum_{v=1}^{n_{Gv}} \left( \frac{G_C(v) \cdot G_S(v)}{N_{Gv}} \times 100 \right)$$

$$FN = \sum_{v=1}^{n_{Gv}} \left( \frac{G_C(v) - G_C(v) \cdot G_S(v)}{N_{Gv}} \times 100 \right)$$

$$FP = \sum_{v=1}^{n_{Gv}} \left( \frac{G_S(v) - G_C(v) \cdot G_S(v)}{N_{Gv}} \times 100 \right)$$

where  $v$  is a voxel index,  $N_{Gv}$  is an average number of GM voxels, and  $G_C(v)$  and  $G_S(v)$  are GM voxel maps (values of GM voxels are

1 and values of other voxels are 0) of classified volume and surface volume, respectively. For surface masking, a scan line filling algorithm was used (Hearn and Baker, 1997). To increase the accuracy on the boundary of the cortex, we subdivided each voxel into 64 subvoxels ( $4 \times 4 \times 4$ ) before the process of filling. The final GM map was then calculated from the filled map counting the number of filled subvoxels of each corresponding voxel on the final GM map. Thus, every voxel on the surface was classified as GM, which could cause FN errors.

*Surface-based validation.* We then evaluated each cortical surface reconstruction method using a surface-based procedure. This evaluation method had two aspects: (i) to compare the surfaces estimated by each method with a known phantom surface and (ii) to examine the reproducibility of the whole procedure including classification and surface extraction. Since there is no readily available “ground truth” with which to assess the performance of individual surface extraction algorithms, we approached the problem with an MRI simulator (Kwan et al., 1996; Collins et al., 1998). This simulator generates a realistic MRI incorporating the calculation of noise and partial volume effects. For each algorithm, the following steps were performed:

- 1) Pial and white surfaces were first extracted from the real MRI volume (Fig. 3a).
- 2) A digital phantom including four tissue types (GM, WM, CSF, and background) was created from the surfaces. WM voxels were defined inside the white surface, and GM voxels were inserted between the pial and white surfaces. To create partial volume effects, voxels on the pial surface were given probabilities of 70% for GM and 30% for CSF. Voxels between the exterior brain mask and the pial surface were labeled as CSF. All other voxels were labeled as background 1 (Fig. 3b).

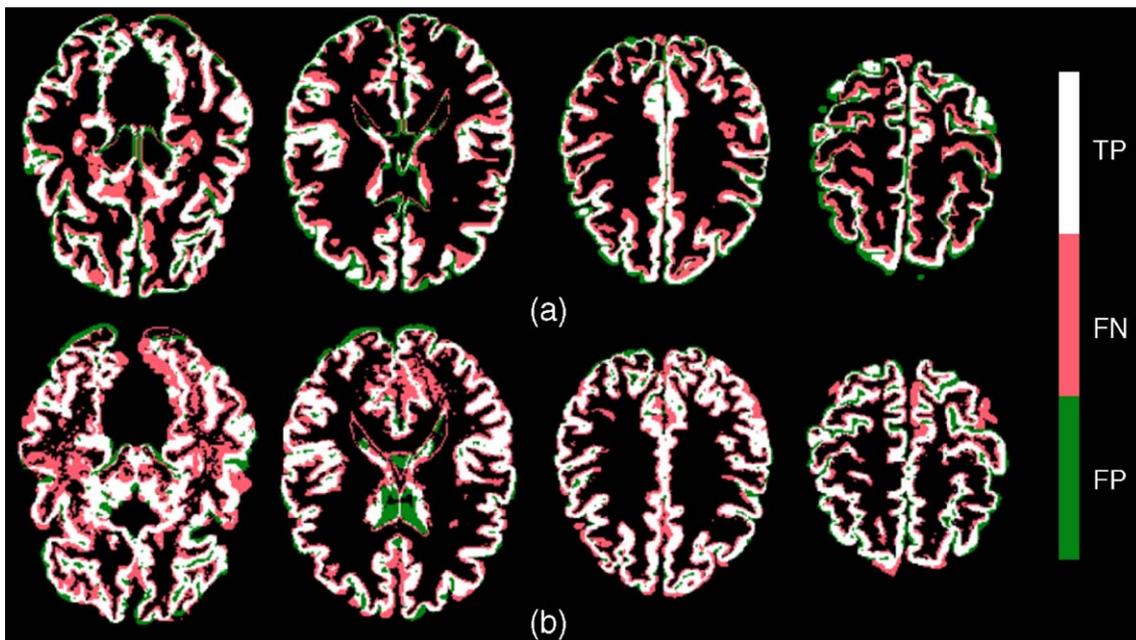


Fig. 2. Map of geometric errors (by volume-based validation). Map of geometric errors between the classified image and the surface-masked image by (a) BrainVISA and (b) CLASP. Voxels of TP (True-Positive) means that they belong both to GM voxels of the classified image and to the GM voxels of the surface-masked image. Voxels of FN (False-Negative) means they belong to GM voxels of the classified image but not to GM voxels of the surface-masked image. Voxels of FP (False-Positive) mean they belong to GM voxels of the surface-masked image but not to GM voxels of the classified image.

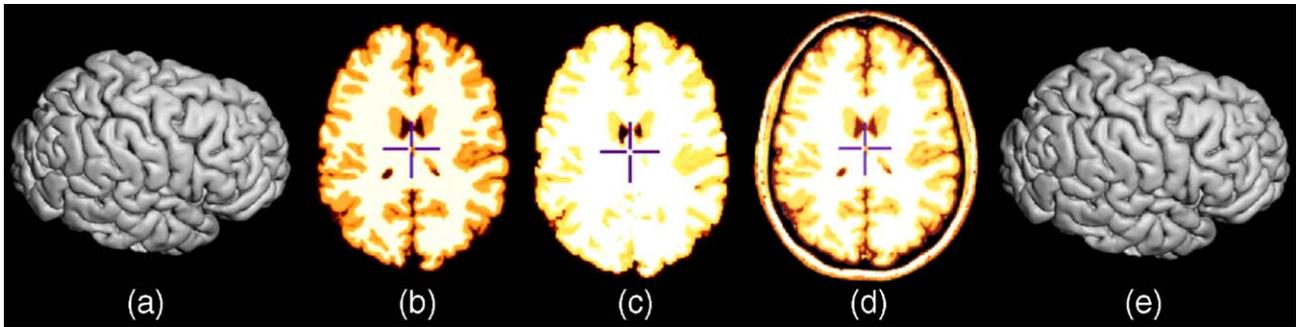


Fig. 3. Process of the evaluation using phantom. (a) Created cortical surface, (b) surface-masked volume, (c) digital brain phantom, (d) phantom including skull, (e) recreated cortical surface from phantom.

- 3) A T1 MR image was simulated from the phantom using the same parameters as the real data acquisition (TR = 18 ms, TE = 10 ms, slice thickness = 1 mm) (Fig. 3c).
- 4) Additional substructures (skull, basal ganglia) were added from the real MRI (Fig. 3d).
- 5) Pial and white surfaces were then extracted from the simulated MRI volume (Fig. 3e).
- 6) Differences between each surface obtained from real or simulated MRI were measured.

In this process, the surface extracted from the real data is regarded, by definition, as “true”. The experiment was designed to assess how well the surface extraction algorithm could re-capture the true surface by operating upon a simulated MRI volume derived from the true surface. To ensure that the process was not biased toward any particular algorithm, the experiment was repeated with each algorithm providing the “true” surface. Then, each algorithm was applied to the simulated MRI to generate a “test” surface. The root mean square (RMS) error between “true” and “test” surfaces (averaged over all surface vertices) then provided a measure of accuracy in surface extraction. To measure RMS error between the “true” and “test” surfaces, we calculated the distances from the vertex of “true” surfaces to the nearest triangular face of the “test” surface. Grand averaging across the  $3 \times 3$  permutations of “true” and “test” algorithm provided an unbiased measure of algorithm performance.

#### Validation of mesh characteristics

**Surface topology.** Since the cerebral cortex has the topology of a 2-D sheet, a topologically correct surface model should have no holes and bridges. Topological errors in the surface model, for example, can lead to erroneous short-cuts from one part of the surface to another. A topologically correct surface allows for better estimation of geodesic distances and for the analysis that is not appropriately performed in the volumetric embedding space (Fischl et al., 1999a,b). It also allows for correct cortical unfolding and mapping to a sphere, which could be used for visualization, measurement, and the establishment of a global coordinate system on the cortex (Fischl et al., 1999a,b; Han et al., 2001). The topologies of the extracted surfaces were validated by computing the Euler number of the surface (Dale et al., 1999; Worsley, 1995).

Euler–Lhulier’s formula

$$v - e + f = 2 - 2g \quad (1)$$

where  $v$ ,  $e$ , and  $f$  are the number of vertices, edges, and faces, respectively (Pont, 1974; Preparata and Shamos, 1985). Since the cerebral cortex has the topology of a 2-D sheet, a topologically correct surface model should have no holes and bridges ( $g = 0$ ) and the Euler number should be ‘2’. Sixteen cortical surfaces were evaluated which were generated by each tool respectively to evaluate the topological accuracies.

**Fractal dimension.** Fractal dimension (FD) is an extremely compact measure of shape complexity, condensing all the details into a single numeric value that summarizes the irregularity of the object (Sarraille and Myers, 1994). For each algorithm and for pial or white surface, we compared the FD of the “true” surface with that of the “test” surface (using the definitions of the previous section). We calculated FD of each cortical surface model using the box-counting method (Sarraille and Myers, 1994). In the box-counting method, the shape of interest is initially mapped onto a rectangular grid or lattice, the edges of each box in the grid being of equal length, and the number of grid boxes occupied by one or more voxels of the image is counted. This box-counting step is repeated several times: the same image is mapped onto a series of rectangular grids of increasing box size, and the number of occupied boxes in each grid is counted (Sarraille and Myers, 1994). In practice, we estimated FD for finite data sets using

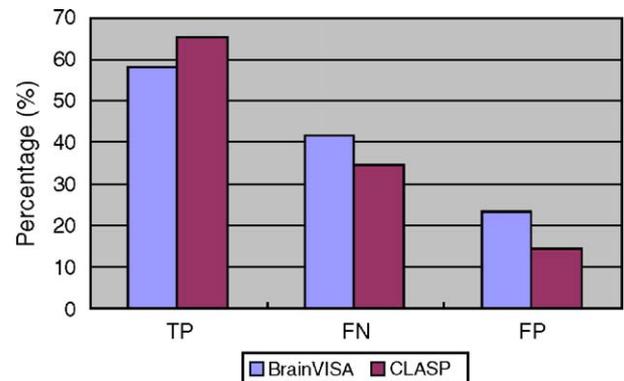


Fig. 4. Geometric errors (by volume-based validation). Note: y axis means percent of each statistical volume compared to the whole volume of gray matter in the classified volume. x axis is statistical volumes. Voxels of TP mean that they belong both to GM voxels of the classified image and to the GM voxels of the surface-masked image. Voxels of FN mean they belong to GM voxels of the classified image but not to GM voxels of the surface-masked image. Voxels of FP mean they belong to GM voxels of the surface-masked image but not to GM voxels of the classified image.

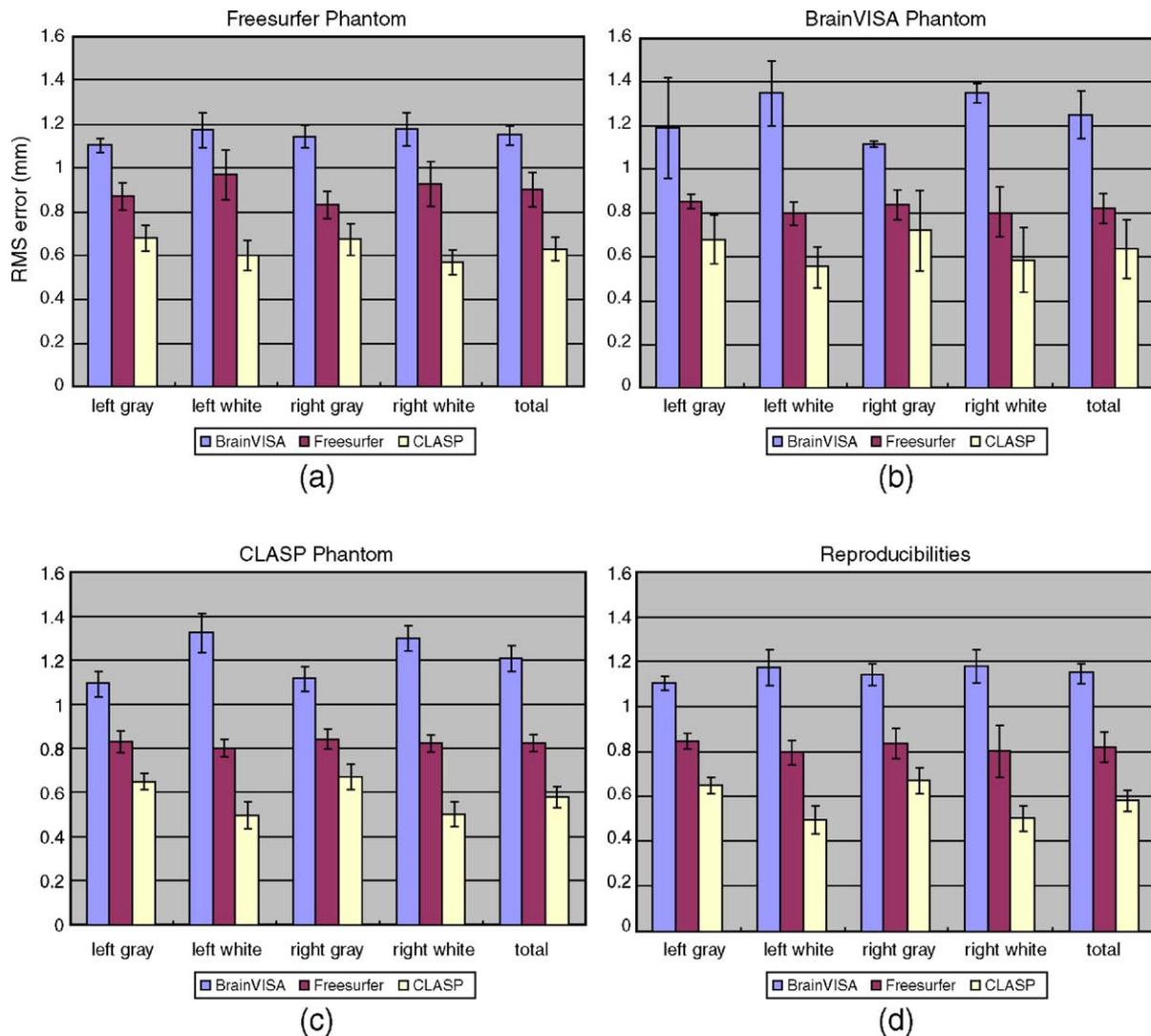


Fig. 5. Geometric errors (by surface-based validation). Geometric errors and reproducibilities of cortical surfaces were reconstructed by 4 tools. It was measured by calculating RMS distance between surfaces of “ground truth” and surfaces made from MR phantom images. (a), (b), and (c) show average RMS errors of cortical surfaces created from (a) Freesurfer phantom, (b) BrainVISA phantom, and (c) CLASP phantom respectively. (d) represents reproducibilities measuring RMS distance between the surface of “ground truth” and the surface made from MR phantom images by the tool which was used to create phantom itself. Note: y axis means RMS errors (mm).

“FD3” originally developed by Sarraïlle and DiFalco (Sarraïlle and Myers, 1994). The data sets which were used as inputs of “FD3” were the set of vertices of the cortical surface. In general, since the value of FD is not dependent on the number of vertices (Sarraïlle and Myers, 1994), we therefore used the original number of vertices created by each tool.

*Local density of points, total surface area.* The local density of points on the surface model is an important factor which describes a cortical surface. High density gives more accuracy, but computation takes longer in post-processing such as inflation and surface registration. Thus, to represent a cortical surface reasonably, the points should be distributed on the surface with the densities according to spatial gradients on the surfaces and at the same time the geometric accuracies also should be preserved. In this study, we measured the point distributions on the cortical surfaces by calculating the standard deviations (SD) of the edge lengths, and areas of triangle meshes on the surfaces. These

measurements give the general point distribution pattern on the surface. We normalized SDs by the average values of the edge lengths and triangle areas respectively. The total surface area has to also be considered. Generally, the reconstruction methods underestimate the geometrical information of original cortical surface, which could lead to the less surface area of reconstructed

Table 2

The results of topologic accuracies which were validated by Euler–Lhulier’s formula

Topology (Euler–Lhulier’s formula)			
	CLASP	Surfer	VISA
Accurate	16	12	7
Inaccurate	0	4	9

The Euler number should be 2 to be “accurate”.

The other cases are regarded as “inaccurate”.

Dale et al., 1999; Kriegeskorte and Goebel, 2001; Worsley, 1995.

model compared to that of the original cortical surface. We compared total surface areas of “test” surfaces to those of the “true” surfaces.

**Results**

*Validation of geometrical accuracy*

*Volume-based validation*

In volume-based validation, we validated only BrainVISA and CLASP. The validation was performed on the whole brain because the classified volume of CLASP is not separated into two hemispheres in the reconstruction process. Fig. 4 shows the results of validation. We compared the classified GM volume and surface-masked GM volume. In the volume-based evaluation, CLASP achieved more accurate results than BrainVISA (Fig. 4). The result showed the percentage of the each statistical volume including TP, FN, and FP compared to the whole volume of GM in the classified volume. CLASP showed a larger TP value, but a smaller FN and FP compared to BrainVISA.

*Surface-based validation*

The geometric errors and reproducibilities of the extracted cortical surfaces were measured by calculating the RMS distance between “true” and “test” surfaces. Figs. 5a, b, and c show average RMS errors when the “true” surface was created with Freesurfer, BrainVISA, and CLASP respectively. Fig. 5d shows the RMS surface distances when “true” and “test” surfaces were obtained with the same algorithm. CLASP showed the smallest errors compared to the other methods in both geometric accuracy and reproducibility validation. BrainVISA showed relatively large geometric errors compared to the other tools.

*Validation of mesh characteristics*

*Surface topology*

The topologies of cortical surface were validated by Euler–Lhulier’s formula (Eq. (1)). We used 16 cortical surfaces including pial/white and left/right surfaces from 4 ICMB MR data set. CLASP achieved results which are topologically correct because it is based on deformable surface model which preserves an initially correct spherical topology. Freesurfer and BrainVISA sometimes showed

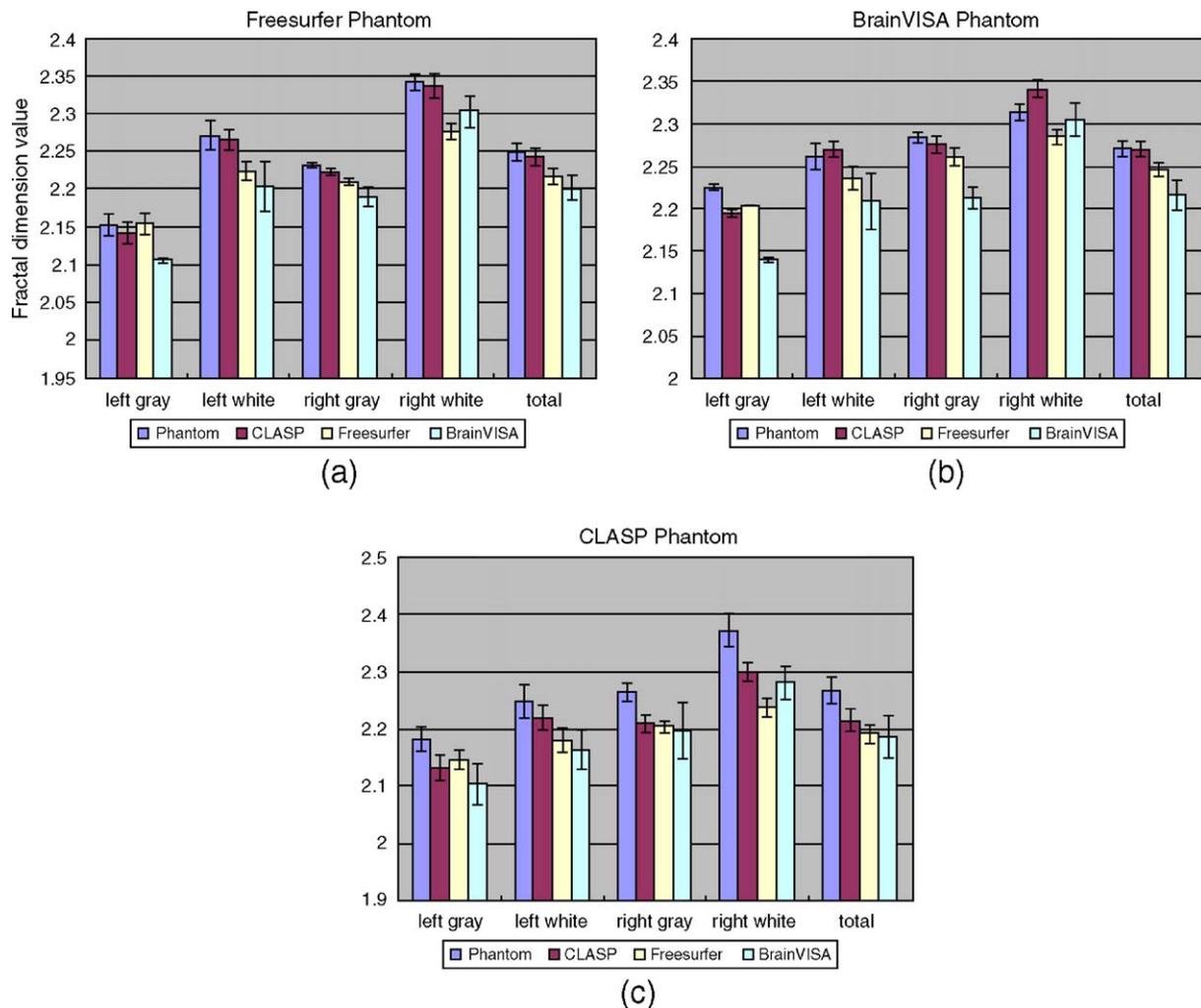


Fig. 6. Comparison of fractal dimension. This shows the FDs of cortical surfaces reconstructed by 4 tools. It was measured by calculating capacity values of FD (Sarraille and Myers, 1994). The image shows FD values of cortical surfaces created from (a) Freesurfer phantom, (b) BrainVISA phantom, and (c) CLASP phantom respectively. Note: y axis means the value of FD (mm).

topological errors on the surface model since these methods tessellate the segmented surface directly in parts and do not preserve topology. In this study, if the Euler number is 2, we considered the result to be “accurate”. The other cases were considered to be “inaccurate”. Table 2 shows the results of topological evaluation.

*Fractal dimension*

The FD values for the “true” surfaces were compared with those of the “test” surfaces. The results in Fig. 6 show the FD values of cortical surfaces created with the “true” surface obtained with (a) Freesurfer, (b) BrainVISA, and (c) CLASP respectively. It indicates that CLASP generally achieved the closest match between “true” and “test” FDs, regardless of which algorithm generated the “true” surface. The surface area study (Fig. 7) showed similar results to the FD validation, although there was a tendency for CLASP to overestimate the surface area of the pial boundary when using BrainVISA to generate the “true” surface. BrainVISA consistently produced the weakest agreement between “true” and “test” surfaces, even if BrainVISA was used to generate the “true” surface (Figs. 5–7).

*Local density of points, total surface area*

Fig. 7 shows surface areas obtained with each algorithm when the “true” surface was obtained using (a) Freesurfer, (b) BrainVISA, or (c) CLASP. The measurement was performed on left white, left gray, right white, right surface, and whole brain cortical surfaces respectively. When using the CLASP or Freesurfer “true” surface, CLASP achieved the closest agreement between “true” and “test”. However, when using the BrainVISA “true” surface, CLASP tended to overestimate the true area. In Fig. 7, the BrainVISA phantom panel at the bottom is showing white matter surface areas in excess of the corresponding gray matter surface areas in the phantom surfaces. This result may be due to the underestimation of the reconstructed pial surface especially on the area of deep sulci in BrainVISA. Fig. 8 shows the standard deviation (SD) in triangle areas when the “true” surface was obtained using (a) Freesurfer, (b) BrainVISA, or (c) CLASP. Fig. 9 shows a similar graph of SD in triangle edge length. These two indices showed that there is a greater variability in triangle dimensions with CLASP than the other methods.

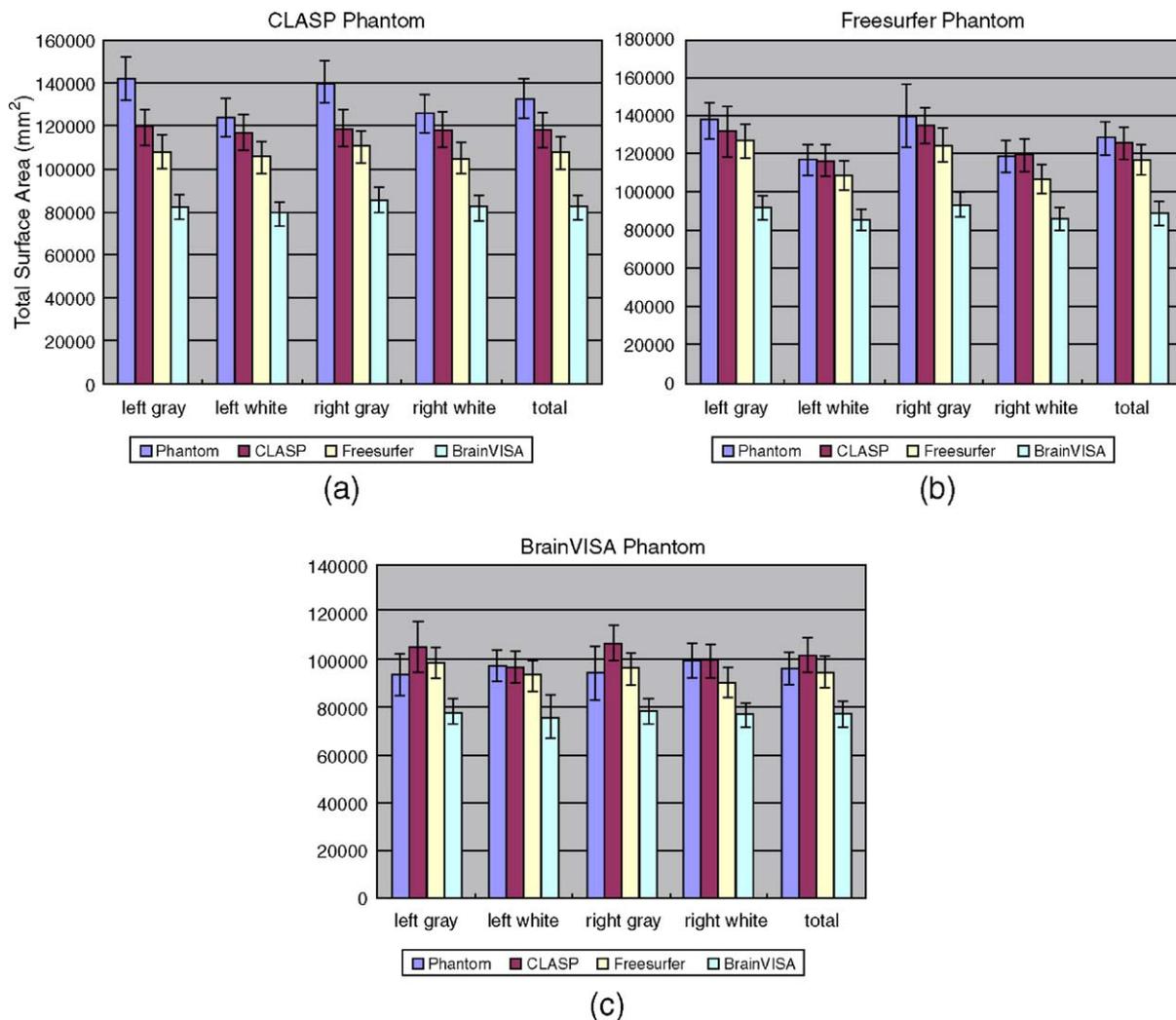


Fig. 7. Comparison of total areas. Areas of cortical surfaces were measured and evaluated. The image shows areas of surfaces created from (a) CLASP phantom, (b) Freesurfer phantom, and (c) BrainVISA phantom respectively. The measurements were performed on left white, left gray, right white, right gray, and whole brain cortical surfaces respectively. Note: y axis means the total area (mm<sup>2</sup>).

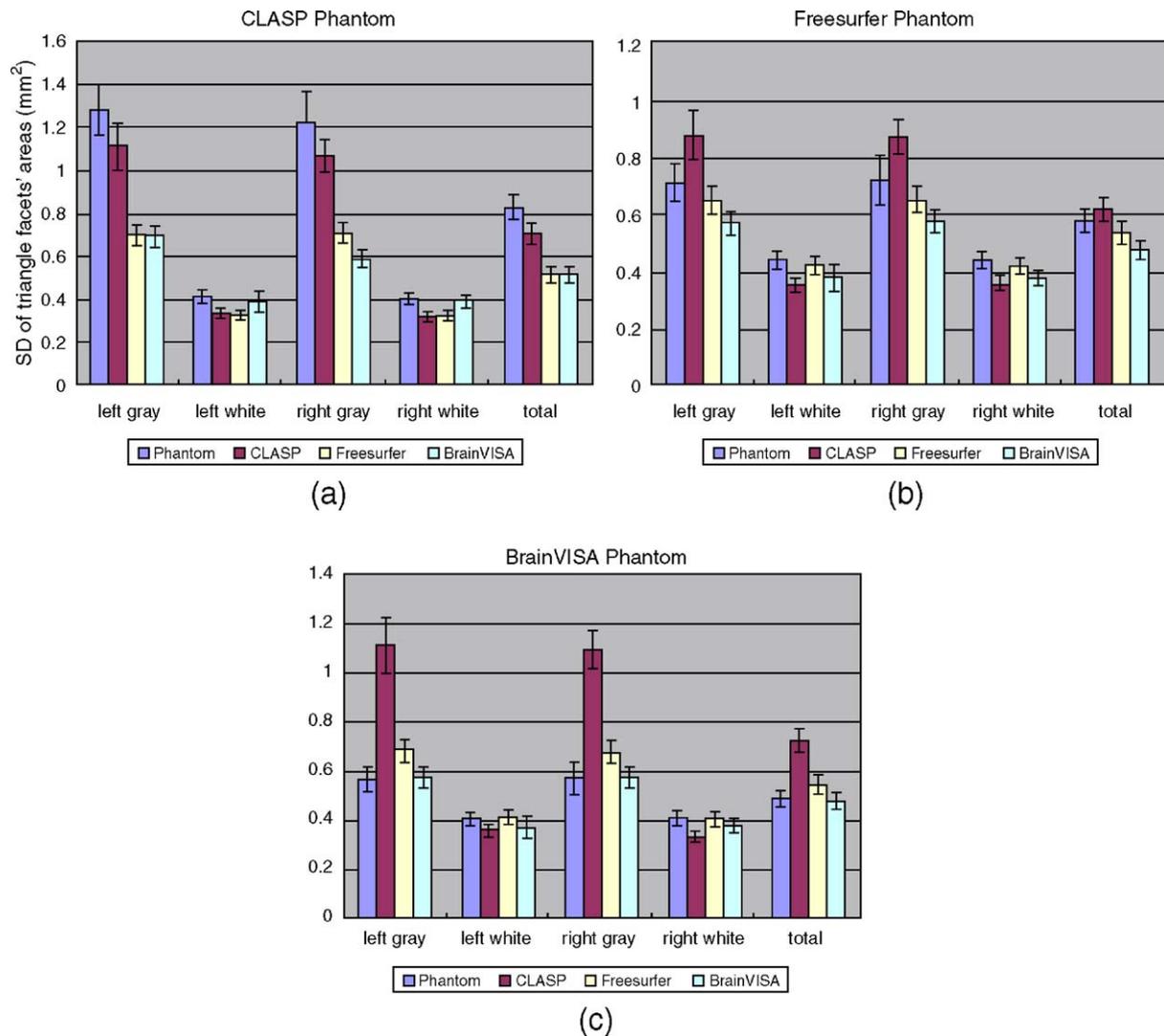


Fig. 8. SD of triangle facets' areas. SDs of triangle facets' areas were measured and evaluated. The image shows surface areas' SD created from (a) CLASP phantom, (b) Freesurfer phantom, and (c) BrainVISA phantom respectively. The measurements were performed on left white, left gray, right white, and right gray cortical surface respectively. Note: y axis means the SD of triangle facets' areas (mm<sup>2</sup>).

#### Visual inspection of cortical surfaces generated through MR phantom

Fig. 10 represents cortical surfaces reconstructed from (a) CLASP phantom, (b) Freesurfer phantom, and (c) BrainVISA phantom which are generated from a single MR image. The first column is composed of phantom surfaces created from a single MR image through each algorithm which are then used for the generation of MR phantom. The other columns are recreated cortical surfaces from the MR phantom simulated from the phantom surfaces (Kwan et al., 1996; Collins et al., 1998). The pial and white surfaces generated from MR phantoms by each tool look very identical to the corresponding phantom surfaces. The pial surface generated by Freesurfer appears more realistic than the BrainVISA and CLASP surfaces in the sense that it looks more like what a real brain looks like. Although CLASP delivered the most accurate geometry of the reconstructed surface, they look less realistic than Freesurfer surface having some creases on its surface. The cortical surfaces generated by BrainVISA appear simpler than the surfaces generated by two other tools especially in sulcal region. This implies that the geometric inaccuracies in BrainVISA

surfaces described above are related with a lack of ability to detect the deep sulci.

#### Computation time

As for the computation time, Freesurfer takes about 12 h for the cortical surface reconstruction including pre-processing. CLASP takes about 20 h including pre-processing. BrainVISA takes about 30 min for the reconstructions. The processing was performed on the Pentium 4 processor (2.0 GHz) of the PC.

#### Discussion

In this study, we conducted a quantitative comparison of the performance of three prominent cortical surface reconstruction tools (BrainVISA, Freesurfer, CLASP) using an MRI simulator. They have been chosen for the evaluation presented here according to some rules of accessibility to a tool, number of users, possibility of file format transformation, or coordinate normalization for the

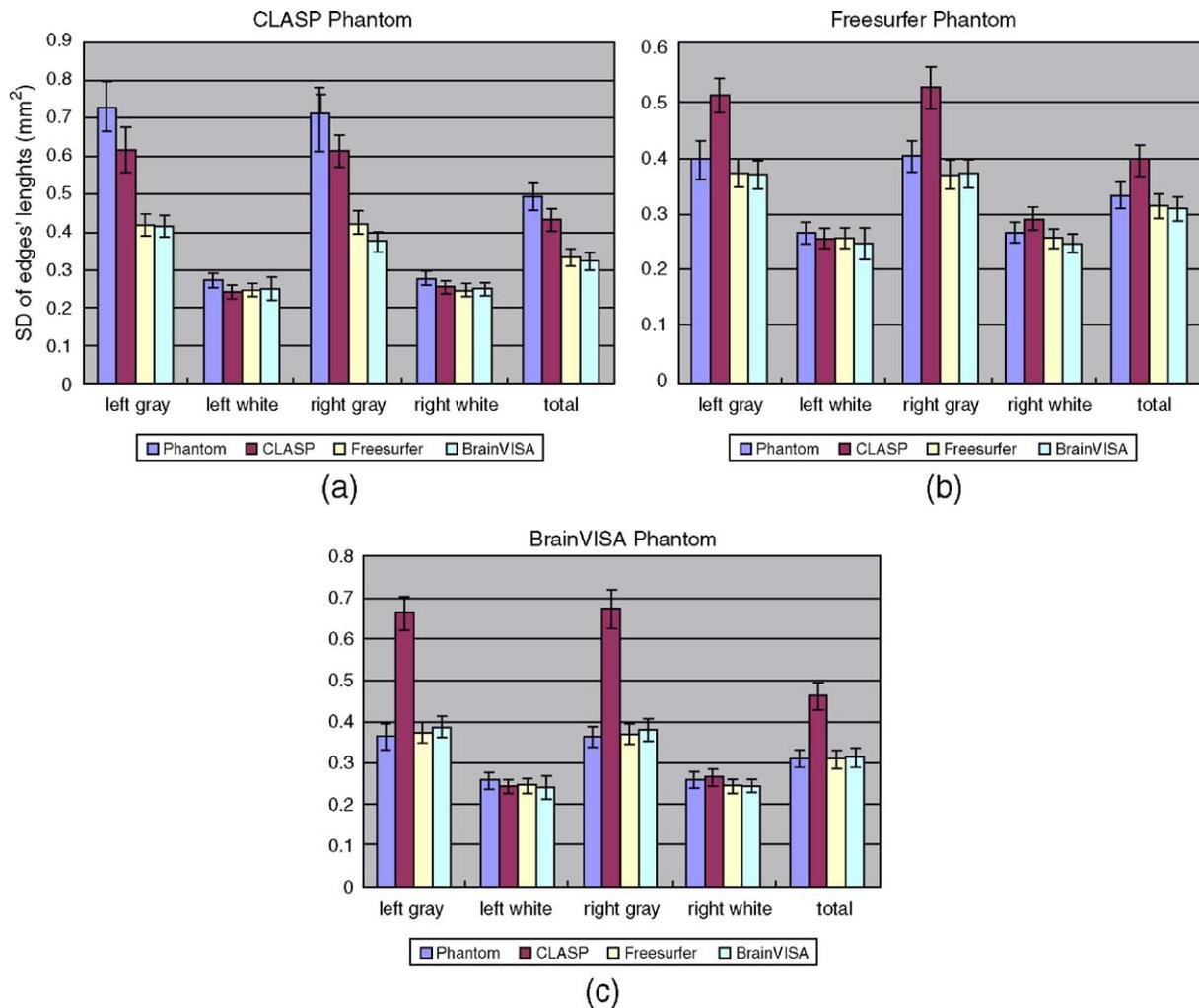


Fig. 9. SD of edges' lengths. SDs of edges' lengths were measured and evaluated. The image shows SD of edge lengths composing cortical surfaces created from (a) CLASP phantom, (b) Freesurfer phantom, and (c) BrainVISA phantom respectively. The measurements were performed on left white, left gray, right white, and right gray cortical surfaces respectively. Note: y axis means the SD of edge lengths (mm).

evaluation. CLASP is not a freely available tool and does not have large number of users, but we selected it because of its well establishment. We also tried to validate other packages such as BrainVoyager and SurfRelax, but there were some difficulties in file format transformation and coordinate normalization of brain volume or surface data. Thus, we remained validations of other packages including BrainVoyager and SurfRelax for future works. The evaluation strategy presented in this paper using MR phantom provides “gold standard” with which to access the performance of cortical surface reconstruction algorithms. The validation included quantitative assessment of geometrical accuracy and mesh characteristics such as surface topology, fractal dimension, surface area, and local surface sampling density.

#### Range of validation

In general, the pre-processing steps such as intensity inhomogeneity correction, skull stripping, and tissue classification are essential to the cortical surface reconstruction. We were mainly interested in the performance of the final surface reconstruction step out of whole reconstruction procedure including pre-processing. However, the validation was performed using the whole

procedure for each tool because it was difficult to separate the procedure into each individual step for the reconstruction (Dale et al., 1999; Kim et al., 2005; Mangin et al., 1995).

#### Geometric/topologic accuracy

Two forms of geometric evaluation were conducted: (i) a volume-based approach which compared the GM map produced by tissue classification with that produced by labeling all voxels between pial and white surfaces as GM and (ii) a simulation study in which “true” surfaces, initially extracted from real MR images, were used to generate simulated MRI volumes which were then analyzed by each algorithm to re-capture the original surfaces.

In volume-based validation, it is possible that any error could be due to an error in classified volume. In general, cortical surface reconstruction algorithms use the information not only of classified brain image but also of characteristics of cortical surface which are already known. Cortical surface has known characteristics such as continuity, smoothness, and a topology of the surface which are used as constraints in the reconstruction process. Therefore, if the reconstructed surface meets these known characteristics, it should

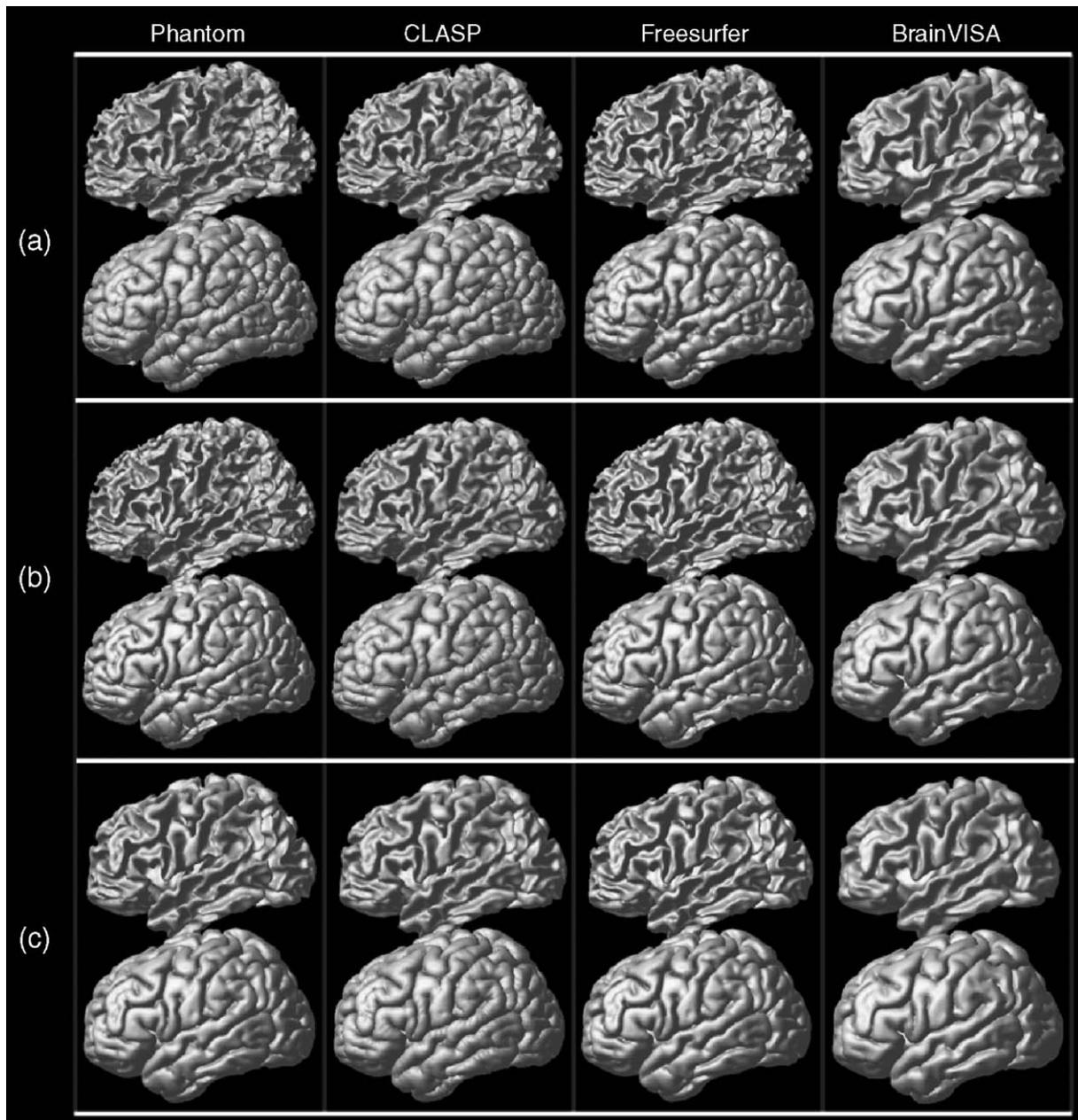


Fig. 10. Reconstructed surfaces using MR phantoms. Cortical surfaces reconstructed from (a) CLASP phantom, (b) Freesurfer phantom, and (c) BrainVISA phantom which are generated from a single MR image. The first column is composed of phantom surfaces created from a single MR image through each algorithm. The other columns are recreated cortical surfaces from the phantom surfaces.

follow the boundary of the input classified volume as possible as it can. The volume-based comparison evaluated fidelities based only on input classified volumes of BrainVISA and CLASP. Although inaccurate voxel-based classification could induce any error in the reconstructed surface and lead to the wrong validation, volume-based validation presents a measure for the automatic validation of entire cortical surface without additional intensive processing such as manual drawing.

While relatively straightforward, the volume-based approach is a global index of agreement. It does not provide any information on the detailed accuracy of surface extraction. This is a difficult problem because there is, in general, no “ground truth” with which the extracted cortical surface can be compared. In our surface-

based evaluation, we used surfaces extracted from real MRI volumes to generate simulated MR images, thereby defining these original surfaces as “true”. The ability of each algorithm to recapture this “true” surface from the simulated MRI volume, which now includes many confounding data acquisition factors (noise, loss of contrast, partial volume effects, inhomogeneity, etc.), could now be quantified. Although a simulator does not incorporate every aspect of real data, this strategy provides quantitative lower-bound performance metrics with which to assess algorithm performance.

The cortical surface must be topologically equivalent to a sphere because topological errors in the reconstructed cortical surface can lead to incorrect results of the post-processing such as

flattening, cortical surface registration, and thickness measure. These erroneous results of post-processing would reduce analytic power of morphological or functional mapping study. For topological validation, we used a simple and automatic approach which assesses the topology of the whole surface by calculating its Euler characteristic (Dale et al., 1999; Worsley, 1995). This does not show the degrees of errors caused by topological inaccuracies in detail. However, it indicates the general pattern of topological accuracies.

#### *Local density of mesh*

In the point distribution study (Figs. 8 and 9), CLASP showed the greatest distribution in surface triangle area and edge length. In some cases, this resulted in closest agreement with the true surface characteristics (Figs. 8a and 9a), but, in others, CLASP test surfaces tended toward greater heterogeneity than the true surface. This was true even if CLASP was used to generate the true surface (Figs. 8c and 9c). Only the pattern of point distribution cannot explicitly explain how great the reconstructed surface is, however, it indicates some surface characteristics which could be used in post-processing such as surface registration, inflation, and flattening.

#### *Differences between the algorithms*

BrainVISA uses a kind of bottom–up approach which is based on a sequence of topologically simple points additions or deletions. In general, bottom–up approach is very fast but more inaccurate for the cortical surface reconstruction than top–down approach like deformable surface which is computationally heavy. The reconstruction method in BrainVISA has been mainly developed for the automatic detection and recognition of the main cortical sulci. Thus, geometrical accuracy of cortical surface could be less important than the other tools which are developed mainly for the shape analysis such as cortical thickness measure. Surface processed by Freesurfer had a good geometric estimation of the cortical surface compared to BrainVISA. It also appears more realistic than the BrainVISA and CLASP surfaces in the sense that it looks more like what a real brain looks like. However, the estimation of the cortical surface in its automatic procedure usually has some topologic defects because it uses tessellation approach in the reconstruction of white surface. It should be noted that Freesurfer allows for manual correction of topological errors generated during the automated surface extraction procedure. But, manual intervention becomes less practical as the number of surfaces requiring correction increases.

CLASP delivered the most accurate geometry of the reconstructed surface. In visual inspection, it looks less realistic than Freesurfer surface having some creases on its surface. These visual defects may be due to the insufficiency of local smoothness. However, in the point of view of the morphological analysis, it is not a concern that cortical surface has some visual defects like small creases on its surface caused by a lack of local smoothness. CLASP was only one among three tools which generates the cortical surface topologically equivalent to a sphere without additional process such as manual editing because it uses a deformable surface model in its whole reconstruction process.

In summary, of the three algorithms studied, the CLASP algorithm demonstrated the most accurate surfaces after fully

automated analysis while BrainVISA produced the least accurate. On the other hand, BrainVISA is much faster in achieving its result. The evaluation strategy presented in this paper using MR phantom provides “gold standard” with which to access the performance of cortical surface reconstruction algorithms and enables the validation of the specific performance which is dependent on applications such as thickness measure, surface area, fractal dimension, and sulcal depth. In future work, we will investigate in more detail the local differences between cortical surfaces reconstructed by each tool. This will allow us to identify in which cortical areas each tool is most susceptible to errors of geometric accuracies and cortical thickness.

#### **Acknowledgment**

This research was supported by a grant (M103KV0100-1403K220101420) from Brain Research Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology of the Republic of Korea.

#### **References**

- Cachier, P., Mangin, J.F., Pennec, X., Rivière, D., Papadopoulos-Orfanos, D., Régis, J., Ayache, N., 2001. In: Niessen, W., Viergever, M. (Eds.), *Multisubject Non-Rigid Registration of Brain MRI Using Intensity and Geometric Features*. Springer Verlag, Berlin Heidelberg, pp. 734–742.
- Carman, G.J., Drury, H.A., Van Essen, D.C., 1995. Computational methods for reconstructing and unfolding the cerebral cortex. *Cereb. Cortex* 5 (6), 506–517.
- Chung, M.K., Worsley, K.J., Robbins, S., Paus, T., Taylor, J., Giedd, J.N., Rapoport, J.L., Evans, A.C., 2003. Deformation-based surface morphometry applied to gray matter deformation. *NeuroImage* 18 (2), 198–213.
- Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18 (2), 192–205.
- Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C., 1998. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.* 17 (3), 463–468.
- Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J. Cogn. Neurosci.* 5 (2), 162–176.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 9 (2), 179–194.
- Essen, D.C., Van Lewis, J.W., Drury, H.A., Hadjikhani, N., Tootell, R.B.H., Bakircioglu, M., Miller, M.I., 2001. Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res.* 41, 1359–1378.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9 (2), 195–207.
- Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8 (4), 272–284.
- Gordon, D., Udupa, J., 1989. Fast surface tracking in three-dimensional binary images. *Comput. Vis. Graph. Image Process.* 45 (6), 196–214.
- Grimson, W.E.L., Ettinger, G., Kapur, T., Leventon, M., Wells, W., Kikinis, R., 1998. Utilizing segmented MRI data in image-guided surgery. *Int. J. Pattern Recogn. Artif. Intell.* 11, 1367–1397.
- Han, X., Xu, C., Prince, J.L. 2001. A topology preserving deformable model using level sets, pp. 765–770.

- Hearn, D., Baker, M.P., 1997. Computer Graphics, C Version. Prentice Hall, Inc., New Jersey, pp. 117–126.
- Jaume, S., Macq, B., Warfield, S.K., 2002. In: Dohi, T., Kikinis, R. (Eds.), Labeling the Brain Surface Using a Deformable Multiresolution Mesh. Springer Verlag, Berlin Heidelberg, pp. 451–458.
- Kim, J.J., Crespo-Facorro, B., Andreasen, N.C., O'Leary, D.S., Zhang, B., Harris, G., Magnotta, V.A., 2000. An MRI-based parcellation method for the temporal lobe. *NeuroImage* 11 (4), 271–288.
- Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I., Evans, A.C., 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage* 27 (1), 210–221.
- Kriegeskorte, N., Goebel, R., 2001. An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical mr volumes. *NeuroImage* 14 (2), 329–346.
- Kruggel, F., Bruckner, M.K., Arendt, T., Wiggins, C.J., von Cramon, D.Y., 2003. Analyzing the neocortical fine-structure. *Med. Image Anal.* 7 (3), 251–264.
- Kwan, R.K.S., Evans, A.C., Pike, G.B., 1996. An extensible MRI simulator for post-processing evaluation. *Lect. Notes Comput. Sci.* 1131, 135–140.
- Ma, C.-M., Wan, S.-Y., 2001. A medial-surface oriented 3-d two-subfield thinning algorithm. *Pattern Recogn. Lett.* 22, 1439–1446.
- MacDonald, D., Kabani, N., Avis, D., Evans, A.C., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage* 12 (3), 340–356.
- Magnotta, V.A., Andreasen, N.C., Schultz, S.K., Harris, G., Cizadlo, T., Heckel, D., Nopoulos, P., Flaum, M., 1999. Quantitative in vivo measurement of gyrification in the human brain: changes associated with aging. *Cereb. Cortex* 9 (2), 151–160.
- Manceaux-Demiau, A., Bryan, R.N., Davatzikos, C., 1998. A probabilistic ribbon model for shape analysis of the cerebral sulci: application to the central sulcus. *J. Comput. Assist. Tomogr.* 22 (6), 962–971.
- Mangin, J.-F., Frouin, V., Bloch, I., Regis, J., López-Krahe, J., 1995. From 3D MR images to structural representations of the cortex topography using topology preserving deformations. *J. Math. Imaging Vis.* 5 (4), 297–318.
- Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *NeuroImage* 2 (2), 89–101.
- Pont, J.C., 1974. *La Topologie Algébrique: Des Origines à Poincaré*. Presses Universitaires de France, Paris.
- Preparata, F.P., Shamos, M.I., 1985. *Computational Geometry: an Introduction*. Springer-Verlag, Berlin.
- Sandor, S., Leahy, R., 1997. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Trans. Med. Imag.* 16, 41–54.
- Sarraile, J.J., Myers, L.S., 1994. FD3: a program for measuring fractal dimension. *Educ. Psychol. Meas.* 54, 94–97.
- Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Trans. Med. Imag.* 17, 87–97.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain*. New York, Thieme Medical Publishers.
- Thompson, P.M., MacDonald, D., Mega, M.S., Holmes, C.J., Evans, A.C., Toga, A.W., 1997. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *J. Comput. Assist. Tomogr.* 21 (4), 567–581.
- Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield, C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W., 2001. Cortical changes in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb. Cortex* 11 (1), 1–16.
- Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Dordrell, D.M., et al., 2003. Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci.* 23 (3), 994–1005.
- Tosun, D., Prince, J.L., 2001. *Hemispherical Map for the Human Brain Cortex*. SPIE Press, Bellingham, WA, pp. 290–300.
- Van Essen, D.C., Drury, H.A., 1997. Structural and functional analyses of human cerebral cortex using a surface-based atlas. *J. Neurosci.* 17 (18), 7079–7102.
- Worsley, K.J., 1995. Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. Appl. Probab.* 27, 943–959.
- Yezzi, A.J., Prince, J.L., 2003. An evolution PDE approach for computing tissue thickness. *IEEE Trans. Med. Imag.* 22 (10), 1332–1339.
- Zijdenbos, A.P., Evans, A.C., Riahi, F., Sled, J.G., Chui, J., Kollokian, V., 1996. Automatic quantification of multiple sclerosis lesion volume using stereotaxic space. *Proc. 4th Intl. Conf. on Visualization in BioMed. Computing VBC*, pp. 439–448.
- Zijdenbos, A.P., Forghani, R., Evans, A.C., 1998. Automatic quantification of MS lesions in 3D MRI brain data sets: validation of INSECT. *MICCAI*, 439–448.
- Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.J., 1988. The human pattern of gyrification in the cerebral cortex. *Anat. Embryol. (Berl.)* 179 (2), 173–179.