

Interplay Between Intensity Standardization and Inhomogeneity Correction in MR Image Processing

Anant Madabhushi* and Jayaram K. Udupa

Abstract—Image intensity standardization is a postprocessing method designed for correcting acquisition-to-acquisition signal intensity variations (nonstandardness) inherent in magnetic resonance (MR) images. Inhomogeneity correction is a process used to suppress the low frequency background nonuniformities (inhomogeneities) of the image domain that exist in MR images. Both these procedures have important implications in MR image analysis. The effects of these postprocessing operations on improvement of image quality in isolation has been well documented. However, the combined effects of these two processes on MR images and how the processes influence each other have not been studied thus far. In this paper, we evaluate the effect of inhomogeneity correction followed by standardization and *vice-versa* on MR images in order to determine the best sequence to follow for enhancing image quality. We conducted experiments on several clinical and phantom data sets (nearly 4000 three-dimensional MR images were analyzed) corresponding to four different MRI protocols. Different levels of artificial nonstandardness, and different models and levels of artificial background inhomogeneity were used in these experiments. Our results indicate that improved standardization can be achieved by preceding it with inhomogeneity correction. There is no statistically significant difference in image quality obtained between the results of standardization followed by correction and that of correction followed by standardization from the perspective of inhomogeneity correction. The correction operation is found to bias the effect of standardization. We demonstrate this bias both qualitatively and quantitatively by using two different methods of inhomogeneity correction. We also show that this bias in standardization is independent of the specific inhomogeneity correction method used. The effect of this bias due to correction was also seen in magnetization transfer ratio (MTR) images, which are naturally endowed with the standardness property. Standardization, on the other hand, does not seem to influence the correction operation. It is also found that longer sequences of repeated correction and standardization operations do not considerably improve image quality. These results were found to hold for the clinical and the phantom data sets, for different MRI protocols, for different levels of artificial nonstandardness, for different models and levels of artificial inhomogeneity, for different correction methods, and for images that were endowed with inherent standardness as well as for those that were standardized by using the intensity standardization method. Overall, we conclude that inhomogeneity correction followed by intensity standardization is the best sequence to follow from the perspective of both image quality and computational efficiency.

Index Terms—Image analysis, image processing, inhomogeneity correction, MRI, standardization.

I. INTRODUCTION

MAGNETIC resonance (MR) imaging (MRI) is a noninvasive method for imaging the human body (and other beings) and has revolutionized medical imaging. Although not commonly needed in daily clinical practice that utilizes MR images, MR image processing, particularly segmentation, and analysis [1]–[7], are used extensively in medical and clinical research for advancing our understanding of the various diseases of the human body, for their diagnosis, and for developing strategies to treat them. These efforts face two major difficulties – the first due to image intensity inhomogeneity present as a background variation component, and the second due to the nonstandardness of the MR image intensity gray scale.

MR signal intensity measured from homogeneous tissue regions is seldom uniform; rather it varies smoothly across the image. This intensity nonuniformity may be caused by a number of different factors, including poor radio frequency (RF) coil uniformity, static field inhomogeneity, RF penetration, gradient-driven eddy currents, and patient anatomy. Although these intensity variations may usually have little impact on visual diagnosis, they can significantly affect the performance of many image processing, segmentation, and analysis techniques. A robust and inexpensive means of correcting this artifact is essential for such processing techniques to have acceptable precision, accuracy, and efficiency. Further, correcting intensity nonuniformity may benefit quantitative measurements such as those used in tissue metabolite studies. Most image processing and analysis methods are affected by these variations, especially if quantitative analysis based on image intensities is the final goal [1]–[7].

Another major difficulty in MR image analysis has been that intensities do not have a fixed tissue-specific numeric meaning, not even within the same MRI protocol, for the same body region, and even for images of the same patient obtained on the same scanner. This implies that MR images cannot be displayed at preset windows for the same protocol, body region, and application; one may have to adjust the window settings per case. (This also leads to wasted films due to improper exposure in many hospitals where digital handling of images has not entered routine radiological practice.) More importantly, for most postprocessing applications such as image segmentation and quantification, this lack of a standard and quantifiable interpretation of image intensities is a major drawback which compromises their precision, accuracy, and efficiency.

Many methods have been published [8]–[15] to address the problem of inhomogeneity since the publication of the first

Manuscript received June 8, 2004; revised December 15, 2004. This work was supported in part by the Department of Health and Human Services (DHHS) under Grant NS37072. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Sonka. *Asterisk indicates corresponding author.*

*A. Madabhushi is with the Department of Biomedical Engineering, Rutgers University, 617 Bowser Road, Rm. 102, BME Bldg., Piscataway, NJ 08854 USA (e-mail: anantm@rci.rutgers.edu).

J. K. Udupa is with the Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: jay@mipg.upenn.edu).

Digital Object Identifier 10.1109/TMI.2004.843256

paper on the subject [8]. Lai and Fang [9] describe a shape from orientation formulation in a regularization framework for correcting intensity inhomogeneities. Likar *et al.* [14] describe a model-based correction method that makes the assumption that an image corrupted by intensity inhomogeneity contains more information than the corresponding un-corrupted image. The image degradation process is described by a linear model, consisting of a multiplicative and an additive component which are modeled by a combination of smoothly varying basis functions. Similar to this work [14], a model-based method for fully automated correction is proposed in [11]. The MR signal is modeled as a realization of a random process with a parametric probability distribution that is corrupted by a smooth polynomial bias field. A nonparametric normalization technique is described in [10] to correct intensity nonuniformity. An iterative approach is employed to estimate both the multiplicative bias field and the distribution of the true tissue intensities. Styner *et al.* [12] describe a parametric model for inhomogeneity correction. The method comprises of a parametric model of tissue class statistics, and a polynomial model of the inhomogeneity field. In the present paper, we use two techniques [15], [26] based on the idea of local scale [19] to correct background intensity variations in the acquired images. Local scale is a fundamental image property [19] that is derivable entirely from the image and that does not require any prior knowledge about the imaging protocol or object material property distributions. In [19] local scale was defined as the radius of the largest ball centered at every voxel, such that all voxels within the ball satisfy a predefined homogeneity criterion. This definition of local scale (referred to as ball-scale or b-scale from here on) was applied to inhomogeneity correction in [15]. In [26], an alternative definition of local scale, referred to as generalized scale (abbreviated as g-scale from here on) was proposed. The g-scale of any voxel v in the scene was defined as the largest connected set of voxels associated with v such that all voxels within this set satisfy a predefined homogeneity criterion. The principal difference between the two scale-based correction algorithms described in [15] and [26] is the way in which the largest scale region is determined. In both correction methods the largest scale region (as determined by the two scale definitions) is used to estimate the scene inhomogeneity. A second degree polynomial is then fit to the intensity variations within the largest scale region and correction is affected. This procedure is repeated iteratively for the corrected image until the size of the segmented homogeneous regions does not change significantly from that in the previous iteration.

The only papers that we are aware of that address the second problem of standardization of image intensities explicitly are [16]–[18]. Most image analysis methods, particularly segmentation algorithms, have free parameters. Setting values for these parameters becomes very difficult without the same MRI protocol-specific intensity meaning in all images acquired as per a given protocol and for a given body region. Other papers that have attempted to deal with this problem have done so from the standpoint of image segmentation and inhomogeneity correction [12], [13]. Guimond *et al.* [13] describe a methodology to perform three-dimensional (3-D) multimodal brain warping by using adaptive intensity correction. The core of the algorithm is about finding a transformation that maps the intensities of one

image to those of another. However, this method is geared toward standardizing the intensities of only the two images that are to be registered. In [12], the authors describe the use of a parametric model of the tissue class statistics and a polynomial model of the inhomogeneity field to achieve intensity normalization and histogram adjustment. In the present paper, we use a technique devised by Nyul and Udupa [16], which transforms images nonlinearly so that there is a significant gain in the similarity of the resulting images. This is a two step method wherein all images (independent of patients and the specific brand of MR scanner used) are transformed in such a way that, for the same protocol and body region, similar intensities will have similar tissue-specific meaning. In the first step, the parameters of the standardizing transformation are learnt from a set of images. In the second step, for each MRI study, these parameters are used to map their intensity gray-scale into a new gray-scale. It has been shown [16]–[18] that standardization significantly minimizes the variation of the overall mean intensity of the MR images within the same tissue region across different studies obtained on the same or different scanners.

Although it has been established that both inhomogeneity correction and intensity standardization are important from the viewpoint of automated quantitative MR image analysis, thus far, their effects have been studied only in isolation. Several fundamental questions arise in correcting for these effects: In what order these operations should be performed for best results? Should they be repeated alternately? Does one of them have an effect on the other? How should the effects be quantified for these operations? The aim of this research is to seek answers to these questions and to find the sequence of operations that will produce the best overall *image quality*. We evaluate *image quality* in terms of the degree of inhomogeneity and the level of nonstandardness present in the image. Images which have lower levels of nonstandardness and inhomogeneity are considered to have higher image quality.

Our methodology, described in Section II, starts with three sets of MR images consisting of clinical brain studies acquired under routine clinical conditions under different protocols, as well as three sets of realistic phantom data sets of the brain corresponding to different MRI protocols. From these six sets, a total of 63 sets of over 700 3-D image data are created by initially applying inhomogeneity correction and standardization in different orders and subsequently artificially introducing different levels of inhomogeneity and nonstandardness in different orders. The images in the sets so created are subsequently corrected for inhomogeneity and standardized following different sequences of these operations. The effects of these sequences of operations on correcting for inhomogeneity and for nonstandardness are analyzed by evaluating certain measures that are based on image intensities in the corrected images in presegmented tissue regions. In all, we analyzed nearly 4000 3-D MR scenes. The results of these analyses are described in Section III. We present our concluding remarks in Section IV.

II. METHODOLOGY

A. Terminology

We denote the set of all protocols used in MR imaging by \mathcal{P} and the set of all body regions by \mathcal{D} . We represent a volume

TABLE I
DESCRIPTION OF THE SCENE DATA SETS USED IN THIS STUDY

Data set	No. of scenes	Protocol	Acquisition Parameters	Scene domain	Voxel size (mm ³)
S^1	12	PD	$TR/TE_{eff}=2500/18$, FOV=22 cm ²	$256 \times 256 \times \mathcal{N}$ ($45 \leq \mathcal{N} \leq 56$)	$0.86 \times 0.86 \times 3$
S^2	12	T2	$TR/TE_{eff}=2500/90$, FOV=22 cm ²	$256 \times 256 \times \mathcal{N}$ ($45 \leq \mathcal{N} \leq 56$)	$0.86 \times 0.86 \times 3$
S^3	10	MT	$TR/TE_{eff}=106/5$, FOV=22 cm ²	$256 \times 256 \times \mathcal{N}$ ($45 \leq \mathcal{N} \leq 56$)	$0.86 \times 0.86 \times 5$
S^4	5	T1	Mathematical Phantom	$181 \times 217 \times \mathcal{N}$ ($20 \leq \mathcal{N} \leq 181$)	$1 \times 1 \times \mathcal{T}$ ($1 \leq \mathcal{T} \leq 9$)
S^5	5	T2	Mathematical Phantom	$181 \times 217 \times \mathcal{N}$ ($20 \leq \mathcal{N} \leq 181$)	$1 \times 1 \times \mathcal{T}$ ($1 \leq \mathcal{T} \leq 9$)
S^6	5	PD	Mathematical Phantom	$181 \times 217 \times \mathcal{N}$ ($20 \leq \mathcal{N} \leq 181$)	$1 \times 1 \times \mathcal{T}$ ($1 \leq \mathcal{T} \leq 9$)

image, called *scene* for short, by a pair $\mathcal{C} = (C, f)$, where C is a finite 3-D rectangular array of voxels, called *scene domain* of \mathcal{C} , covering a body region of the particular patient for whom scene \mathcal{C} is acquired, and f is a function that assigns an integer intensity value $f(u)$ for each $u \in C$. We denote by \mathcal{S}^{PD} the set of all scenes that can possibly be generated as per a given protocol $P \in \mathcal{P}$ and for a given body region $D \in \mathcal{D}$. For short, we refer to the process of correction associated with magnetic field inhomogeneities as simply *correction* and to the process of standardizing the MR intensity scale to account for scanner-dependent gray scale variations as *standardization*. We will use super-scripts c , s , \bar{c} , and \bar{s} to denote, respectively, the scenes resulting from an *initial process* of correction, standardization, and the initial introduction of artificial inhomogeneity and non-standardness. Sub-scripts c and s will be used to denote the *subsequently corrected and standardized* scenes, respectively. As an example, $X_{sc}^{cs\bar{c}\bar{s}}$ denotes the set of scenes resulting from an initial set X of scenes after the application of the following sequence of operations on X : an initial correction followed by standardization (X^{cs}), introduction of a known level of inhomogeneity followed by a known level of nonstandardness ($X^{cs\bar{c}\bar{s}}$), and finally applying a standardization method followed by correction ($X_{sc}^{cs\bar{c}\bar{s}}$). We use c and c_g to differentiate between b-scale and g-scale-based correction methods. The operations for correction are denoted by κ (for b-scale) and by κ_g (for g-scale), and those for standardization are denoted by ψ . We shall denote by $\bar{\kappa}$ and $\bar{\psi}$ the operations of introducing inhomogeneity and nonstandardness. With this notation, in the previous example, if $\mathcal{C} \in X_{sc}^{cs\bar{c}\bar{s}}$, then the corresponding scene in $X_{sc}^{cs\bar{c}\bar{s}}$ resulting from applying ψ followed by κ will be denoted by $\mathcal{C}_{sc} = \psi\kappa(\mathcal{C})$. Similarly, if $\mathcal{C} \in X^{cs}$, then $\mathcal{C}^{\bar{c}\bar{s}} = \bar{\kappa}\bar{\psi}(\mathcal{C})$ will be an element of $X^{cs\bar{c}\bar{s}}$.

B. Outline of Approach

Our approach comprises of the following five steps.

Step 1) Data Acquisition: Six sets of scenes, $S^i \subset \mathcal{S}^{P_i D}$, for $1 \leq i \leq 3$, from clinical studies, and $S^i \subset \mathcal{S}^{P_i D}$, for $4 \leq i \leq 6$, consisting of realistic simulated phantom data sets, corresponding to the same body region D (brain) but different

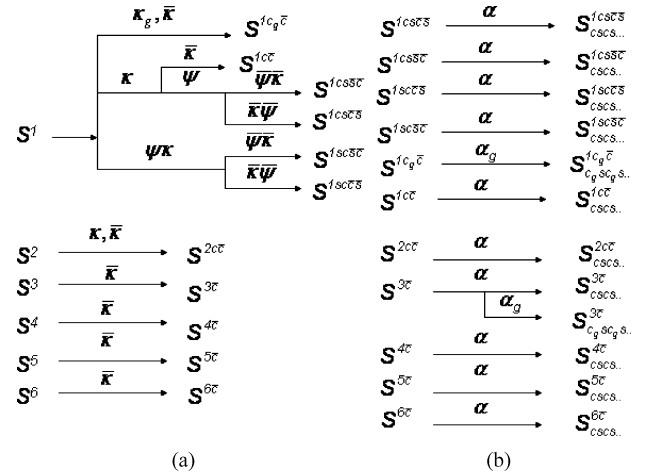


Fig. 1. (a) Preparing data sets with known levels of inhomogeneity and nonstandardness. (b) The various processing operations applied to the data sets generated in (a). The operations expressed by the operator sequences denoted by β and β_g are also applied in a manner similar to that depicted in (b).

MRI protocols P_i , for $1 \leq i \leq 6$, are obtained. Table I summarizes the description of the data sets used in this study.

Step 2) Adding Inhomogeneity and Nonstandardness: From the six sets S^i , $1 \leq i \leq 6$, eleven sets of scenes are generated by mixing in various orders, the processes of initial inhomogeneity correction and standardization and the processes of artificially introducing different levels of inhomogeneity and nonstandardness, as depicted in Fig. 1(a).

Step 3) Correction Followed by Standardization: The operation expressed by the two alternating sequences of correction followed by standardization, denoted by $\alpha = \kappa\psi\kappa\psi \dots \kappa\psi$ and $\alpha_g = \kappa_g\psi\kappa_g\psi \dots \kappa_g\psi$, are applied to the eleven sets of scenes of Fig. 1(a) as depicted in Fig. 1(b) to produce the 12 sets listed in Fig. 1(b) for given sequences α and α_g .

Step 4) Standardization Followed by Correction: The operations expressed by another two sequences of standardization followed by correction, denoted by $\beta = \psi\kappa\psi\kappa \dots \psi\kappa$ and $\beta_g = \psi\kappa_g\psi\kappa_g \dots \psi\kappa_g$, are applied to the eleven sets of scenes listed in Fig. 1(a) in a manner similar to the illustration

TABLE II
DESCRIPTION OF THE NUMBER OF LEVELS OF $\bar{\kappa}$ AND $\bar{\psi}$ THAT HAS BEEN USED FOR THE DIFFERENT DATA SETS

Data Sets	Models for $\bar{\kappa}$	No. of levels for $\bar{\kappa}$	No of levels for $\bar{\psi}$	Total no. of scenes
S^{1c}, S^{1c_g}	Gaussian	5	-	120
S^{1cs}, S^{1sc}	Gaussian	3	3	432
S^{2c}	Inverted Gaussian	5	-	60
S^3	Gaussian,	3	-	30
	Inverted Gaussian			30
S^4, S^5, S^6	via simulation	2	-	30

in Fig. 1(b) for sequences α and α_g . These operations result in another group of 12 sets for given sequences of β and β_g .

Step 5) Varying Levels of Inhomogeneity and Nonstandardness: Different models and levels of inhomogeneity, and different levels of nonstandardness are added in Step 2, and for each group of resulting sets, Steps 3 and 4 are repeated. The effectiveness of correction and standardization are assessed and compared for the sequences α , β , α_g , and β_g , and the dependence of the results on the levels of inhomogeneity and nonstandardness is evaluated.

The above steps are described in detail in the next section.

C. Detailed Description of the Approach

Step 1) Acquisition of Scene Data Sets: We have utilized two kinds of data sets – clinical patient data sets S^1 , S^2 , and S^3 obtained from routine MRI scans, and phantom MRI data sets S^4 , S^5 and S^6 , obtained from the Montreal Neurological Institute (MNI) simulated brain image database [20]–[22]. Table I gives a summary of the details pertaining to these data sets. S^1 , S^2 , and S^3 are taken randomly from a database of approximately 1500 such scenes obtained by the routine scanning of the head of multiple sclerosis (MS) patients in our hospital. The MTR scenes are created from the two Magnetization Transfer image acquisitions obtained corresponding to the magnetization saturation pulse on and off. The MTR scene in each acquisition is computed by dividing the MT “on” scene intensity value by the corresponding MT “off” scene intensity value in a voxel-by-voxel manner [24], [25]. The five scenes within each protocol in S^4 , S^5 , and S^6 correspond to the same normal brain atlas but having varying slice thicknesses of 1 mm, 3 mm, 5 mm, 7 mm, and 9 mm.

Our rationale for considering the six sets of scene data is as follows. S^1 and S^2 represent real clinical images. Since they are from two different MRI protocols, we can ascertain that the phenomenon we are studying is not specific to a particular protocol. Since background intensity inhomogeneity is generally considered to be multiplicative, the process of division in determining MTR scenes from the MT “on” and “off” scenes eliminates its effect in the MTR scenes. Further, MTR values in specific tissue regions are known to be highly reproducible among MRI scanners [24], [25]. Thus, among clinically acquired scenes, perhaps MTR scenes are as close as we can get in terms of real scenes with the least degree of inhomogeneity and nonstandardness. Therefore, S^3 would allow us to study how the correction process may introduce nonstandardness, in addition to providing a third protocol among clinical scenes. S^4 , S^5 , and S^6 constitute realistic, simulated scenes under different MRI protocols, and they provide a “clean” basis for controlling the process

of introducing inhomogeneity and for determining the various true factors since truth is known to begin with.

We do not claim that the various permutations and combinations of operations κ , κ_g , ψ , $\bar{\kappa}$, and $\bar{\psi}$ we have considered are exhaustive. However, we believe that they constitute a very reasonable, extensive, and realistic strategy within the scope of a single paper.

Step 2) Adding Known Levels of Inhomogeneity and Nonstandardness: As depicted in Fig. 1(a), to keep the number of scenes to be procured manageable, and since nonstandardness is a less studied and less understood phenomenon compared to inhomogeneity, only for S^1 , all possible combinations and orders of applying κ , κ_g , ψ , $\bar{\kappa}$, and $\bar{\psi}$ have been considered. Since S^3 , S^4 , S^5 , and S^6 do not have an issue of inhomogeneity and nonstandardness, an initial correction and standardization is not applied. For $\bar{\kappa}$, we considered two different models and five different degrees of inhomogeneity for the clinical scenes, and two different levels of inhomogeneity for the phantom scenes. For $\bar{\psi}$, we employed three different levels of nonstandardness. This gave us a total of 63 sets of scene data and with a total of 702 scenes considering all models and all levels of inhomogeneity and nonstandardness. Our rationale for initially correcting and standardizing and subsequently adding different levels of inhomogeneity was to study in a controlled manner, as a function of the degree of these artifacts, how the correction and standardization operations and their order of application influence the quality of the resulting scenes. Table II lists the number of models and levels utilized for controlling $\bar{\kappa}$ and $\bar{\psi}$ for the various data sets. In the rest of this section, the methods we have used for κ [15], κ_g [26], ψ [16], [17] are first briefly described, followed by the methods for $\bar{\kappa}$ and $\bar{\psi}$.

For inhomogeneity correction, we utilize two methods (κ , κ_g) both of which use the concept of local scale [15], [26]. In [15], object scale, called *ball scale*, or *b-scale*, at any voxel was defined as the radius of the largest ball centered at that voxel, and which lies entirely in the same object region, or equivalently, which is homogeneous under an appropriately defined criterion of homogeneity. In [26], we defined another local scale concept called *generalized scale*, or *g-scale*, as the largest connected set of voxels associated with every voxel in the scene such that all voxels in this set satisfy a predefined homogeneity criterion. The fundamental difference between the two scale-based inhomogeneity correction methods is the manner in which the local scale for every voxel is estimated. The first step for both correction methods is the same and comprises of computing scale (b-scale or g-scale) for all foreground voxels. For the b-scale correction method [15], the union of all balls centered at all voxels having

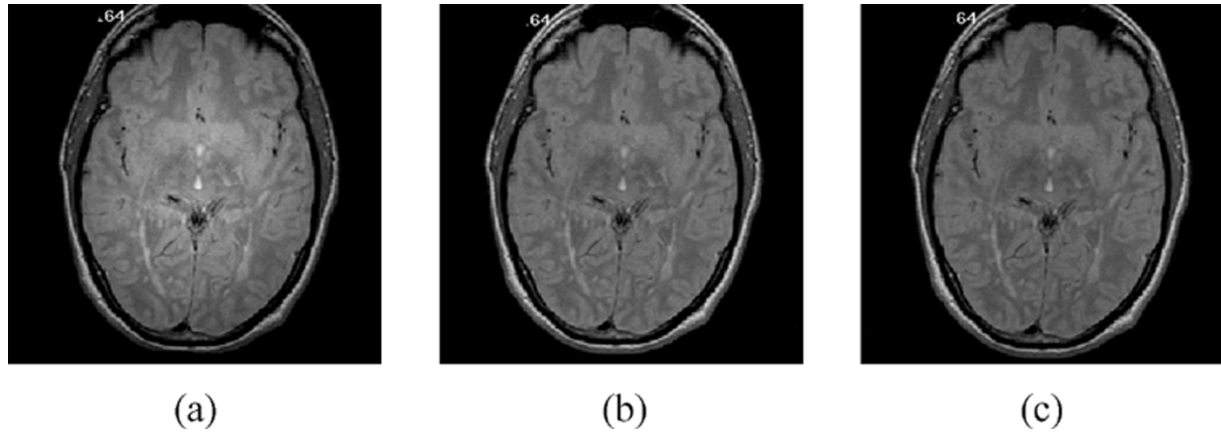


Fig. 2. (a) A slice of a clinical PD-weighted scene of the brain. The corresponding slice of the b-scale corrected (b) and g-scale corrected scene (c).

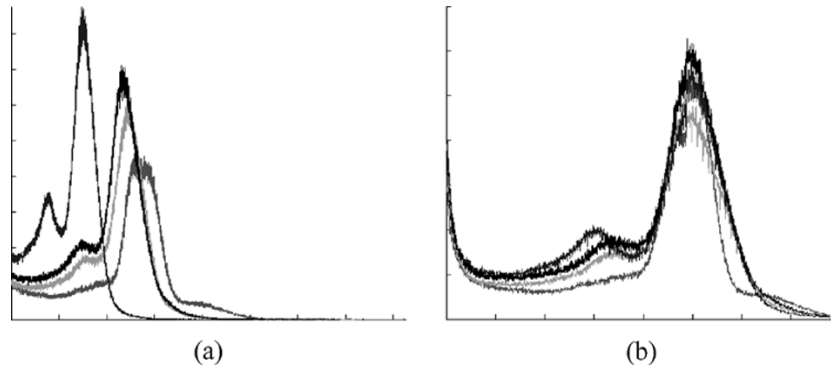


Fig. 3. (a) Histograms of 4 clinical PD scenes from set S^{1c} . (b) Histograms of corresponding standardized scenes.

the largest scale value is constructed. The mean and standard deviation of scene intensities within this region are computed and are used to threshold the original scene into a binary scene. From this binary scene, the background variation is estimated within the largest connected component of the binary scene. For the g-scale correction method, the background variation is estimated within the largest g-scale region. A second-order polynomial is fitted to the estimated variation for both methods, and by using this polynomial, a correction is applied to the entire scene. For both correction methods, the entire process is applied to the resulting scene and repeated iteratively. The basic premise of both algorithms is that, as correction is effected, the largest scale region increases, as demonstrated in [15], [26], as iterations continue. If this increase is insignificant, the iterative procedure stops. A slice of a PD-weighted scene from S^1 , and the corresponding slice obtained by using the b- and g-scale correction methods (on the 3-D scene) are shown in Fig. 2.

To standardize the scenes in a set S^{PD} for a given $P \in \mathcal{P}$ and $D \in \mathcal{D}$, in the training step, the intensities corresponding to certain landmarks obtained from the histogram of each scene in a subset of S^{PD} are mapped to a properly selected standard intensity scale by a linear mapping (see [17] for details). The mean locations for these mapped landmark locations are obtained. In the transformation step, the histogram of each scene \mathcal{C} in S^{PD} is first computed, and the intensity values corresponding to the landmark locations are determined. Sections of the intensity scale of \mathcal{C} are mapped to the corresponding sections of the

standard scale linearly to match the corresponding landmarks. This generally induces a nonlinear mapping from the intensity scale of \mathcal{C} to the standard scale. The standardization process is illustrated in Fig. 3, wherein histograms of four scenes from the set S^{1c} are shown before [Fig. 3(a)] and after [Fig. 3(b)] standardization.

The known background variation component which we wish to introduce into a given scene \mathcal{C} can itself be expressed as a scene $\mathcal{C}^a = (\mathcal{C}, f^a)$ which can be subsequently applied to \mathcal{C} to simulate $\bar{\kappa}$ as described below. For the clinical data sets, we used two different models for simulating background inhomogeneity. In the first model, applied to the scenes in S^{1c} , S^{1c_g} , S^{1cs} , S^{1sc} , $S^{1cs\bar{s}}$, $S^{1sc\bar{s}}$, and S^3 , we used Gaussian variations as a function of the x, y, z coordinates of voxels, with the Gaussian centered at the center of the scene domain and with a variance of σ^2 . In the second model, applied to the scenes in S^{2c} , and S^3 we used an inverted Gaussian in a similar manner, centered at the center of the scene domain. The first model makes the resulting scene brighter at the center of the scene domain, and the second model makes scene intensities higher at the periphery of the scene domain. In both cases, from a display of the slices of the scenes and of intensity profiles along lines selected on the slices, we ascertain that the effect of inhomogeneity is as in the original acquired scenes or more severe. It is generally considered that background inhomogeneity in MR images is multiplicative [12]. For any scene $\mathcal{C} = (\mathcal{C}, f)$, the scene $\mathcal{C}^{\bar{c}} = (\mathcal{C}, f^{\bar{c}})$ resulting from introducing inhomogeneity into it via \mathcal{C}^a is given, for any

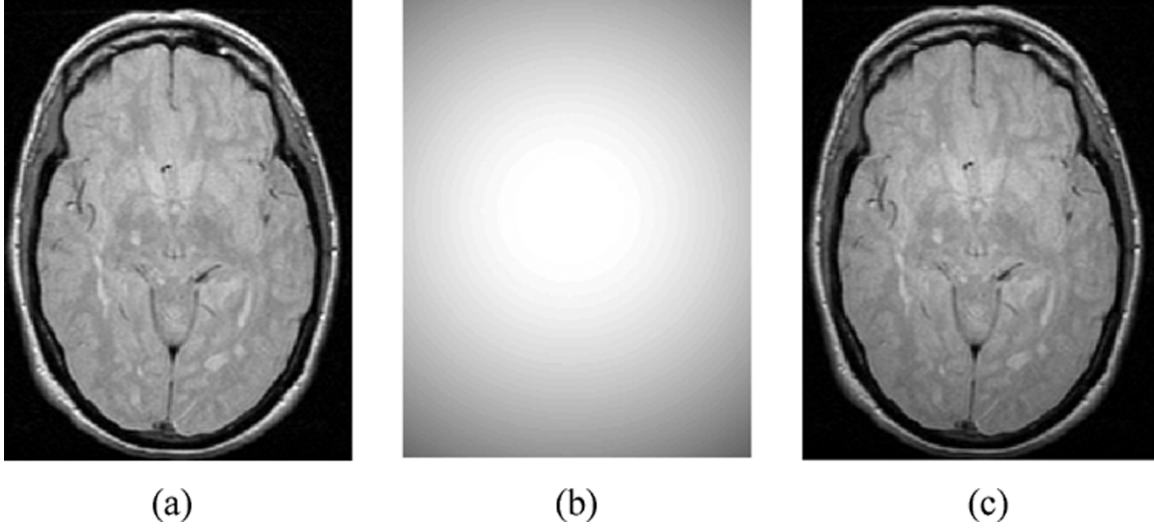


Fig. 4. (a) A slice from a clinical PD-weighted scene from the set S^{1c} , (b) corresponding slice from the Gaussian ($\sigma = 160$) inhomogeneity scene, and (c) corresponding slice from the inhomogeneity-added scene in the set $S^{1c\bar{c}}$.

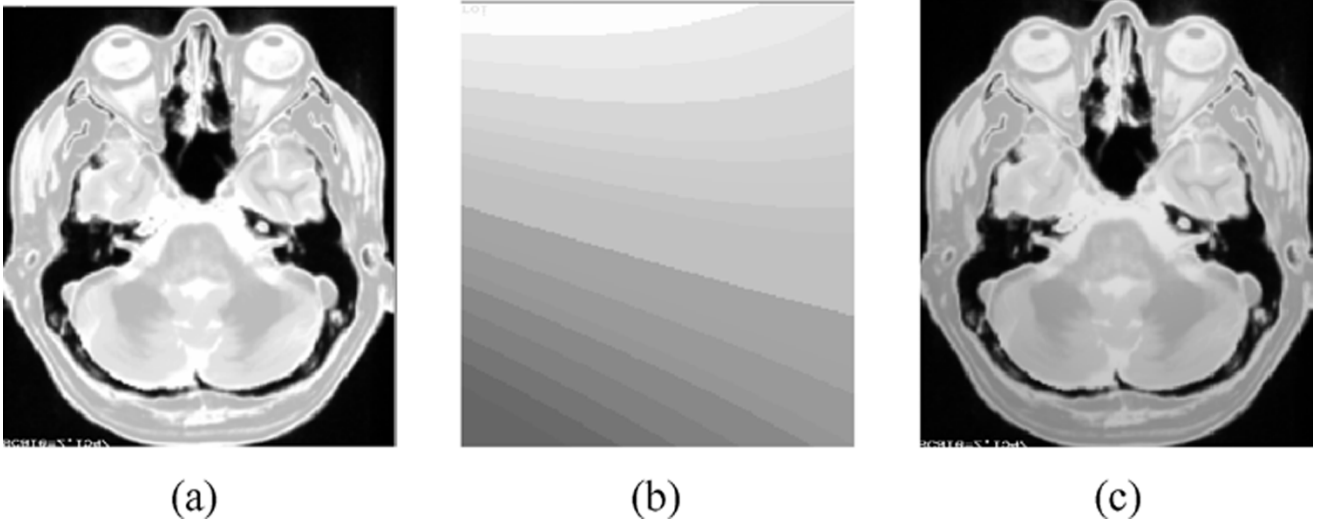


Fig. 5. Corresponding slices from a scene C in S^4 (a), the simulated inhomogeneity scene (b), and C^c in S^{4c} .

$u \in C$, by (1)–(2) at the bottom of the page, σ represents the standard deviation of the Gaussian which is the same along the x , y and z directions, $\|u - u_0\|$ is the Euclidean distance between u and u_0 , and u_0 is the center of the scene domain. For the phantom brain scenes, two levels of artificial inhomogeneity corresponding to 20% and 40% were introduced. This model of inhomogeneity was constructed based on real inhomogeneity patterns observed in clinical MR brain images (for more details, refer to [20]–[22]). The phantom scenes with and without inhomogeneity were created by the researchers at MNI and are available via <http://www.bic.mni.mcgill.ca/brainweb>. Figs. 4 and 5

show an example of scenes C , C^a , and C^c for C selected from each of S^{1c} and S^4 .

We now come to the process of introducing nonstandardness. In [17], several methods of standardization that use multiple histogram landmarks are described, but for simplicity, we focus on the method that uses three landmarks – the intensity p_1 corresponding to a certain lower percentile of the foreground of the scene, the median intensity (μ) of the foreground, and the intensity p_2 corresponding to a certain upper percentile of the foreground; see [16], [17] for details. In this approach, p_1 is taken to be 0, p_2 is taken to be the intensity corresponding to the 99.8

$$f^c(u) = f^a(u)f(u) \quad (1)$$

$$\text{where } f^a(u) = \begin{cases} e^{-1/2[\|u-u_0\|^2/\sigma^2]}, & \text{for scenes in } S^{1c}, S^{1c_g}, S^{1c_s}, S^{1sc}, S^{1cs\bar{s}}, S^{1sc\bar{s}}, S^3 \\ e^{-1/2[(\|u-u_0\|-\sqrt{3}\sigma)^2/\sigma^2]}, & \text{for scenes in } S^{2c}, S^{2c_g}, S^{2c_{gB}}, S^3. \end{cases} \quad (2)$$

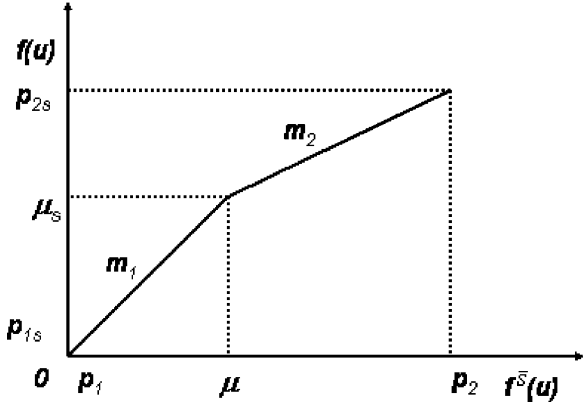


Fig. 6. The standardization mapping with the various parameters indicated.

percentile of the foreground, and a typical standardization mapping appears as shown in Fig. 6. In this figure, the horizontal axis indicates the input (nonstandard) scene intensity, and the vertical axis denotes the standardized intensity. Our approach to realize $\bar{\psi}$ is to vary the slopes m_1 and m_2 within the range of variations we have observed in our standardization mappings of clinical MR scenes, and then to apply the inverse mapping to generate the nonstandardized scenes. Let $C = (C, f)$ be a scene (from one of the sets S^{1cs} , $S^{1cs\bar{c}}$, S^{1sc} , and $S^{1sc\bar{c}}$) to which nonstandardness is to be introduced, and let $C^s = (C, f^s)$ be the resulting scene. Since C has already been standardized, from the point of view of nonstandardization, the vertical axis can be considered to represent $f(u)$ and the horizontal axis to represent $f^s(u)$ for any $u \in C$. With this, for a given value of the slopes m_1 and m_2

$$f^s(u) = \begin{cases} \left\lceil \frac{f(u)}{m_1} \right\rceil, & \text{if } f(u) \leq \mu_s \\ \left\lceil \frac{(f(u) - \mu_s)}{m_2} \right\rceil + \mu, & \text{if } f(u) > \mu_s \end{cases} \quad (3)$$

Here, $\lceil \cdot \rceil$ denotes the *ceiling* operation, and μ_s represents the median intensity on the standard scale. As an example, in Fig. 7 we display the intensity histograms of four scenes from the set S^{1cs} in Fig. 7(a) and of the scenes resulting from nonstandardization from the set $S^{1cs\bar{c}}$ in Fig. 7(b).

Step 3) Correction Followed by Standardization: This step requires no further explanation, and Fig. 1(b) illustrates the operations done.

Step 4) Standardization Followed by Correction: This step is analogous to Step 3 but with the sequences β and β_g . We argue that, at the outset, the approaches represented by both α , α_g and β , β_g sequences are plausible, and there is no clear and overwhelming recommendation of one over the other. Note that the goal of this work was not to compare the performance of individual α (α_g) or β (β_g) sequences, but rather to compare the performance of α sequences against β sequences in general. α , α_g make sense because inhomogeneity is an intra scene phenomenon, and its correction prior to standardization seemingly would facilitate standardization. β , β_g also appear reasonable because the correcting function that is fit to the estimated inhomogeneity has to, and will, be dependent on the mean scene intensity in some tissue region; this can introduce a bias depending on the particular tissue region utilized. (This phenom-

enon is not characteristic of only the method of correction we have employed. In Section III-B2, we offer an explanation for the reasons for this bias.) Therefore, standardization to begin with, which could account for inter scene variations, can potentially overcome this bias effect.

Step 5) Varying Levels of Inhomogeneity and Nonstandardness: For the clinical data sets, five different levels of inhomogeneity were simulated by varying σ in the definition of $f^a(u)$ in (2). Note that, smaller values of σ cause more severe inhomogeneity in the resulting scene. The choice of σ values was dictated by the dimensions of the scene domain $- \mathcal{L} \times \mathcal{M} \times \mathcal{N}$, where \mathcal{L} , \mathcal{M} , and \mathcal{N} are the scene dimensions in the x , y , and z directions. If we express the level of the introduced inhomogeneity by

$$\rho_\sigma = 2\sqrt{\frac{\sigma^2}{\mathcal{L}\mathcal{M}}} \quad (4)$$

then a value of $\rho_\sigma = 1.9$ is roughly the inhomogeneity level we observed in the acquired clinical images. If $\sigma = \mathcal{L}/2 = \mathcal{M}/2$, it implies that the intensity from the center of the scene domain to the boundary in the x - y plane has changed by 39.34%, which corresponds to very severe inhomogeneity. We decided to use levels of inhomogeneity corresponding to $1.25 \leq \rho_\sigma \leq 1.85$, representing moderate to severe inhomogeneity. With $\sigma = 160, 180, 200, 220$, and 240 for our experiments, ρ_σ fell in the above range for the scenes considered in our experiments. For the phantom data sets, two levels of inhomogeneity corresponding to 20% and 40% (representing moderate and high levels) were used.

Three ranges of values $-2.0 \leq m_1, m_2 \leq 2.5$, $2.0 \leq m_1, m_2 \leq 3.0$, and $2.0 \leq m_1, m_2 \leq 3.5$ – were used to artificially introduce three levels of nonstandardness. These were arrived at based on the levels of nonstandardness observed in the clinical MR scenes in our database.

To recall, our aim is to determine: which among the many possible sequences of operations represented by α and β produces the best possible improvement in terms of correction and standardization, if the phenomenon observed will depend on the particular method of correction, and if there is any dependency on the actual levels of nonstandardness and inhomogeneity. To fulfill our aim, our strategy is to define a measure for evaluating each of correction and standardization efficacy, and by utilizing these measures to perform paired t -tests to compare corresponding sets of scenes produced by the various sequences of operations depicted in Fig. 1(b). In the rest of this section, we first define the measures used and subsequently indicate the pairing that is done in our comparative tests.

Several different measures have been used for evaluating the effectiveness of correction procedures. In [15], the normalized standard deviation within segmented tissue regions was shown to be reduced significantly after correction. In [12], the square root of the mean squared difference between the modeled and estimated bias fields was used. The most popular choice of measure, however, seems to be the percent coefficient of variation (%CV) [10], [11], [14] within tissue regions, defined by

$$\%CV_t = \frac{\sigma_t}{\mu_t} \times 100 \quad (5)$$

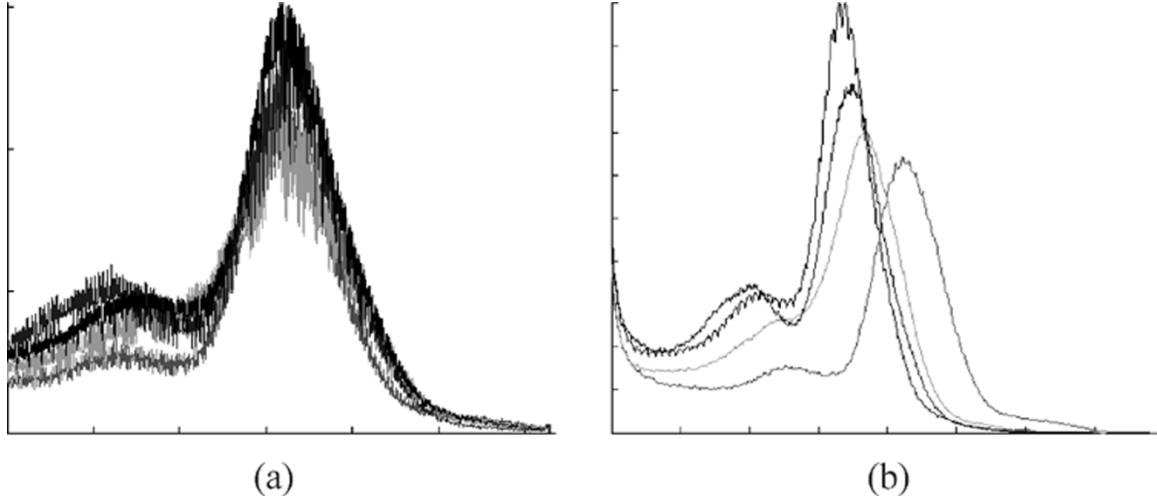


Fig. 7. (a) Histograms of four clinical PD scenes from set S^{1cs} . (b) Histograms of corresponding scenes, after the introduction of nonstandardness.

where μ_t and σ_t are the mean and standard deviation of scene intensities within a region corresponding to tissue t in a given scene. The premise here is that $\%CV_t$ will be lower after correction than before within the region corresponding to the same tissue t .

As we pointed out earlier, nonstandardness in scene intensity is an inter scene phenomenon (unlike inhomogeneity which is an intra scene process). The standard deviation of the mean intensity within the region corresponding to a given tissue t obtained from scenes corresponding to different acquisitions under the same MRI protocol has been used previously to characterize the effectiveness of standardization [16], [17]. Instead of utilizing directly the mean intensity within the region corresponding to tissue t (for determining the standard deviation), it was demonstrated in [16] that it is better to normalize the mean intensity as follows:

$$NMI_t = \frac{\mu_t}{H_{99.8} - L} \quad (6)$$

where μ_t is the mean intensity in the region of tissue t in a given scene C , L is the minimum scene intensity in C , and $H_{99.8}$ is the scene intensity in C corresponding to the 99.8 percentile value of the histogram of C . The reason for choosing $H_{99.8}$ rather than the maximum intensity in C is to leave out outlier intensities, and the choice of the number 99.8 is justified in [16, Fig. 3, p. 1075]. The standard deviation σ_{NMI_t} of NMI_t values for the different scenes in a set X is utilized to determine the nonstandardness present in X or the effectiveness of the standardization operation that gave rise to X .

To obtain the same tissue region, the scenes in set S^i for $1 \leq i \leq 3$, were segmented for white matter (WM) by using the fuzzy connectedness method [5] and were subsequently corrected by an expert (neuroradiologist) where needed. All segmentations, thus, passed the scrutiny for accuracy of a trained expert. WM was utilized since it is the largest tissue region in the brain and since the interior of this tissue region can be ascertained more reliably than the other brain tissue regions such as grey matter and cerebro-spinal fluid. The latter two regions have far more voxels in the tissue interface region (compared to their interior) than WM, and these are subjected to partial

volume effects, which will interfere with σ_{NMI_t} values considerably. To get voxels in the interior of the WM region, we use an erosion operation on the above segmented binary scenes so that a layer of two voxels from the boundary is removed. The WM masks for the MNI phantom scenes in the sets S^4 , S^5 , and S^6 were available from the MNI website.

To determine the effectiveness of correction, those pairs of sets in each of the following 11 collections of sets which are most likely to exhibit the phenomenon we are studying are compared: (1), (2) $S^{ic\bar{c}}$, $S_c^{ic\bar{c}}$, $S_s^{ic\bar{c}}$, $S_{cs}^{ic\bar{c}}$, $S_{sc}^{ic\bar{c}}$, $S_{csc}^{ic\bar{c}}$, $S_{scs}^{ic\bar{c}}$, $S_{scsc}^{ic\bar{c}}$ for $i = 1, 2$; (3)–(5) as in (1) and (2) but for MNI phantom data ($S^{i\bar{c}}$, $i = 4, 5, 6$); (6) $S^{1c_g\bar{c}}$, $S_{c_g}^{1c_g\bar{c}}$, $S_s^{1c_g\bar{c}}$, $S_{c_g s}^{1c_g\bar{c}}$, $S_{s c_g}^{1c_g\bar{c}}$, $S_{c_g s c_g}^{1c_g\bar{c}}$, $S_{s c_g s}^{1c_g\bar{c}}$, $S_{c_g s c_g s}^{1c_g\bar{c}}$; (7) as in (1) but for $S^{3\bar{c}}$; (8) $S^{1cs\bar{c}}$, $S_c^{1cs\bar{c}}$, $S_s^{1cs\bar{c}}$, $S_{cs}^{1cs\bar{c}}$, $S_{sc}^{1cs\bar{c}}$, (9)–(11) as in (8) but for $S^{1cs\bar{c}}$, $S^{1sc\bar{c}}$, and $S^{1sc\bar{c}}$.

To determine the effectiveness of standardization, those pairs of sets in each of the following 10 collections of sets which are most likely to exhibit the phenomenon we are studying are compared: (1), (2) $S^{ic\bar{c}}$, $S_c^{ic\bar{c}}$, $S_s^{ic\bar{c}}$, $S_{cs}^{ic\bar{c}}$, $S_{sc}^{ic\bar{c}}$, $S_{csc}^{ic\bar{c}}$, $S_{scs}^{ic\bar{c}}$, $S_{scsc}^{ic\bar{c}}$ for $i = 1, 2$; (3)–(5) as in (1) and (2) but for MNI phantom data ($S^{i\bar{c}}$, $i = 4, 5, 6$); (6) $S^{1c_g\bar{c}}$, $S_{c_g}^{1c_g\bar{c}}$, $S_s^{1c_g\bar{c}}$, $S_{c_g s}^{1c_g\bar{c}}$, $S_{s c_g}^{1c_g\bar{c}}$, $S_{c_g s c_g}^{1c_g\bar{c}}$, $S_{s c_g s}^{1c_g\bar{c}}$, $S_{c_g s c_g s}^{1c_g\bar{c}}$; (7) $S^{1cs\bar{c}}$, $S_c^{1cs\bar{c}}$, $S_s^{1cs\bar{c}}$, $S_{cs}^{1cs\bar{c}}$, $S_{sc}^{1cs\bar{c}}$, (8)–(10) as in (7) but for $S^{1cs\bar{c}}$, $S^{1sc\bar{c}}$, and $S^{1sc\bar{c}}$.

III. RESULTS AND DISCUSSION

In this section, we first present results demonstrating qualitatively the differences in the different orders of processing. Subsequently, we shall compare different sequences quantitatively by utilizing $\%CV_t$ and σ_{NMI_t} as the measures of effectiveness for correction and standardization, respectively.

A. Qualitative

One effective way of demonstrating the nonstandardness that exists in the scenes in a given set X is to display their histograms as illustrated in [16]. Fig. 8 shows histograms of 4 of the 12 scenes in the clinical PD data set $S^{1c\bar{c}}$ and in the different sets resulting from various sequences of processing operations done on $S^{1c\bar{c}}$. As can be expected, the histograms are spread out in

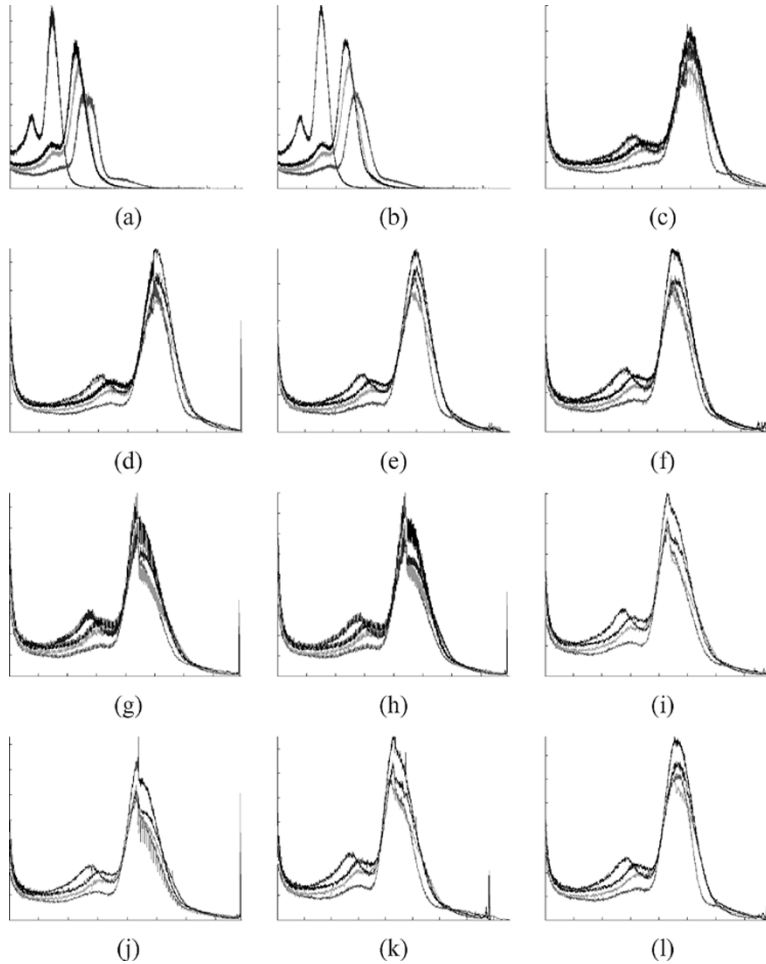


Fig. 8. Plots of histograms of four clinical PD scenes from the set $S^{1c\bar{c}}$ for various sequences of processing. (a) $S^{1c\bar{c}}$, (b) $S^{1c\bar{c}}$, (c) $S^{1c\bar{c}}$, (d) $S^{1c\bar{c}}$, (e) $S^{1c\bar{c}}$, (f) $S^{1c\bar{c}}$, (g) $S^{1c\bar{c}}$, (h) $S^{1c\bar{c}}$, (i) $S^{1c\bar{c}}$, (j) $S^{1c\bar{c}}$, and (k) $S^{1c\bar{c}}$. In (a) to (k) the starting set $S^{1c\bar{c}}$ is obtained by applying a high level of inhomogeneity ($\sigma = 160$). In (l), the histograms of the set $S^{1c\bar{c}}$ are shown resulting from a starting set $S^{1c\bar{c}}$ of scenes with a lower level of inhomogeneity ($\sigma = 240$).

Fig. 8(a) while they converge in Fig. 8(c) and in all other sets with a standardization operation in them. The histograms of $S^{1c\bar{c}}$ are of particular interest. The subsequent b-scale correction operation done after standardization seems to introduce a nonstandardness as we surmised earlier. Since it is difficult to visually appreciate this nonstandardness in Fig. 8(e), we magnified the same region of interest (ROI) on the histograms corresponding to sets $S_s^{1c\bar{c}}$, $S_{sc}^{1c\bar{c}}$ and $S_{csc}^{1c\bar{c}}$ and displayed them in Fig. 9. Fig. 9(a) corresponds to the histograms of scenes in $S_s^{1c\bar{c}}$ for $\sigma = 160$. As expected the histograms are all well aligned with one another. The images in Fig. 9(b) and (c) correspond to a magnified version of the same ROI in the histograms of the clinical PD scenes in $S_{sc}^{1c\bar{c}}$ and $S_{csc}^{1c\bar{c}}$, also for $\sigma = 160$. The peaks of the curves have moved away from one another, displaying the nonstandardness introduced by the b-scale correction procedure. Similar displays are shown in Fig. 9(d)–(f) for $\sigma = 240$, demonstrating a similar phenomenon.

Fig. 10(a) shows the histograms of 10 MTR scenes from the set $S^{3\bar{c}}$ containing artificial inhomogeneity, and Fig. 10(b), (c) shows the histograms of the corresponding scenes after the application of the b-scale and g-scale correction techniques, respectively. Since the MTR scenes have an inherent standardness property, the histograms in Fig. 10(a) are aligned with one another. However, Fig. 10(b), (c) shows a mis-alignment due to

the nonstandardness introduced by the correction methods. This suggests that the nonstandardness introduced by correction is not dependent on the specific correction method used. We shall come back to this point in Section III-B2 and present an argument as to why any other method of correction is also likely to introduce nonstandardness as demonstrated in Figs. 9 and 10. Further, from Figs. 8–10 it appears that the bias introduced into the standardized images by correction is independent of whether the images were naturally standardized (like in MTR images), or whether a standardization method [16], [17] had been employed post-hoc.

The larger the degree of inhomogeneity, the greater seems to be the level of nonstandardness introduced in this manner; compare the corresponding magnified ROI's in the histograms for the sets $S_{sc}^{1c\bar{c}}$, $S_{csc}^{1c\bar{c}}$ for $\sigma = 160, 240$ in Fig. 9(b), (c), (e), and (f). Once inhomogeneity is corrected the first time, any subsequent correction seems to have a negligible effect on standardization as borne out by Fig. 8(f), (i)–(k). This also leads us to conclude that long sequences of operations are not needed from the viewpoint of standardization if effective correction can be achieved in one or two repetitions of the correction operation.

Conversely, does standardization have an effect on inhomogeneity? We do not believe so, and this phenomenon is difficult to analyze qualitatively through display examples. We hence proceed to a discussion of quantitative analysis.

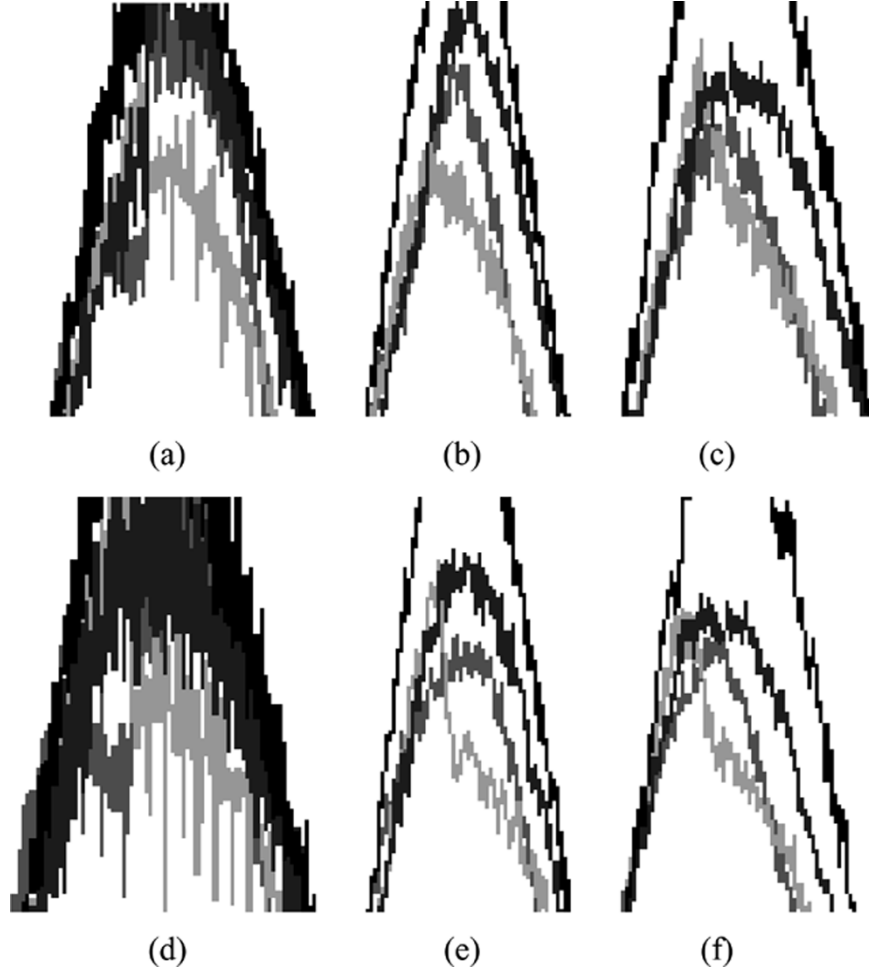


Fig. 9. Magnification of the same ROI in histograms from four scenes in the set $S^{1c\bar{c}}$ of clinical PD data for (a) $S_s^{1c\bar{c}}$, (b) $S_{sc}^{1c\bar{c}}$, (c) $S_{csc}^{1c\bar{c}}$ all for $\sigma = 160$, and (d) $S_s^{1c\bar{c}}$, (e) $S_{sc}^{1c\bar{c}}$, and (f) $S_{csc}^{1c\bar{c}}$ for $\sigma = 240$.

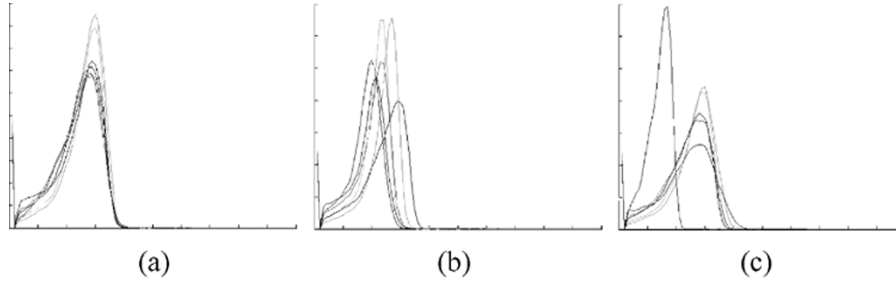


Fig. 10. (a) Histograms of 10 MTR scenes with artificially introduced inhomogeneity ($S^{3\bar{c}}$). (b), (c) Histograms of corresponding b-scale ($S_c^{3\bar{c}}$) and g-scale corrected scenes ($S_g^{3\bar{c}}$).

B. Quantitative

1) *Effects of Standardization:* Since the differences for most sets were insignificant in α , α_g and β , β_g sequences with more than four operations, we limit our discussion to the situations involving utmost four operations. For similar reasons, for the sets derived from $S^{1c\bar{c}\bar{s}}$, $S^{1c\bar{c}\bar{c}}$, $S^{1sc\bar{c}\bar{s}}$, and $S^{1sc\bar{c}\bar{c}}$ from α and β sequences, we limit our discussion for these sets to utmost two operations. Also we present tables only for some sample sets since the behavior for other sets were all similar.

In Tables III and IV are listed the σ_{NMI_t} values for the clinical data sets for different MRI protocols and for different models

and levels of applied inhomogeneity. The σ_{NMI_t} values for the clinical PD and T2 scenes at different inhomogeneity levels were compared for each pair of the sets $S^{ic\bar{c}}$, $S_c^{ic\bar{c}}$, $S_s^{ic\bar{c}}$, $S_{cs}^{ic\bar{c}}$, $S_{sc}^{ic\bar{c}}$, $S_{csc}^{ic\bar{c}}$, $S_{scs}^{ic\bar{c}}$, and $S_{scsc}^{ic\bar{c}}$, for $i = 1, 2$, by using a paired t-test under the null hypothesis that there is no difference in σ_{NMI_t} values between the sets. The comparisons were made for 8 among the 36 possible pairs for each i . These pairs are those which are most likely to exhibit the phenomena we are studying in this subsection. The p values for these pairs are listed in Table V. Along similar lines, in Tables VI–VIII are listed the σ_{NMI_t} values for the MNI phantom data sets for the different protocols and for the two levels of applied inhomogeneity. Sim-

TABLE III

 σ_{NMI_t} VALUES FOR A GAUSSIAN MODEL OF INHOMOGENEITY FOR $\sigma = 160, 180, 200, 220$, AND 240 FOR CLINICAL PD SCENES FROM THE SET $S^{1c\bar{c}}$

σ	S^{1cc}	S_c^{1cc}	S_s^{1cc}	S_{cs}^{1cc}	S_{sc}^{1cc}	S_{csc}^{1cc}	S_{scs}^{1cc}	S_{cscs}^{1cc}	S_{scsc}^{1cc}
160	.032	.029	.007	.006	.024	.021	.006	.006	.017
180	.034	.026	.007	.006	.016	.017	.006	.006	.015
200	.034	.031	.007	.006	.019	.027	.006	.006	.019
220	.034	.031	.007	.006	.015	.015	.006	.005	.016
240	.033	.031	.007	.006	.016	.017	.006	.006	.017

TABLE IV

 σ_{NMI_t} VALUES FOR AN INVERTED GAUSSIAN MODEL OF INHOMOGENEITY FOR $\sigma = 160, 180, 200, 220$, AND 240 FOR CLINICAL T2 SCENES FROM THE SET $S^{2c\bar{c}}$

σ	S^{2cc}	S_c^{2cc}	S_s^{2cc}	S_{cs}^{2cc}	S_{sc}^{2cc}	S_{csc}^{2cc}	S_{scs}^{2cc}	S_{cscs}^{2cc}	S_{scsc}^{2cc}
160	.017	.018	.008	.008	.020	.019	.009	.009	.019
180	.019	.019	.008	.008	.017	.035	.008	.008	.031
200	.018	.020	.008	.008	.018	.017	.008	.008	.016
220	.019	.020	.008	.008	.021	.027	.008	.008	.026
240	.018	.021	.008	.008	.019	.018	.008	.008	.026

TABLE V

 p -VALUES FOR PAIRED t -TESTS FOR COMPARING σ_{NMI_t} VALUES FOR DIFFERENT PAIRS OF SETS ($S^{1c\bar{c}}$, $S^{2c\bar{c}}$) CORRESPONDING TO CLINICAL PD AND T2 SCENES OBTAINED FOR TWO DIFFERENT INHOMOGENEITY MODELS AND FIVE LEVELS OF INTENSITY VARIATIONS

	$i=1$	$i=2$		$i=1$	$i=2$
$S_s^{ic\bar{c}}-S_{sc}^{ic\bar{c}}$.001	$3.334e^{-5}$	$S_{scs}^{ic\bar{c}}-S_{csc}^{ic\bar{c}}$.001	.006
$S_{scs}^{ic\bar{c}}-S_{cscs}^{ic\bar{c}}$.061	.317	$S_{csc}^{ic\bar{c}}-S_{scsc}^{ic\bar{c}}$	$1.841e^{-5}$.004
$S_{cs}^{ic\bar{c}}-S_{sc}^{ic\bar{c}}$.001	$2.901e^{-5}$	$S_{csc}^{ic\bar{c}}-S_{scsc}^{ic\bar{c}}$.142	.337
$S_c^{ic\bar{c}}-S_{cs}^{ic\bar{c}}$	$1.312e^{-5}$	$2.391e^{-5}$	$S_{sc}^{ic\bar{c}}-S_{csc}^{ic\bar{c}}$.234	.211

TABLE VI

 σ_{NMI_t} FOR TWO LEVELS OF INHOMOGENEITY FOR PHANTOM T1 SCENES FROM THE SET $S^{4\bar{c}}$

	$S^{4\bar{c}}$	$S_c^{4\bar{c}}$	$S_s^{4\bar{c}}$	$S_{cs}^{4\bar{c}}$	$S_{sc}^{4\bar{c}}$	$S_{csc}^{4\bar{c}}$	$S_{scs}^{4\bar{c}}$	$S_{cscs}^{4\bar{c}}$	$S_{scsc}^{4\bar{c}}$
20%	.025	.039	.020	.020	.030	.072	.026	.027	.031
40%	.065	.053	.049	.015	.071	.112	.022	.041	.051

TABLE VII

 σ_{NMI_t} VALUES FOR TWO LEVELS OF INHOMOGENEITY FOR PHANTOM T2 SCENES FROM THE SET $S^{5\bar{c}}$

	$S^{5\bar{c}}$	$S_c^{5\bar{c}}$	$S_s^{5\bar{c}}$	$S_{cs}^{5\bar{c}}$	$S_{sc}^{5\bar{c}}$	$S_{csc}^{5\bar{c}}$	$S_{scs}^{5\bar{c}}$	$S_{cscs}^{5\bar{c}}$	$S_{scsc}^{5\bar{c}}$
20%	.008	.009	.007	.007	.015	.007	.006	.006	.008
40%	.005	.008	.005	.005	.007	.008	.006	.004	.009

TABLE VIII

 σ_{NMI_t} VALUES FOR TWO LEVELS OF INHOMOGENEITY FOR PHANTOM PD SCENES FROM THE SET $S^{6\bar{c}}$

	$S^{6\bar{c}}$	$S_c^{6\bar{c}}$	$S_s^{6\bar{c}}$	$S_{cs}^{6\bar{c}}$	$S_{sc}^{6\bar{c}}$	$S_{csc}^{6\bar{c}}$	$S_{scs}^{6\bar{c}}$	$S_{cscs}^{6\bar{c}}$	$S_{scsc}^{6\bar{c}}$
20%	.008	.026	.007	.008	.025	.035	.009	.009	.035
40%	.016	.013	.010	.009	.011	.011	.010	.009	.012

ilarly, in Tables IX and X are listed the σ_{NMI_t} values for the clinical data sets $S^{1cs\bar{c}\bar{s}}$ and $S^{1sc\bar{c}\bar{c}}$, which were compared for each pair of the sets in $S_c^{1cs\bar{c}\bar{s}}$, $S_s^{1cs\bar{c}\bar{s}}$, $S_{cs}^{1cs\bar{c}\bar{s}}$, $S_{sc}^{1cs\bar{c}\bar{s}}$ and in $S_c^{1sc\bar{c}\bar{c}}$, $S_s^{1sc\bar{c}\bar{c}}$, $S_{cs}^{1sc\bar{c}\bar{c}}$, $S_{sc}^{1sc\bar{c}\bar{c}}$ for different levels of applied inhomogeneity and applied nonstandardness. From the results for the clinical data sets (Tables III–V), we note that the σ_{NMI_t} values for $S_s^{ic\bar{c}}$, $S_{cs}^{ic\bar{c}}$, $S_{scs}^{ic\bar{c}}$ and $S_{cscs}^{ic\bar{c}}$ are lower than those for $S_{sc}^{ic\bar{c}}$, $S_{csc}^{ic\bar{c}}$, and $S_{scsc}^{ic\bar{c}}$ for $i = 1, 2$. This difference was statistically significant for all sets in which correction was the last operation that

was performed (sets $S_s^{ic\bar{c}}$, $S_{cs}^{ic\bar{c}}$, and $S_{cscs}^{ic\bar{c}}$ for $i = 1, 2$) when compared to those in which standardization was the last operation that had been applied (sets $S_s^{ic\bar{c}}$, $S_{cs}^{ic\bar{c}}$, $S_{scs}^{ic\bar{c}}$, and $S_{cscs}^{ic\bar{c}}$ for $i = 1, 2$). Similar results were observed for the MNI phantom data sets (Tables VI–VIII). From the results for the clinical data sets $S^{1cs\bar{c}\bar{s}}$ and $S^{1sc\bar{c}\bar{c}}$ (Tables IX and X), we note that the σ_{NMI_t} values for $S_s^{1cs\bar{c}\bar{s}}$, $S_{cs}^{1cs\bar{c}\bar{s}}$ and $S_s^{1sc\bar{c}\bar{c}}$, $S_{cs}^{1sc\bar{c}\bar{c}}$ are lower than those for $S_c^{1cs\bar{c}\bar{s}}$, $S_{sc}^{1cs\bar{c}\bar{s}}$ and $S_c^{1sc\bar{c}\bar{c}}$, $S_{sc}^{1sc\bar{c}\bar{c}}$, respectively. Further, the same behavior was observed across different pro-

TABLE IX
 σ_{NMI_t} VALUES FOR A GAUSSIAN MODEL OF INHOMOGENEITY FOR $\sigma = 160, 180, 200$ FOR CLINICAL PD SCENES FROM THE SET S^{1cscs}

m_1, m_2	σ	S^{1cscs}	S_c^{1cscs}	S_s^{1cscs}	S_{cs}^{1cscs}	S_{sc}^{1cscs}
$2.0 \leq m_1, m_2 \leq 2.5$	160	.018	.088	.010	.009	.091
	180	.018	.076	.010	.009	.108
	200	.017	.100	.010	.009	.092
$2.0 \leq m_1, m_2 \leq 3.0$	160	.022	.031	.011	.005	.044
	180	.026	.032	.008	.006	.044
	200	.029	.037	.008	.006	.041
$2.0 \leq m_1, m_2 \leq 3.5$	160	.027	.042	.012	.005	.045
	180	.036	.041	.011	.005	.044
	200	.039	.043	.010	.005	.044

TABLE X
 σ_{NMI_t} VALUES FOR A GAUSSIAN MODEL OF INHOMOGENEITY FOR $\sigma = 160, 180, 200$ FOR CLINICAL PD SCENES FROM THE SET S^{1scsc}

m_1, m_2	σ	S^{1scsc}	S_c^{1scsc}	S_s^{1scsc}	S_{cs}^{1scsc}	S_{sc}^{1scsc}
$2.0 \leq m_1, m_2 \leq 2.5$	160	.018	.131	.011	.013	.080
	180	.017	.040	.008	.005	.039
	200	.017	.034	.007	.005	.032
$2.0 \leq m_1, m_2 \leq 3.0$	160	.020	.025	.010	.011	.066
	180	.026	.035	.008	.005	.032
	200	.027	.041	.007	.008	.091
$2.0 \leq m_1, m_2 \leq 3.5$	160	.029	.090	.013	.008	.089
	180	.037	.043	.010	.005	.035
	200	.039	.047	.009	.005	.039

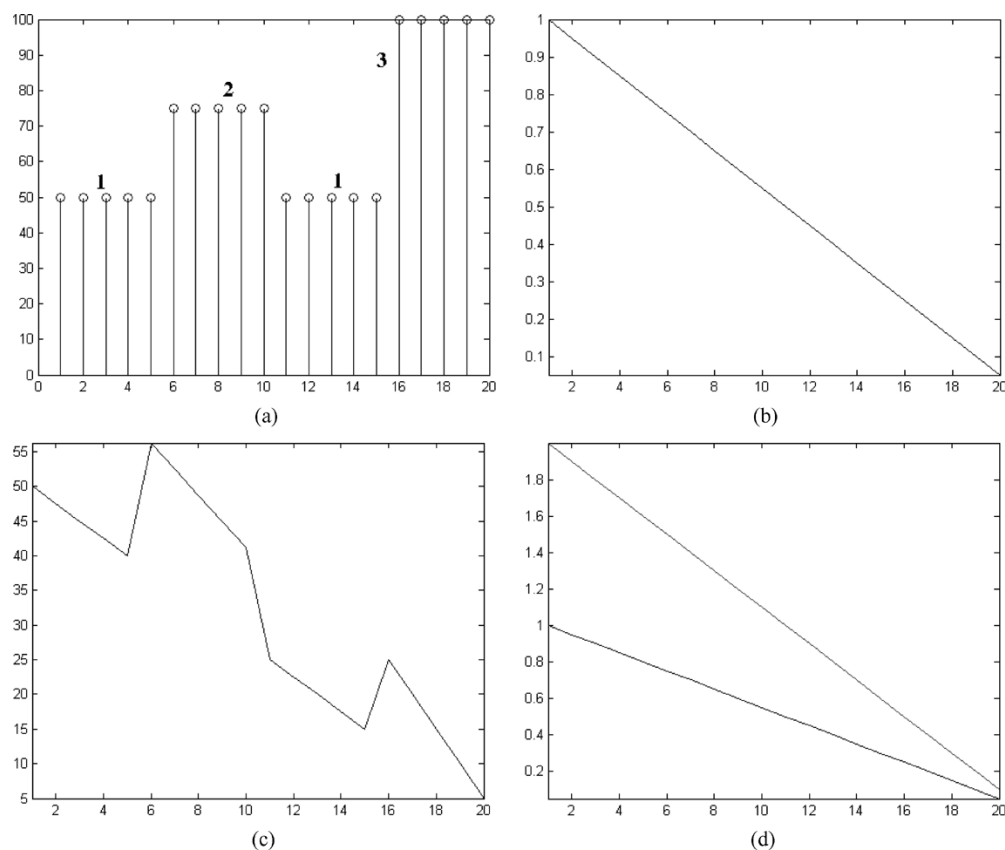


Fig. 11. (a) A 1-D scene representing three different tissue regions. (b) The ramp inhomogeneity introduced into the scene in (a). (c) Scene intensities after introduction of the ramp inhomogeneity. (d) Inhomogeneity estimated by using different regions of the same tissue (Tissue 1).

ocols, different models and levels of inhomogeneity, different levels of applied nonstandardness, for different inhomogeneity

correction methods (b-scale and g-scale), and for both naturally standardized (MTR) images as well as for those that had been

TABLE XI
%CV_t VALUES FOR DIFFERENT VALUES OF σ FOR THE GAUSSIAN INHOMOGENEITY MODEL FOR CLINICAL PD SCENES FROM THE SET S^{1cc}

σ	S^{1cc}	S_c^{1cc}	S_s^{1cc}	S_{cs}^{1cc}	S_{sc}^{1cc}	S_{csc}^{1cc}	S_{scs}^{1cc}	S_{scsc}^{1cc}	S_{scsc}^{1cc}
160	8.661	8.095	8.804	8.145	8.262	8.014	7.983	8.240	8.641
180	8.221	7.921	8.333	8.012	7.980	7.910	7.931	7.980	7.962
200	7.953	7.811	8.001	7.861	7.882	7.842	7.883	7.857	7.853
220	7.780	7.715	7.846	7.754	7.795	7.804	7.794	7.804	7.823
240	7.674	7.683	7.685	7.713	7.735	7.784	7.726	7.782	7.791

TABLE XII
%CV_t VALUES FOR TWO LEVELS OF INHOMOGENEITY FOR PHANTOM T1 SCENES FROM THE SET S^{4e}

	S^{4e}	S_c^{4e}	S_s^{4e}	S_{cs}^{4e}	S_{sc}^{4e}	S_{csc}^{4e}	S_{scs}^{4e}	S_{scsc}^{4e}	S_{scsc}^{4e}
20%	9.618	8.450	9.708	8.441	8.464	8.244	8.245	8.481	9.099
40%	14.781	6.800	8.675	6.952	6.824	7.649	7.816	6.980	7.653

TABLE XIII
p-VALUES FOR PAIRED t-TESTS FOR COMPARING %CV_t VALUES FOR DIFFERENT PAIRS OF SETS FOR CLINICAL PD SCENES FROM THE SET S^{1cc}

	$\sigma=$ 160	$\sigma=$ 180	$\sigma=$ 200	$\sigma=$ 220	$\sigma=$ 240		$\sigma=$ 160	$\sigma=$ 180	$\sigma=$ 200	$\sigma=$ 220	$\sigma=$ 240
$S^{1cc}-S_c^{1cc}$	$2.112e^{-4}$.013	.085	.141	.416	$S_{csc}^{1cc}-S_{scs}^{1cc}$.394	.064	.154	.378	.065
$S_c^{1cc}-S_{sc}^{1cc}$.050	.128	.115	.054	.139	$S_{csc}^{1cc}-S_{scsc}^{1cc}$.190	.240	.380	.103	.400
$S_s^{1cc}-S_{cs}^{1cc}$	$6.131e^{-4}$.022	.106	.149	.383	$S_c^{1cc}-S_{scs}^{1cc}$.133	.099	.095	.057	.098
$S_{cs}^{1cc}-S_{sc}^{1cc}$.092	.340	.204	.126	.189	$S^{1cc}-S_{csc}^{1cc}$.044	.021	.172	.394	.111
$S_s^{1cc}-S_{sc}^{1cc}$	$3.412e^{-4}$.006	.108	.296	.284	$S_{cs}^{1cc}-S_{scsc}^{1cc}$.327	.104	.404	.170	.052
$S_c^{1cc}-S_{cs}^{1cc}$.187	.066	.163	.182	.288	$S_{sc}^{1cc}-S_{scsc}^{1cc}$.221	.259	.317	.173	.051

standardized by using the method in [16] and [17]. This quantitatively demonstrates the effect of the correction procedure on nonstandardness, earlier alluded to in Figs. 8–10.

We also argue that this phenomenon is not an aberration of the correction methods we have employed. Since there is no guarantee that the same homologous tissue region across patients will be used to estimate the inhomogeneity, since the shape of these regions vary among patients, and since there is insufficient knowledge as to how to anchor the correction curve in exactly the same geographical location in the body for all data sets, this bias *will* be introduced regardless of the type of correction method used. Fig. 11 schematically illustrates these arguments. Fig. 11(a) is a one-dimensional (1-D) scene containing three different tissues (denoted by the labels 1–3). After the introduction of a ramp inhomogeneity [Fig. 11(b)], the observed intensities are altered [Fig. 11(c)]. By using the maximum intensities within the two regions corresponding to Tissue 1 as a surrogate for the true intensity of Tissue 1 (any other form of estimate, such as the mean, will also work in a similar manner), we obtain the estimated background inhomogeneities displayed in Fig. 11(d). The offset between the two curves has to be clearly independent of the correction method utilized and accounts for the standardization bias that we have observed in our experiments.

From Tables III–V, we may further make the following observation for the clinical data sets. For those sets ($S_{sc}^{ic\bar{c}}$, $S_{csc}^{ic\bar{c}}$, and $S_{scsc}^{ic\bar{c}}$, for $i = 1, 2$) in which standardization had been performed and in which correction was the last operation performed, there was no statistically significant difference in the σ_{NMI_t} values across protocols and across different models and levels of inhomogeneity. This suggests that, for a given level of inhomogeneity,

TABLE XIV
p-VALUES FOR PAIRED t-TESTS FOR COMPARING %CV_t VALUES FOR DIFFERENT PAIRS OF SETS FOR THE PHANTOM T1 SCENES FROM THE SET S^{4e}

	20%	40%		20%	40%
$S^{4e}-S_c^{4e}$.012	.009	$S_{csc}^{4e}-S_{scs}^{4e}$.322	.065
$S_c^{4e}-S_{sc}^{4e}$.402	.071	$S_{csc}^{4e}-S_{scsc}^{4e}$.232	.055
$S_s^{4e}-S_{cs}^{4e}$.012	.035	$S_c^{4e}-S_{scs}^{4e}$.352	.087
$S_{cs}^{4e}-S_{sc}^{4e}$.354	.055	$S^{4e}-S_{csc}^{4e}$.023	.005
$S_s^{4e}-S_{sc}^{4e}$.012	.032	$S_{cs}^{4e}-S_{scsc}^{4e}$.337	.088
$S_c^{4e}-S_{cs}^{4e}$.173	.019	$S_{sc}^{4e}-S_{scsc}^{4e}$.059	.228

generality, a standardization operation after correction accounts completely for the bias introduced by the preceding correction operation. Hence, the bias introduced by correction in the scenes in set $S_{scsc}^{ic\bar{c}}$ for $i = 1, 2$, is only on account of the second correction operation, since the second standardization operation corrects for the bias introduced by the first correction. The results from $S_{sc}^{ic\bar{c}}$, $S_{csc}^{ic\bar{c}}$, $S_{scsc}^{ic\bar{c}}$ for $i = 1, 2$, $S_{sc}^{1c\bar{c}\bar{s}}$, and $S_{sc}^{1sc\bar{c}\bar{s}}$ further suggest that the bias introduced by the last correction operation is independent of the correction and standardization operations that preceded the introduction of the artifacts. The results for $S_{sc}^{1c\bar{c}\bar{s}}$ and $S_{sc}^{1sc\bar{c}\bar{s}}$ suggest that the bias introduced by the final correction operation is independent of the order in which the artifacts were initially introduced.

The only notable difference in results between the clinical and phantom data sets is that the σ_{NMI_t} values for the latter sets are generally much lower than those for the former sets. This is understandable because the phantom data sets are all derived from the same single normal brain atlas.

TABLE XV
%CV_t FOR A GAUSSIAN MODEL OF INHOMOGENEITY FOR $\sigma = 160, 180, 200$ FOR CLINICAL PD SCENES FROM THE SET $S^{1sc\bar{s}\bar{c}}$

m_1, m_2	σ	$S^{1sc\bar{s}\bar{c}}$	$S_c^{1sc\bar{s}\bar{c}}$	$S_s^{1sc\bar{s}\bar{c}}$	$S_{cs}^{1sc\bar{s}\bar{c}}$	$S_{sc}^{1sc\bar{s}\bar{c}}$
$2.0 \leq m_1, m_2 \leq 2.5$	160	12.013	11.733	12.157	10.604	10.411
	180	9.178	7.898	9.144	7.823	7.934
	200	8.123	7.715	8.235	7.697	7.687
$2.0 \leq m_1, m_2 \leq 3.0$	160	12.043	9.081	12.067	8.718	9.076
	180	9.109	7.633	9.142	7.632	7.578
	200	8.267	7.645	8.234	7.611	7.596
$2.0 \leq m_1, m_2 \leq 3.5$	160	12.089	8.091	12.832	7.978	9.541
	180	9.107	7.651	9.323	7.644	7.788
	200	8.266	7.566	8.283	7.598	7.665

TABLE XVI
%CV_t FOR A GAUSSIAN MODEL OF INHOMOGENEITY FOR $\sigma = 160, 180, \text{ AND } 200$ FOR CLINICAL PD SCENES FROM THE SET $S^{1cs\bar{s}\bar{c}}$

m_1, m_2	σ	$S^{1cs\bar{s}\bar{c}}$	$S_c^{1cs\bar{s}\bar{c}}$	$S_s^{1cs\bar{s}\bar{c}}$	$S_{cs}^{1cs\bar{s}\bar{c}}$	$S_{sc}^{1cs\bar{s}\bar{c}}$
$2.0 \leq m_1, m_2 \leq 2.5$	160	11.879	10.192	12.090	9.712	10.495
	180	11.911	9.324	12.089	9.256	10.637
	200	11.983	10.336	12.080	10.087	10.079
$2.0 \leq m_1, m_2 \leq 3.0$	160	9.016	7.645	9.133	7.689	7.628
	180	9.029	7.645	9.102	7.787	7.649
	200	9.128	7.729	9.122	7.793	7.682
$2.0 \leq m_1, m_2 \leq 3.5$	160	8.192	7.658	8.329	7.673	7.751
	180	8.238	7.647	8.382	7.601	7.596
	200	8.192	7.596	8.233	7.622	7.739

TABLE XVII
p-VALUES FOR PAIRED t-TESTS FOR COMPARING %CV_t VALUES FOR DIFFERENT PAIRS OF SETS FOR CLINICAL PD SCENES FROM THE SET $S^{1cs\bar{s}\bar{c}}$

	$2.0 \leq m_1, m_2 \leq 2.5$	$2.0 \leq m_1, m_2 \leq 3.0$	$2.0 \leq m_1, m_2 \leq 3.5$
σ	160	180	200
$S^{1cs\bar{s}\bar{c}} - S_c^{1cs\bar{s}\bar{c}}$.011	$6.412e^{-4}$.071
$S_c^{1cs\bar{s}\bar{c}} - S_s^{1cs\bar{s}\bar{c}}$.010	$5.971e^{-4}$.069
$S_{cs}^{1cs\bar{s}\bar{c}} - S_{sc}^{1cs\bar{s}\bar{c}}$.159	.063	.492
$S_c^{1cs\bar{s}\bar{c}} - S_{cs}^{1cs\bar{s}\bar{c}}$.327	.049	.360
$S_c^{1cs\bar{s}\bar{c}} - S_{sc}^{1cs\bar{s}\bar{c}}$.270	.051	.244
$S_s^{1cs\bar{s}\bar{c}} - S_{cs}^{1cs\bar{s}\bar{c}}$.012	$5.001e^{-4}$.015
	160	180	200
$S^{1cs\bar{s}\bar{c}} - S_c^{1cs\bar{s}\bar{c}}$	$1.651e^{-5}$	$1.855e^{-5}$	$2.441e^{-5}$
$S_c^{1cs\bar{s}\bar{c}} - S_s^{1cs\bar{s}\bar{c}}$	$2.571e^{-5}$	$2.874e^{-5}$	$3.927e^{-5}$
$S_{cs}^{1cs\bar{s}\bar{c}} - S_{sc}^{1cs\bar{s}\bar{c}}$.346	.291	.163
$S_c^{1cs\bar{s}\bar{c}} - S_{cs}^{1cs\bar{s}\bar{c}}$.368	.274	.060
$S_c^{1cs\bar{s}\bar{c}} - S_{sc}^{1cs\bar{s}\bar{c}}$.377	.363	.425
$S_s^{1cs\bar{s}\bar{c}} - S_{cs}^{1cs\bar{s}\bar{c}}$	$7.655e^{-5}$	$9.085e^{-5}$	$1.023e^{-5}$
	160	180	200
$S^{1cs\bar{s}\bar{c}} - S_c^{1cs\bar{s}\bar{c}}$.002	.002	.002
$S_c^{1cs\bar{s}\bar{c}} - S_s^{1cs\bar{s}\bar{c}}$	$7.962e^{-4}$	$8.262e^{-4}$	$9.224e^{-4}$
$S_{cs}^{1cs\bar{s}\bar{c}} - S_{sc}^{1cs\bar{s}\bar{c}}$.243	.294	.361
$S_c^{1cs\bar{s}\bar{c}} - S_{cs}^{1cs\bar{s}\bar{c}}$.192	.101	.094
$S_c^{1cs\bar{s}\bar{c}} - S_{sc}^{1cs\bar{s}\bar{c}}$.241	.178	.280
$S_s^{1cs\bar{s}\bar{c}} - S_{cs}^{1cs\bar{s}\bar{c}}$.003	.003	.003

2) *Effects of Inhomogeneity Correction:* To evaluate the effects of correction independently, we compared %CV_t values for different models and levels of applied inhomogeneity for sets $S^{i\bar{c}}$, for $i = 1, 2, S^{1c_g\bar{c}}, S^{1cs\bar{c}\bar{s}}, S^{1sc\bar{c}\bar{s}}, S^{1cs\bar{s}\bar{c}}, S^{1sc\bar{s}\bar{c}}$, and $S^{i\bar{c}}$, for $3 \leq i \leq 6$, and the corresponding sets of scenes derived by the application of α, β or α_g , and β_g sequences of operations. As in Section III-B1, we present only some example tables since the behavior in other cases was similar. In Tables XI and XII are listed the %CV_t values for the sets derived from $S^{1c\bar{c}}$ and $S^{4\bar{c}}$ for the different protocols and for different models and levels of inhomogeneity. In Tables XV and XVI, we also list the %CV_t values for the sets derived from $S^{1sc\bar{s}\bar{c}}$ and $S^{1cs\bar{s}\bar{c}}$ for different levels of applied inhomogeneity and nonstandardness. These values were compared for every pair of sets for each level of inhomogeneity and nonstandardness (where applicable) by using a paired t-test under the null hypothesis that there is no difference in these values among the sets. The comparisons were made for 12 among the 36 possible pairs for each

collection of sets derived from $S^{i\bar{c}}$, for $i = 1, 2, S^{1c_g\bar{c}}$, and $S^{i\bar{c}}$, for $3 \leq i \leq 6$. For the sets derived from each of $S^{1cs\bar{c}\bar{s}}, S^{1sc\bar{c}\bar{s}}, S^{1cs\bar{s}\bar{c}}$, and $S^{1sc\bar{s}\bar{c}}$, the tests for statistical significance were made for 6 of the 10 possible pairs. These pairs are those which are most likely to exhibit the phenomena we are studying in this subsection. The p values for these pairs are listed in Tables XIII and XIV for the sets $S^{1c\bar{c}}$ and $S^{4\bar{c}}$, and in Tables XVII and XVIII for the sets $S^{1cs\bar{s}\bar{c}}$ and $S^{1sc\bar{s}\bar{c}}$. The pattern of behavior was found to be similar for other clinical and phantom data sets, for different methods of inhomogeneity correction, and for both naturally standardized (MTR) images, as well as those that had been standardized by using the method in [16] and [17], and hence the p -values for other comparisons are not listed here.

We make the following observations from these and other similar tables. All those sets in which correction had been performed at least once (sets $S_c^{1c\bar{c}}, S_{cs}^{1c\bar{c}}, S_{sc}^{1c\bar{c}}, S_{scs}^{1c\bar{c}}, S_{scs}^{1c\bar{c}}$, and $S_{scs}^{1c\bar{c}}$, for the set $S^{1c\bar{c}}$) had lower %CV_t than sets for which no correction had been performed (sets $S^{1c\bar{c}}$ and $S_s^{1c\bar{c}}$ for the set

TABLE XVIII

 p -VALUES FOR PAIRED t -TESTS FOR COMPARING $\%CV_t$ VALUES FOR DIFFERENT PAIRS OF SETS FOR CLINICAL PD SCENES FROM THE SET $S^{1sc\bar{c}}$

	$2.0 \leq$	m_1, m_2	≤ 2.5	$2.0 \leq$	m_1, m_2	≤ 3.0	$2.0 \leq$	m_1, m_2	≤ 3.5
σ	160	180	200	160	180	200	160	180	200
$S^{1sc\bar{c}} - S_c^{1sc\bar{c}}$.043	$1.112e^{-4}$.006	.021	$3.355e^{-6}$	$4.624e^{-4}$	$3.675e^{-7}$	$3.572e^{-6}$	$4.965e^{-4}$
$S^{1sc\bar{c}} - S_s^{1sc\bar{c}}$.042	$1.971e^{-4}$.008	.021	$9.423e^{-6}$	$8.917e^{-4}$	$1.912e^{-7}$	$3.322e^{-6}$	$3.524e^{-4}$
$S_{cs}^{1sc\bar{c}} - S_{sc}^{1sc\bar{c}}$.395	.428	.473	.385	.108	.087	.081	.363	.294
$S_c^{1sc\bar{c}} - S_{sc}^{1sc\bar{c}}$.210	.366	.373	.454	.127	.082	.103	.138	.174
$S_c^{1sc\bar{c}} - S_{cs}^{1sc\bar{c}}$.180	.342	.378	.244	.342	.395	.142	.412	.407
$S_s^{1sc\bar{c}} - S_{cs}^{1sc\bar{c}}$.011	$3.501e^{-4}$.011	$3.855e^{-4}$	$1.110e^{-5}$.001	$3.954e^{-8}$	$3.781e^{-6}$	$4.567e^{-4}$

TABLE XIX

FOR THE SET $S^{1c\bar{c}}$, THE BEST RESULTS FROM THE STANDPOINT OF STANDARDIZATION, CORRECTION, IMAGE QUALITY, AND THE SPEED OF OPERATION ARE LISTED

Standardization	Correction	Overall Image Quality	Time Required
$S_s^{1c\bar{c}}, S_{cs}^{1c\bar{c}}, S_{scs}^{1c\bar{c}}, S_{cscs}^{1c\bar{c}}$	$S_c^{1c\bar{c}}, S_{sc}^{1c\bar{c}}, S_{cs}^{1c\bar{c}}, S_{csc}^{1c\bar{c}}$ $S_{scs}^{1c\bar{c}}, S_{cscs}^{1c\bar{c}}, S_{scsc}^{1c\bar{c}}$	$S_{cs}^{1c\bar{c}}, S_{scs}^{1c\bar{c}}, S_{cscs}^{1c\bar{c}}$	$S_c^{1c\bar{c}}, S_s^{1c\bar{c}}, S_{cs}^{1c\bar{c}}, S_{sc}^{1c\bar{c}}$

$S^{1c\bar{c}}$). For the clinical data sets, this phenomenon was observed at higher and moderate levels of inhomogeneity ($\sigma = 160, 180, 200$). For low inhomogeneity levels (meaning higher values of σ), no statistically significant difference was found for the clinical data sets. A similar behavior was observed for the sets derived from $S^{1cs\bar{c}}$, $S^{1sc\bar{c}}$, $S^{1cs\bar{c}}$, and $S^{1sc\bar{c}}$, despite the fact that, in addition to inhomogeneity, nonstandardness had also been previously introduced into these scenes. This suggests that nonstandardness in the scene does not affect the performance of the correction operation. For the phantom data sets, the applied inhomogeneity corresponding to 20% and 40% were more severe than the inhomogeneity corresponding to the Gaussian and inverted Gaussian models for $\sigma = 200, 220$, and 240. Hence, these sets showed a statistically significant difference between all correction/no-correction pairs across protocols and levels of inhomogeneity.

The pairs of sets $S_{cs}^{1c\bar{c}}$, $S_{cscs}^{1c\bar{c}}$ and $S_{sc}^{1c\bar{c}}$, $S_{scsc}^{1c\bar{c}}$ did not have any statistically significant difference between them in terms of $\%CV_t$ for any level of inhomogeneity for any $i = 1, 2$. This pattern was similarly observed for the sets derived from $S^{1c_g\bar{c}}$, $S^{i\bar{c}}$, for $3 \leq i \leq 6$. This leads us to conclude that long sequences of operations are not needed from the viewpoint of correction if most of the background inhomogeneity can be removed in one or two repetitions of the correction operation. From this we infer that repeated correction followed by standardization and repeated standardization followed by correction do not improve image quality considerably from an inhomogeneity correction perspective. This pattern was seen across protocols, for both the clinical and phantom data sets, for different models and levels of inhomogeneity, for different levels of nonstandardness, for different methods of inhomogeneity correction, and for scenes that were naturally standardized (MTR) and were standardized with the intensity standardization method [16], [17].

In Table XIX, we list the best resulting sets derived from $S^{1c\bar{c}}$ in terms of residual nonstandardness, inhomogeneity, overall image quality, and the computational time required. Similar results hold for the sets derived from $S^{2c\bar{c}}$, $S^{1c_g\bar{c}}$, $S^{1cs\bar{c}}$, $S^{1sc\bar{c}}$, $S^{1cs\bar{c}}$, $S^{1sc\bar{c}}$, $S^{i\bar{c}}$, for $3 \leq i \leq 6$.

IV. CONCLUDING REMARKS

In this paper, we described a methodology for determining the effects of standardization on correction and *vice-versa*, in order to find the best sequence to be followed in processing an MR image from the perspective of producing the best image quality.

Two sequences of operations were performed on clinical and phantom images. In the first sequence, inhomogeneity correction followed by intensity standardization (and further alternate repetition of correction and standardization) was applied. In the second sequence, the order was reversed, with standardization followed by correction (and alternate repetition). The effects of the two sequences of operations on correcting for inhomogeneity and nonstandardness were determined by evaluating certain measures that were based on image intensities in the corrected images in presegmented tissue regions.

Based on the results obtained for the clinical and synthetic data (nearly 4000 3-D images in all), across different MRI protocols, for different models and levels of inhomogeneity, for different levels of artificial nonstandardness, for different methods of inhomogeneity correction, and for both naturally standardized (MTR) images as well as scenes that had been standardized by using an intensity standardization method [16], [17], we may make the following conclusions, with the only reservation that these are based on image data all pertaining to the human brain.

- If correction is the final operation performed on the scene, it will likely introduce some nonstandardness.
- From an overall image quality perspective, a sequence of operations in which intensity standardization is preceded by correction and in which the final operation applied on the scene is not correction, is the best sequence (sets $S_{cs}^{1c\bar{c}}$, $S_{scs}^{1c\bar{c}}$, and $S_{cscs}^{1c\bar{c}}$, for the set $S^{1c\bar{c}}$).
- No statistically significant improvement in terms of residual nonstandardness and inhomogeneity seems to ensue from sequences $\psi\kappa\psi\kappa, \dots, \psi\kappa$ and $\kappa\psi\kappa\psi, \dots, \kappa\psi$ over $\kappa\psi$, or from $\psi\kappa_g\psi\kappa_g, \dots, \psi\kappa_g$ and $\kappa_g\psi\kappa_g\psi, \dots, \kappa_g\psi$ over $\kappa_g\psi$.
- Overall, in terms of correcting for nonstandardness, background inhomogeneity, and the time required for

performing the different operations, the sequence $\kappa\psi$ or $\kappa_g\psi$ of operations seems to be the candidate of choice.

Thus far, we have looked at only the interaction between intensity standardization and inhomogeneity correction. A third important form of artifact, in addition to inhomogeneity and nonstandardness, is image noise. It is known that correction enhances noise, and we have shown that it enhances nonstandardness also. Consequently, taking noise into account, a new set of questions arise. How does correction affect noise filtering? Does filtering affect standardization? Is there a certain sequence of operations of filtering, standardization, and correction which when performed on the image produces the best overall image quality? Some preliminary results obtained in this direction are reported in [23]. In our future work, we shall investigate the interaction among these three phenomena carefully. The results presented in this paper should also be evaluated by using other inhomogeneity correction techniques [8]–[14].

ACKNOWLEDGMENT

The authors would like to acknowledge the comments and suggestions made by J. Liu, A. Souza, G. Grevera, and P. Saha.

REFERENCES

- [1] J. C. Bezdek, L. O. Hall, and L. P. Clarke, "Review of MR image segmentation techniques using pattern recognition," *Med. Phys.*, vol. 20, no. 4, pp. 1033–1048, 1993.
- [2] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–338, 2000.
- [3] M. Kamber, R. Shingal, D. L. Collins, S. Francis, and A. C. Evans, "Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images," *IEEE Trans. Med. Imag.*, vol. 14, no. 3, pp. 442–453, Sep. 1995.
- [4] B. Johnston, M. S. Atkins, B. Mackiewicz, and M. Anderson, "Segmentation of multiple sclerosis lesions in intensity corrected multi-spectral MRI," *IEEE Trans. Med. Imag.*, vol. 15, no. 2, pp. 154–169, Apr. 1996.
- [5] J. K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M. A. Van Buchem, and R. Grossman, "Multiple sclerosis lesion quantification using fuzzy-connectedness principles," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 598–609, Oct. 1997.
- [6] 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 Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [7] D. Shen, E. H. Herskovits, and C. Davatzikos, "An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures," *IEEE Trans. Med. Imag.*, vol. 20, no. , pp. 257–270, 2001.
- [8] L. Axel, J. Constantini, and J. Listerud, "Intensity correction in surface-coil MR imaging," *Amer. J. Radiol.*, vol. 148, pp. 418–420, 1987.
- [9] S.-H. Lai and M. Fang, "A new variational shape-from-orientation approach to correcting intensity inhomogeneities in MR images," in *Proc. Workshop Biomedical Image Analysis*, 1998, pp. 56–63.
- [10] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
- [11] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based bias field correction of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 885–896, Oct. 1999.
- [12] M. Styner, C. Brechbuhler, G. Szekely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 153–165, Mar. 2000.
- [13] A. Guimond, A. Roche, N. Ayache, and J. Meunier, "Three-dimensional multi-modal brain warping using the demons algorithm and adaptive intensity corrections," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 58–69, Jan. 2001.
- [14] B. Likar, M. A. Viergever, and F. Pernus, "Retrospective correction of MR intensity inhomogeneity by information minimization," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1398–1410, Dec. 2001.
- [15] Y. Zhuge, J. K. Udupa, J. Liu, P. Saha, and T. Iwanaga, "A scale-based method for correcting background intensity variation in acquired images," *Proc SPIE (Medical Imaging)*, vol. 4684, pp. 1103–1111, 2002.
- [16] L. G. Nyul and J. K. Udupa, "On standardizing the MR image intensity scale," *Magn. Reson. Med.*, vol. 42, pp. 1072–1081, 1999.
- [17] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.
- [18] Y. Ge, J. K. Udupa, L. G. Nyul, L. Wei, and R. I. Grossman, "Numerical tissue characterization in MS via standardization of the MR image intensity scale," *J. Magn. Reson. Med.*, vol. 12, pp. 715–721, 2000.
- [19] P. K. Saha, J. K. Udupa, and D. Odhner, "Scale-based fuzzy connectedness image segmentation: Theory, algorithms, and validation," *Comput. Vis. Image Understanding*, vol. 77, pp. 145–174, 2000.
- [20] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–468, Jun. 1998.
- [21] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," *NeuroImage*, pt. 2/4, vol. 5, no. 4, p. S425, 1997.
- [22] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation," *Visualization Biomed. Computing*, vol. 1131, pp. 135–140, 1996.
- [23] A. Montillo, J. Udupa, L. Axel, and D. Metaxas, "Interaction between noise suppression and inhomogeneity correction," *Proc. SPIE*, vol. 5032, pp. 1025–1036, 2003.
- [24] V. Dousset, R. I. Grossman, K. N. Ramer, M. D. Schnall, L. H. Young, S. F. Gonzalez, E. Lavi, and J. A. Cohen, "Experimental allergic encephalomyelitis and multiple sclerosis: Lesion characterization with magnetization transfer imaging," *Radiology*, vol. 182, pp. 483–492, 1992.
- [25] R. N. Muller, M. J. Marsh, M. L. Bernardo, and P. C. Lauterbur, "True 3-D imaging of limbs by NMR zeugmatography with off-resonance irradiation," *Eur. J. Radiol.*, vol. 3, pp. 286–290, 1983.
- [26] A. Madabhushi, J. K. Udupa, and A. Souza, "Generalized scale: Theory, algorithms, and application to inhomogeneity correction," *Proc. SPIE*, vol. 5370, pp. 765–776, 2004.