

Evaluating Reliability of Performance Metrics for Bias Field Correction in MR Brain Images

V. Zagorodnov¹, Z. Chua¹, and W. Zheng¹

¹Computer Engineering, Nanyang Technological University, Singapore, Singapore

Introduction. Performance of nonuniformity correction approaches is usually evaluated indirectly, on the basis of remaining tissue intensity variability rather than the actual estimated bias field. Common indirect measures include coefficient of variation of white matter CV(WM) and gray matter CV(GM), and coefficient of joint variation CJV between WM and GM [1]. However, disagreements between indirect measures on what is the best performing method, reported in several recent studies [2,3], suggest that indirect measures might not reliably reflect the true nonuniformity correction performance. Here we examined reliability of several common performance measures using simulated brain data.

Methods. We used a set of synthetic 1x1x1mm T1W MR data of a normal brain from the Montréal Simulated Brain Database (SBD) [4], corrupted by various degrees of noise and intensity nonuniformity (overall 9 volumes with all possible combinations of 0%, 20%, and 40% bias fields and 1%, 3%, and 5% noise). Each volume was processed by two common nonuniformity correction algorithms. Some of the parameters of these algorithms (the amount of smoothing, brain mask, and estimation domain) were varied to obtain 621 different instances of nonuniformity correction. The quality of the correction in each instance was evaluated using several indirect measures, correlating the results with a fully reliable direct measure, which was based on the Euclidian distance between the true and estimated bias fields. Due to observed non-linear relationships between the metrics and importance of rank preservation, we used Spearman rank correlation coefficient as the measure of reliability.

We hypothesized that the quality of GM/WM masks used for computation of indirect measures can have substantial effect on reliability. For example, the presence of partial volume voxels or voxels from different tissue classes can increase tissue intensity variability, which can be wrongly attributed to inferior nonuniformity correction. To test this hypothesis, we used three types of masks - original, corresponding to WM/GM tissue segmentation provided by SBD, conservative, obtained by excluding partial volume voxels from the original mask, and erroneous, obtained by random perturbation of the boundary of the original mask by 1-2 voxels in order to simulate potential segmentation errors that may arise in expert-guided or automatic segmentation. We also hypothesized that contribution of image noise might reduce reliability of indirect measures and investigated whether noise suppression (by smoothing using a 3x3x3mm filter) prior to evaluation of indirect measures would improve their reliability.

Results.

Table 1. Spearman correlation coefficients between direct and indirect metrics. Correlation of 85% and higher are highlighted in bold, underlined values correspond to correlations above 95%.

Mask	Metric	Noise(%)			Average
		1	3	5	
Original	CV(WM)	46.2%	38.5%	43.4%	65.9%
	CV(GM)	67.2%	77.2%	77.8%	
	CJV	89.0%	89.0%	65.2%	
	mod CV(WM)	47.0%	42.5%	38.8%	66.8%
	mod CV(GM)	73.9%	83.7%	86.1%	
	mod CJV	65.9%	78.8%	84.5%	
Conservative	CV(WM)	94.4%	83.9%	71.0%	76.8%
	CV(GM)	89.2%	90.6%	88.9%	
	CJV	94.6%	66.2%	12.7%	
	mod CV(WM)	95.6%	96.6%	95.7%	93.8%
	mod CV(GM)	87.4%	89.5%	90.4%	
	mod CJV	97.1%	96.9%	94.8%	
Erroneous	CV(WM)	37.6%	31.8%	25.2%	10.1%
	CV(GM)	17.4%	22.7%	10.5%	
	CJV	-2.6%	-12.8%	-39.1%	
	mod CV(WM)	49.9%	44.0%	49.2%	29.5%
	mod CV(GM)	29.8%	24.8%	12.3%	
	mod CJV	30.0%	19.1%	6.7%	

Table 1 shows the observed Spearman correlation coefficients for all noise levels, masks and metrics. Metrics that were applied on smoothed data are designated with 'mod' prefix. Overall, as expected, conservative masks and smoothing had a large positive effect on reliability. Erroneous masks led to the worst reliability, with CJV's correlation coefficients even reaching negative values. Using original masks somewhat improved reliability to an average level of 65%; this value was not improved even after smoothing. Conservative masks resulted in the best performance. CV(GM) evaluated over conservative GM mask had the best reliability over all metrics when no smoothing was used. However, the best reliability (exceeding 95% correlation for all noise levels) was obtained using CV(WM) and CJV when conservative masking and smoothing were combined.

Despite equal performance of CV(WM) and CJV on conservative masks and smoothed data, CV(WM) appears to be a safer choice, because it uses only WM mask, while CJV relies on both WM and GM masks. Generally, due to complex structure of the brain cortex we can expect that obtaining accurate GM segmentation (by expert-guided manual delineation or through an automatic segmentation algorithm) would be more problematic. These would potentially diminish reliability of CJV to the level of the erroneous mask.

Conclusion.

Our study examined reliability of several common performance measures and revealed that their low reliability can lead to misleading conclusions when performing parameter optimization or choosing a superior correction approach. Improved reliability can be obtained by evaluating the same metrics over conservative masks and by applying slight smoothing to suppress image noise. On real brain MR data, due to potential difficulty in obtaining a conservative GM mask, we recommend using coefficient of variation (CV) measured on smoothed conservative WM mask.

Acknowledgements. This work is supported by SBIC C-012/2006 grant provided by A*STAR, Singapore (Agency for Science and Technology and Research).

References

- [1] J. Luo et al. Correction of bias field in MR images using singularity function analysis. *IEEE Trans Med Imaging*, 24(8):1067-1085, 2005
- [2] J.D. Gispert et al. Method for bias field correction of brain T1-weighted magnetic resonance images minimizing segmentation error. *Hum Brain Mapp*, 22(2):133-144, 2004
- [3] B. Likar et al. Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE Trans Med Imaging*, 20(12):1398-1410, 2001
- [4] D.L. Collins et al. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*, 17(3):463-468, 1998