Physica Medica 50 (2018) 26–36

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth

Fei Yang, Nesrin Dogan, Radka Stoyanova, John Chetley Ford*

Department of Radiation Oncology, University of Miami, Miami, FL 33136, United States

ARTICLE INFO

Keywords:
Radiomics
Texture
MRI
Repeatability

ABSTRACT

The purpose of this study was to examine the dependence of image texture features on MR acquisition parameters and reconstruction using a digital MR imaging phantom. MR signal was simulated in a parallel imaging radiofrequency coil setting as well as a single element volume coil setting, with varying levels of acquisition noise, three acceleration factors, and four image reconstruction algorithms. Twenty-six texture features were measured on the simulated images, ground truth images, and clinical brain images. Subtle algorithm-dependent errors were observed on reconstructed phantom images, even in the absence of added noise. Sources of image error include Gibbs ringing at image edge gradients (tissue interfaces) and well-known artifacts due to high acceleration; two of the iterative reconstruction algorithms studied were able to mitigate these image errors. The difference of the texture features from ground truth, and their variance over reconstruction algorithm and parallel imaging acceleration factor, were compared to the clinical “effect size”, i.e., the feature difference between high- and low-grade tumors on T1- and T2-weighted brain MR images of twenty glioma patients. The measured feature error (difference from ground truth) was small for some features, but substantial for others. The feature variance due to reconstruction algorithm and acceleration factor were generally smaller than the clinical effect size. Certain texture features may be preserved by MR imaging, but adequate precautions need to be taken regarding their validity and reliability. We present a general simulation framework for assessing the robustness and accuracy of radiomic textural features under various MR acquisition/reconstruction scenarios.

1. Introduction

Radiomics, which regards images as data rather than pictures [1], involves, in part, the exploitation of information that cannot necessarily be discerned on an image, or set of images, by even an expert observer. These data are designed to be mined from standard radiologic images; thus, a shared multi-institutional database can lead to a large subject pool from which image-based predictive models of patient outcome can be built. Texture features are an important subset of the quantitative image characteristics that can be extracted, in addition to first-order intensity histogram statistics, shape-based features, and higher order statistical methods such as fractal and wavelet analysis. As a radiomics tool, texture analysis (TA) is increasingly being applied to diverse imaging modalities to provide anatomical segmentation, cancerous lesion delineation, and prediction of response of normal and pathologic tissue to radiation therapy [2–6]. TA of magnetic resonance (MR) images has been applied to diverse clinical sites in the context of radiation treatment (RT), for example: brain [7,8], head and neck [9,10], breast [11], kidney [12], bladder [13], prostate [14–17], and extremities [18]. The results of those studies, as well as those of ongoing clinical research, indicate great potential for the incorporation of information gleaned from quantitative texture features into initial RT planning as well as adaptive RT of cancer patients.

Successful translation of quantitative imaging research into the clinic will depend in part on the ability to reliably repeat (e.g., multiple scans on the same subject) and reproduce (e.g., on various MR scanners) extracted texture features from MR images [19]. As noted by Mayerhoefer et al. [20], an impediment to widespread clinical application of MR-based TA is its sensitivity to the choice of MRI scanner and imaging protocol; several studies [14,20–24] have investigated the dependence of TA on MRI field strength, scanner manufacturer, and MRI acquisition parameters in both living subjects and phantoms. While evaluation of repeatability and reproducibility of texture features measured by MRI are very important, in this paper we endeavor to address the issue of accuracy as well. There have been, to our knowledge, no attempts at TA validation by direct comparison of measured TA features to known ground truth texture of the object of interest. In the absence of absolute knowledge of the texture features of an object,

---

* Corresponding author at: Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Department of Radiation Oncology, 1475 NW 12thAvenue, Suite C123, Miami, FL 33136, United States.

E-mail address: jcf137@miami.edu (J.C. Ford).

https://doi.org/10.1016/j.ejmp.2018.05.017
Received 21 December 2017; Received in revised form 4 April 2018; Accepted 16 May 2018

only relative comparisons can be made between MRI-derived features extracted using various imaging protocols or multiple MR scanners. Moreover, if two MR imaging methodologies produce different texture features, then in the absence of a ground truth, one cannot determine which method is more accurate.

Our overall goal is to overcome the deficiencies of the aforementioned relative TA comparisons by creation of a ground truth digital MRI phantom. By utilizing the digital phantom as input to an appropriate MR image simulator, we are able to create images inside a wide universe of MR imaging scenarios, and establish the dependence of the absolute accuracy of extracted TA features on MR field strength, MR pulse sequence, arrangement of receive coils, presence of image artifacts, and choice of image reconstruction algorithm. The purpose of this paper is: i) to lay the basic methodological framework for determination of absolute texture feature dependencies, using a two-dimensional digital ground truth phantom as input to an MR simulator; ii) to evaluate the sensitivity of texture feature accuracy and variance to noise level, acceleration factor and image reconstruction algorithms; and, iii) to compare the magnitude of texture feature error and variance due to acquisition details and/or reconstruction algorithm choice to the magnitude of clinically relevant texture features differences (high versus low grade glioma) in clinical brain MR images.

2. Methods

2.1. Study design

Texture features were examined in i) simulated MRI datasets, based on a digital phantom of the brain, and ii) clinical brain MR images of glioma patients. For the first part of the study, an idealized 2D digital phantom of an axial slice of human brain, developed by Guerquin-Kern et al. [25] served as a ground truth. We employed an MRI simulator [26], developed by the same group that created the idealized brain phantom, that produces complex k-space data from the digital phantom. This simulator is open-source software consisting of code written in Matlab (Mathworks, Natick, MA) for performing MRI simulation and reconstruction. We utilized the script 'DemoSimuAndRecon.m', which defines a parallel MRI experiment setting and Cartesian k-space sampling. Several sets of noise were added and the MR image was reconstructed using the following algorithms, appropriate for parallel imaging with phased-array coils, which are included in the open source MRI simulator software: conjugate gradient (CG), total variation (TV), and wavelet regularization (WL). A fourth algorithm, conventional inverse Fast Fourier Transform (iFFT), appropriate for non-accelerated (non-parallel) imaging, was also implemented by straightforward modification of the Matlab code. The clinical image data consisted of T1- and T2-weighted MRI scans from glioma patients, included in the BRATS 2015 challenges (Multimodal Brain Tumor Segmentation) [27]. Regions of interest (ROIs) were selected both on simulated and clinical data and texture features were extracted. Absolute feature error and variance among the reconstruction algorithms, using images with similar signal-to-noise ratio to the clinical brain images, were then compared to the magnitude and clinically relevant tumor feature differences (high versus low grade) in the clinical brain images.

2.2. Simulated data

2.2.1. MR signal simulation

MRI simulation was considered in two dimensions. The in-plane excited spins were modeled as radio-frequency emitters featuring position-dependent precession frequency and phase. The signal $m_i(\boldsymbol{k}_l)$ received by a coil of sensitivity $S_i(\boldsymbol{r})$ for a given frequency location $\boldsymbol{k}_l$ in $\boldsymbol{k}$ space is the Fourier transform of the coil sensitivity-modulated object function $\rho(\boldsymbol{r})$:

$$m_i(\boldsymbol{k}_l) = \int_{R^2} S_i(\boldsymbol{r})\rho(\boldsymbol{r})\boldsymbol{e}^{-2j\pi\boldsymbol{k}_l\cdot\boldsymbol{r}}dr. \tag{1}$$

For an array of receiving coils the measurement $\boldsymbol{m}$ is formed by concatenating signals received from each individual coil. The signal m_i can be calculated analytically in the MRI simulator because the object $\boldsymbol{\rho}(\boldsymbol{r})$ is comprised of regions (i.e., connected and bounded sets) of constant intensity; the k-space signal m_i is obtained by summing the signal over all regions [25,26]. Alternatively, the simulator can utilize the traditional procedure of sampling the analytical object with a grid of a given size, then resampling the inverse Fourier transform of this discrete image according to the desired k-space matrix size, a so-called "rasterized" simulation. Sampling the object with infinite resolution will theoretically give equivalent results to direct calculation using the regions defined by the analytical object. We chose the rasterized simulation for its great reduction in computation time, at the sacrifice of some (negligibly small) accuracy due to finite object sampling grid sizes, as detailed previously [25]. To investigate the influence of sampling grid size, we recorded the signal-to-error ratio of reconstructed images as a function of grid size. Based on this investigation, described in the Results section, a grid refinement factor of 5 was chosen for all subsequent k-space generation.

The noise-free measurement $\boldsymbol{m}$ can be expressed as

$$\boldsymbol{m} = \boldsymbol{E\rho}, \tag{2}$$

where $\boldsymbol{E}$ denotes the MRI system matrix and $\boldsymbol{\rho}$ is the object. When taking into account the presence of noise along with other uncertainties, a more realistic object estimation model to consider is as follows:

$$\boldsymbol{m} = \boldsymbol{E\rho} + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{n}$ represents the noise perturbation. This formulates the MRI reconstructions as an inverse problem that attempts to recover the object from the corrupted measurement $\boldsymbol{m}$.

Shown in Fig. 1a is a 256 × 256 rasterization of the object $\boldsymbol{\rho}(\boldsymbol{r})$, a brain slice, that was adopted as the ground truth image. The perturbation signal was modeled as a centered multivariate complex Gaussian process, resulting in realistic MR image noise [28]. The MR signal was modeled under the following scenarios: a parallel imaging (phased array) coil consisting of eight circular RF coils of radius 7 cm placed equiangular and equidistant at 17 cm from the isocenter, and a single conventional birdcage coil of 17 cm radius. All utilized a FOV = 28 cm, and 256 × 256 acquisition matrix, and the parallel coil used acceleration factor of one, two or four.

2.2.2. MRI reconstruction

Three iterative image reconstruction methods, based respectively on conjugate gradient (CG) [29], total variation (TV) [30], and wavelet (WL) regularization [31], were considered. These three were chosen not only because of their ready accessibility as part of the MRI simulator package, but also because they have been thoroughly evaluated including quantification of reconstruction error in several phantoms as well as the brain phantom used in this work [25]. We utilized the default reconstruction algorithm parameters that were used in those evaluations. In addition, we implemented a conventional inverse fast Fourier transform (iFFT) method as a fourth, non-iterative reconstruction algorithm for comparison.

The conjugate gradient method pursues the reconstruction solution $\tilde{\boldsymbol{x}}$ by solving the following minimization problem,

$$\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}(\|\boldsymbol{Ex}-\boldsymbol{m}\|_{l_2}^2 + \lambda\|\boldsymbol{x}\|_{l_2}^2) \tag{4}$$

where $\|\cdot\|_{l_2}$ denotes the l_2 norm. Its cost function is a trade-off between the fidelity term, which enforces consistency of the solutions with the measurements, and a regularization term, which penalizes non-regular solutions. The parameter λ is a positive constant that balances the influence of the two terms. The total variation method favors particularly for piecewise-constant reconstructions and formulates the solutions as