Bayesian Framework Inspired No-Reference Region-of-Interest Quality Measure for Brain MRI Images

Michael Osadebey, Marius Pedersen, Douglas Arnold, and Katrina Wendel-Mitoraj, The Alzheimer's Disease Neuroimaging Initiative*

Abstract

We describe a new post-acquisition quality assessment method for brain MRI images. It is based on Bayesian theory. Entropy is the event of interest. The low and the high entropy regions of a slice are partitions of image quality sample space. Three separate observations of image quality are contrast, standard deviation and details feature images of the original slice. Two separate quality measures are defined as the posterior probability of an entropy region given corresponding region in an observation of image quality. Prior belief in each entropy region is determined by the normalized total clique potential (TCP) energy of the slice. For TCP below predefined threshold the prior probability is determined from a quality model described by normal distribution of percentage composition of low and high entropy regions. The quality model was built from 250 MRI volume data. The data was made available by the Alzheimers disease neuroimaging initiative (ADNI). For TCP above threshold the prior is computed using a TCP-Noise level model. Total and global quality measures are weighted sums of quality measures from each regions and observations, respectively. Experimental results demonstrate good correlation with subjective opinions of radiologists for different types and levels of quality distortion.

Index Terms

Magnetic Resonance Imaging (MRI), Image Quality, Entropy, Contrast, Standard deviation, Total Clique Potential (TCP) and Bayes Formula

I. INTRODUCTION

VIRTUALLY all spheres of medicine need information contained in medical images. Structural information based on the chemical and physical properties that distinguish different anatomical structures are highly desired in medical images. This requirement makes magnetic resonance imaging (MRI) system a popular imaging modality for the study of human anatomy, diagnosis of diseases and the clinical trials of drugs for the treatment of neurological diseases [1], [2], [3]. MRI can display images in three different perpendicular planes and has the potential to discriminate the constituent soft anatomical structures with high spatial and contrast resolution.

The quality of a medical image is strongly dependent on the acquisition procedures [4]. During acquisition there are several factors relating to the imaging system, the actions of the operator and the subject under investigation that limit the attainment of an ideal quality image. Image quality in radiation-based imaging systems is dependent on radiation dose [5]. There is trade-off between image quality and

M. Osadebey is with NeuroRx Research Inc, 3575 Parc Avenue, Suite # 5322, Montreal, QC, H2X 3P9, Canada e-mail: mosadebey@neurorx.com.

M. Pedersen is with the Faculty of Computer Science and Media Technology, Norwegian University of Science and Technology, Teknologivegen 22, N-2815 Gjøvik, Norway email: marius.pedersen@ntnu.no.

D. Arnold is with NeuroRx Research Inc, 3575 Parc Avenue, Suite # 5322, Montreal, QC, H2X 3P9, Canada e-mail: douglas.arnold@mcgill.ca.

K. Wendel-Miitoraj is with BrainCare Oy, Finn-Medi 1, PL 2000, 33521 Tampere, Finland, e-mail: katrina@braincare.fi.

patient safety [6]. The antiscatter grids determines the level of contrast in digital mammography images [7]. Quality of reconstructed SPECT images are influenced by the number of projection angles [8]. In breast ultrasound images artifacts are caused by improper positioning of the nipple relative to the breast mass and loose contact between the breast mass and the transducer [4]. In MRI images noise is the result of trade-off between signal-to-noise-ratio, image resolution and length of scan time [9]. Bias fields is the combined effects of non-uniform sensitivity of radio-frequency coils and non-uniform sensitivity of static fields [10]. Motion of patient, respiration, blood flow and patient position relative to the iso-center of the magnetic bore introduces blur and extraneous features into the image [11], [12]. Chemical shift and partial volume artifacts are the result of improper parameter settings [13]. The numerous factors that influence the acquisition procedures makes quality evaluation a non-trivial and complex task [14], [15].

Quality evaluation of medical images is integrated into the imaging workflows of routine clinical practice and research. Quality measure evaluates the integrity of the anatomical information contained in a scan. Optimal transmission of information from a medical image, the discernment and the confidence in the interpretation of an image by a physician, radiologist or a trained reader and the efficacy of post-acquisition processing tools is dependent on the outcome of image quality evaluation. Recently, the European collaborative research network that studies MRI in multiple sclerosis (MS) recommend high quality brain MRI images for clinical research [16]. The recommendation will prevent inaccurate diagnosis by enforcing the use of only MRI data of acceptable quality for clinical research. The recommendation also outlined the criteria for establishing multiple sclerosis dissemination in space and time. They are the presence of at least one lesion in at least two of four different regions of T2 brain MRI images [16]. The regions are juxtacortical, periventricular, infrantentorial and the spinal cord [16]. MRI system is also a strong biomarker for the diagnosis of Alzheimer's disease (AD) [17]. In response to these recommendations MRI-based metrics such as [18], [19], [20] have been formulated to extract information from different structural regions of the brain. The metrics include atrophy measurements of the brain parenchyma, cortical gray matter, white mater, ventricle and the thalamus.

Popular objective quality evaluation methods such as root-mean-square error (RMSE), signal-to-noise ratio (SNR) and structural similarity index (SSIM) [21] were designed to solve research problems that were not related to image interpretation but on efficient compression, storage and transmission of images. Their formulation assume that all classes of medical images and all other natural images have same descriptive features. Medical images exhibit characteristics such as texture, gray-scale color, noise and homogeneity which distinguish them from each other and other classes of natural images. There are many definitions of PSNR, SNR and MSE which makes it difficult to compare quality measures from different imaging system, modalities and researchers [8]. Quality indices derived from these popular methods does not always correlate with the performance of observers using the imaging system on the task for which they are intended [8].

Some of the problems associated with the popular quality assessment methods were addressed when task-based quality assessment methods was introduced to the medical imaging and computer vision community [22], [23], [24]. The philosophy behind task-based quality assessment is the believe that a rigorous quantification of image quality should be defined by specifying a clearly defined task and the observer who will be performing the task. In medical imaging the task can be classification or estimation task. An example of classification task is the detection of lesions in brain MRI images of patients diagnosed with multiple sclerosis disease. Quantification of brain atrophy in patients diagnosed with Alzheimer's disease is an example of estimation task. The observer can be human, human model or Bayesian ideal observer [25], [26]. The efficacy of task-based quality assessment reported in [27], [25], [28] is the motivation behind pioneering research in the field of task-based adaptive imaging. Task-based adaptive imaging has been applied for the optimization of imaging systems parameters [29], [8], [5], optimization of image quality in imaging systems [26], [30], [31], evaluation of cardiac ejection fraction estimation algorithms [32] and the evaluation of diffusion weighted MRI segmentation algorithms [33].

The mainstream approach to quality evaluation methods for brain MRI images focus on the acquisition stage. The several parameters associated with MRI system acquisition process are exploited to evaluate and

optimize image quality. Proposed methods in this category include [34], [35], [36], [37], [38], [39], [40]. There are few contributions on post-acquisition quality assessment of brain MRI images. Post-acquisition quality evaluation is an important step in the quality control procedures of clinical research organizations (CRO). Quality evaluation ensures that the variations in the quality of MRI images from different MRI system manufacturers, different clinical trial sites and different acquisition protocols are assessed and standardized before they are fed to automated image analysis systems.

We identified and reviewed five contributions in the literature on post-acquisition quality evaluation of brain MRI images. The report in [41] apply analysis of variance (ANOVA) algorithm to assess the variation of several quality measures with different levels of distortions. The authors in [42] combine the detection of artifacts and estimation of noise level to measure image quality. In [43] null space analysis and just noticeable difference scanning method was proposed as a better quality metric compared to root mean square error. The popular signal-to-noise ratio is the quality metric adopted in [44]. Subjective quality assessment was reported in [45].

The contribution by [44] which adopt signal-to-noise ratio cannot discriminate the quality of two images that are perceptually dissimilar [43], [46]. A significant contribution by [42] is the adoption of artifacts as quality attributes. Several types of distortions combine with ideal features to manifest as an image attribute [47]. Thus the adoption of only artifacts and noise in [42] are too few attributes to evaluate the quality of an image. There is risk of ambiguity in quality measures that are based on the use of ANOVA models [48]. ANOVA limits the performance of the technique proposed in [41] to the detection of distortion. It cannot transform the different levels of distortion into a quality index. Quality evaluation by a trained MRI reader which requires negotiation around several types and levels of distortion [39] and the need for consensus with other readers makes the subjective method proposed in [45] inefficient to manage the large volume of MRI data in clinical research organizations [49].

The different distortion processes have different effects on the different homogeneous regions of the brain. Noise introduce excessive and possibly spurious details into the smoothly varying white matter and ventricle regions. It lowers the density of edges that characterize the cortical gray matter region and blur the boundaries between the different homogeneous regions. The smoothen effect of motion blur result in loss of texture features that distinguish the white matter and the thalamus regions. Motion blur reduce the contrast between the white matter and other homogeneous regions. Bias fields corrupt the natural homogeneity of the major anatomical structures by introducing new smoothly varying intensity levels.

Current post-acquisition quality assessment for brain MRI images have rigid designs. Their techniques regard the brain as a single homogeneous region. They are not flexible and rigorous enough to account for the effect of all possible distortions on the different homogeneous regions. Few or most relevant quality attributes are applied across the entire image to predict image quality [50]. They will be inefficient for quality assessments in region-based measurements such as normalized cortical gray matter volume, normalized white matter volume, cortical gray matter atrophy, white matter atrophy and thalamus atrophy. In the study of disease progression in MS and AD the focus of the physician, radiologist, a trained reader and automated image analysis systems varies with the type of MRI-based metric. The white matter region is the focus for T2 lesion count. The focus shift to the cortical gray matter region during gray matter atrophy measurement, to the ventricle in ventricular volume measurement and to the global view of the image in whole brain atrophy measurement. Thus global quality measure increase the risk of non-optimal quality measure in specific region-of-interest.

This report describe a new post-acquisition no-reference, region-of-interest objective quality assessment method for brain MRI images. The proposed method is inspired by Bayesian theory of probability and statistics [51]. There are two separate events of interest. They are the low and the high entropy regions of a MRI slice. These entropy regions are regarded as partitions of image quality sample space. There are three separate observations for each partition of the sample space. They are the local contrast, local standard deviation and local details feature images. The feature images are extracted from the original slice using appropriate local filters. Quality score for each entropy region is defined as the posterior probability of a region in the local entropy feature image given corresponding region in the feature image that describes the

observation of image quality. Thus there are three quality measures for each entropy region. Each quality measure is regarded as a quality attribute. The total quality score of an entropy region is the weighted sum of each quality attribute. The global quality score is the wighted sum of the quality scores from each entropy region. The prior probability of each entropy region is determined by the normalized total clique potential (TCP) energy of the image. When TCP is below a predefined threshold the prior belief is computed from a quality model. The quality model is based on the philosophy adopted in [52], [53]. We adopt the principle of similarity in the geometry of human brain across age, gender and race [52]. Furthermore we invoke the central limit theorem on the random variables generated from MRI volume data of 250 subjects [53]. The data was made available by the Alzheimer's disease neuroimaging initiative (ADNI). For TCP above the threshold the prior is computed using a TCP-Noise level model.

This paper is organized as follows. The next section describe the theory behind our proposed method. Materials and methods are described in Section 3. Results from quality measure experiments are displayed in Section 4 and discussed in Section 5. Section 6 introduces future research direction. Section 7 concludes this report.

II. THEORY

A. Local Entropy and Classical Quality Attributes

Let $I(x, y)$ denote an image cast on a 2D grid of size $N \times N$. The spacing between pixel locations (x, y) is h and there are Q number of gray levels in the image. A local region $i(x, y)$ has size $h \times h$ and n_q is the number of each gray level q within the region. The local Shannon entropy $s(x, y)$ [54]:

$$s(x, y) = -\sum_{q=0}^{Q-1} p_q \log p_q \tag{1}$$

where p_q, the probability of a gray level is:

$$p_q = \frac{n_q}{h^2} \tag{2}$$

Thus the local entropy $s(x, y)$ express the diversity G of pixel gray levels

$$s(x, y) = G(p_q), \quad q = \{0, 1, \cdots, Q\} \tag{3}$$

In the following sections we will show the relationship between local entropy and the classical image quality attributes.

1) Brightness: The mean $\mu_{i(x,y)}$ of the gray levels [55]:

$$\mu_{i(x,y)} = \sum_{q=0}^{Q-1} q p_q \tag{4}$$

is a measure of the local brightness attribute. The mean brightness can also be expressed as functions Y_{11}, Y_1 of gray level probability and entropy ;

$$\begin{aligned} \mu_{i(x,y)} &= Y_{11}(p_q) \\ \mu_{i(x,y)} &= Y_1(s(x, y)) \end{aligned} \tag{5}$$

2) Noise: Given that the deviation d_q of any pixel q from the mean is:

$$d_q = \|q - \mu_{i(x,y)}\| \tag{6}$$

the local variance σ^2 [56]:

$$\sigma^2 = \sum_{q=0}^{Q-1} p_q d_q^2 \tag{7}$$

is one of the variables considered in the estimation of noise level in the local region. The local variance can also be expressed as functions Y_{22}, Y_2 of gray level probability and entropy:

$$\begin{aligned} \sigma^2 &= Y_{22}(p_q) \\ \sigma^2 &= Y_2(s(x,y)) \end{aligned} \tag{8}$$

3) Contrast: For a local region dominated by single gray level $d_q = 0$. Thus intuitively the deviation expressed in Eq. 6 is a measure of the local contrast attribute t [57]:

$$t = \frac{1}{(N^2 - 1)} \sum_{q=0}^{Q} d_q^2 \tag{9}$$

The local contrast can also be expressed as functions Y_{33}, Y_3 of gray level probability and entropy:

$$\begin{aligned} t &= Y_3(p_q) \\ t &= Y_3(s(x,y)) \end{aligned} \tag{10}$$

4) Details: The deviations d_x, d_y of any pixel from the center pixel in the x and y directions are:

$$\begin{aligned} d_{q_x} &= \|i(x+h, y) - i(x, y)\| \\ d_{q_y} &= \|i(x, y+h) - i(x, y)\| \end{aligned} \tag{11}$$

The local gradient is expressed by the forward difference in the x and y:

$$\nabla i_{q_x} = \frac{d_{q_x}}{h} \tag{12}$$

$$\nabla i_{q_y} = \frac{d_{q_y}}{h} \tag{13}$$

The gradients determines the sharpness [58]

$$\frac{1}{N^2} \sum_x \sum_y \left| \frac{\nabla I(x,y)}{I(x,y)} \right|^2 I(x,y) \tag{14}$$

and sharpness determines the amount of details in an image. For a local region dominated by single gray level $\nabla i_{q_x} = 0$ and $\nabla i_{q_y} = 0$. Thus intuitively the gradients expressed in Eq. 12 and Eq. 13 are measures of the local details attribute. The local details attribute g can be expressed as functions Y_{44} , Y_4 of gray level probability and entropy:

$$\begin{aligned} g_{q_x} &= Y_{4x}(p_q) \\ g_{q_y} &= Y_{4y}(p_q) \\ g_{(x,y)_q} &= Y_4(s(x,y)) \end{aligned} \tag{15}$$

5) Local Entropy: Combining the expressions in Eq. 3, Eq. 5, Eq. 8, Eq. 10 and Eq. 15 we can now express the local entropy as a multivariate function Y where the variables describe the quality attributes:

$$s(x,y) = Y(\mu_{\mu(x,y)}, \sigma^2_{(x,y)}, t_{(x,y)}, g_{(x,y)}) \tag{16}$$

Our conclusion is in agreement with the contributions in [59], [58] that demonstrate entropy as the aggregate of image attributes. Furthermore, entropy maximization has been adopted in [60], [61] to improve the quality of tomographic images.

B. The Central Limit Theorem

The central limit theorem says that, as the number n of independent and identically distributed random variables $\mathcal{X}$ tend towards infinity, the probability distribution of their sum or their average will be approximately normal with mean $\mu_\mathcal{X}$ and variance $\sigma_\mathcal{X}^2$ regardless of the underlying condition [62]:

$$\lim_{n \to \infty} P(\mathcal{X}) = \frac{1}{\sigma_\mathcal{X} \sqrt{2\pi}} \exp - \left(\frac{(\mathcal{X} - \mu_\mathcal{X})^2}{2\sigma_\mathcal{X}^2} \right) \tag{17}$$

There is no closed form expression for the normal cumulative density function. Thus, all continuous distributions probabilities cannot be computed from the area under the curve. A method to overcome this setback is to standardize the random variable. The standard value of a normally distributed random variable is called a z score;

$$z = \frac{\mathcal{X} - \mu_\mathcal{X}}{\sigma_\mathcal{X}} \tag{18}$$

The standard normal distribution $P(z)$ is:

$$P(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{z^2}{2} \right) \tag{19}$$

The probability is calculated from the table of cumulative standard normal distribution function $P(Z \leq z) = \Phi(z)$:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{y^2}{2} \right) dy \tag{20}$$

Since the standard normal distribution is symmetric about $\mu = 0$,

$$\Phi(-z) = 1 - \Phi(z) \tag{21}$$

C. Markov Random Field Energy - Noise Level Relationship

The Markov random field energy U of an image I is dependent on pixel configuration f. The pixel configuration is quantized into local cliques c in a clique system S that describes spatial coherence or clusters of similar pixels. The Markov random field energy is the sum of local clique potentials V_c [63]:

$$U(f) = \sum_{c \in C} V_c(f) \tag{22}$$

For the clique system we adopt second order neighborhood of size two with neighboring pixels indexed as $(i, i^{'})$. In this system the Markov random field energy is the sum of potential function contributions from single site and pair-site cliques, [64]:

$$U(f) = \sum_{i \in S} \alpha_1 V_1(f_i) + \sum_{i \in S} \sum_{i' \in N} \alpha_2 V_2(f_i, f_{i'}) \tag{23}$$

where α_1 and α_2 are interaction coefficients. For computational convenience we set $\alpha_1 = \alpha_2 = 1$. At each local clique the contribution of each neighboring pixel to the clique potential energy is determined according to the expression [64];

$$V_{c_i}(f_i) = \begin{cases} \xi_r & \text{if } f_i = f_{i'} \\ \xi_p & \text{otherwise} \end{cases} \tag{24}$$

in such a way as to reward ξ_r conformity ($f_i = f_{i'}$) with the smoothness constraint and penalize ξ_p violations ($f_i \neq f_{i'}$) of the smoothness constraint.

D. Bayes Theorem

Bayes theorem is a rule applied in a random experiment where the event of interest A forms $m = \{1, 2, \cdots, M\}$ partitions of the sample space. Bayes rule expresses the probability of the event A as posterior probability based on condition B that is related to event A. The posterior probability of event A_j given that event B has occurred is [51]:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{m=1}^{M} P(B|A_m)P(A_m)} \tag{25}$$

where $P(A_j)$ is the prior probability and $P(B|A_j)$ is the observation. The prior probability is the belief in the event A_j before the actual observation B derived from performing the experiment. The term $P(B|A_j)$ is the probability of observing event B given that event A_j is true. The posterior probability $P(A_j|B)$ is the probability that the event A_j is true supposing that our new piece of evidence B is true. Given that there are $k = \{1, 2, \cdots, K\}$ separate observations for each partition of the sample space, the posterior probability of event A_j given that event B_k has occurred is:

$$P(A_j|B_k) = \frac{P(B_k|A_j)P(A_j)}{\sum_{m=1}^{M} P(B_k|A_m)P(A_m)} \tag{26}$$

The denominator in Eq. 25 and Eq. 26 is referred to as the the total probability or normalizing constant. Disregarding the normalizing constant the posterior probability can be expressed as proportional to the product of the prior and the likelihood [65]:

$$P(A_j|B_k) \propto P(B_k|A_j)P(A_j) \tag{27}$$

We consider an experiment where the event of interest is the entropy of an image. In the experiment the local entropy A of the image represent the image quality. The local entropy image is partitioned into $M = 2$ separate regions $m = \{1, 2\}$, the low entropy region A_1 and the high entropy region A_2. For each local region there are $K = 3$ possible tests or observations $k = \{1, 2, 3\}$ of image quality. They are the local contrast B_1, the local standard deviation B_2 and the local details B_3. The likelihood is defined as matching the structural information of an entropy region to corresponding region in an observation image:

$$P(B_1|A_m) = \frac{P(B_1 \bigcap A_m)}{P(A_m)} = \left(\frac{\left(\frac{n_{B_1,A_m}}{n_{fg}} \right)}{\left(\frac{n_{A_m}}{n_{fg}} \right)} \right) = \frac{n_{B_1,A_m}}{n_{A_m}} \tag{28}$$

$$P(B_2|A_m) = \frac{P(B_2 \bigcap A_m)}{P(A_m)} \left(\frac{\left(\frac{n_{B_2,A_m}}{n_{fg}} \right)}{\left(\frac{n_{A_m}}{n_{fg}} \right)} \right) = \frac{n_{B_2,A_m}}{n_{A_m}} \tag{29}$$

$$P(B_3|A_m) = \frac{P(B_3 \bigcap A_m)}{P(A_m)} \left(\frac{\left(\frac{n_{B_3,A_m}}{n_{fg}} \right)}{\left(\frac{n_{A_m}}{n_{fg}} \right)} \right) = \frac{n_{B_3,A_m}}{n_{A_m}} \tag{30}$$

where n_{B_k,A_m} are the number of pixels derived from matching the local entropy image to each observation image and n_{fg} is the number of foreground pixels.

An image quality attribute is defined as the posterior probability of a local entropy region given any of local contrast, local standard deviation and local details. The three image quality attributes are:

contrast quality attribute

$$P(A_m|B_1) \tag{31}$$

standard deviation quality attribute

$$P(A_m|B_2) \tag{32}$$

details quality attribute

$$P(A_m|B_3) \tag{33}$$

Accurate measurement of entropy is limited by the presence of noise [66]. Analytically entropy increases with sharpness but there is no fair correlation between entropy and noise [59]. The report in [59] suggest that entropy can be considered as a quality index only if noise can be considered a unique type of information which can be distinguished from Shannon entropy. We adopt this suggestion in order to find the best estimate of the prior probability $P(A_j)$. We define a global threshold $E_t = T_r$ to classify an image into energy bands. Lower energy band is below the threshold. In the lower energy band there is a general increase of entropy with sharpness. Noise is significant above the threshold. Given a normalized total clique potential E_t measured from the test image and a pre-defined global threshold T_r, the prior belief for both partitions of the sample space is defined as:

$$P(A_j) = \begin{cases} \left(\frac{\Phi(z_j)}{0.5}\right) & \text{if } E_t \leq T_r \\ P(E) & \text{otherwise} \end{cases} \tag{34}$$

where z_j is the z score for A_j and 0.5 is the maximum possible z score.

There are two methods to determined $P(E)$. The first method requires an estimate of the noise level in the slice. The contribution in [67] adopt the theoretical principles stated in Eq. 22, Eq. 23 and Eq. 24 to propose a power model that express the relationship between the normalized total clique potential E_t and the standard deviation $\hat{\sigma}$ of Gaussian noise for brain MRI images:

$$P(E) = a\hat{\sigma}^b + c, \quad \{E_t : 0 \leq E \leq 1\} \tag{35}$$

$$a_{b_g} = -1.67, a_{f_g} = -0.6863 \tag{36}$$

$$b_{b_g} = -0.6764, b_{f_g} = -0.3663 \tag{37}$$

$$c_{b_g} = 1.053, c_{f_g} = 1.105 \tag{38}$$

where $a_{b_g}, a_{f_g}, b_{b_g}, b_{f_g}, c_{b_g}, c_{f_g}$ are the model parameters for the foreground (f_g) and background (b_g) modes. Plots of the power models are shown in Fig. 1a for the foreground mode and Fig. 1b for the background mode.

In the second method to determine $P(E)$ we define D_r as the distance between T_r and the maximum possible total clique potential $E_{max} = 1$. Furthermore let D_r be linearly spaced into d number of points where d correspond to the operating range of noise level for the model.

$$D_r = \left(\frac{E_{max} - T_r}{d}\right) \tag{39}$$

The probability $P(E)$ of the total clique potential energy:

$$P(E) = E_t - \left(\frac{1}{d}\right)\left(\frac{E_t - T_r}{D_r}\right) + \left(\frac{E_t - T_r}{1 - T_r}\right) T_L \tag{40}$$

where T_L is a user defined lower limit of the probability scale. The default value of $T_L = 0$. In Eq. 40, for increasing total clique potential energy, the first term $E_t \rightarrow E_{max} = 1$, the second term $\left(\frac{1}{d}\right)\left(\frac{E_t - T_r}{D_r}\right) \rightarrow \left(\frac{1}{d}\right)\left(\frac{E_{max} - T_r}{D_r}\right) \rightarrow 1$, the third term $\left(\frac{E_t - T_r}{1 - T_r}\right) T_L \rightarrow \left(\frac{E_{max} - T_r}{1 - T_r}\right) T_L \rightarrow T_L$. Thus with increasing MRF energy the degree of spatial coherence approaches the lower limit of the probability scale.

The total conditional probability P_m for each m partition of the sample space is:

$$P_m = \sum_{k=1}^{K} (P(A_m|B_k)) w_k \tag{41}$$

where w_k are weights assigned to conditional probability derived from each observation, such that:

$$\sum_{k=1}^{K} w_k = 1 \tag{42}$$

We define the global conditional probability P as the total conditional probability of all the partitions in the sample space:

$$P = \sum_{m=1}^{M}(P_m)\lambda_m \tag{43}$$

where λ_m are weights assigned to conditional probability derived from each partition of the sample space, such that:

$$\sum_{m=1}^{M}\lambda_m = 1 \tag{44}$$

III. MATERIALS AND METHODS

A. Sources of Data

Data used for the performance evaluation of our proposed method were obtained from three sources. They are from NeuroRx research Inc. (https://www.neurorx.com), BrainCare Oy. (http://braincare.fi/) and the Alzheimer's disease neuroimaging initiative (ADNI) database (www.adni.loni.usc.edu).

NeuroRx research Inc. is an international clinical research organization dedicated to working with the pharmaceutical industry to facilitate clinical trials of new drugs for multiple sclerosis (MS) and other neurological diseases. BrainCare Oy (http://braincare.fi/) is a Tampere University of Technology spin-off company founded in 2013 to deliver personalized solutions to improve the quality of life of epilepsy patients. The organization recently concluded clinical trials for a novel mobile application and supporting solutions for long-term monitoring for epileptic patients. The (ADNI) was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD).

B. Setup of Experiment

1) Quality Model: We build an image quality model from 250 MRI volume data. The quality model is a normal probability distribution of random variables. The data was obtained from ADNI database. The random variable is the percentage composition of the low and the high entropy region in each volume data. Each volume data was acquired using high resolution T1 magnetization-prepared rapid gradient echo (MPRAGE) pulse sequence. Images acquired using MPRAGE pulse sequence was our choice to build a quality model because they exhibit superior gray-white matter contrast compared to the conventional T1 and other structural brain MRI images [68], [69], [70]. Each slice in a volume data has thickness of 1.2 mm and dimension of 190×160 pixels. Details of ADNI acquisition protocol and the initial processing steps is available in [71]. We seek the opinion of radiologists in the selection of MRI volume data on ADNI website. This is to ensure that only MPRAGE MRI volume data that meets the expected high quality attributes are selected for the quality model. Furthermore we seek MRI data of patients with healthy and normal brains that were without lesions or with very mild lesions. For each MRI data, slices towards the most inferior and most superior sections are discarded because they highlight more of scalp and bone structures than brain structures. The number of useful slices for each subject vary between 50 and 85. A total of 12005 slices were derived from the 250 MRI volume data. For each patient data the index of useful slices were coded in a special function which can be called up by the algorithm during the modeling experiment.

2) Test Data: The performance of our proposed method was evaluated on 350 MRI slices extracted from 31 MRI volume data. The test data were from two sources; 20 from NeuroRx and 11 from BrainCare. Each slice from NeuroRx has 2.4 mm thickness with dimension of 256×256 pixels in a 60-slice volume data. BrainCare data has thickness of 7.4 mm with dimension of 448×390 pixels in a 24-slice volume data. NeuroRx data consist of 10 conventional T1 images that were originally acquired with bias fields. The remaining data from NeuroRx were without any perceived distortion. They are 3 T2, 3 T1, 3 PD and 2 fluid attenuation inversion recovery (FLAIR) images. All the 11 MRI volume data from BrainCare had no perceived distortion. They are 3 T2, 3 T1, 2 PD and 2 FLAIR images. Description of the test data are displayed in Table I.

The test data were evaluated in their original state of acquisition. Different levels of motion blur and Rician noise were artificially induced on 2 T2 and 2 T1 images that are without perceived distortion. Motion blur was induced on a slice by convolving it with a special filter which approximates the linear motion of a camera. The linear motion is described by two parameters of the motion blur point spread function $H(x, y : L, \theta)$ [72]:

$$H(x, y : L, \theta) = \begin{cases} \frac{1}{L} & if \sqrt{x^2 + y^2} \leq \frac{L}{2} and \quad \frac{x}{y} = -\tan\theta \\ 0 & \text{otherwise} \end{cases} \tag{45}$$

where L is the linear distance in pixels and θ is the angular distance in degrees. The range of the linear and angular distances are $\{1 \leq L \leq 30\}$ and $\{1 \leq \theta \leq 60\}$, respectively. Both linear and angular distances were linearly spaced into 20 data points so that the filter generated 20 increasing levels of motion blur. Rician noise was generated in a three-step process. In the first step we generate Gaussian noise. The noise level is quantified as a percentage of the maximum pixel intensity level in the test image. For a m percent Rician noise level the standard deviation of the equivalent normal distribution is given by

$$\sigma \approx \mathcal{N}\left(0, \frac{\tau m}{100}\right) \tag{46}$$

where τ is the maximum pixel intensity [73]. In the second step we simulate the real and imaginary components in the complex plane of MRI acquisition process by adding two separate and identical Gaussian noise levels to the test image. The third step computes the magnitude of the complex data. The Rician noise level was scaled from 1 percent to 20 percent in steps of 1 percent.

C. Modeling Experiment

The foreground is segmented from the background region of each slice. Foreground are the pixels representing the anatomical structures. Local entropy image of each slice is computed using a 3×3 local entropy filter defined according to Eq. 1. The mean of the local entropy image is determined and used as a global threshold for transformation to the binary domain. Binary transformation allows classification into low and high entropy regions. The low energy and the high energy regions contain the zero pixels and ones pixels, respectively within the region enclosed by the foreground of the slice. The number of pixels in the low and high entropy regions are calculated and divided by the number of pixels in the foreground to determine the percentage composition of each region in a slice. The percentage composition of each region in all the slices in a volume data are summed and averaged to obtain the first random variable for the low $\mathcal{X}_{low,1}$ and the high entropy region $\mathcal{X}_{high,1}$. From the 250 MRI volume data we generate 250 random variables for the low entropy region $\{\mathcal{X}_{low,1}, \cdots, \mathcal{X}_{low,250}\}$ and the high entropy region $\{\mathcal{X}_{high,1}, \cdots, \mathcal{X}_{high,250}\}$. The mean derived by averaging the percentage composition of each region in all the slices of a volume data are finite. According to the central limit theorem, $\mathcal{X}_{low}$ and $\mathcal{X}_{high}$ will tend to a Gaussian distribution [74],[62],[75]. The mean of the normally distributed random variable for the low entropy region $\mu_{low} = 0.4734$ and the high entropy region $\mu_{high} = 0.5471$ are in agreement with the results reported in [76] on the percentage volume of white matter, gray matter and cerebrospinal fluids in a healthy adult brain. We build two normally distributed quality models for each normally distributed

random variables. Each quality model is symmetrical about $\mu_{low} = 0.4734$ and $\mu_{high} = 0.5471$. The standard deviation σ_{low}, σ_{high} for each quality model was computed from the principle of three-sigma rule [77].

$$P(\mu_{low} - 3\sigma_{low} \leq \mathcal{X} \leq \mu_{low} + 3\sigma_{low}) \approx 0.9973$$
$$P(\mu_{high} - 3\sigma_{high} \leq \mathcal{X} \leq \mu_{high} + 3\sigma_{high}) \approx 0.9973$$
(47)

D. How the Algorithm Works

The algorithm was implemented in the Matlab computing environment. The flow chart in Fig. 2 and the images in Fig. 3 explains how our proposed method works. Foreground extractor **FRX** segments the test image **TIM** shown in Fig. 3a into two regions; the foreground **FRG** region and the background region. The foreground region contain only the pixels that describe the anatomical structures in the slice. The foreground extractor is based on Otsu method for global image threshold [78]. Otsu method is computationally efficient and fully automated. It operates directly on the gray level pixels and automatically chooses the threshold that minimizes the weighted within-class variance. The foreground extractor also incorporates image filling operation and region property analysis. The image filling operation improve the accuracy of the extracted foreground by filling holes which mimic background pixels within the foreground region. Area threshold eliminate small regions such as extraneous and ghosting artifacts that mimic foreground pixels in the background region.

Feature extractor **FEX** consist of four separate local filters; entropy, contrast, standard deviation and details. These local filters extracts four image features from the original slice. They are local entropy **ENT** in Fig. 3b, local contrast **CON** shown in Fig. 3c and local standard deviation **STD** in Fig. 3d. The choice of filter size is determined by trade-off between spatial accuracy and computational efficiency. Generally, efficient performance of a filter is determined by the number of neighbouring pixels enclosed by its window. Larger filter size tend to dilate the original edge thickness that demarcates boundaries. This will cause loss of fine details in the image during the filtering process [79]. Filter size that is too small relative to the image dimension will result in loss of spatial coherence in the filtered image. Loss of spatial coherence manifest as discontinuous edges in the filtered image. For this reasons and based on our experience during the performance evaluation we recommend filter sizes of 3×3 and 5×5 for images with dimensions comparable to 256×256 and 512×512, respectively. Haar wavelet transform was applied [51] to extract the local details feature image shown in Fig. 3e.

A binary image transformer **BIT** computes the mean of each feature image. The mean is adopted as global threshold to transform the feature image into a binary image. The binary image is a cluster of each feature image into low energy region labeled with dark pixels and high energy region labeled with the bright pixels.

The operation of the algorithm is based on Bayes theory. Image quality is the event of interest. Entropy represents the event. The low entropy region (**EL**) and the high entropy region (**EH**) are regarded as the partitions of image quality sample space. Three separate tests or observations of each partitions of the sample space are the corresponding regions in the local contrast (**CL**, **CH**), local standard deviation (**SL**, **SH**) and local details (**DL**, **DH**) feature images. The feature images are extracted from the original slice. Thus the postprocessing of these feature images does not require additional resources such as computationally intensive image registration algorithm.

The first step in the execution of Bayes rule is the computation of the prior belief (**PBL**) in the low entropy region and the prior belief (**PBH**) in the high entropy region). The normalized total clique potential of the image is computed. We set a threshold of $T_r = 0.5$ for the normalized total clique potential energy of the image. The threshold separates noise from entropy so that entropy can be regarded as the aggregate of image quality attributes [59]. The threshold classifies an image into the low and the high energy bands. It is expected that good quality images, images with insignificant level of noise, blurry images and images degraded with bias fields will fall into the category of lower energy band. Images with significant level of noise will fall into the higher energy band.

For $T_r < 0.5$ the algorithm counts the number n_{A_m} of bright pixels in each entropy region. The ratio of the number of bright pixels in each region to the number n_{fg} of foreground pixels is the percentage composition $P(A_m)$ of each region. The percentage composition is inserted into Eq. 18 to standardize the quality model and determine the z score. The probability from the standard normal table is determined according to Eq. 21. The prior belief in each region is computed according to Eq. 34.

When $T_r \geq 0.5$ there are two methods to compute the prior belief. Where a noise estimation algorithm is available the prior belief is computed according to Eq. 35, otherwise the prior belief is computed according to Eq. 40.

The likelihood is computed for each entropy region. The bright pixels in each of the low contrast (**CL**), low standard deviation (**SL**) and low details (**DL**) feature images are separately matched, without registration, with the pixels in the low entropy region (**EL**). The ratio of the number n_{B_k, A_m} of their common elements to the number n_{fg} of foreground pixels is expressed by $P(B_k \bigcap A_m)$. The ratio of the number of bright pixels in the low entropy region to the number of foreground pixel gives $P(A_m)$. Both $P(B_k \bigcap A_m)$ and $P(A_m)$ are inserted into Eq. 28, Eq. 29 and Eq. 30. to derive three separate likelihood (**LL**) for the low entropy region. Three posterior probabilities corresponding to each likelihood are defined as the contrast, standard deviation and details attributes quality scores are computed according to Eq. 31, Eq. 32 and Eq. 33, respectively. The total quality score (**PTL**) for the low entropy region is the weighted sum of quality scores for quality attributes computed according to Eq. 41. Each quality attribute had equal weight. For the high entropy region (**EH**), similar steps are followed to compute three separate likelihood (**LH**), corresponding posterior probabilities and total quality score (**PTH**). The global quality score (**PG**) for the slice is the weighted sum of quality scores for each region according to Eq. 43.

E. Validation of Results

Our proposed method was validated with subjective experiments conducted with a group of human observers. The group consist of four radiologists and one MRI reader. MRI reader is a trained professional with experience working on MRI images that are affected by pathology [44]. The experiment was conducted with the aid of **QuickEval**, a web-based tool for psychometric image evaluation provided by the Norwegian Colour and Visual Computing Laboratory (www.colourlab.no/quickeval)at the Norwegian University of Science and Technology, Gjøvik, Norway. We choose the mean opinion score (MOS) subjective experiment method because of its popularity and simplicity. Mean opinion score is the average of the quality scores assigned to an image by multiple viewers [80]. There are four categories of the experiment. They are MRI volume data without perceived distortion, MRI volume data originally acquired with bias fields, MRI volume data artificially degraded with motion blur and MRI volume data artificially degraded by Rician noise. Table 1 display the categories of the experiment and the description of the MRI volume data utilized for the experiment. The observer assigns a score from 101 possible quality scores to each low and high energy regions of a slice. The possible scores are between 0 and 100, in steps of 1. Each region have equal weights, thus the global quality score is the average of the low and high entropy regions. In the category of MRI volume data with artificially induced distortion, each observer was first presented with an undistorted version of an MRI slice. This was followed by four different increasing levels of distortion of the original slice. The distorted levels are 5, 10, 15 and 20. The relationship between our objective results and the score assigned by human observers was determined using the spearman's rank correlation coefficient ρ [81]:

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n} \tag{48}$$

where n, the number of observations is the total number of slices contained in all the volume data in each category of the experiment, d is the difference between the two ranks of each observation.

IV. RESULTS

In this section we provide results of the experiments on five MRI volume data. One T2 and a conventional T1 volume data are from NeuroRx Inc. The T1 data was originally degraded by bias fields during acquisition. The data from BrainCare Oy are two T2 and a conventional T1 volume data.

1) MRI Volume Data Without Perceived Distortion: Two slices in a T2 volume data from BrainCare are shown in Fig. 4a, and Fig. 4b. The slices are without perceived distortion. The plots of the posterior probability for the low energy region, high energy region and total probability of the contrast, standard deviation and details quality attributes are displayed in Fig. 4c, Fig. 4d and Fig. 4e, respectively. Each quality attribute had average low energy quality score of 0.85. The high energy region of the contrast and standard deviation quality attribute had average of 0.8 compared to corresponding details quality score of 0.7. The global quality score for ten successive slices in the volume data is in Fig. 4f

Another two slices are shown in Fig. 5a and 5b. They are in a T1 volume data from BrainCare and are without perceived distortion. The plots of the posterior probability displayed in Fig. 5c, Fig. 5d and Fig. 5e have average of 0.7 and 0.6 for the low and the high energy regions of the contrast, standard deviation and details quality attributes, respectively. The average total score for each quality attribute is 0.55. Global quality scores for 10 successive slices in the volume data is displayed in Fig. 5f.

Figure 6a and Fig. 6b are two slices in a NeuroRx T2 volume data considered to be without perceived distortion. The average of the quality scores displayed in Fig. 6c, Fig. 6d and Fig. 6e for the low and high entropy regions is 0.9. The average quality score for 14 successive slices in the volume data (see Fig. 6f) is 0.8.

2) MRI Volume Data Degraded by Motion Blur: The image in Fig. 7a was degraded by 20 different levels of motion blur. Figure 7b is the degraded version of the image in Fig. 7a with motion bur level of 20. The plots of the posterior probability corresponding to the low and the high entropy regions for each quality attributes are shown for Fig. 7c, Fig. 7d and Fig. 7e. These plots and the plot in Fig. 7f show decrease in quality scores for different levels of degradation by motion blur. At insignificant level of blur the quality score for the low entropy region is close to 0.9 for the three quality attributes. The quality scores decreases in proportion to the successive increase in blur level to quality score of 0.4 for motion blur level of 20. In the high entropy region the three quality attributes exhibit different profiles. Contrast and standard deviation decrease from 0.6 to 0.4 and 0.6 to 0.35, respectively. There is slight increase from 0.5 to 0.55 in the level of details for motion blur increase from 1 to 5. Subsequently there is progressive decrease in quality score from 0.55 to 0.3 for motion blur level decrease from 6 to 20. The plot in Fig. 7f show a general decrease in image quality with increasing levels of motion blur.

3) MRI Volume Data Degraded by Noise: Figure 8a and Fig. 8b are the original image and its degraded version, respectively. The Rician noise level varies from 1 percent to 20 percent. The image on Fig. 8b is degraded by 20 percent Rician noise. Plots of the posterior probabilities of the three quality attributes for the low and the high entropy regions are displayed in Fig. 8c, Fig. 8d and Fig. 8e. The global quality score is shown in Fig. 8f. The plots of the contrast and standard deviation quality attributes shown in Fig. 8c and Fig. 8d have very close corresponding quality scores. The zero distortion level quality score for the high entropy region of the details quality attribute is a lower value of 0.5 compared to 0.6 for the other quality attributes. Corresponding maximum distortion level quality score is 0.3 compared to 0.35 for the other quality attributes. Loss of details is clearly evident when the visual quality of the original image in Fig. 8a is compared to its degraded version in Fig. 8b. Figure 8f shows a general decrease of image quality with increasing noise level.

4) MRI Volume Data Degraded by Bias Fields: Figure 9a, Fig. 9b, Fig. 9c, Fig. 9d and Fig. 9e are five slices in a 60-slice T1 MRI volume data. The volume data was degraded by bias fields during its acquisition and the slices exhibit different configurations of bias fields. The three quality attributes of 21 successive slices in the volume data are displayed in Fig. 9f , Fig. 9g and Fig. 9h. There are variations in the quality scores for the different slices. The global quality scores of slice numbers 1, 6, 8, 14 and 19 displayed in Fig. 9 are 0.2, 0.2, 0.25, 0.3 and 0.5, respectively.

5) Validation of Results: The plots in Fig. 10 show the results of the validation experiments on MRI volume data with no perceived distortion. The objective scores recorded by our proposed method and the scores recorded by human observers for T2, T1, FLAIR and PD images are displayed in Fig. 10a, Fig. 10b, Fig. 10c and Fig. 10d, respectively. In the T2 volume data shown in Fig. 10a our proposed method recorded 0.85, 0.7 and 0.8 for the low, high and global quality scores. The scores recorded by human observers are lower by an average of 30 percent. The corresponding scores are 0.7, 0.6 and 0.65. For conventional T1 MRI volume data shown in Fig. 10b there is an overage of 20 percentage difference between our proposed method and human observers. Our proposed method recorded 0.7, 0.65 and 0.6 for the low, high and global quality scores compared to 0.60, 0.55 and 0.6, the corresponding quality scores recorded by the human observers. The objective quality scores are 0.85, 0.7 and 0.8 for the low, high and global regions, respectively for FLAIR images (see Fig. 10c). The scores corresponding to human observers are lower by 30 percent; 0.75, 0.55 and 0.6. The plot in Fig. 10d shows that there is 30 percent difference in the quality scores recorded by our proposed method and human observers. The low, high and global quality scores recorded by our proposed method on PD images are 0.9, 0.6 and 0.8, respectively compared to 0.7, 0.6 and 0.6, the corresponding scores recorded by human observers.

In Fig. 11a, Fig 11b and Fig. 11c we compare the quality scores recorded by our proposed method on T2 images degraded by motion blur with the scores assigned by human observers. In the absence of motion blur (zero level) our proposed method recorded 0.9, 0.7 and 0.8 for the low, high and global regions. Human observers recorded 0.8 for the low energy region. This is 10 percent lower than corresponding region recorded by our proposed method. The scores recorded by our proposed method and human observers for the high energy and global regions are very close to each other. The scores are 0.7 and 0.65 for the high entropy region and 0.8 and 0.75 for the global region. The plots show that our proposed method and human observers recorded gradual decrease in quality scores for increasing level of motion blur. At motion blur level of 20 our proposed method recorded 0.3, 0.2 and 0.25 for the low, high and global regions, respectively. Quality scores corresponding to human observers are 0.2, 0.1 and 0.2.

The plots in Fig. 11d, Fig. 11e and Fig. 11f show that our proposed method is comparable to quality scores recorded by human observers on T2 images degraded by different levels of Rician noise. In the absence of noise our proposed method and human observers predicted quality scores of 0.6 for the high energy region. In the low energy region our proposed method recorded 0.9 compared to 0.75, a 20 percent decrease with respect to the score recorded by human observers. Comparable quality scores of 0.75 and 0.7 were recorded by our proposed method and human observers, respectively for the global region. For increasing level of Rician noise there is a general decrease in the quality score recorded by our proposed method and human observers. At Rician noise level of 20 percent our proposed method predict quality scores of 0.3, 0.2 and 0.25 for the low, high and global regions. Corresponding quality scores by human observers were lower by 30 percent; 0.3, 0.25 and 0.25.

Validation of our proposed method by human observers on conventional T1 images degraded by bias fields is displayed in Fig. 11g. Our proposed method recorded 0.45, 0.4 and 0.4 for the low, high and global regions. Quality scores of 0.4, 0.3 and 0.35 assigned by human observers for the low, high and global regions are lower by 10 percent, 30 percent and 15 percent, respectively.

6) Comparative Performance Evaluation: It is difficult to carry out comparative performance evaluation on current methods for post-acquisition quality evaluation of brain MRI images. Existing methods adopt different distortion models. There is currently no ground truth data available for clinical brain MRI images. Evaluation results reported are data dependent because the different techniques are evaluated on different types of image data. The use of phantom is not an effective approach for comparative performance evaluation because it lacks the natural anatomical variability and image acquisition artifacts that are usually encountered in real images [82]. Furthermore the algorithms for these proposed methods are not readily available from the authors and it is difficult to faithfully implement the techniques without direct interaction with the authors [41]. Having regard to these shortcomings we describe three characteristics that distinguish our proposed methods from existing post-quality assessment methods for brain MRI images. The comparative performance evaluation is based on sharpness and noise levels quality measures. The

global sharpness was defined in [58] and implemented as a quality measure in [41]. Noise level estimation was according to [83].

The image in Fig. 12a is a slice in a T1 MRI volume data. The data was originally acquired as degraded with bias fields. The plots in Fig. 12d and Fig. 12g are the measures of global sharpness and noise levels from successive 21 slices in the volume data. The image in Fig. 12b is a slice from a T2 volume data. The variation of global sharpness measured from the image and the estimated noise levels for 20 different levels of artificially induced noise are displayed in Fig. 12e and Fig. 12h, respectively. The image in Fig. 12c is another slice in a T2 MRI volume data. It is degraded by 20 different levels of motion blur. Variation of sharpness and estimated noise levels for the different levels of motion blur are shown in Fig. 12f and Fig. 12i, respectively. Below we describe the three characteristics that distinguish our proposed method from existing post-acquisition quality evaluation methods.

1) Identification and Management of Quality Attributes
   In real-world scenarios it is not always possible to have prior knowledge of the specific type of distortion that degrades an image [84]. The most important step to achieve an efficient and manageable evaluation of image quality is to identify the attributes that essentially captures all the possible distortions in an image [15]. The plots in Fig. 12e demonstrate how global sharpness attribute of an image can efficiently discriminate between the different levels of noise in an image. The plot in Fig. 12h demonstrate accurate estimation of the different levels of noise in an image. The global sharpness in Fig. 12d and the estimated noise levels in Fig. 12g for the 21 successive slices degraded by bias fields are accurate estimates of quality attributes but are erroneous quality indicators for the image degraded by bias fields (see Fig. 12a). Single quality attribute or too few relevant attributes such as the combination of global sharpness and noise level cannot account for all the possible distortions present in brain MRI images [50].

2) Transformation of Distortion Levels to Quality Indices
   Image quality evaluation is the transformation of the different distortion levels into quality indices that correlates with the characteristics of human visual system. The plot in Fig. 12e shows how sharpness attribute decrease with increasing levels of noise. In the absence of noise the image sharpness is 700. With increasing levels of noise the image sharpness gradually decrease to 400 for noise level of 20 percent. Detection of distortion is the sharpness-noise level variations recorded for different brain MRI images as shown in Fig12d, Fig. 12e and Fig. 12f. There is no fixed scale that defines the lower and upper limit for the different levels of distortion recorded for different images. This makes detection of distortion useless quality index. The absence of a quality index makes it impossible to compare image quality from the different plots of sharpness-noise level variations. Quality measures such as global sharpness will be a useful quality indicator when it is transformed to a quality index.

3) Globalization
   The frontal and posterior regions of white matter structure of the T1 MRI image in Fig. 12a is demarcated by bias fields into two different homogeneous intensity levels. The pixel intensity levels of the posterior peripheral cortex is comparable to the intensity levels of the surrounding white matter. The presence of bias fields makes it difficult to visually distinguish the posterior peripheral cortex from its surrounding white matter. This form of distortion can cause even robust automated image analysis systems to give erroneous results. Sharpness and the estimated noise levels (Fig. 12d and Fig. 12g) in the image are global quality measures. They were computed by summing and averaging through the entire image. The average global sharpness of 6×10^4 and the average global estimate of noise level $\sigma \approx 7$ recorded for the slices in the volume data are erroneous quality indicators which does not reflect the perceived distortions in specific regions of the brain.

V. Discussion

In the experiment we evaluate the performance of our proposed method on different types of original brain MRI volume data. The proposed method compute image quality index for the low energy region, high

energy region and the whole-brain. The low energy region is dominated by the white matter structure. The high energy region include the cerebral cortex and the boundaries that demarcate the different anatomical structures. The whole-brain is the brain regarded as a single homogeneous region.

In all the categories of the experiment the results show very good correlation (≥ 0.9) between our proposed method and the subjective evaluation by human observers. This indicates that our proposed method correlates with human visual system. The high correlation can be attributed to four factors. They are the efficient performance our proposed algorithm, expertise of the human observers, the wide interval between distortion levels and the number of observations. The observers, by their expertise, have knowledge of successive slices extracted from the same MRI volume data. Their visual perception of the small variations in the image quality between the different but successive slices will be reflected in the subjective evaluation. The interval between the different levels of distortion is reasonably large for the human eye to efficiently discriminate between the differen levels of perceived distortions in the images. The number of data points used for computation of the correlation coefficient is reasonably high. The minimum number of data points in each category is 30.

The performance evaluation results indicate that the different regions of the brain may not have the same quality index. Our proposed method demonstrate very good evaluation of image quality at different regions of a slice and across slices of a MRI volume data. Very good objective evaluation was also recorded for MRI volume data with same and different acquisition protocols as well as different types and levels of perceived distortion.

The plots in Fig. 10 show that our objective quality assessment closely agrees with the subjective evaluation on MRI volume data without perceived distortion. There is a general decrease in image quality with increasing levels of motion blur and noise as displayed in Fig. 7 and Fig. 8. Our proposed method demonstrate excellent performance on images degraded by bias fields. The plots in Fig. 9f, Fig. 9g and Fig. 9h demonstrate accurate quantification of the variations in the configurations of bias fields across different slices in a MRI volume data.

Comparative performance evaluation displayed in Fig. 12 demonstrate that few relevant quality attributes such as noise and sharpness may be useful only for predicting noise level and global sharpness attributes but are inefficient for the prediction of image quality when bias fields is present in the images. Image quality is the aggregate of the effects of all possible distortions on an ideal image. Different types of distortions have different effects on the different homogeneous regions. Thus quality measure based on a single attribute or few relevant quality attributes cannot exhibit robust quality evaluation. Our proposed method is a robust and efficient quality evaluation method for structural brain MRI images. It identifies and adopt quality attributes that essentially captures all possible distortions with a strategy that optimally evaluates image quality in the different homogeneous regions.

There is increasing interest in the use of MRI imaging systems in the clinical trials of drugs for monitoring the progression of multiple sclerosis, Alzheimers disease and other neurological diseases. This report will encourage reliable results in many areas of brain volumetric analysis such as whole brain volume measurement, ventricular volume measurement, ventricular atrophy measurement, cortical gray matter atrophy measurement and cortical gray matter thickness measurement. Our proposed method can directly evaluate image quality of the corresponding regions-of-interest in a longitudinal data before input to an automated image analysis system.

In clinical research organizations quality control procedures requires physicians, radiologists and trained readers to subjectively evaluate the outputs of automated image analysis systems. Subjective image quality evaluation can be a cumbersome task for a trained reader. In the evaluation of MRI-derived atrophy metrics the reader's visual attention changes with the different metrics. The reader's visual attention is on the white matter region during white matter atrophy measurement. The focus changes towards the cortical gray matter region during gray matter atrophy measurement and to the whole-brain for whole-brain atrophy measurement. The good correlation between the outputs of our proposed method and expert human observers indicates that our proposed method can contribute to the efficient management of the large volume of data in clinical research organizations and meet the deadline to deliver image analysis

reports to the sponsoring pharmaceutical organizations.

Some image analysis tasks such as semi-automated identification of lesions requires the consensus of at least two experts. These tasks can be time consuming because of variability in the visual judgement of the experts. Objective method such as our proposed method can improve the discernment of a trained MRI reader, help find a common ground between two trained MRI readers and speed up consensus quality assessment. The inter-expert variability is much higher when it comes to image qualities that are considered borderline cases because the human visual system has no clearly defined threshold of quality index. Our proposed method can be used as a tool to define a quality index threshold that classifies an image into acceptable quality image and poor quality image.

Daily very large volumes of MRI data are transferred from clinical trial sites around the globe to clinical research organizations (CRO) that manage clinical trials for pharmaceutical organizations. This new quality evaluation method will find useful application in CRO environments to enforce the standardization of variations in the quality attributes of MRI images from different manufacturers, different clinical trial sites and different acquisition protocols.

VI. FUTURE RESEARCH DIRECTION

The goal attained in the current research work is a general framework to measure image quality at specific regions of the brain. In this work the quality model was built from MRI data of healthy subjects. In real-world scenarios the MRI volume data are derived from subjects with and without brain atrophy. Future work will consider building the quality model from MRI data that includes subjects with atrophy in the different regions of the brain. This approach will induce a trade-off which will improve the confidence in the prior belief computed by our algorithm. Current work regard the white matter, thalamus and the ventricle as a single region which we termed the low entropy region. In the future the low entropy region will be segmented into three separate regions consisting of the white matter, thalamus and ventricle. This will extend the application of our proposed method to quality evaluation of the thalamus and the ventricle regions.

VII. CONCLUSIONS

We hereby propose a new post-acquisition quality assessment method for structural brain MRI images. Our proposed method is significant because of the increasing interest in the use of MRI system for monitoring disease progression in multiple sclerosis, Alzheimers disease and other neurological diseases. We exploit the relationship between entropy and the classical image quality attributes to develop new quality measures for brain MRI images. Entropy is regarded as the aggregate of image quality attributes. Local contrast, local standard deviation and local details are the tests of quality attribute. Quality measure is formulated as a probability problem with focus on the different homogeneous regions of the brain. Bayes theorem is applied to compute the quality scores. Experimental results demonstrate that our proposed method gave good quality measures across images with different acquisition protocols, different types and levels of distortion and correlates with subjective evaluation by human observers. It will be suitable for automated environments and in applications where specific regions of the the brain are required for image analysis. This new method will encourage the use of MRI images of acceptable quality in MRI-based clinical trials, ensure accurate diagnosis and improve the performance of a trained reader in the performance evaluation of image analysis systems. The algorithm does not require image registration. It operates on binary images and thus has the potential of real-time operation.

REFERENCES

[1] D. Arnold, X. You, S. Shang, B. Sperling, and V. Evilevitch, "Long-term efficacy in mri and no evidence of disease activity outcomes in patients with relapsing-remitting multiple sclerosis treated with peginterferon beta-1a (p7.266)," *Neurology*, vol. 84, no. 14 Supplement, 2015.

[2] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural mri in alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010.

[3] R. H. Hashemi, W. G. Bradley, and C. J. Lisanti, *MRI: the basics.* Lippincott Williams & Wilkins, 2012.

[4] J. Schwaab, Y. Diez, A. Oliver, R. Mart, J. v. Zelst, A. Gubern-Mrida, A. B. Mourri, J. Gregori, and M. Gnther, "Automated quality assessment in three-dimensional breast ultrasound images," *Journal of Medical Imaging*, vol. 3, no. 2, p. 027002, 2016.

[5] H. H. Barrett, K. J. Myers, C. Hoeschen, M. A. Kupinski, and M. P. Little, "Task-based measures of image quality and their relation to radiation dose and patient risk," *Physics in medicine and biology*, vol. 60, no. 2, p. R1, 2015.

[6] E. Abouei, S. Lee, and N. L. Ford, "Quantitative performance characterization of image quality and radiation dose for a cs 9300 dental cone beam computed tomography machine," *Journal of Medical Imaging*, vol. 2, no. 4, p. 044002, 2015.

[7] H. Chen, M. Danielsson, C. Xu, and B. Cederstrm, "On image quality metrics and the usefulness of grids in digital mammography," *Journal of Medical Imaging*, vol. 2, no. 1, p. 013501, 2015.

[8] M. A. Kupinski and H. H. Barrett, *Small-animal SPECT imaging.* Springer, 2005, vol. 233.

[9] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 323–331, 2003.

[10] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.

[11] L. Qin, E. J. Schmidt, Z. T. H. Tse, J. Santos, W. S. Hoge, C. Tempany-Afdhal, K. Butts-Pauly, and C. L. Dumoulin, "Prospective motion correction using tracking coils," *Magnetic Resonance in Medicine*, vol. 69, no. 3, pp. 749–759, 2013.

[12] Z. Caramanos, V. S. Fonov, S. J. Francis, S. Narayanan, G. B. Pike, D. L. Collins, and D. L. Arnold, "Gradient distortions in mri: characterizing and correcting for their effects on siena-generated measures of brain volume change," *NeuroImage*, vol. 49, no. 2, pp. 1601–1611, 2010.

[13] P. Kaur, S. S. Kumaran, R. Tripathi, S. Khushu, and S. Kaushik, "Protocol error artifacts in mri: sources and remedies revisited," *Radiography*, vol. 13, no. 4, pp. 291–306, 2007.

[14] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4. IEEE, 2002, pp. IV–3313.

[15] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen, "Attributes of image quality for color prints," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 016–011 016–13, 2010.

[16] M. Filippi, M. A. Rocca, O. Ciccarelli, N. De Stefano, N. Evangelou, L. Kappos, A. Rovira, J. Sastre-Garriga, M. Tintorè, J. L. Frederiksen *et al.*, "Mri criteria for the diagnosis of multiple sclerosis: Magnims consensus guidelines," *The Lancet Neurology*, vol. 15, no. 3, pp. 292–303, 2016.

[17] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux *et al.*, "The diagnosis of dementia due to alzheimers disease: Recommendations from the national institute on aging-alzheimers association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's & dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[18] S. M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P. Matthews, A. Federico, and N. De Stefano, "Accurate, robust, and automated longitudinal and cross-sectional brain change analysis," *Neuroimage*, vol. 17, no. 1, pp. 479–489, 2002.

[19] K. Nakamura, N. Guizard, V. S. Fonov, S. Narayanan, D. L. Collins, and D. L. Arnold, "Pairwise jacobian integration method to measure grey matter atrophy in multiple sclerosis," in *International Society for Magnetic Resonance in Medicine Workshop on Multiple Sclerosis as a Whole-Brain Disease*, 2013.

[20] K. Nakamura, N. Guizard, V. Fonov, S. Narayanan, L. Collins, and D. Arnold, "Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis," *NeuroImage: Clinical*, vol. 4, pp. 10–17, 2014.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[22] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," *Proceedings of the National Academy of Sciences*, vol. 90, no. 21, pp. 9758–9765, 1993.

[23] H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. ii. fisher information, fourier crosstalk, and figures of merit for task performance," *J. Opt. Soc. Am. A*, vol. 12, no. 5, pp. 834–852, May 1995.

[24] H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A*, vol. 7, no. 7, pp. 1266–1278, Jul 1990.

[25] C. K. Abbey and H. H. Barrett, "Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *JOSA A*, vol. 18, no. 3, pp. 473–488, 2001.

[26] M. A. Kupinski, J. W. Hoppin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging with use of markov-chain monte carlo techniques," *JOSA A*, vol. 20, no. 3, pp. 430–438, 2003.

[27] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *JOSA A*, vol. 4, no. 12, pp. 2447–2457, 1987.

[28] S. Park, A. Badano, B. D. Gallas, and K. J. Myers, "Incorporating human contrast sensitivity in model observers for detection tasks," *IEEE Transactions on medical imaging*, vol. 28, no. 3, pp. 339–347, 2009.

[29] H. H. Barrett, L. R. Furenlid, M. Freed, J. Y. Hesterman, M. A. Kupinski, E. Clarkson, and M. K. Whitaker, "Adaptive spect," *IEEE transactions on medical imaging*, vol. 27, no. 6, pp. 775–788, 2008.

[30] M. Freed, M. A. Kupinski, L. R. Furenlid, D. W. Wilson, and H. H. Barrett, "A prototype instrument for single pinhole small animal adaptive spect imaging," *Medical physics*, vol. 35, no. 5, pp. 1912–1925, 2008.

[31] C. P. Favazza, K. A. Fetterly, N. J. Hangiandreou, S. Leng, and B. A. Schueler, "Implementation of a channelized hotelling observer model to assess image quality of x-ray angiography systems," *Journal of Medical Imaging*, vol. 2, no. 1, pp. 015 503–015 503, 2015.

[32] A. K. Jha, M. A. Kupinski, J. J. Rodríguez, R. M. Stephen, and A. T. Stopeck, "Task-based evaluation of segmentation algorithms for diffusion-weighted mri without using a gold standard," *Physics in medicine and biology*, vol. 57, no. 13, p. 4425, 2012.

[33] M. A. Kupinski, J. W. Hoppin, J. Krasnow, S. Dahlberg, J. A. Leppo, M. A. King, E. Clarkson, and H. H. Barrett, "Comparing cardiac ejection fraction estimation algorithms without a gold standard," *Academic radiology*, vol. 13, no. 3, pp. 329–337, 2006.

[34] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger *et al.*, "Sense: sensitivity encoding for fast mri," *Magnetic resonance in medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[35] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (grappa)," *Magnetic resonance in medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.

[36] C. Mistretta, O. Wieben, J. Velikina, W. Block, J. Perry, Y. Wu, and K. Johnson, "Highly constrained backprojection for time-resolved mri," *Magnetic resonance in medicine*, vol. 55, no. 1, pp. 30–40, 2006.

[37] J. Tsao, P. Boesiger, and K. P. Pruessmann, "k-t blast and k-t sense: Dynamic mri with high frame rate exploiting spatiotemporal correlations," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1031–1042, 2003.

[38] F. A. Breuer, P. Kellman, M. A. Griswold, and P. M. Jakob, "Dynamic autocalibrated parallel imaging using temporal grappa (tgrappa)," *Magnetic resonance in medicine*, vol. 53, no. 4, pp. 981–985, 2005.

[39] J. Miao, F. Huang, S. Narayan, and D. L. Wilson, "A new perceptual difference model for diagnostically relevant quantitative image quality evaluation: A preliminary study," *Magnetic resonance imaging*, vol. 31, no. 4, pp. 596–603, 2013.

[40] J. Miao, D. Huo, and D. L. Wilson, "Quantitative image quality evaluation of mr images using perceptual difference models," *Medical physics*, vol. 35, no. 6, pp. 2541–2553, 2008.

[41] J. Woodard and M. Carley-Spencer, "No-reference image quality metrics for structural mri," *Neuroinformatics*, vol. 4, no. 3, pp. 243–262, 2006.

[42] B. Mortamet, M. A. Bernstein, C. R. Jack, J. L. Gunter, C. Ward, P. J. Britson, R. Meuli, J.-P. Thiran, and G. Krueger, "Automatic quality assessment in structural brain magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 62, no. 2, pp. 365–372, 2009.

[43] F. Prieto, M. Guarini, C. Tejos, and P. Irarrazaval, "Metrics for quantifying the quality of mr images," in *Proceedings of the 17th Annual Meeting of ISMRM*, 2009, p. 4696.

[44] E. L. Gedamu, D. Collins, and D. L. Arnold, "Automated quality control of brain mr images," *Journal of Magnetic Resonance Imaging*, vol. 28, no. 2, pp. 308–319, 2008.

[45] F. Heckel, F. Arlt, B. Geisler, S. Zidowitz, and T. Neumuth, "Evaluation of image quality of mri data for brain tumor surgery," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 97 871L–97 871L.

[46] B. Liu, "Parallel magnetic resonance imaging: Theory, algorithm and applications," Ph.D. dissertation, Milwaukee, WI, USA, 2008, aAI3318277.

[47] A. Mittal, A. Moorthy, and A. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[48] G. McCarthy and C. C. Wood, "Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models," *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 62, no. 3, pp. 203–208, 1985.

[49] N. De Stefano, A. Giorgio, M. Battaglini, M. Rovaris, M. Sormani, F. Barkhof, T. Korteweg, C. Enzinger, F. Fazekas, M. Calabrese *et al.*, "Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes," *Neurology*, vol. 74, no. 23, pp. 1868–1876, 2010.

[50] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.

[51] H. H. Barrett and K. J. Myers, *Foundations of image science*. John Wiley & Sons, 2013.

[52] K. E. Wendel, "The influence of tissue conductivity and head geometry on eeg measurement sensitivity distributions," Ph.D. dissertation, Tampereen teknillinen yliopisto. Julkaisu-Tampere University of Technology., 2010.

[53] H. Barrett and K. Myers, "Statistical characterization of radiological images: basic principles and recent progress," in *Medical Imaging*. International Society for Optics and Photonics, 2007, pp. 651 002–651 002.

[54] B. Monica, *Fundamentals in Information Theory and Coding*.   Springer, 2011.

[55] S.-D. Chen and A. R. Ramli, "Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 1301–1309, 2003.

[56] R. A. Johnson, D. W. Wichern *et al.*, *Applied multivariate statistical analysis*.   Prentice hall Upper Saddle River, NJ, 2002, vol. 5, no. 8.

[57] E. Peli, "Contrast in complex images," *JOSA A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[58] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar, "Multiresolution histograms and their use for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 831–847, July 2004.

[59] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *JOSA A*, vol. 24, no. 12, pp. B42–B51, 2007.

[60] H. Zhu, H. Shu, J. Zhou, X. Dai, and L. Luo, "Conditional entropy maximization for pet image reconstruction using adaptive mesh model," *Computerized Medical Imaging and Graphics*, vol. 31, no. 3, pp. 166–177, 2007.

[61] C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Transactions on image processing*, vol. 2, no. 1, pp. 96–103, 1993.

[62] A. Leon-Garcia, *Probability, Statistics and Random Processes for Electrical Engineering*.   3rd edition, Chapter 3: Pearson Education Inc., Prentice Hall, 2008.

[63] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[64] S. Li, *Markov Random Field Modeling in Image Analysis*.   3rd edition, Chapter 2: Springer Publishing Company, Inc., 2009.

[65] P. M. Lee, *Bayesian statistics: an introduction*.   John Wiley & Sons, 2012.

[66] A. Jayawardena, P. Xu, and W. Li, "Modified correlation entropy estimation for a noisy chaotic time series," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 2, p. 023104, 2010.

[67] M. Osadebey, N. Bouguila, and D. Arnold, "The clique potential of markov random field in a random experiment for estimation of noise levels in 2d brain mri," *International Journal of Imaging Systems and Technology*, vol. 23, no. 4, pp. 304–313, 2013.

[68] N. Shiee, P.-L. Bazin, J. L. Cuzzocreo, C. Ye, B. Kishore, A. Carass, P. A. Calabresi, D. S. Reich, J. L. Prince, and D. L. Pham, "Reconstruction of the human cerebral cortex robust to white matter lesions: method and validation," *Human brain mapping*, vol. 35, no. 7, pp. 3385–3401, 2014.

[69] J. P. Mugler and J. R. Brookeman, "Three-dimensional magnetization-prepared rapid gradient-echo imaging (3d mp rage)," *Magnetic Resonance in Medicine*, vol. 15, no. 1, pp. 152–157, 1990.

[70] P. Reimer, J. F. Meaney, P. M. Parizel, and F. A. Stichnoth, *Clinical MR imaging*.   Springer, 2010.

[71] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L Whitwell, C. Ward *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.

[72] A. M. Deshpande and S. Patnaik, "A novel modified cepstral based technique for blind estimation of motion blur," *Optik-International Journal for Light and Electron Optics*, vol. 125, no. 2, pp. 606–615, 2014.

[73] P. Coupe, J. Manjon, E. Gedamu, D. Arnold, M. Robles, and D. Collins, "Robust rician noise estimation for mr images," *Physics in Medicine and Biology*, vol. 14, no. 4, pp. 483–493, Aug. 2010.

[74] D. L. Ruderman, "The statistics of natural images," *Network: computation in neural systems*, vol. 5, no. 4, pp. 517–548, 1994.

[75] G. Norman and D. Streiner, *Biostatistics: The Bare Essentials*.   3rd edtion, Chapter 4: People's Medical Publishing House.

[76] E. Lüders, H. Steinmetz, and L. Jäncke, "Brain size and grey matter volume in the healthy human brain," *Neuroreport*, vol. 13, no. 17, pp. 2371–2374, 2002.

[77] E. W. Grafarend, *Linear and nonlinear models: fixed effects, random effects, and mixed models*.   Walter de Gruyter, 2006.

[78] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[79] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 2, pp. 165–168, March 1980.

[80] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (mos) revisited: methods and applications, limitations and alternatives," *Multimedia Systems*, vol. 22, no. 2, pp. 213–227, 2016.

[81] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach*.   John Wiley & Sons, 2014.

[82] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in mri," *IEEE transactions on medical imaging*, vol. 26, no. 3, pp. 405–421, 2007.

[83] J. Immerkaer, "Fast noise variance estimation," *Computer vision and image understanding*, vol. 64, no. 2, pp. 300–302, 1996.

[84] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3129–3138, 2012.

AUTHORS' BIOGRAPHY

Michael Osadebey obtained a masters degree in biomedical engineering from Tampere University, Finland in 2009 and a PhD in engineering and computer science from Concordia University, Montreal, Canada in 2015. His PhD study was focused on the processing of MRI images of the brain. Michael is currently a MRI Reader at NeuroRx Research Inc, a Montreal-based clinical research organization. Michael Osadebey duties at NeuroRx include application of advanced image analysis software in the reading of longitudinal MRI data of neurological diseases undergoing clinical trial drug treatment.

Marius Pedersen received his BsC in Computer Engineering in 2006, and MiT in Media Technology in 2007, both from Gjvik University College, Norway. He completed a PhD program in color imaging in 2011 from the University of Oslo, Norway, sponsored by Oce. He is currently employed as an associate professor at NTNU Gjøvik, Norway. He is also the director of the Norwegian Colour and Visual Computing Laboratory (Colourlab). His work is centered on subjective and objective image quality.

Dr. Douglas Arnold, is the director Magnetic Resonance Spectroscopy Lab, McGill University, Montreal, Canada and the President/CEO NeuroRx research Inc., a Montreal-based clinical research organization. Dr. Arnold is a neurologist with special expertise in MRI. His personal research interests are centered on the use of advanced neuro-imaging techniques to assess the pathological evolution of multiple sclerosis and Alzheimer's disease and to quantify the effects of therapy on these diseases.

Katrina Wendel-Mitoraj obtained her PhD in biomedical engineering from Tampere University of Technology in 2010. Her PhD study was focused on electroencephalography (EEG) electrode sensitivity distributions. Katrina is the CEO and founder BrainCare Oy. BrainCare Oy (http://braincare.fi/) is a Tampere University of Technology spin-off company founded in 2013 to deliver personalized solutions to improve the quality of life of epilepsy patients. The organization recently concluded clinical trials for a novel mobile application and supporting solutions for long-term monitoring for epileptic patients.

CONFLICT OF INTEREST STATEMENT

Dr. Michael Osadebey reports no potential conflicts of interest.

Dr. Marius Pedersen reports no potential conflicts of interest.

Dr. Douglas Arnold reports no potential conflicts of interest.

Dr. Katrina Wendel-Mitoraj reports no potential conflicts of interest.

TABLE I
DESCRIPTION OF TEST DATA UTILIZED FOR THE PERFORMANCE EVALUATION OF OUR PROPOSED METHOD FOR THE QUALITY EVALUATION OF AN MRI SLICE

Categories of Experiment	Type of MRI Volume Data	Number of MRI Volume Data	Number of Slices
MRI volume data without perceived distortion	T2	2	40
	T1	2	30
	PD	5	30
	FLAIR	4	30
MRI volume data degraded by bias fields	T1	10	100
MRI volume data degraded by motion blur	T2	2	30
	T1	2	30
MRI volume data degraded by noise	T2	2	30
	T1	2	30
Total		**31**	**350**
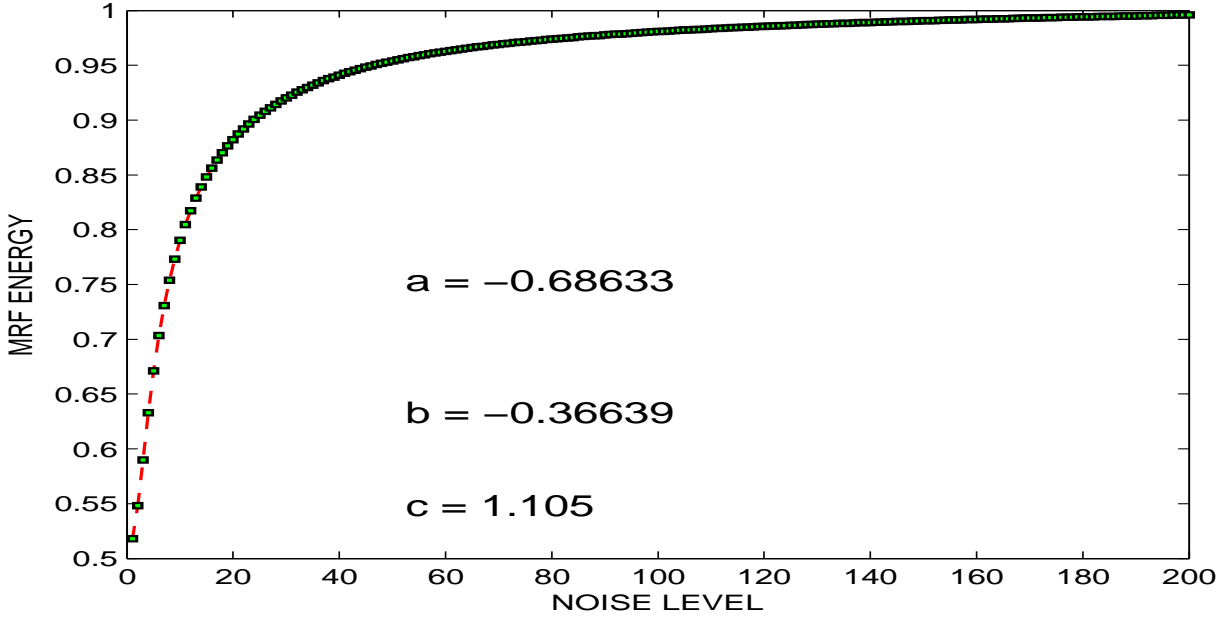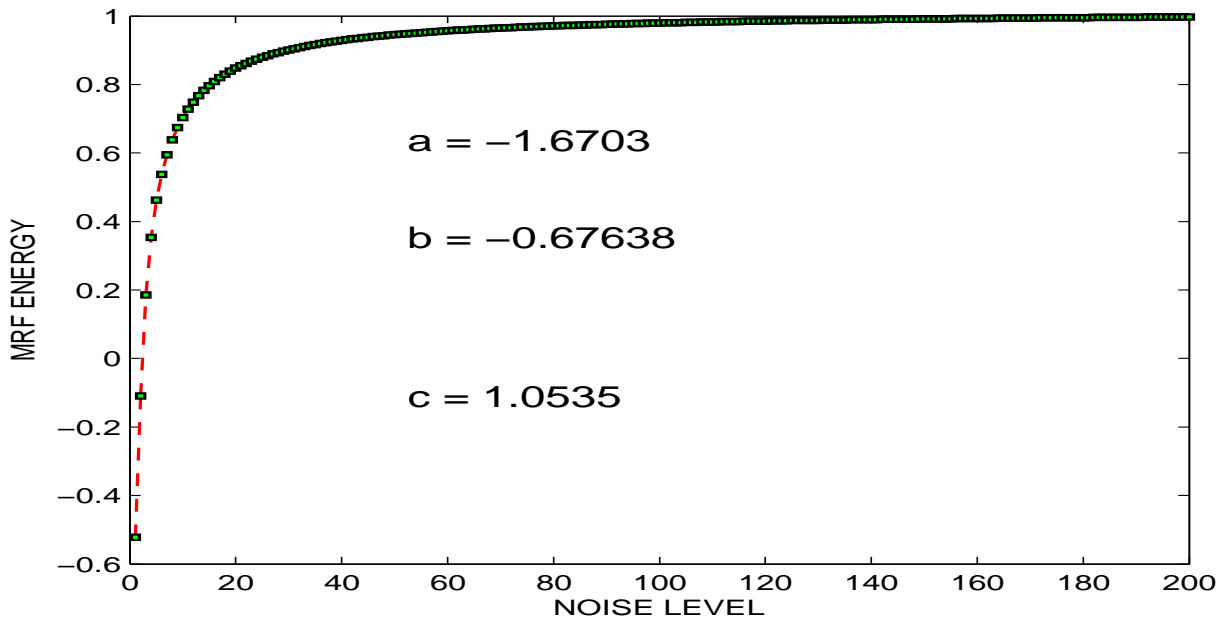
Fig. 1. The plots of the proposed generalized mathematical models for describing the relationship between TCP energy and noise level in the (a) foreground and (b) background modes.
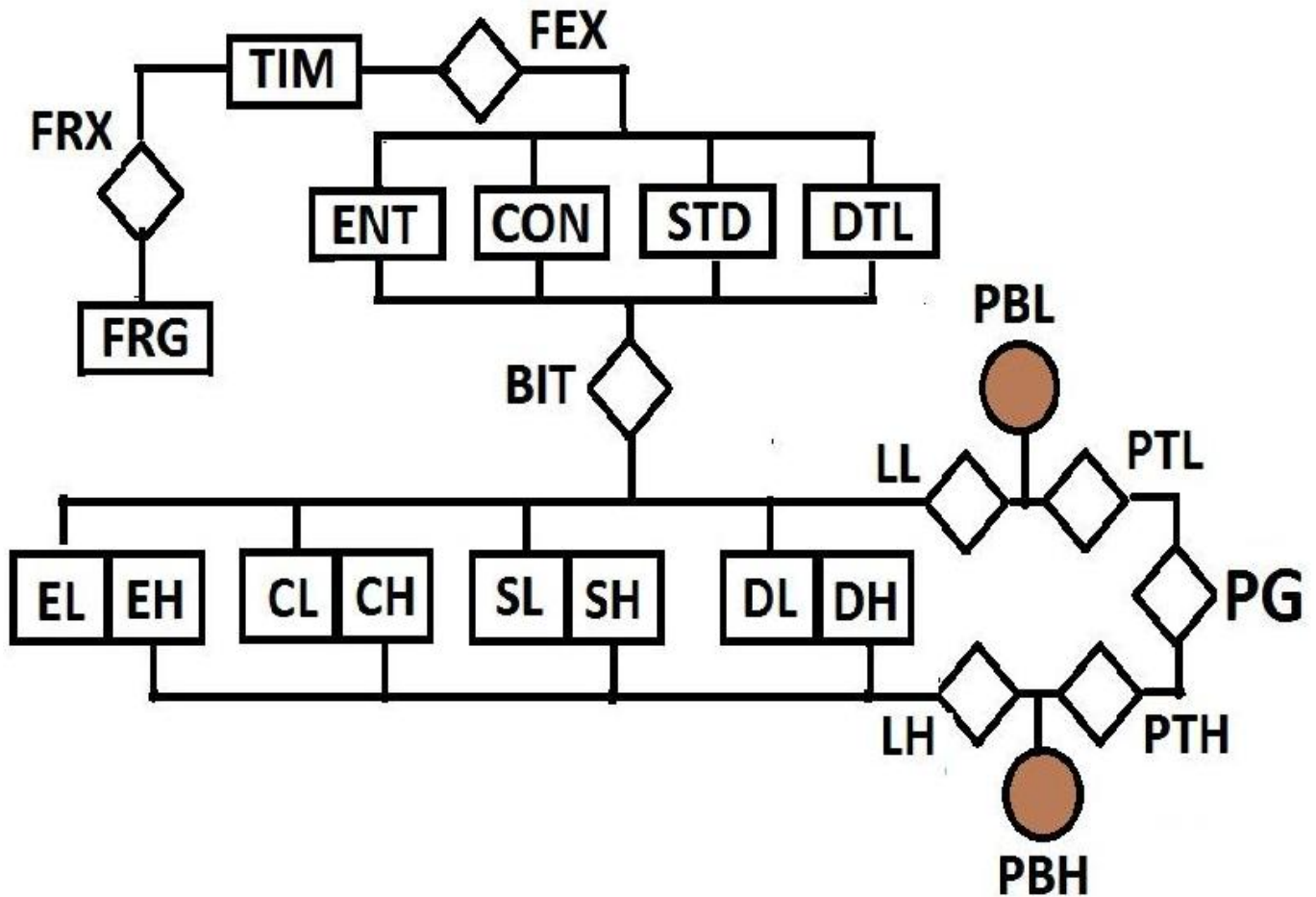
Fig. 2. The flow chart of post-acquisition quality evaluation for a brain MRI image. The event of interest, local entropy (**ENT**), is extracted (**FEX**) from the test image, Three separate observations; local contrast (**CON**), local standard deviation (**STD**) and local details (**DTL**) feature images are also extracted (**FEX**) from the same test image. These images, except foreground (**FRG**), are transformed (**BIT**) to the binary domain for classification into low and high energy regions. Each region (**EL**, **EH**) in the local entropy feature image (**ENT**) is combined with corresponding region in the contrast (**CL**, **CH**), standard deviation (**SL**, **SH**) and details (**DL**, **DH**) feature images to obtain the likelihood (**LL**, **LH**) for each observation. For each region, Bayes rule combines the prior belief (**PBL**, **PBH**) with the likelihood (**LL**, **LH**) corresponding to each feature image to compute quality score for each quality attribute. The total quality score for each region (**PTL**, **PTH**) is the weighted sum of the quality score of each quality attribute. The global quality score (**PG**) of the slice is the weighted sum of the total quality score (**PTL**, **PTH**) for each region.
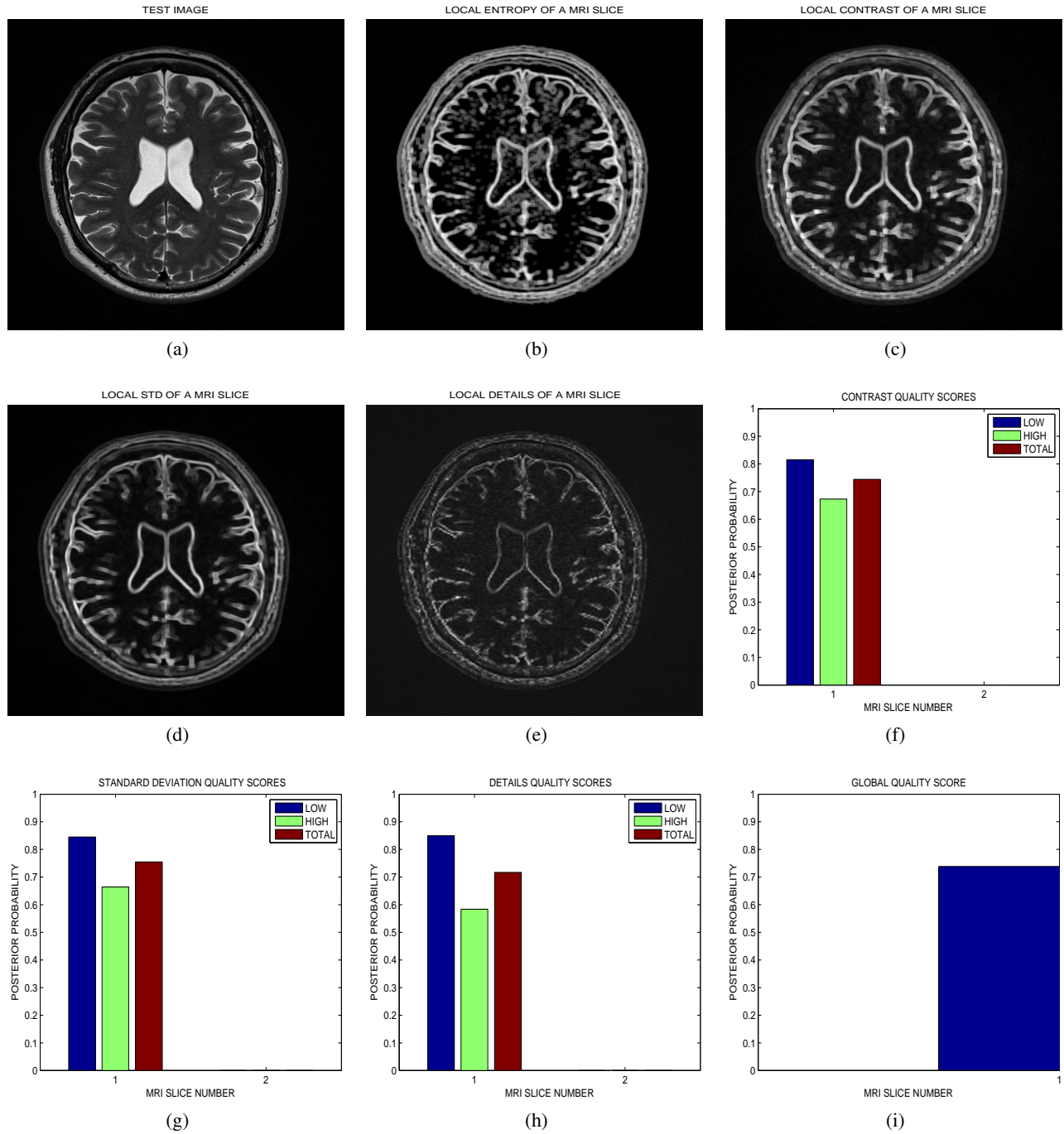
Fig. 3. Bayesian framework inspired post-acquisition quality assessment of (a) a brain MRI slice (**TIM**) in a MRI volume data. The event of interest, (b) the local entropy (**ENT**) which represents image quality is extracted from the test image. Three separate observations of image quality are (c) local contrast (**CON**) (d) local standard deviation (**STD**) (e) and local details (**DTL**) feature images. The four feature images are transformed to the binary domain and classified into the low and the high energy regions. For each region, the posterior probability of each region given any of the observations gives quality score for (f) contrast quality attribute, (g) standard deviation quality attribute and (h) details quality attribute. The total quality score for each region is the weighted sum of quality scores for each quality attribute. The weighted sum of the total quality score for each region is the ((i) global quality score.
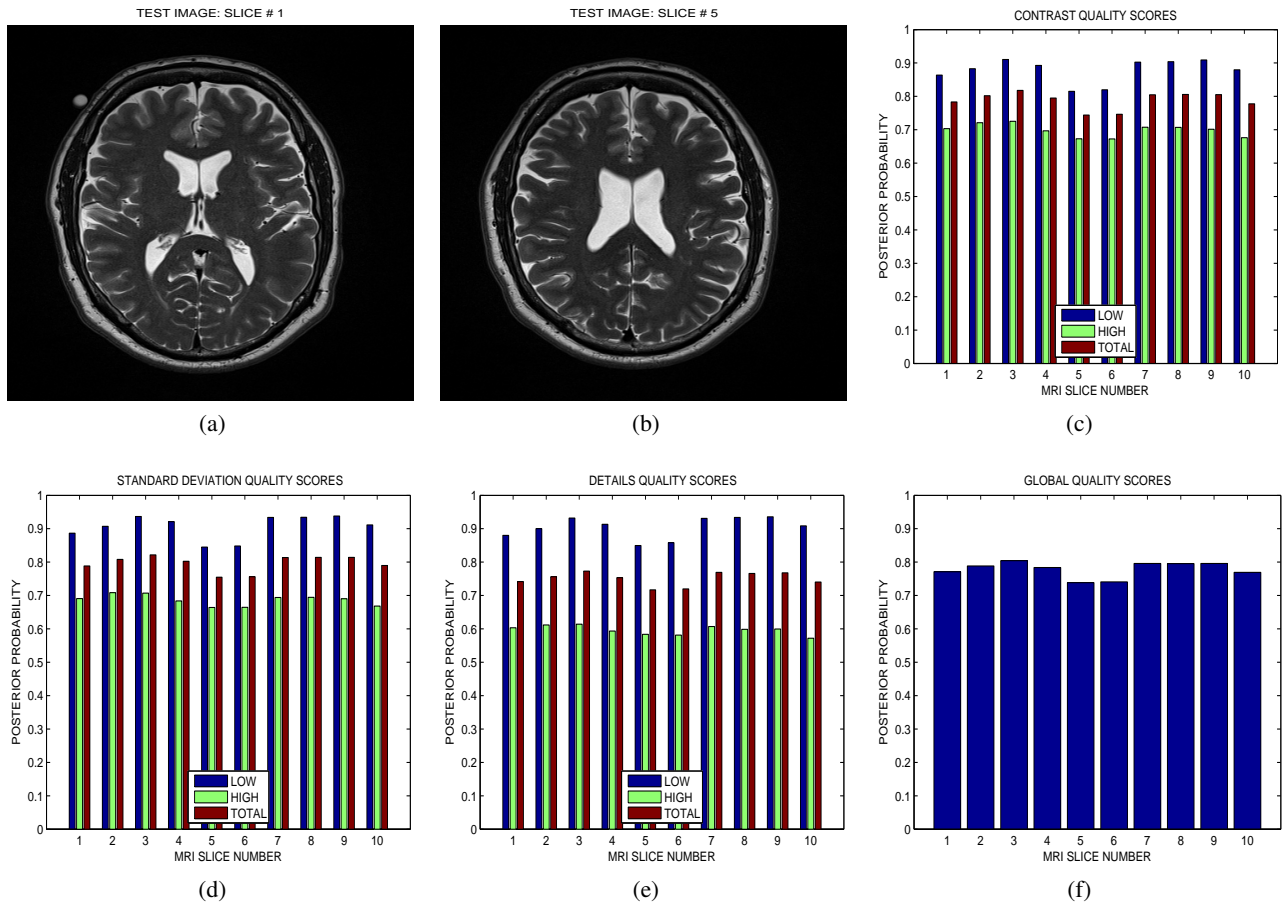
Fig. 4. Two slices (a) and (b) in a T2 MRI volume data from BrainCare Oy, (c) contrast attribute quality scores, (d) standard deviation attribute quality scores, (e) details attribute quality scores and (f) global quality scores for 10 successive slices in the MRI volume data
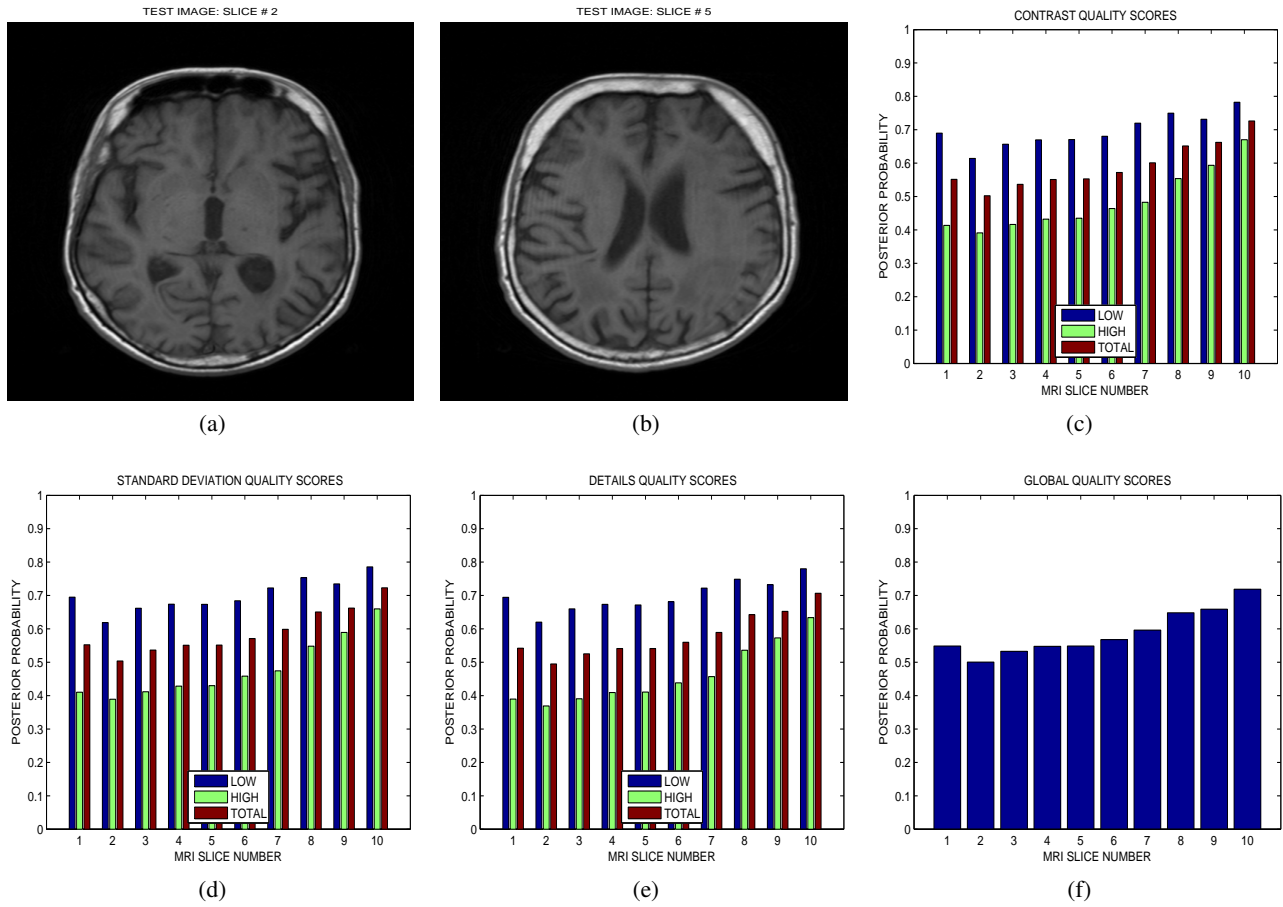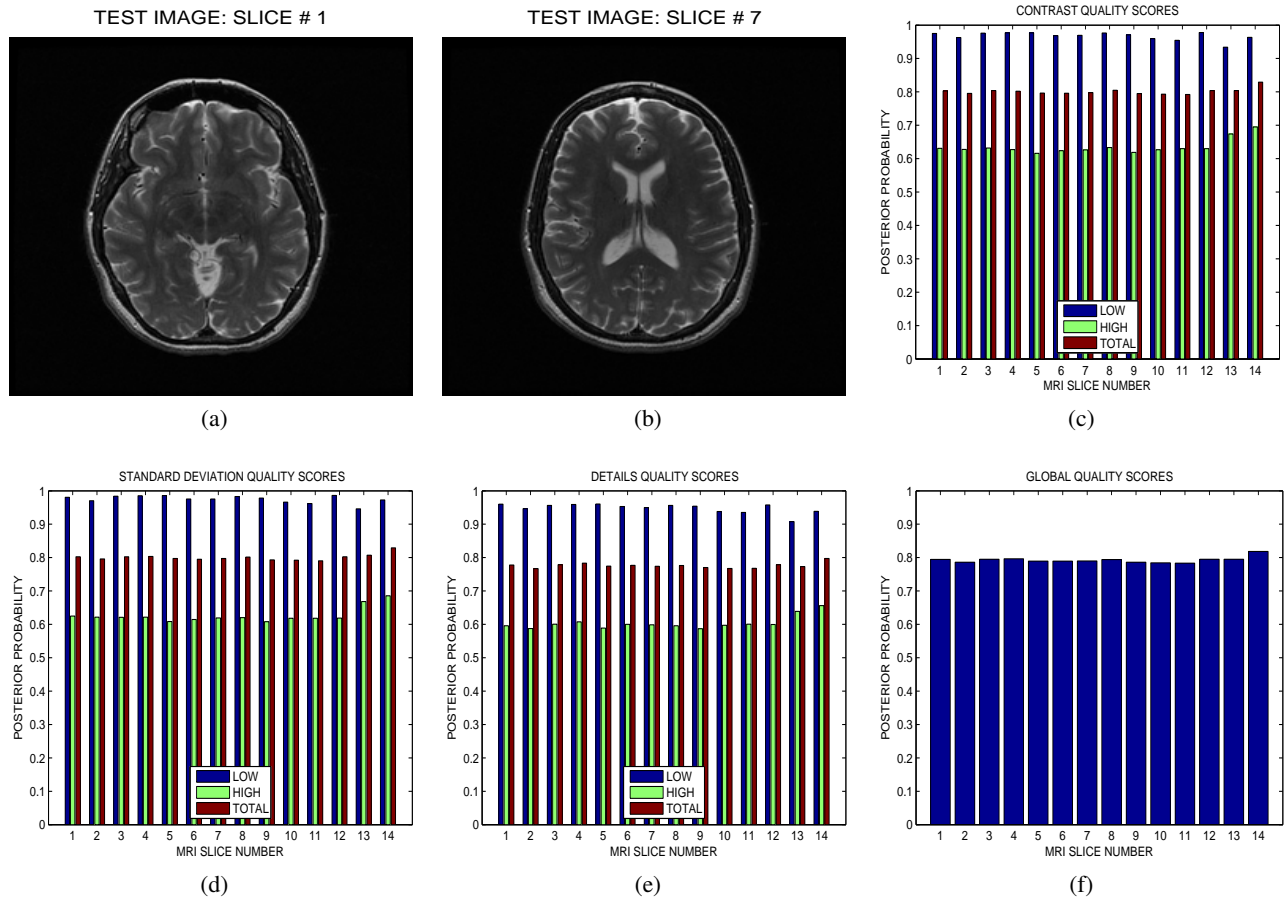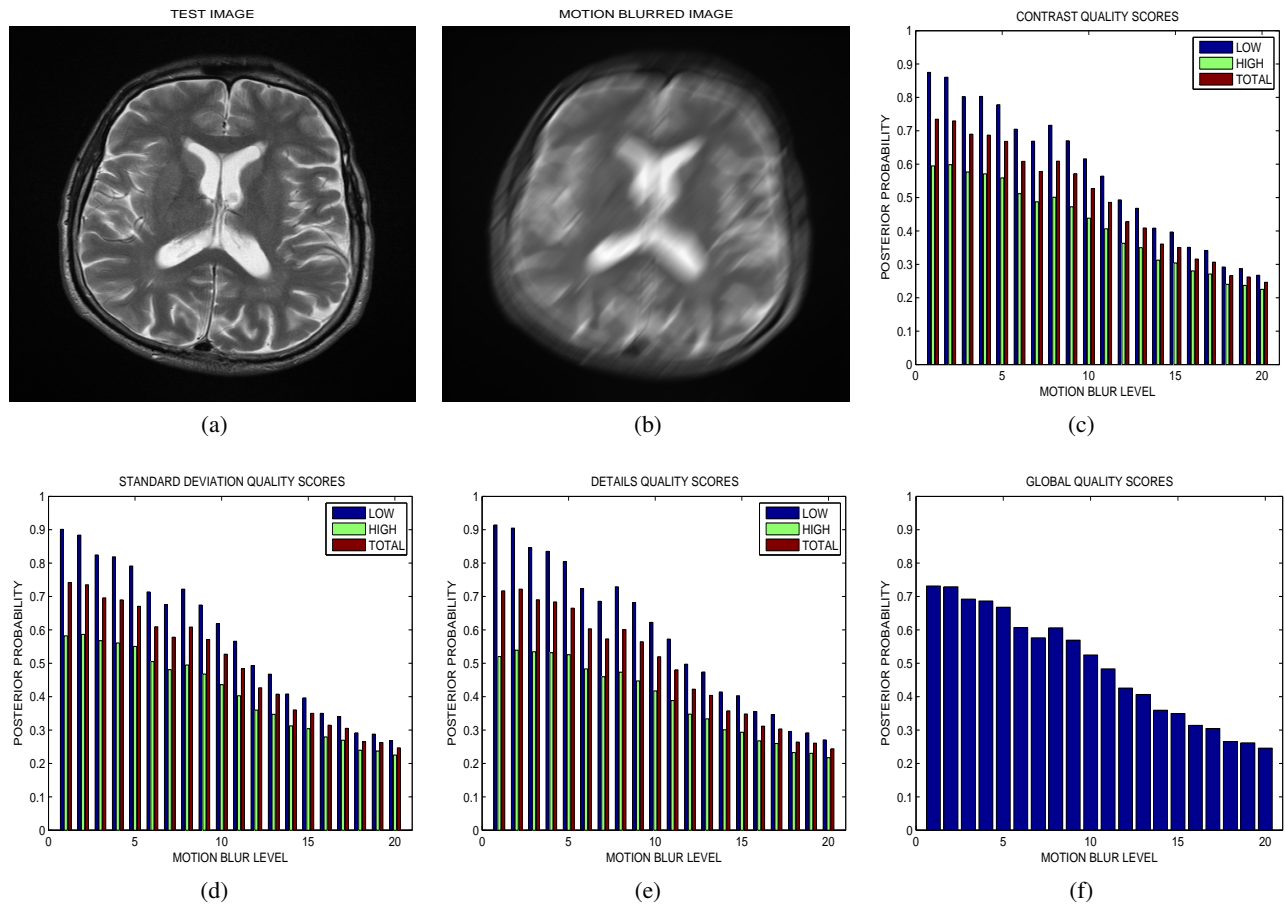
Fig. 5.   Two slices (a) and (b) in a T1 MRI volume data from BrainCare Oy, (c) contrast attribute quality scores (d) standard deviation attribute quality scores, (e) details attribute quality scores and (f) global quality scores for 10 successive slices in the MRI volume data

Fig. 6. Two slices (a) and (b) in a T2 MRI volume data from NeuroRx Research Inc, (c) contrast attribute quality scores, (d) standard deviation attribute quality scores, (e) details attribute quality scores and (f) global quality score for 10 successive slices in the MRI volume data
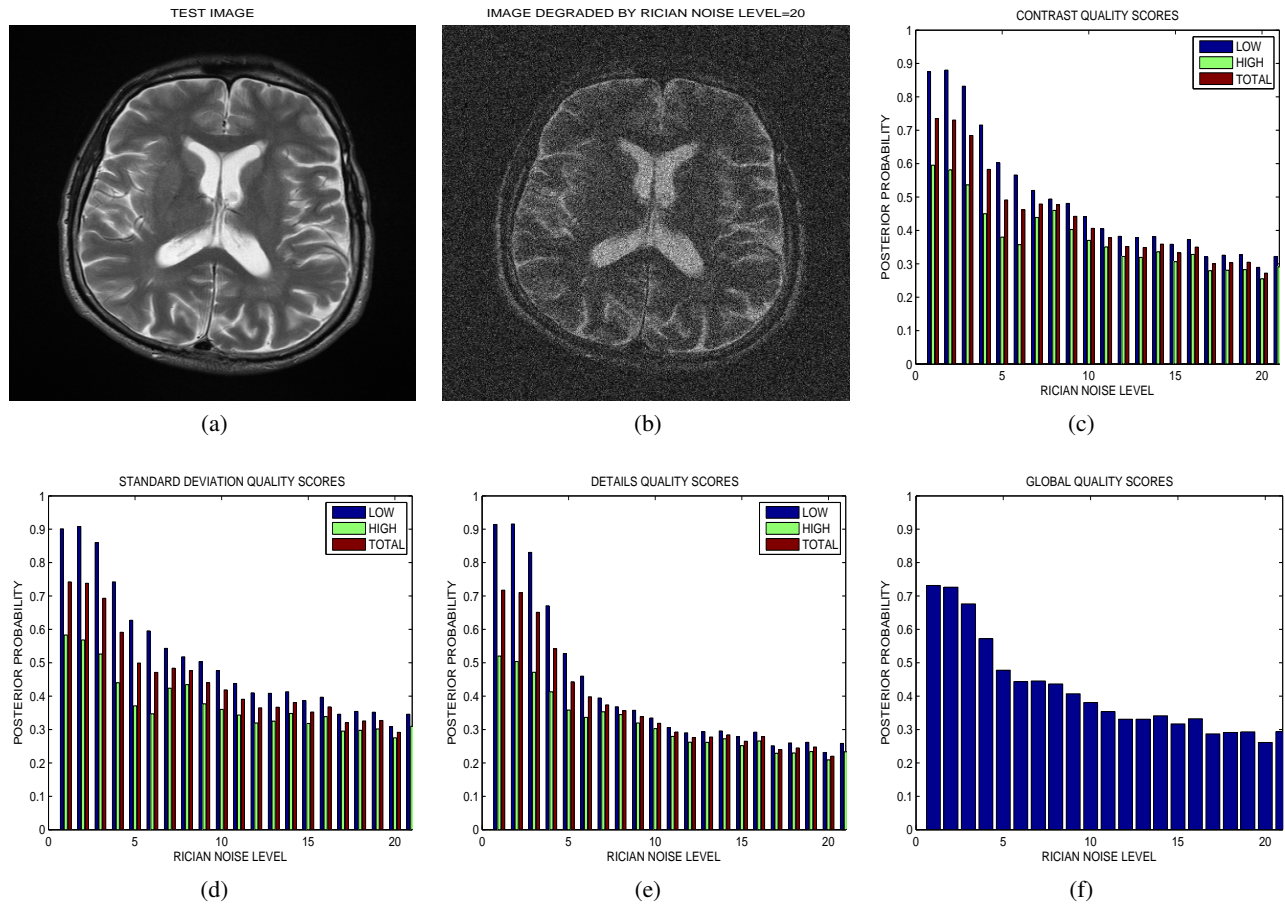
Fig. 7. (a) A slice in a T2 MRI volume data from BrainCare Oy is artificially degraded by motion blur. The degradation was scaled from 0 to 20. (b) Degraded version of the image corresponding to motion blur level of 20. (c) The contrast attribute quality scores, (d) standard deviation attribute quality scores, (e) details attribute quality scores and (f) global quality scores for the different levels of motion blur

Fig. 8. (a) A slice in a T2 MRI volume data from BrainCare Oy is artificially degraded by different levels of Rician noise. The Rician noise level was scaled from 0 to 20. (b) Degraded version of the image corresponding to Rician noise level of 20. (c) The contrast attribute quality scores, (d) standard deviation attribute quality scores (e) details attribute quality scores and (f) global quality scores for the different levels of Rician noise.
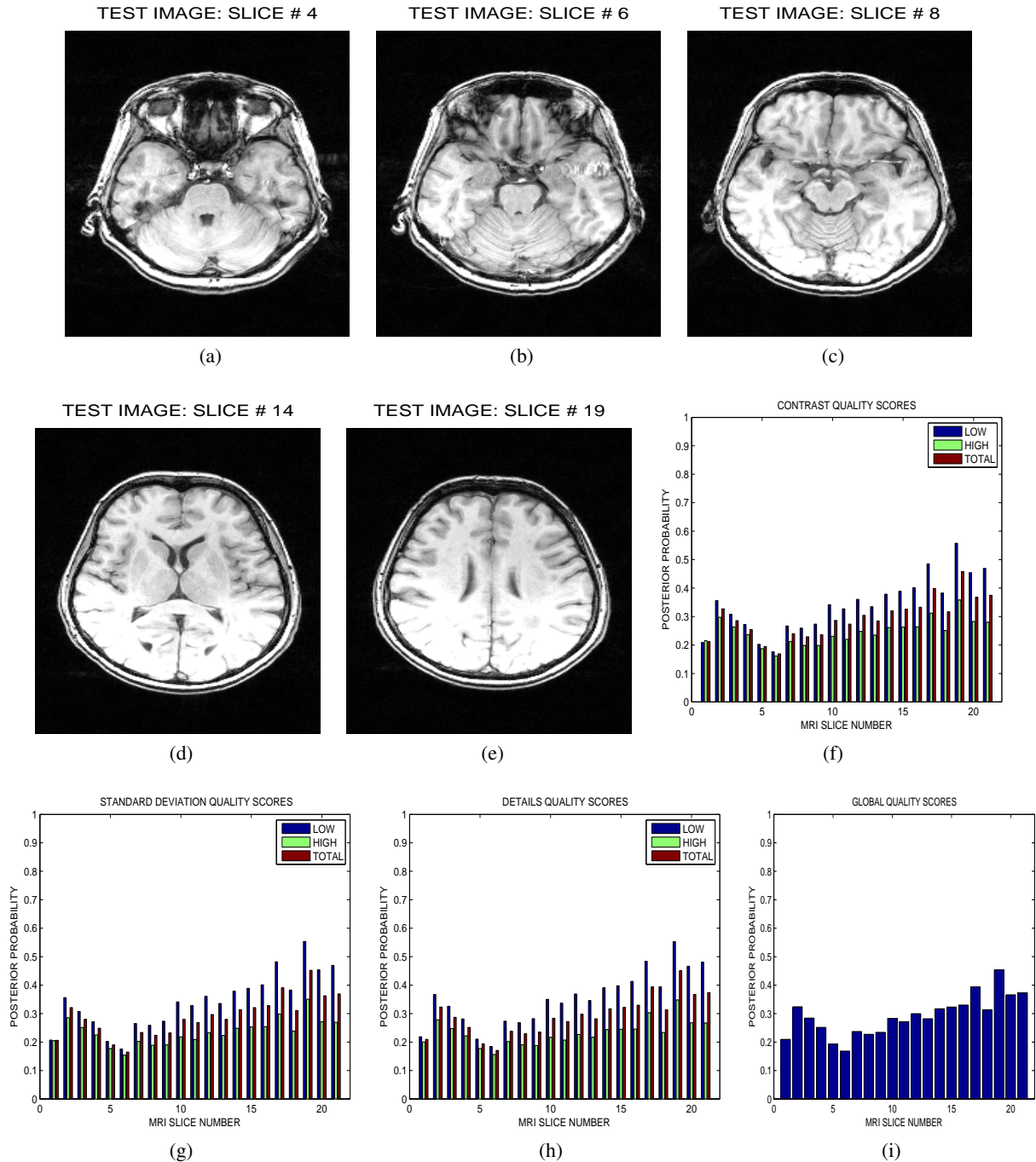
Fig. 9. Five slices (a), (b), (c), (d) and (e) from a T1 MRI volume data from NeuroRx Research Inc. The data was originally acquired with bias fields. (f) contrast attribute quality scores, (g) standard deviation attribute quality scores, (h) details attribute quality scores and (i) global quality scores for 21 successive slices in the MRI volume data.
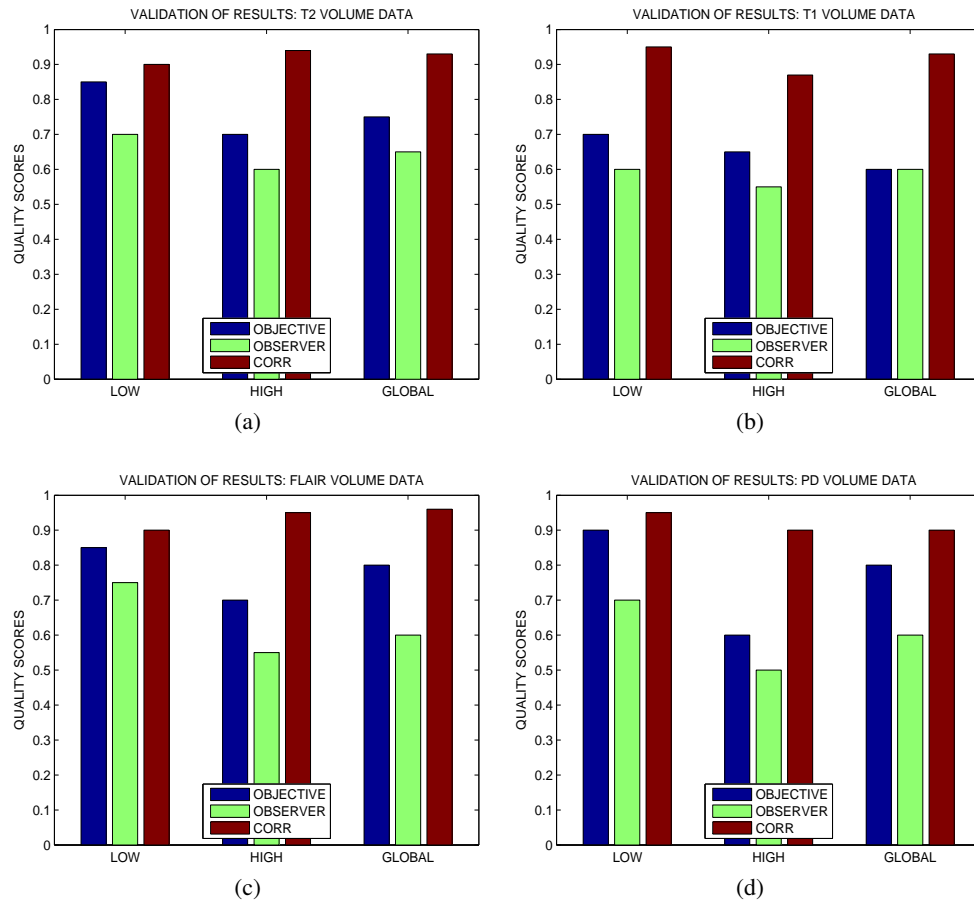
Fig. 10. Validation of the objective quality scores for MRI volume data that are without perceived distortion. (a) T2 MRI volume data, (b) T1 MRI volume data, (c) FLAIR MRI volume data and (d) PD MRI volume data.
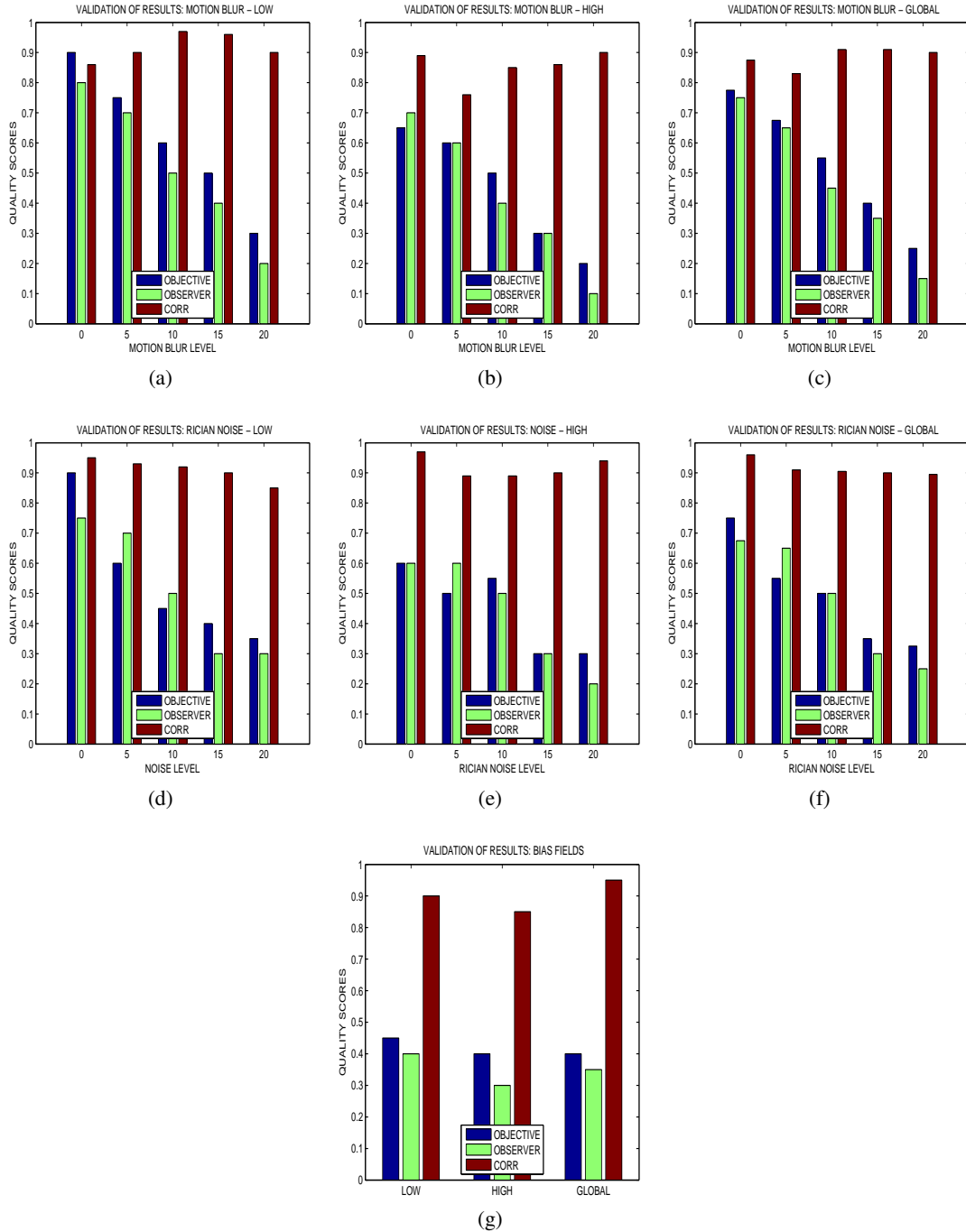
Fig. 11. The mean quality scores for (a) low energy, (b) high energy and (c) global energy regions of T2 MRI volume data degraded by different levels of motion blur. The mean quality scores for (d) low energy, (e) high energy and (f) global regions of T2 MRI volume data degraded by different levels of Rician noise. (g) The mean quality scores for the low energy, high energy and global regions of T1 MRI volume data originally acquired with bias fields
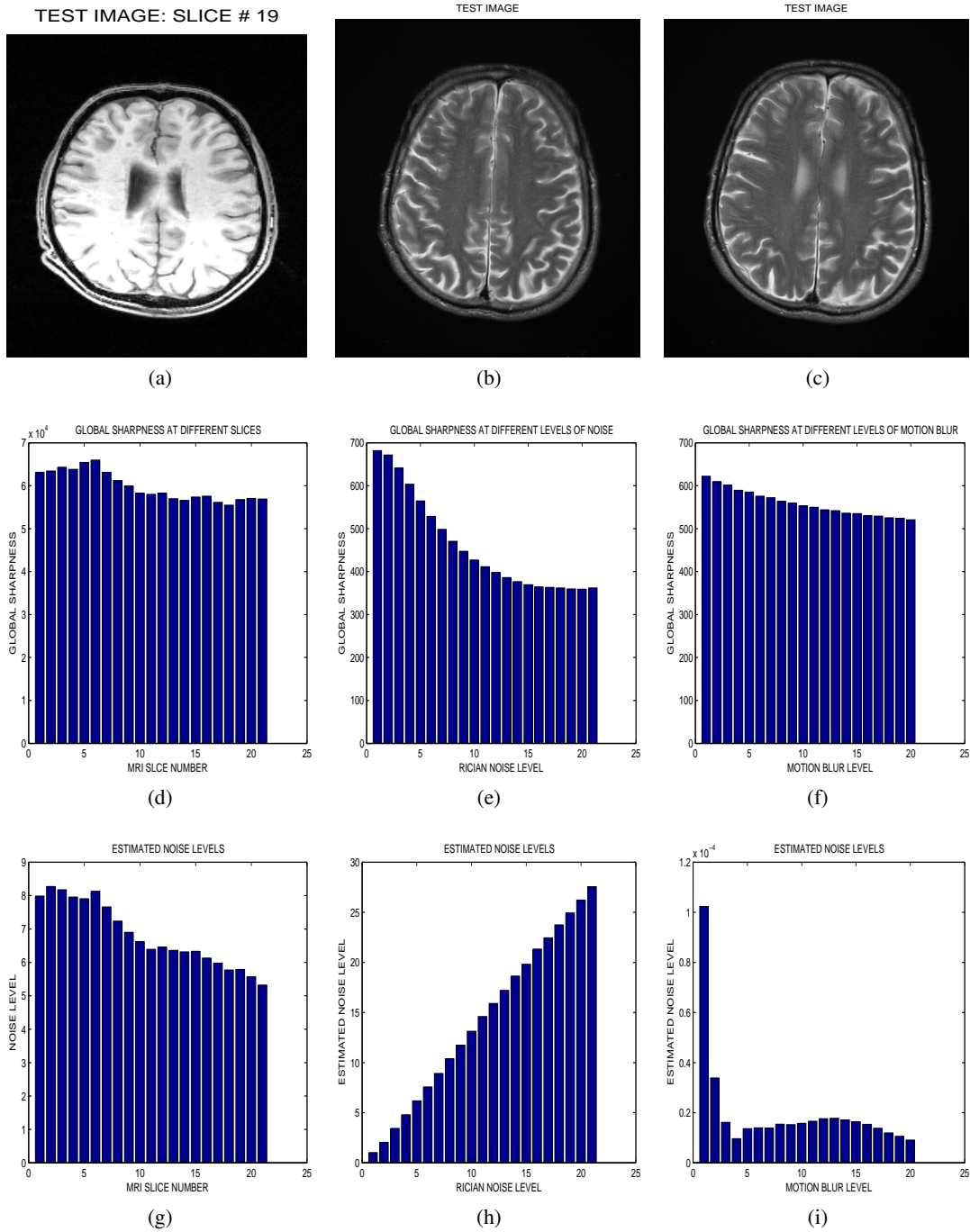
Fig. 12.   (a) A slice from a MRI volume data degraded with bias fields. Two different MRI slices degraded with 20 different levels of (b) Rician noise and (c) motion blur. The global sharpness and the noise level of the different slices in the MRI volume data in (a) are displayed in (d) and (g), respectively. The global sharpness and the noise level of the different levels of Rician noise for the slice in (b) are displayed in (e) and (h), respectively. The global sharpness and the noise level of the different levels of motion blur for the slice in (c) are displayed in (f) and (i), respectively.