Quantitative Evaluation of non-linear Reconstruction Methods in MRI

Matthias Schloegl[1], Florian Knoll[1], Katharina Gruber[1], Franz Ebner[2], and Rudolf Stollberger[1]

[1]Institute for medical Engineering, TU Graz, Graz, Austria, [2]Universitätsklinik für Radiologie, Medizinische Universität Graz, Graz, Austria

Introduction

Non-linear reconstruction methods pose a rapidly evolving field in MR imaging in the context of compressed sensing[1], opening up for many new possibilities. Yet with this class of reconstruction we are facing the situation that important information, e.g. about pathology, can get lost even though overall quality seems to improve.
This is a crucial and yet sparsely studied point in the field on non-linear reconstruction. Therefore it is highly desirable to have access to a suitable tool to objectively evaluate reconstruction quality close to human expert-rating. Common measures such as mean-squared-error (MSE) or peak-signal-to-noise-ratio(PSNR) may not be sufficient to rate model-dynamic information. The Perceptual Difference Model (PDM)[2] is a well-investigated metric, that showed promising results for parallel image reconstruction but has not been evaluated in this context. The following study systematically investigates the use of PDM along with five other metrics including MSE and PSNR together with evaluations from six experienced radiologists.

Methods

A total of 60 reconstructions for each dataset were gained using conjugate-gradient Sensitivity Encoding (CGSENSE)[3], Total Generalized Variation constrained reconstruction(TGV)[4], iteratively regularized Gauss-Newton (IRGN-TV)[5] method and iterative self-consistent parallel imaging (l1-SPIRiT)[6] on sub-sampled Cartesian, radial and randomized trajectories with acceleration factors of 2,3,4,6 and 8. 8 coil-sensitivities were simulated with Biot-Savarts law and modulated to the fully sampled standard reconstruction. For CGSENSE and TGV these were re-estimated for input to the reconstruction while for IRGN-TV are jointly estimated with the image content. For l1-SPIRiT no explicit sensitivity maps are used similar to GRAPPA. The datasets were preselected according to diagnostic value, leaving 42 reconstructions that were presented to the radiologists under standard clinical evaluation conditions. The radiologists were trained to focus on the identified lesions in the original image that could be reviewed at any time. Each image was then evaluated according to the DSQCS proposal of the International Telecommunication Union (ITU)[7] for impression of overall quality (OQ) first, then for recognisability of anatomy (AN), severity of imaging artefacts (AR) and variance of the known pathology (PT). The PDM model was implemented according to Huo[8] with the non-rotational invariant contrast sensitivity function described by Daly[9] normalized to a produce a peak at 8 cyc/degree. Structural Similarity Index (SSIM)[10] and Universal Image Quality Index (UIQI)[11] were implemented according to their references. These three metrics attempt to model the human visual system. In comparison to those models we also used mutual information (MI), a statistical approach, PSNR and mean-squared-error (MSE), being pure statistical resp. mathematical measures. Scores were also calculated for both a large region of interest (ROI) centred around the lesion and the whole image (see Figure 1). Correlations were calculated using Spearmans method.

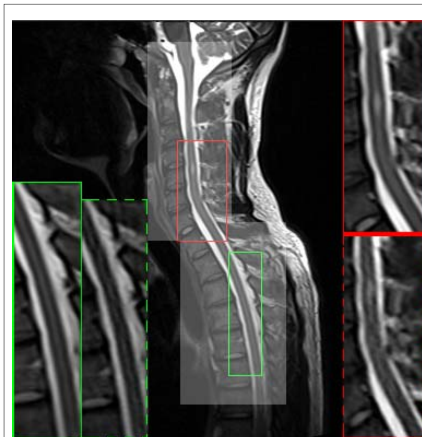

Fig.1 gold-standard reconstruction for one database under investigation with ROI (brighter part in center) and enlarged details of the lesion in the medulla of the reference image (solid red and green frames) and a reconstruction example with radial IRGN-TV with acceleration factor 6 (dashed red and green frames)

Results

Figure 2 exemplary shows for OQ the averaged expert-rating (red dots) linearly fitted to the respective metric scores among with the individual expert ratings (crosses) for the spinal dataset (see Figure 1). For the AN rating, UIQI yields highest value with 83%. Concerning PT, MI scores highest with 77% and finally for AR, best score yields MSE (79%). With no ROI used, metrics modelling the human visual system show correlations below 50 % for all rating types, while MI, PSNR and MSE prove to be more stable (50% < corr < 71%). For the correlations between OQ, AN, PT and AR no correlation coefficient was below 91%. For inter-expert correlation no value was below 88%.

Discussion and Conclusion

The first point to note is that evaluation of OQ, PT, AN and AR exhibits good correlation, justifying the use of metrics for medical imaging where in the end the focus lies on detectability of the pathology. This has strongly to be taken into account as, even though OQ could improve, the pathology can still be disturbed by lesion-like artifacts as in the case for many reconstructions with random sampling pattern, producing cloudy artifacts. Secondly, the study shows that the human-visual-based metrics PDM and UIQI have – when a ROI is used - potential for evaluation of non-linear reconstruction methods, while failing for the whole image assessment due to overestimation of artifacts in the background. MI, PSNR and MSE proved to be more stable, yet lacking performance. For the PDM model this could be further improved using an updated version as in Miao[2]. Further evaluations of other anatomical studies, namely brain, pelvis, knee and abdomen are in progress. We conclude that quantitative image metrics can be suitable applied on reconstruction evaluation with non-linear methods. Furthermore the use for automated parameter calibration or stopping criteria during iterations needs to be explored and is expected to improve reconstruction- quality and time.

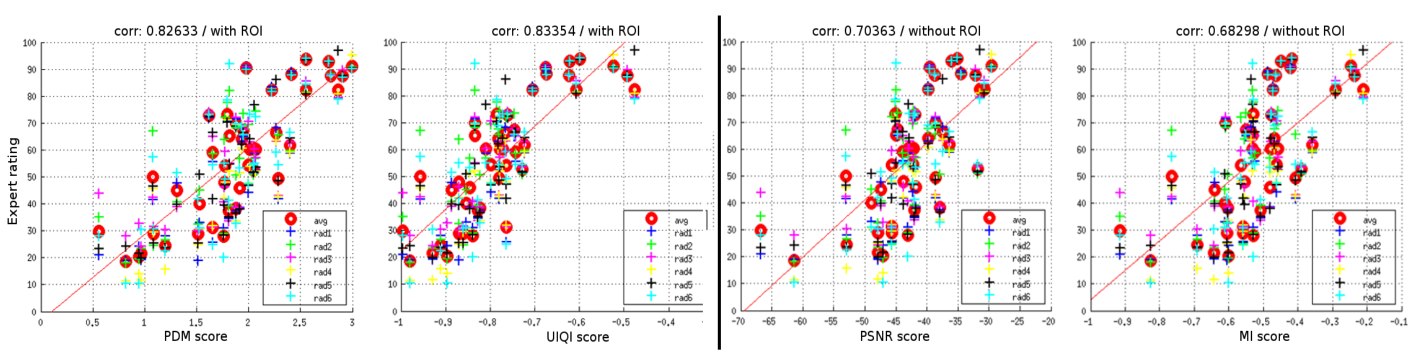

Fig. 2 selected results from individual human observer ratings (crosses) together with best metric scores with- (two images on the left) and without ROI (two images on the right) for overall quality rating for a set of 42 reconstructions. Red lines mark linear fit of averaged expert (red circles) ratings with corresponding metric score. Correlation values are displayed in title.

References

[1] Lustig M., MRM 2007, 58:1182-1195 [2] Miao J., Medical Physics, 2008, 35:2541-2553 [3] Pruessmann K., MRM 2001, 46:638-651 [4] Knoll F., MRM 2011, 65:480-491 [5] Knoll F., MRM 2012, 67:34-41 [6] Lustig M.b., MRM 2010, 64:457-471 [7] International Telecommunication Union, Rec. ITU-R BT.500-11 [8] Huo D., Ph.D. Thesis 2007 [9] Daly S., Digital images and human vision, MIT Press 1993, 179:206 [10] Wang Z., IEEE Transactions on Image Processing 2004, 13:600-612 [11] Wang Z., IEEE Signal Processing Letters 2002, 9:81-84