# A Novel Quantitative Validation of the Cortical Surface Reconstruction Algorithm Using MRI Phantom: Issues on Local Geometric Accuracy and Cortical Thickness

Junki Lee[1], Jong-Min Lee[1], Jae-Hun Kim[1], In Young Kim[1], Alan C. Evans[2], and Sun I. Kim[1]

[1] Dept. Biomedical Engineering, Hanyang University,
17 Haengdang-dong Sungdong-gu, Seoul, 133-791, Korea
[2] McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada
`jklee@bme.hanyang.ac.kr`

**Abstract.** 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 describe the true cortical surface in both geometry and topology. However there has been no explicit method for the quantitative evaluation of the whole-cortical-surface models. In this study, we present a novel phantom-based evaluation method of the cortical surface reconstruction algorithm and quantitatively validated the local morphometric accuracy of CLASP which is one of the well-established reconstruction methods. The evaluation included local geometrical accuracy and performance of cortical thickness measure. The validation study revealed that there were some underestimations of cortical thickness measure using CLASP in the ventral and sulcal areas of the cortex and overestimations in the gyral areas and inferior temporal lobe. This study could present a generic metric for the quantitative evaluation of cortical surface reconstruction algorithm.

## 1 Introduction

The relationship between the cortical morphometry and the cytoarchitectonic and functional organization of the underlying cortex is a subject of much current interest and debate. Reconstructed cortices enable the visualization and the study of the sulcal and gyral patterns of an individual subject [1] and allow morphometric measurements such as cortical thickness [2]. To fulfill these needs, it is important to faithfully represent true cortical surface in terms of geometry. This task is difficult, because of artifacts such as image noise, partial volume effects, and intensity inhomogeneities [2]. 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.

For the evaluation of the cortical surface reconstruction, there have been various methods such as a visual validation [3], a repeatability test of the procedure [4], and a

landmark-based validation [5]. In visual validation, the intersections of the reconstructed surface with the underlying MRI intensity data are inspected. Although this is very intuitive and easy to perform, it cannot assess the whole cortical surface in objective and quantitative way. In repeatability test, it could answer whether the procedure results in similar reconstructed surface models from different scans of the same subject. However, the accuracies of reconstructions could not be measured. In the landmark-based validation, experts select several landmark points or draw some landmark areas on the MR brain volume image. Then the closest point from each landmark to the corresponding surface is found and the distance between them computed. This distance from each landmark to the estimated cortical surface serves as a measure of accuracy. This is a quantitative validation approach. However, when experts select or draw the landmarks, some biases could be induced. In addition, it is difficult to validate the whole boundaries of cerebral cortex with this approach.

In this study, we present a novel method for the quantitative evaluation of the cortical surface reconstruction algorithm using an MRI simulator generating a realistic MRI incorporating the calculation of noise and partial volume effects. The evaluation strategy provides "gold standard" with which to access the performance of cortical surface reconstruction algorithms In previous studies, we developed a method called CLASP [3], an enhanced version of the iterative-morphing method first developed by MacDonald et al. and validated this method using a phantom-based approach. In this paper, we validated CLASP with phantom-based approach focused on the local geometric accuracy and the performance of cortical thickness measure.

## 2    Methods

In this phantom-based evaluation, an MRI simulator [6, 7] was used to create a "gold standard" with which to access the performance of cortical surface reconstruction algorithms directly. The accuracies of cortical geometries and thickness measures were evaluated using phantom-based comparison detailed below. T1-weighted MR images (n=12) with 1.0 mm × 1.0 mm × 1.0 mm resolution and 181 × 217 × 181 voxel dimensions were selected randomly from the datasets of the International Consortium for Brain Mapping (ICBM) [8] for this evaluation.

### 2.1    Cortical Surface Reconstruction and Thickness Measure

In each brain, the white (i.e. GM/WM boundary) and pial (i.e. GM/CSF boundary) surface was extracted by CLASP [3]. The CLASP algorithm consists of several stages as follows:

Acquired T1 MR images are preprocessed by intensity inhomogeneity correcti on [9] and spatial normalization to stereotaxic space [10]. Preprocessed images are classified into GM, WM, and CSF tissues [11]. The classified volumes are divided into left and right hemispheres for reconstructing two hemispheric cortical surfaces. The WM surface is reconstructed by deforming a spherical polygon model to the white matter boundary. A Laplacian field is generated between the WM surface resampled

to voxel space and a skeletonized CSF fraction image. The GM surface is initiated from the WM surface and is expanded to the boundary between GM and CSF along the Laplacian field. These stages are described in greater detail in the original paper [3]. In order to measure the cortical thickness, several preprocessing algorithms were required.
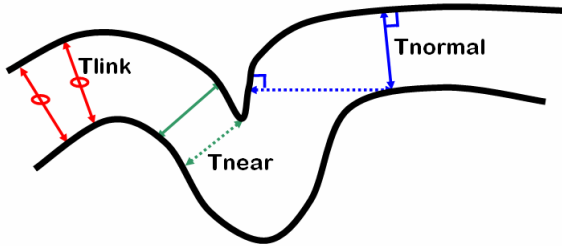


**Fig. 1.** Three metrics for cortical thickness measure: Tlink is the distance between corresponding points. Tnear is the distance from a vertex on the outer surface to the nearest point on the inner surface. Tnormal is the distance from a vertex on the outer surface to the nearest point on the inner surface in the direction of the surface normal. The dot lines represent improper estimations of thicknesses.

Given a surface representing the gray/CSF boundary and another representing the gray/white boundary, the cortical thickness measure could be performed [2]. Among various metrics for thickness measure, Tlink was used for the evaluation of thickness measure, which means the distance from a vertex on the outer surface to the corresponding vertex on the inner surface, as defined by the linkages used in the two-surface deformation of the CLASP algorithm (Fig. 1).The Tlink measure attempts to use the correspondence between points on the surface, which provides a measure of thickness that is less sensitive to fluctuations in surface normal and areas of high curvature. White and pial surfaces reconstructed by CLASP had the same vertex number, and the correspondence of each vertex between surfaces was defined. Thus, the cortical thickness was easily measured using the Tlink method [2]. Tnear is another approach for thickness measure which indicates the distance from a vertex on the outer surface to the nearest point on the inner surface (Fig. 1). We used Tnear metric for the measure of geometric differences between two surfaces, because it provides robust performance without the explicit correspondence of each vertex between two surfaces.

## 2.2  Phantom-Based Evaluation

We evaluated the cortical surface reconstruction method using a phantom-based procedure. Since there is no readily-available "gold standard" with which to assess the performance of the surface extraction algorithm, we approached the problem with an MRI simulator [6, 7]. This simulator generates a realistic MRI incorporating the calculation of noise and partial volume effects. For the validation, the following steps were performed:

1) pial and white surfaces were first extracted from the real MRI volume (Fig. 2 (a)). 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 (Fig. 2 (b)). 3) A T1 MR image was simulated from the phantom using the same parameters as the real data acquisition (TR=18ms, TE=10ms, slice thickness=1mm) (Fig. 2 (c)). 4) Additional substructures (skull, basal ganglia) were added from the real MRI (Fig. 2 (d)). 5) Pial and white surfaces were then extracted from the simulated MRI volume (Fig. 2 (e)). 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. CLASP was applied to the simulated MRI to generate a "test" surface. The root mean square (RMS) error between "true" and "test" surfaces then provided a measure of accuracy in surface extraction. To measure RMS error between the "true" and "test" surfaces, we calculated the distances using Tnear metric which are detailed in the previous section [2]. To validate thickness measure, we calculated the differences between "true" thicknesses and "test" thicknesses measured by Tlink. The validation of geometric accuracy and thickenss measure were performed on regional areas as well as on whole brain. To measure the local errors, we used non-rigid registration of 2-D cortical surfaces which is to find the corresponding regions of the surface model between subjects. This 2-D registration method is described in detail in next section.
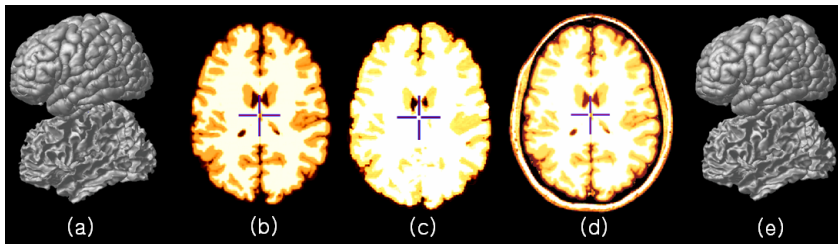


**Fig. 2.** Process of the evaluation using phantom (a) created cortical surface (b) surface masked volume (c) digital brain phantom (d) phantom including skull (e) recreated cortical surface from phantom

## 2.3   Statistical Analysis

To find the thickness of corresponding regions of the surface model between the groups, the thickness value was spatially normalized using surface based 2-D registration [12]. In the CLASP algorithm, since the cortical surfaces start from a spherical polygon model, the vertices are easily transformed to the spherical model. Vertices of each subject are nonlinearly registered to an average template on the sphere by matching crowns of gyri between subjects using a geodesic distance map [12]. Using the transformation,

thickness information and local geometric errors on the vertices were transformed to a template. Diffusion smoothing, which generalizes Gaussian kernel smoothing, with 20 mm FWHM (full-width half-maximum) was used to increase the signal-to-noise ratio of the measured cortical thickness and local geometric accuracy [13]. We then calculated mean difference between and error rate which is a difference value normalized by a thickness of "true" surface.

## 3   Results

The global geometric errors of the extracted cortical surfaces were measured by calculating the mean RMS distance between "true" and "test" surfaces. Fig. 3 shows mean RMS errors. The mean RMS errors was measured by Tnear ($0.42 \pm 0.096$ mm) and Tnormal ($2.73 \pm 0.94$ mm) respectively (Fig. 3). The local geometric errors were measured after finding correspondence between subjects. Fig. 4 shows maps of local geometric accuracies which were measured by RMS distance. RMS distance was measured by Tnear.  Incorrect estimations were found in the ventral and sulcul areas of the cortex.
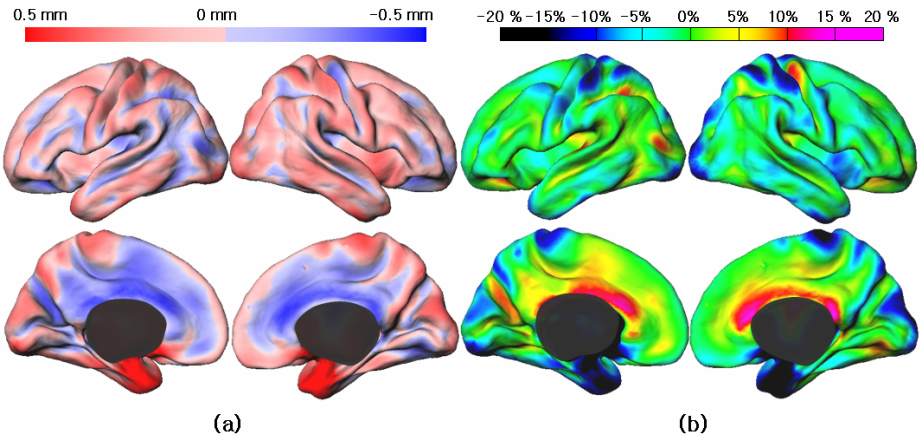
Cortical thicknesses using Tlink were measured and evaluated. Global/local mean differences between cortical thicknesses were measured from "true" and "test" surfaces respectively. Global mean difference of cortical thicknesses were measured on the left hemisphere ($-0.061\pm0.24$ mm), the right hemisphere ($-0.080\pm0.20$ mm), and the whole brain ($-0.07\pm0.22$ mm). Validation of local mean difference of the thickness revealed some errors in the specific areas (Fig. 5). There were some overestimations of cortical thickness measure from the "test" surface (i.e. the surface reconstructed from MRI phantom) in the ventral area and sulcus of the cortex. Underestimations were also found in the gyrus, occipital lobe, and inferior parts of temporal lobe. We measured also error rate of the cortical thickness. These values are normalized mean differences by the thicknesses of "true" surface. The overestimations of the cortical thicknesses were up to about 10% in the cingulated gyrus and some sulcal parts of the cortex. The underestimations of the cortical thicknesses were up to about 15% in the gyrus, occipital lobe, and inferior parts of temporal lobe.
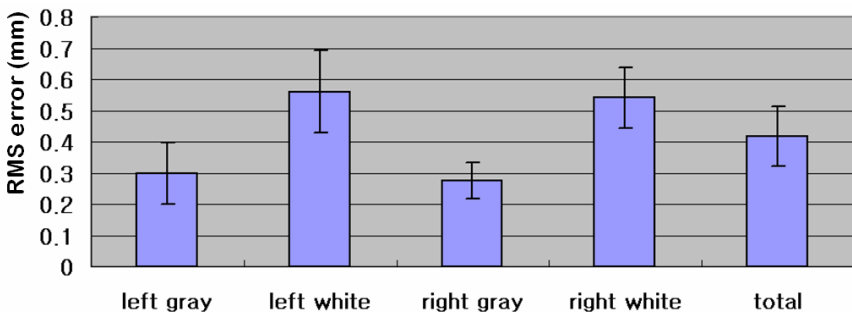


**Fig. 3.** Geometric errors of cortical surfaces. It was measured by calculating RMS distance between surfaces of "gold standard" and surfaces made from MR phantom images. Each column shows average RMS errors of cortical surfaces (pial/white surfaces of left/right hemispheres) created from CLASP.
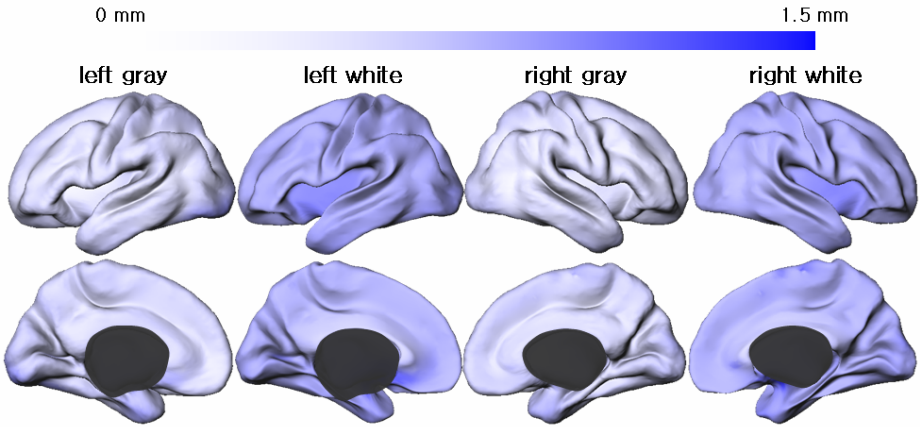
**Fig. 4.** Local geometric accuracies which mean the RMS distances between "true" and "test" surfaces. RMS distance was measured by Tnear. Incorrect estimations were found in the ventral and sulcul areas of the cortex.



**Fig. 5.** (a) mean difference and (b) error rate between cortical thicknesses measured from "true" and "test" surfaces respectively. Error rate is normalized mean difference by the thicknesses of "true" surface. There were some overestimations of cortical thickness measure from the "test" surface (i.e. the surface reconstructed from MRI phantom) in the ventral area and sulcus of the cortex. The overestimations of the cortical thicknesses were up to about 10% in the cingulated gyrus and some sulcal parts of the cortex. Underestimations were also found in the gyrus, occipital lobe, and inferior parts of temporal lobe. The underestimations of the cortical thicknesses were up to about 15% in the gyrus, occipital lobe, and inferior parts of temporal lobe.

## 4   Discussion and Conclusions

Given the variety of cortical surface reconstruction methods, choosing an appropriate algorithm for an existing or a new problem can be quite a challenging task. Therefore,

an objective validation is necessary to provide the intrinsic characteristics of the methods, evaluate their performances and limitations. Moreover, while developing a new method, validation is essential in order to compare new and existing methods and estimate of the optimal processing parameters. However, since MR imaging, like many other medical modalities, is an in vivo study, validation becomes even more challenging and its issues are often overlooked. In this study, we conducted a quantitative validation of the performance of the cortical surface reconstruction tool using an MRI simulator. The evaluation strategy presented in this paper using MR phantom provides "gold standard" with which to access the performance of cortical surface reconstruction algorithms. We performed simulation study in which "true" surfaces, initially extracted from real MR images, were used to generate simulated MRI volumes which were then analyzed by the cortical surface reconstruction tool to re-capture the original surfaces. The ability of an algorithm to re-capture 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.

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

Despite of many advantages, a limitation to this phantom-based approach is that the MR phantom generated from any particular algorithm could be biased toward that method. This bias would underestimate the errors measured from the phantom-based validation. In order to overcome this limitation, it is possible that we could use another algorithm to reconstruct the phantom surfaces and cross-validate between various algorithms. The evaluation strategy presented in this paper using MR phantom provides "gold standard" with which to access the performance of cortical surface reconstruction algorithm and enables the evaluation of the specific performance which is dependent on applications such as thickness measure, surface area, fractal dimension and sulcal depth.

# References

[1] D. C. Van Essen, J. W. Lewis, H. A. Drury, N. Hadjikhani, R. B. Tootell, M. Bakircioglu, and M. I. Miller, "Mapping visual cortex in monkeys and humans using surface-based atlases," *Vision Res*, vol. 41, pp. 1359-78, 2001.

[2] D. MacDonald, N. Kabani, D. Avis, and A. C. Evans, "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI," *Neuroimage*, vol. 12, pp. 340-56, 2000.

[3]  J. S. Kim, V. Singh, J. K. Lee, J. Lerch, Y. Ad-Dab'bagh, D. MacDonald, J. M. Lee, S. I. Kim, and A. C. Evans, "Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification," *Neuroimage*, vol. 27, pp. 210-21, 2005.

[4]  N. Kriegeskorte and R. Goebel, "An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical mr volumes," *Neuroimage*, vol. 14, pp. 329-46, 2001.

[5]  X. Han, D. L. Pham, D. Tosun, M. E. Rettmann, C. Xu, and J. L. Prince, "CRUISE: cortical reconstruction using implicit surface evolution," *Neuroimage*, vol. 23, pp. 997-1012, 2004.

[6]  D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans Med Imaging*, vol. 17, pp. 463-8, 1998.

[7]  R. K. S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for postprocessing evaluation." *Lecture Notes in Computer Science*, vol. 1131, pp. 135-140, 1996.

[8]  J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM)," *Neuroimage*, vol. 2, pp. 89-101, 1995.

[9]  D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Transaction on Medical Imaging*, vol. 17, pp. 463-468, 1998.

[10] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *J Comput Assist Tomogr*, vol. 18, pp. 192-205, 1994.

[11] A. P. Zijdenbos, R. Forghani, and A. C. Evans, "Automatic quantification of MS Lesions in 3D MRI brain data sets: Validation of INSECT," presented at Medical Image Computing and Computer Aided Intervention, 1998.

[12] S. Robbins, "Anatomical Standardization of the Human Brain in Euclidean 3-space and on the Cortical 2-Manifold.," *Ph.D. Thesis*, vol. School of Computer Science, McGill University, Montreal, Que., Canada, 2003.

[13] M. K. Chung, K. J. Worsley, S. Robbins, T. Paus, J. Taylor, J. N. Giedd, J. L. Rapoport, and A. C. Evans, "Deformation-based surface morphometry applied to gray matter deformation," *Neuroimage*, vol. 18, pp. 198-213, 2003.