Automatic detection of cortical thickness measurement errors using Support Vector Regression

V. Zagorodnov[1], and K. H. Reddy[2]
[1]Nanyang Technological University, Singapore, NA, Singapore, [2]Nanyang Technological University

Introduction. Brain **c**ortical thickness can be measured without any human intervention, using a variety of existing fully automated segmentation algorithms [1-3]. However, despite steady improvements in these algorithms over recent years, segmentation errors still frequently occur and require tedious manual review and editing of the segmentation results [4]. Most of such errors are caused by dura attachments, which have not been removed by skull stripping and cause the overgrowth of the pial surface [5], or inadequately corrected intensity nonuniformity [6]. We propose a framework based on Support Vector Regression (SVR) to automatically highlight the errors in cortical thickness measurements, obtained using FreeSurfer segmentation pipeline [1]. Our approach exploits high correlations between regional cortical thicknesses [5], which allows "guessing" what the correct measurement should be for a specific region, based on the rest of the brain measurements.

Methods. FreeSurfer labels the brain into approximately 100 structures (exact number varies depending on the software version) and provides aggregate regional measures, such as volumes and average thicknesses, for these structures. Regional brain measurements are highly correlated with each other [7], a property has recently been interpreted in the context of structural connectivity. But high correlations also imply that the cortical thickness in a specified region can be predicted from the rest of the brain measurements. While standard regression can be used for prediction, due to high dimensionality of the data this will likely to result in overfitting and hence poor generalization. Hence we used SVR for prediction, which is known to be robust to overfitting [8]. Separate SVR predictors were built for each region labeled by FreeSurfer. We used the difference between predicted and actual measurements as a proxy for potential segmentation error. This difference is never equal to zero due to prediction error. Hence the regions were highlighted as wrongly segmented only when the difference exceeded the two standard deviations of the prediction error.

For evaluation we used a data set containing 261 normal subjects (age 56-88) acquired on a Siemens Tim Trio 3T scanner using a standardized imaging procedure (TR = 2530, TE=1.64, 3.5, 5.36, 7.22ms (multi-echo acquisition); 1200ms; flip angle = 7°; Bandwidth 651Hz/pixel; FOV 256 x 240mm, 256 x 256 matrix; resulting voxel dimensions: $1.0\times1.0\times1.0mm^3$). The dataset was processed by FreeSurfer 4.2.0 segmentation pipeline and then examined by an expert, who manually edited the most obvious segmentation errors in the estimated pial and white matter (WM) surfaces, and then rerun FreeSurfer pipeline to re-estimate the cortical thicknesses (Fig. 1). This set of corrected regional measurements was then used for training of our regressor and evaluation of prediction errors. Since even the corrected data was not completely error-free, the prediction error was measured as Mean Absolute Deviation (MAD), which is known to be robust to outliers.

Results. Our experiments indicate that some regions can be more accurately predicted than others. The prediction error varied between regions, from MAD=0.05mm for the right superior parietal region to MAD=0.21mm for the left entorhinal and caudal anterior cingulate regions. Large MAD suggests that only large segmentation errors can be detected for these regions. To evaluate the quality of our detection, we compared the set of regions highlighted by our approach with the set of regions that underwent manual editing. Since the manual editing was applied to only the largest (most visible) segmentation errors, we have further limited the automatically highlighted set by selecting only the regions where the difference between predicted and actual measurements was larger than 0.5mm. The overlap between the sets of manually and automatically highlighted regions was lower than expected (30%). Further examination revealed the cause of the low overlap. Many regions in our set were not highlighted in the manually edited set, either due to operator omission (Fig. 2) or wrong parcellation that resulted in the correct pial surface appearance but wrong cortical thickness (Fig. 3). There was also an opposite problem, where some regions in the manually highlighted set were not present in the automatically highlighted set. This was primarily due to large errors in the pial surface which did not affect the measured cortical thickness. This may happen, for example, when the pial surface underestimation is accompanied by similar WM surface underestimation at the same location, with preservation of relative distance between the surfaces. The measured cortical thickness then is close to the true thickness, even though the two are not directly related (Fig. 1). These findings suggest that manually edited results might not be completely appropriate for evaluation of our approach, and that more work needs to be done to establish the validity of our framework.

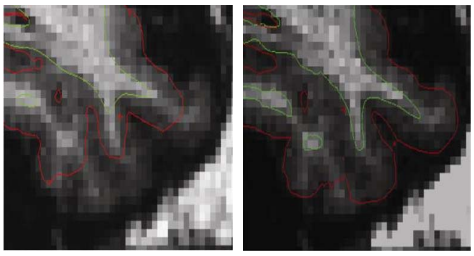

Fig. 1. Mis-estimated cortical surfaces before (left) and after (right) correction. Cortical thickness remains unchanged.

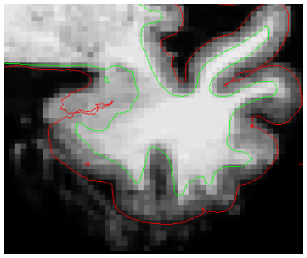

Fig. 2. Uncorrected pial surface overgrowth due to dura attachment

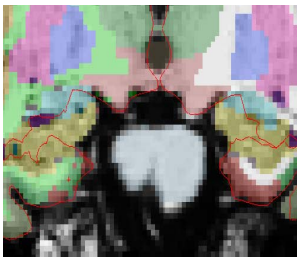

Fig. 3. Wrong labeling of entorhinal region as parahippocampal

Conclusions. We proposed an SVR based framework for automatic detection of cortical thickness measurement errors, and evaluated it against a set of manually corrected data. The overlap between the automatically and manually identified errors was low, which we found to be due to difference between the error detection targets (thickness correctness in automated detection vs. cortical surface correctness in manual detection). In the future work we will attempt to validate the proposed detection approach based on its influence on subsequent statistical analyses, for example, of thickness-age correlations and/or cortical thickness group differences.

References. [1] Dale, A.M. et al, 1999. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage 9, 179-194. [2] Kim, J.S. et al, 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, 210-221. [3] Ashburner, J., Friston, K.J., 2000. Voxel-Based Morphometry--The Methods. NeuroImage 11, 805-821. [4] Levinski, K. et al 2009. Interactive surface-guided segmentation of brain MRI data. Comput Biol Med 39, 1153-1160. [5] Sadananthan, S.A., et al, 2010. Skull stripping using graph cuts. NeuroImage 49, 225-239. [6] Zheng, W. et al., 2009. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. NeuroImage 48, 73-83. [7] Koch, M.A., et al, 2002. An investigation of functional and anatomical connectivity using magnetic resonance imaging. Neuroimage, 16(1), 241-50. [8] Vladimir Vapnik, C.C., 1995. Support-Vector Networks. Machine Leaming, 20.