MR Image Quality Evaluation using Weighted Perceptual Difference Model (Case-PDM)

J. Miao[1], W. C. Wong[1], and D. L. Wilson[1,2]

[1]Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio, United States, [2]Radiology, University Hospital of Cleveland, Cleveland, Ohio, United States

INTRODUCTION

The perceptual difference model (Case-PDM) is being used to quantify image quality of fast, parallel MR acquisitions and reconstruction algorithms by comparing to slower, full k-space, high quality reference images. Case-PDM objectively, quantitatively evaluates image quality, and we have found it to be quite useful in investigations of keyhole, spiral, SENSE, and GRAPPA applications [1-4]. To date, most perceptual difference models average a single scalar image quality metric over a large region of interest. In this paper, we create an alternative metric weighted to image features. Spatial filters were applied to the reference image to create edge and flat region images, then weighted and aggregated to create “structural” images which in turn spatially weighted the perceptual difference maps. We design human subject experiment to calibrate the proposed method and to validate the results in the context of MR image quality evaluation. Five different input MR images (3 brain, 1 cardiac, and 1 phantom images) and over 200 reconstructions and 8 human subject trials were used in experiments.

METHODS

Spatial filters are applied to each reference image to create a structure image with high response values at certain features (e.g. edge and homogenous regions). Region in the maps are then weighted and aggregated to compose a weight map. A feature-weighted PDM score is derived from the visual difference map and the weight map. A modified DSCQS (Double Stimulus Continuous Quality Scale) experiment was designed (GUI is shown in figure a) to obtain human subject ratings and ROIs in the reference image that are relevant to the ratings. The experiment covered 5 different images (3 brain images, a cardiac image and a phantom image) and 4 MR image reconstruction algorithms (SENSE, spiral, GRAPPA, and WGRAPPA). In total, 200 test images covering a wide range of image quality are evaluated (40 test images for each of the 5 datasets). We optimized the spatial filter parameters using a linear correlation coefficient as an objective function calculated from pairs of human ratings and weighted PDM scores among the test images. We also compare visually the weight maps against the average ROI drawn by the subjects in the DSCQS experiments.

RESULTS

We found that the weighted PDM score has better correlation with human subject ratings for all the test images, as compared to the average value in the difference map (aka unweighted average score). For 5 different images (3 different brain, 1 cardiac, and 1 phantom images), r values [weighted PDM, average PDM] were improved ([0.96, 0.94], [0.94, 0.91], [0.95, 0.94], [0.97, 0.91], [0.98, 0.97]) in all cases. We then used the optimal filter parameters of a dataset with images from a reconstruction algorithm to calculate the weighted PDM scores for another dataset of 40 images reconstructed by a different algorithm. We found that, as long as the two datasets share similar anatomical structures, the weighted PDM scores maintains a high correlation with human ratings, indicating that the assessment is predictive (Figure). Visual comparison of our weight maps with the average ROI drawn by the subjects in the DSCQS experiments shows that they are not similar. Nevertheless, the average subjects’ ROI does cover high-value region in the weight map. Despite a large variation among subjects’ ROIs (Dice Similarity Coefficient <0.7), such observation suggests that our weight map is able to identify ROIs that are relevant to image quality evaluation. In addition, we compared our method to the entropy-based visual ROI weighting method proposed by Rao et al. [5]. Experimental results suggest that our method is superior in predicting the rank order of image quality by human subjects.

CONCLUSIONS

We conclude that for image quality ratings, the proposed weighted Case-PDM score can more faithfully represent the human subject evaluation over a wide range of image quality, as compared to the unweighted average Case-PDM and entropy-based scores. Our method is robust across subjects and anatomy; that is, scores maintain a high correlation with human ratings even if the test dataset is different from the training dataset. Re-calibration is only necessary if the dataset has a different anatomical structure. We have now created a framework for evaluating region-specific image quality. By including radiologist specialists in future studies, we can create image quality scores relevant to diagnosis.

ACKNOWLEDGEMENT

This work was supported under NIH grant R01 EB004070 and the Research Facilities Improvement Program Grant NIH C06RR12463-01. We thank the subjects for participating in the experiments.

REFERENCES

[1] KA Salem, Journal of Electronic Imaging, 2002 [2] D Huo, J Magn Reson 2006 [3] D Huo, ISMRM 2006 [4] J Miao, ISMRM 2007 [5] Rao *et al.*, ICCTA'07

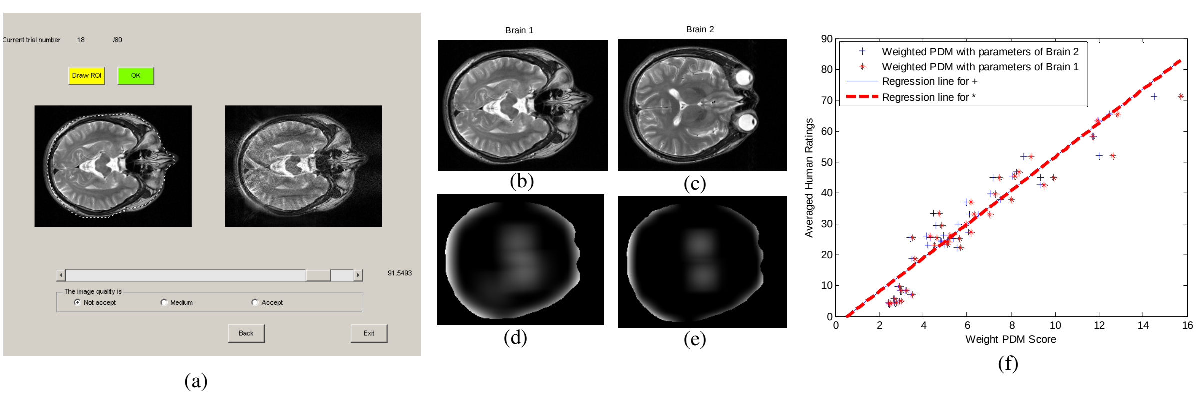

Figure The GUI for an image quality rating and ROI drawing task is shown in (a). In the GUI, the left image is a reference, and the right one is a test image. Subject was asked to rate the overall image quality of the test image as compared to the reference, and draw his/her ROI on the reference image. (b) and (c) are two different MR brain slices acquired from a healthy subject, they are references for Brain 1 and Brain 2 datasets used for the experiment. Both (d) and (e) are weight maps used to calculate weighted PDM scores for Brain 2 test images, with (d) using optimal filter parameters from Brain 1 experiment and (e) using optimal filter parameters from Brain 2 experiment. In (f), regression line (r=0.95) for weighted PDM scores (cross) by weight map (e) is overlapped with regression line (r=0.95) for weighted PDM scores (asterisk) by weight map (d), when evaluating the same Brain 2 image dataset. This shows that optimal filter parameters could be exchanged within different image datasets which share similar anatomy structures.