A new approach for automatic image quality assessment

Thomas Küstner[1,2], Parnia Bahar[2], Christian Würslin[1], Sergios Gatidis[1], Petros Martirosian[3], Nina Schwenzer[1], Holger Schmidt[1], and Bin Yang[2]
[1]*Department of Radiology, University Hospital of Tübingen, Tübingen, Baden-Württemberg, Germany,* [2]*Institute of Signal Processing and System Theory, University of Stuttgart, Stuttgart, Baden-Württemberg, Germany,* [3]*Diagnostic and Interventional Radiology, University Hospital of Tübingen, Tübingen, Baden-Württemberg, Germany*

Purpose: Besides the various advantages of MRI as an essential imaging technique in medicine with its vast variety of different imaging possibilities, a stable, reliable and meaningful image quality evaluation is often very demanding, especially when developing new technologies for which no reference or gold-standards are available yet. Nowadays, evaluation mainly depends on human observers (HO), like trained physicians or experienced radiologists (ER), to determine the underlying image quality with respect to a certain diagnostic question at hand. But this task can be very time consuming and costly. Hence an automatic evaluation is desired. However, most automatic evaluations depend on the existence of a reference image, which is often not available to perform a one-to-one similarity/dissimilarity mapping, like intensity-based metrics. Additionally, these intensity-based methods are not translation- or rotation-invariant and do not reflect any image distortions or MRI artifacts resulting from the acquisition, like motion ghosting, blurring or aliasing. The latter one is of special interest when it comes to the evaluation of the image quality of accelerated acquisitions by means of subsampling the phase-encoding directions via Parallel Imaging or Compressed Sensing (CS) techniques. For the general image quality assessment, it is important to differentiate the images based on the overall achievable image quality and on the useful extractable information depending on the chosen MR sequence and the underlying diagnostic question. Former works try to toggle this problem by evaluating log-derivatives of the image globally without the need of a reference image[1], but do not consider the diagnostic question at hand. Lorente et al.[2] proposed an active learning approach to select the best matching images for building up a reliable training data set to measure lesion detectability in a four-dimensional feature space. We propose a robust, accurate and flexible evaluation system which is independent of the presence of a reference image and is based on a machine-learning approach to evaluate the quality of the complete image. In this work, we focused on determining MRI image quality of subsampled acquisitions which show different kinds of aliasing and noise artifacts.

Material and Methods: The proposed framework is shown in Fig. 1. 3D and 2D multislice MR images are considered as input and are classified into one of five different classes representing the image quality assessed from the reading of the ERs. The classes range from very high quality (1) over high quality (2), medium quality (3) and poor quality (4) to very poor quality (5). A database with a total of 41168 2D transversal image slices of 20 patients and 15 healthy volunteers of the thorax, abdomen and pelvis with different imaging sequences and contrast weights (like T2-HASTE, T2-TIRM, T1-VIBE or T1-FLASH) as well as different subsampling strategies and acceleration factors (like GRAPPA or CS) with corresponding reconstruction techniques, was constructed. We use 2911 2D MR image slices, 2038 as training set and 873 as test set to evaluate the classification, with known image quality labels obtained by a blindfolded reading of 5 ERs. The median of the labels has been used if experts did not agree with each other.

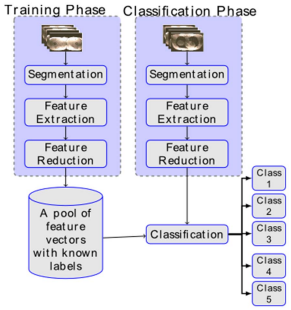

Fig. 1: The proposed framework

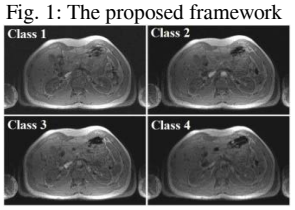

Fig. 2: Exemplary images of the database

Table 1: Extracted features

Name of feature sets	
Gray level co-occurrence [4]	672
Run length [5]	44
Gabor filters [6]	1080
Local binary pattern [7]	1024
Fractal [8]	27
Gradient-based [9]	24

Segmentation: To prune out irrelevant background information, a Chan-Vese segmentation[3] is applied. An initial curve is evolved to fit on the body boundary and hence dividing the background from the foreground.

Feature Extraction: In order to represent the MR images by richer attributes, several statistical features are extracted from the image. Some characteristics such as smoothness, coarseness, regularity and homogeneity are applicable interpretations of MR images and lead to a better understanding. Thus we decide to extract features based on gray level intensity, texture, contrast, gradient and resolution with associated meaningful metrics like energy, correlation, entropy, etc. This results in 2911 features listed in Table 1.

Feature Reduction: In order to preserve just the useful and most meaningful features representing each class and to avoid overfitting, Principle Component Analysis (PCA) is employed to reduce the dimensionality of the feature space. The PCA selects the M most dominant eigenvectors/principal components (PCs) representing the complete feature set in a lower-dimensional subspace. Other feature reduction techniques like Sequential Floating Forward Selection have also been implemented, but are too time-consuming for this task and will not be considered further.

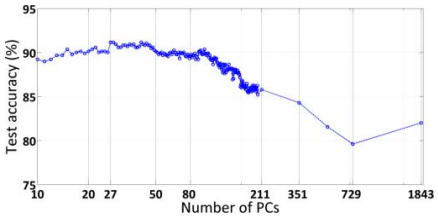

Fig. 3: Classification test accuracy

Classification: A 3D input image is sliced and each slice is represented by a feature vector $\boldsymbol{x}$. The image slice is assigned into a given class y by means of a multi-class Support Vector Machine (SVM) [10] using an one-against-one approach. In order to separate the classes with maximum generalization ability and margins, an optimal separating hyperplane must be determined yielding the minimization problem: $\min_{b,\xi} \frac{1}{2}\|\boldsymbol{w}\| + C\sum_{i=1}^{M}\xi_i$, s.t. $y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, \dots, M$ where b denotes a bias term, $\boldsymbol{w}$ the hyperplane norm vector, ξ the slack variable and C the error penalty term. A Radial Basis Function kernel is used to find a linear separable hyperplane in the augmented feature space by measuring the distance between feature vector $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$: $\exp(-\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)$, where $\gamma \geq 0$ controls the acceptance radius. To generate the best model, γ and C should be precisely specified. For this purpose, for each pair of γ and C, the classifier is trained with the training set and the classification error is calculated by a 10-fold cross-validation. 9 folds are used to train the classifier and the remaining one is used for validation. The final classification model is created by the best γ and C values. Finally, this optimal model is evaluated against the test set to retrieve the test accuracy.

Table 2: Confusion matrix

	estimated class by algorithm				
class. by ER	70	27	0	0	0
	23	170	3	0	0
	0	1	244	5	0
	0	0	14	239	0
	0	0	0	4	73

Results & Conclusions: Fig. 3 illustrates the overall test accuracy of the optimal model against the number of chosen PCs. Selecting the first $M = 27$ PCs achieves an overall test accuracy of 91.2%. The decreasing trend in Fig. 3 implies an increase of irrelevant information and hence a more complex hyperplane is needed to partition the samples. A more detailed insight into the test accuracy of the optimal model with 27 PCs reveals the confusion matrix in Table 2. Most failures are related to class 1, since it is already hard for a HO to distinguish between class 1 and 2. In conclusion, we presented a new machine-learning based automatic MR image quality assessment which is able to deal with different kind of input images (like contrast, body region or acquisition type) and achieves a high accuracy with reasonable computational cost while this would be a challenging and time-demanding task for clinical experts. In the future,we will investige on the stability and applicability of the algorithm regarding different databases and apply an active learning strategy to reduce the number of needed training labels and to further increase the accuracy.

References: [1] Yi, J Electro. Imag., 2013, 22(4), 043025-22 [2] Lorente, ISBI, 2014, 1352-55 [3] Chan, IEEE T. Image Process., 2001, 10(4), 266-277 [4] Haralick, IEEE Proc., 1979, 67(5), 786–804 [5] Tang, IEEE T. Image Process., 1998, 7(11), 1602–1609 [6] Jain, Mach. Vision Appl., 1992, 5(3), 169–184 [7] Hadid, Springer, 2011, 40 [8] Iftekharuddin, Mach. Vision Appl., 2003, 13(5-6), 352–362 [9] McGee, JMRI, 2000, 11(2), 174–181 [10] Chang, ACM T. Intelligent Syst. and Tech., 2011, 2, 1–27.