New methods of MR image intensity standardization via generalized scale

Anant Madabhushi
Department of Biomedical Engineering, Rutgers The State University of New Jersey, 617 Bowser Road, Room 101, Piscataway, New Jersey 08854

Jayaram K. Udupa[a)]
Medical Image Processing Group, Department of Radiology, University of Pennsylvania, 423 Guardian Drive, Blockley Hall, 4th Floor, Philadelphia, Pennsylvania 19104-6021

(Received 22 February 2006; revised 16 June 2006; accepted for publication 13 July 2006; published 30 August 2006)

Image intensity standardization is a post-acquisition processing operation designed for correcting acquisition-to-acquisition signal intensity variations (non-standardness) inherent in Magnetic Resonance (MR) images. While existing standardization methods based on histogram landmarks have been shown to produce a significant gain in the similarity of resulting image intensities, their weakness is that in some instances the same histogram-based landmark may represent one tissue, while in other cases it may represent different tissues. This is often true for diseased or abnormal patient studies in which significant changes in image intensity characteristics may occur. In an attempt to overcome this problem, in this paper, we present two new intensity standardization methods based on two scale concepts developed in Madabhushi *et al.* [Computer Vision Image Understanding **101**, 100–121 (2006)] for image processing applications. These scale concepts are utilized in this paper to accurately determine principal tissue regions within MR images. Landmarks derived from these regions are used to perform intensity standardization. The new methods were qualitatively and quantitatively evaluated on a total of 67 clinical three dimensional (3D) MR images corresponding to four different protocols and to normal, Multiple Sclerosis (MS), and brain tumor patient studies. The new scale-based methods were found to be better than the existing methods, with a significant improvement observed for severely diseased and abnormal patient studies. © *2006 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2335487]

I. INTRODUCTION

A major difficulty in MR image analysis[2] has been that intensities do not have a fixed tissue-specific numeric meaning, even within the same MRI protocol, for the same body region, and even for images of the same patient obtained on the same scanner. For most post-processing applications such as image segmentation and quantification, this lack of a standard and quantifiable interpretation of image intensities is a major drawback that compromises their precision, accuracy, and efficiency. A post-processing technique to automatically adjust the contrast and brightness of MR images (i.e., windowing) for image display has been presented in Ref. 3. However, although such automatic windowing may achieve display uniformity, they may not be adequate for quantitative image analysis, since the intensities still may not have tissue-specific numeric meaning after the windowing transformation. The only papers that we are aware of that address the problem of the standardization of image intensities explicitly are in Refs. 2, 4, and 5. 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. The few papers that have attempted to deal with this problem have done so from the standpoint of image segmentation and inhomogeneity correction,[6,7] and for explicitly creating standardized images that may be further analyzed by using other operations.

In Ref. 2, Nyul and Udupa presented a method that 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 a similar tissue-specific meaning. In the first step, the parameters of the standardizing transformation are learned 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[2,4,5] 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. In Ref. 2, the mode on the histogram was used as the landmark for transforming the scene intensities. In later work,[4] it was shown that the mode was not a robust landmark, and a variant of the original standardization procedure was described, replacing the mode with the median and other quartile locations on the histogram. These methods were shown to be more robust than the original mode-based method.

In cases when a disease is so pervasive that normal tissue image intensities are altered significantly over a significant portion of the image domain, the above histogram based landmark selection techniques are not fully effective in attaining good standardization. In an attempt to overcome these limitations, in this paper, we present a group of meth-

ods that uses a locally adaptive concept of image scale to identify in a robust manner tissue-specific landmarks on the histogram for carrying out standardization. The notion of scale employed in these new methods is a fundamental concept that has been found useful in many image processing and analysis tasks including segmentation, filtering, interpolation, registration, visualization, and quantitative analysis.

To overcome the sensitivity of the existing standardization methods to the landmarks on the histogram, we present two new methods in this paper based on two recently developed scale models called g and g_B-scale. These new methods, described in Sec. II, exploit the ability of the g and g_B-scale to automatically partition the image into homogeneous regions; the latter, in the context of medical images, correspond to different tissue regions. Unlike the existing methods, the new scale-based methods utilize landmarks derived from the individual scale regions in the image to perform the nonlinear mapping of intensities. We demonstrate in Sec. III that this makes the scale-based methods more robust than the existing methods, especially for the cases of patient studies with abnormalities.

II. METHODOLOGY

A. Theory

We represent a 3D volume image $\mathcal{C}$, called *scene* for short, by a pair $\mathcal{C}=(C,f)$, where C is a finite 3D rectangular array of voxels, called the *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* $f(u)$ to each $u \in C$. We denote the set of all protocols used in MR imaging by $\mathcal{P}$, the set of all body regions by $\mathcal{D}$, and the set of all scenes that can possibly be generated as per a given protocol $P \in \mathcal{P}$ for a given body region $D \in \mathcal{D}$ by $\mathcal{S}^{PD}$. The histogram of any scene $\mathcal{C}$ is a pair $\mathcal{H}=(\mathbf{G},h)$, where $\mathbf{G}$ is the set of all possible intensity values (gray values) in $\mathcal{C}$ and h is a function whose domain is $\mathbf{G}$ and whose value for each $x \in \mathbf{G}$ is the number of voxels $u \in C$ for which $h(u)=x$. Let m_1 and m_2 be the minimum and maximum gray values in $\mathcal{C}$, respectively.

Broadly speaking, scale concepts utilized in image processing can be divided into three categories: (1) multiscale or scale-space representation, (2) local scale, and (3) locally adaptive scale. The motivation for the original formulation of scale in the form of *scale-space* theory came from the presence of multiple scales in nature and the desire to represent measured signals at multiple scales.[8] However, since this representation did not suggest how to select the scales appropriately, the notion of local scale was proposed for choosing the *right scale* for a particular application from the multiscale representation of the image.[9–11] Recently, there has been considerable interest in developing *locally adaptive scales*,[12–15] the idea being to consider the local size of object in carrying out whatever local operations that are to be done on the image. In Ref. 1, we proposed a generalized scale (abbreviated as a *g-scale* from now on) model that is adaptive like other local morphometric models, and that possesses the global spirit of multiscale representations. A variant of the g-scale called the generalized ball scale (abbreviated as a g_B-*scale*), which in addition to having the advantages of the g-scale model, has superior noise resistance properties and was also described in Ref. 1. A local scale model, called the *ball scale*[12] or *b-scale*, was previously proposed to determine the size of local structures at every voxel in the scene. The *b-scale* at every voxel was defined as the radius of the largest ball centered at the voxel such that all voxels within the ball satisfy a predefined homogeneity criterion. Thus, for any given scene, the b-scale concept yields a b-scale scene with the scene intensity of a voxel representing the b-scale value. The b-scale model was shown to have excellent noise resistance properties.[16] To remove the shape, size, and anisotropic constraints of the spherical model of the b-scale, the *generalized scale* or *g-scale* $G(c)$ at any voxel c in a scene $\mathcal{C}=(C,f)$ was defined as the largest fuzzily connected[18] subset of C containing c, such that all voxels in $G(c)$ satisfied a predetermined homogeneity criterion.[1] For any voxel c in $\mathcal{C}$, the g_B-*scale* $G_B(c)$ was defined as the largest connected subset of C containing c such that the b-scale of voxels within $G_B(c)$ were greater than a specified tolerance value. The difference between the g- and g_B-scale models is fundamental in their definition. While the g-scale is estimated by the addition of individual voxels into a g-scale set based on the homogeneity criterion, the g_B-scale is determined by the inclusion of voxels that satisfy a homogeneity criterion for their b-scale regions. While both the g- and g_B-scale models share similar properties, the difference in the manner in which they are defined makes the g_B-scale model more resistant to noise than the g-scale. The g-scale corresponds essentially to a fuzzy connected component (based on the homogeneity) of $\mathcal{C}$, and, hence, it is computed via dynamic programming.[18] The g_B-scale requires the computation of the corresponding b-scale scene first. The g_B-scale $G_B(c)$ of c is then determined as the (hard) connected component, containing c, in the binary scene resulting from thresholding the b-scale scene at the tolerance value. The set of all g and g_B-scales associated with $\mathcal{C}$ are denoted by $\mathcal{G}(\mathcal{C})$ and $\mathcal{G}_B(\mathcal{C})$, respectively. Both definitions induce a partitioning on the scene domain C. That is, the elements of $\mathcal{G}(\mathcal{C})$ and $\mathcal{G}_B(\mathcal{C})$ correspond to the elements of their partition; see Ref. 1 for details. We will use subscripts s, s_g, and s_{g_B} to denote the scenes and the sets of scenes resulting from applying the histogram landmark-based,[2] g-, and g_B-scale-based standardization methods, respectively, on scenes and sets of scenes. With this notation, a subset S of $\mathcal{S}^{PD}$ of scenes that have been standardized by using the three standardization methods will be denoted by S_s, S_{s_g}, and $S_{s_{g_B}}$, respectively.

The basic idea of the standardization methods described in Refs. 2 and 4 is to identify a set of landmarks on the gray scale of the scenes via the scene histograms in such a manner that each landmark has the same tissue-specific meaning. To achieve standardization, these landmarks are mapped onto a fixed standard gray scale in a piecewise linear manner. The main departure in this paper from Refs. 2 and 4 is in the

manner in which the landmarks are identified. Subsequently, the mapping is done in exactly the same way as in Refs. 2 and 4.

As described in Refs. 2 and 4, it is desirable to cut off the tails of the histograms of the scenes for arriving at a standardization mapping because they often cause problems. Usually the high-intensity tail corresponds to artifacts and outlier intensities and causes considerable inter- and intrapatient/scanner variations. With this in mind, let pc_1 and pc_2 denote the minimum and maximum cutoff percentile values, respectively, of the histogram $\mathcal{H}$ of a given scene $\mathcal{C}$. Let the actual intensities corresponding to pc_1 and pc_2 be p_1 and p_2. $[p_1,p_2]$ represents the range of *intensities of interest* (IOI) for $\mathcal{C}$. Outside this range, the intensities are not of any consequence. Within $[p_1,p_2]$, additional landmarks are determined. For example, in one of the methods described in Ref. 4, the median intensity p_3 of the foreground of $\mathcal{C}$ is used as a landmark in $[p_1,p_2]$. Subsequently $[p_1,p_3]$ and $[p_3,p_2]$ are mapped linearly onto the standard gray scale. So as not to lose any intensities in the input gray scale, $[m_1,p_1]$ and $[p_2,m_2]$ are mapped onto the standard gray scale to extend the ends of the standard gray scale. The mapping functions for these two segments are assumed to be the same linear mappings as those used on $[p_1,p_3]$ and $[p_3,p_2]$, respectively.

In Refs. 2 and 4, the mode, median, deciles, and quartiles were all used in the following landmark configurations for the histogram based standardization method:

$$L_1=\{pc_1,\mu,pc_2\},$$

$$L_2=\{pc_1,\mu_{50},pc_2\},$$

$$L_3=\{pc_1,\mu_{25},\mu_{50},\mu_{75},pc_2\},$$

$$L_4=\{pc_1,\mu_{10},\mu_{20},\mu_{30},\ldots,\mu_{90},pc_2\}, \tag{2.1}$$

where μ_p for $p\in\{10,20,25,30,\ldots,75,\ldots,90\}$ represents the intensity value corresponding to the pth percentile in the histogram associated with the foreground part of the scene, and μ represents its mode. For the new generalized scale-based standardization methods, we may consider any of these configurations. Since the difference between L_2 and L_3, and between L_2 and L_4 has been found to be insignificant in Ref. 4, and since L_2 is superior to L_1, in this paper we shall focus on L_2. The only difference is that, for the new methods, μ_{50} represents the median intensity within the region that is selected by the scale-based method. Let $L_{2_g}=\{pc_1,\mu_{50_g},pc_2\}$ and $L_{2_{g_B}}=\{pc_1,\mu_{50_{g_B}},pc_2\}$ be the configurations similar to L_2 but used in the g- and g_B-scale methods, where μ_{50_g} and $\mu_{50_{g_B}}$ denote the median value determined from the respective methods.

B. Methods

The method comprises of two separate steps: training, transformation. In the first step (training), a set of scenes of the same body region D and protocol P corresponding to a population of patients is given as input. The scale sets $[\mathcal{G}(\mathcal{C})$ or $\mathcal{G}_B(\mathcal{C})]$ for the training scenes are computed, and the landmark locations on the standard scale, which are required for the intensity transformation process, are learned from these scene data. This step needs to be executed only once for a given D and P. In the second (transformation) step, the scenes are transformed by using the parameters learned in the first step. This transformation is scene dependent and needs to be done for each given scene. These steps are described in more detail below.

1. Training

(i) For a given $P\in\mathcal{P}$ and $D\in\mathcal{D}$, a subset T^{PD} of $\mathcal{S}^{PD}$ of scenes is collected and used for training.

(ii) The upper and lower percentile intensity values p_1 and p_2 on the histogram $\mathcal{H}$ of $\mathcal{C}$ corresponding to pc_1 and pc_2 are determined, as described in Refs. 2 and 4, for each scene $\mathcal{C}\in T^{PD}$.

(iii) The g-/g_B-scale set over each of the training scenes $\mathcal{C}\in T^{PD}$ is then computed by using g- and g_B-scale algorithms,[1] and the largest scale region is determined. The median intensity μ_{50_g} or $\mu_{50_{g_B}}$ within the largest scale region is computed.

(iv) The intensities from the interval $[p_1,p_2]$ are linearly mapped to $[s_1,s_2]$, where s_1 and s_2 are the minimum and maximum intensities on the standard scale. The formula for mapping $x\in[p_1,p_2]$ to $x'\in[s_1,s_2]$ is the following:

$$x'=s_1+\frac{x-p_1}{p_2-p_1}(s_2-s_1). \tag{2.2}$$

In this process, the median tissue intensity, μ_{50_g} or $\mu_{50_{g_B}}$, is transformed to μ'_{50_g} or $\mu'_{50_{g_B}}$ on the standard scale for each scene $\mathcal{C}\in T^{PD}$.

(v) The rounded median intensity μ_{s_g} or $\mu_{s_{g_B}}$ on the standard scale is computed from the average of μ'_{50_g} or $\mu'_{50_{g_B}}$ over all scenes in T^{PD}.

2. Transformation

(i) For any given scene $\mathcal{C}\in\mathcal{S}^{PD}$ to be standardized, its p_1 and p_2 values and the largest g-/g_B-scale region are determined. The median intensity μ_{50_g} or $\mu_{50_{g_B}}$ of the scale region is then computed.

(ii) A piecewise linear mapping is then determined, as described in Refs. 2 and 4, so as to match the upper and lower percentile intensities p_1 and p_2 of $\mathcal{C}$ with s_1 and s_2 and μ_{50_g} or $\mu_{50_{g_B}}$ with μ_{s_g} or $\mu_{s_{g_B}}$. Figure 1 shows a plot of the mapping function. The lower and upper ends of the standard scale are subsequently extended to s'_1 and s'_2, respectively, by mapping $[m_1,p_1]$ to $[s'_1,s_1]$ and $[p_2,m_2]$ to $[s_2,s'_2]$ for scene $\mathcal{C}\in\mathcal{S}^{PD}$, as illustrated in Fig. 1. We call this mapping from the intensities $[m_1,m_2]$ of $\mathcal{C}$ to $[s'_1,s'_2]$ of the standard scale the standardizer of $\mathcal{C}$ and denote it by ψc_g or ψc_{g_B}. The expression for ψc_g for the g-scale-based method (from Fig. 1) is as follows:

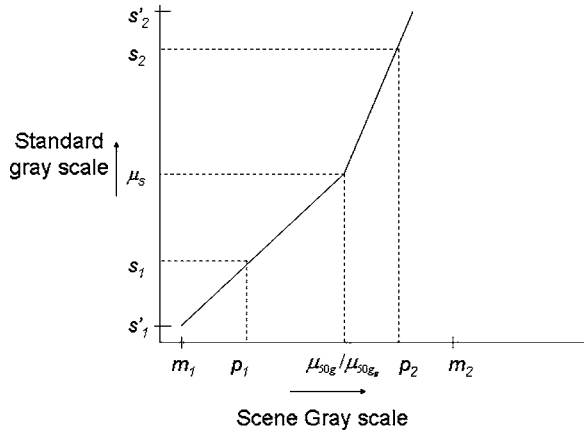

FIG. 1. g-scale-based standardization mapping with the various parameters indicated.

$$\psi c_g(x) = \begin{cases} \left\lceil \mu_{s_g} + (x - \mu_{50_g}) \dfrac{s_1 - \mu_{s_g}}{p_1 - \mu_{50_g}} \right\rceil, & \text{if } m_1 \leqslant x \leqslant \mu_{50_g}, \\ \left\lceil \mu_{s_g} + (x - \mu_{50_g}) \dfrac{s_2 - \mu_{s_g}}{p_2 - \mu_{50_g}} \right\rceil, & \text{if } \mu_{50_g} \leqslant x \leqslant m_2, \end{cases} \tag{2.3}$$

where $\lceil \cdot \rceil$ denotes the *ceiling* operation, (it converts any real number y to the closest integer Y such that $Y \geqslant y$). Instead, a *floor* operator ($Y \leqslant y$) may also be used. Note that $s_1' = \psi c_g(m_1)$, and $s_2' = \psi c_g(m_2)$. The scene $\mathcal{C}_{s_g} = (C, f_{s_g})$ resulting from the g-scale-based standardization mapping of $\mathcal{C}$ is given by, for all $c \in C, f_{s_g}(c) = \psi c_g[f(c)]$ wherein μ_{50_g} is used in (2.3). $\mathcal{C}_{s_{g_B}}$ is similarly defined by replacing μ_{50_g} by $\mu_{50_{g_B}}$ and μ_{s_g} by $\mu_{s_{g_B}}$ in Eq. (2.3). We point out that the free ends characterized by the values of s_1' and s_2' of the standard scale depend on the given scene $\mathcal{C}$. In other words, the range $[s_1', s_2']$ may vary from scene to scene. However, $[s_1, s_2]$ is independent of $\mathcal{C}$ and this is the interval within which a uniformity of tissue-specific meaning is achieved.

Table I shows the different parameter settings that were used for the three methods (histogram landmark-based,[2,4] g-scale-based, and g_B-scale-based methods). We evaluated the three methods listed in Table I both qualitatively and quantitatively by using a set of 67 clinical MR image datasets corresponding to four different protocols [PD-, T2–, T1–, and T1–weighted with gadolinium enhancement (T1E)] and acquired from 22 normal subjects, 33 patients with Multiple Sclerosis, and from 12 brain tumor patients. The seven groups of datasets, denoted by S^1 to S^7, are described in Table II. For all sets, the body region D considered was the head. Prior to intensity standardization, each of the 67 datasets was corrected for bias field intensity variations via the generalized-scale based inhomogeneity correction method described in Ref. 1. As justified in Ref. 17, an inhomogeneity correction was done first because of the fact that any such method can itself introduce non-standardness into the scene data. For qualitative evaluation, we considered (1) plotting the standardized histograms and (2) displaying the binary versions of standardized scenes obtained after thresholding at fixed levels. A quantitative evaluation was performed by computing and comparing statistics within the largest tissue regions for the different standardization methods. The training was done by using five different patient studies under each protocol for each of the standardization methods.

TABLE I. Parameter configurations used for the different standardization methods.

Method	pc_1	pc_2	s_1	s_2	Landmarks
Histogram	0	99.8	0	4095	μ_{50}
g-scale	0	99.8	0	4095	μ_{50_g}
g_B-scale	0	99.8	0	4095	$\mu_{50_{g_B}}$

III. RESULTS

A. Qualitative

We conducted qualitative comparisons for the following MRI protocols: PD, T2, T1, and T1E. Our hypothesis was that the performance of the new scale-based standardization methods would be comparable to that of the existing methods[2] on normal datasets and on datasets wherein scene intensities do not undergo significant changes due to a disease, but would be significantly better in abnormal and severely diseased cases. Within any of the protocols used in our study, the image acquisition parameters were identical for all patient studies. The voxel intensities were represented as 12-bit integers. No additional preprocessing was done on any of these scene data. We have also experimented with studies of different slice thickness and orientation and found no significant differences in the results. Since the method is applied to the whole scene and whole volume histogram and not to the individual slices, the slice orientation and the resolution has negligible effect on the transformation within reasonable limits.

1. Histograms

Histograms of PD, T2, and T1E scenes (selected from datasets S^1, S^4, and S^7) corresponding in turn to normal, MS, and brain tumor patients before and after standardization by using the existing and new methods are shown in Fig. 2. To avoid clutter we have shown only four (and not all) of the scene intensity histograms for each case. The low intensity part of the histogram that corresponds to the background voxels has been removed from the display in order to show the IOI on a better scale. A visual comparison shows that, for all studies and for all protocols, all three methods produce standardized scenes whose histograms are more similar in alignment than those of the original scenes. For the normal patient studies (S^1), the histograms corresponding to the scenes produced by the existing method (S_s^1), the g-scale ($S_{s_g}^1$) and g_B-scale ($S_{s_{g_B}}^1$) appear similar in shape and alignment [Figs. 2(d), 2(g), and 2(j)]. For the MS studies (S^4), the histograms of the scenes in $S_{s_g}^4$ and $S_{s_{g_B}}^4$ seem more closely

TABLE II. A description of the datasets used in an evaluation.

Set	Number of scenes	Protocol	Type	Acquisition parameters	Scene domain	Voxel size (mm^3)
S^1	11	PD	Normal	$TR/TE_{eff}=2500/18$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $40\leq\mathcal{N}\leq44$	$0.86\times0.86\times3$
S^2	11	T2	Normal	$TR/TE_{eff}=2500/90$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $40\leq\mathcal{N}\leq44$	$0.86\times0.86\times3$
S^3	11	PD	MS	$TR/TE_{eff}=2500/18$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $45\leq\mathcal{N}\leq60$	$0.86\times0.86\times3$
S^4	11	T2	MS	$TR/TE_{eff}=2500/90$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $45\leq\mathcal{N}\leq60$	$0.86\times0.86\times3$
S^5	11	T1E	MS	$TR/TE_{eff}=600/27$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $45\leq\mathcal{N}\leq60$	$0.86\times0.86\times3$
S^6	6	T1	Tumor	$TR/TE_{eff}=600/27$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $28\leq\mathcal{N}\leq32$	$0.86\times0.86\times5$
S^7	6	T1E	Tumor	$TR/TE_{eff}=600/27$, $FOV=22\ cm^2$	$256\times256\times\mathcal{N}$ $28\leq\mathcal{N}\leq32$	$0.86\times0.86\times5$

aligned with one another than the histograms in S_s^4. Further, the scenes in $S_{s_{g_B}}^4$ appear to have less residual non-standardness than the scenes in $S_{s_g}^4$. Finally, for the tumor studies (S^7), the histograms of the scenes in $S_{s_g}^7$ and $S_{s_{g_B}}^7$ are clearly better aligned with one another than for the scenes in S_s^7.

2. Binary scenes at fixed thresholds

The first three images in each row of Fig. 3 show a slice from each of three different T1E brain tumor studies. The rows from top to bottom correspond, respectively, to the scenes from sets S^7, S_s^7, $S_{s_g}^7$, and $S_{s_{g_B}}^7$. For each method, the scenes are displayed at a fixed window setting arrived at interactively for the first image in the row. The second set of three images in each row are displayed in binary form by using fixed thresholds to segment approximately the WM region of the brain and correspond exactly to the first three images in each row. The threshold interval was chosen to roughly segment the WM region in the first study by visual inspection, and the same interval was then used for the remaining two studies. In Row 1, it is well demonstrated that the same fixed threshold interval does not highlight the same tissue in different studies. In the third study, the threshold interval falls well below the brain tissue intensities. The problem in using histogram-based landmarks is well illustrated by the displays in the second row. The fixed interval threshold segments WM in the first two studies [Figs. 3(j) and 3(k)], but misses out on most of the WM in the third study [Fig. 3(l)]. The segmentation for the g- and g_B-scale-based methods is clearly more consistent and accurate compared to the histogram-based method [compare Fig. 3(l) with Figs. 3(r) and 3(x)]. This implies that, for the tumor studies, the numeric meaning of intensities on the standard scale is more consistent after g- and g_B-scale-based standardization than that in the histogram landmark-based method. Note that there is no significant visual difference among the results obtained from g- and g_B-scale-based methods for any of the patient studies.

B. Quantitative

In order to assess the effectiveness of the standardization methods objectively, we computed the WM intensity statistics over the population of 67 datasets. The method of fuzzy connectedness[18] was used to segment the different tissue regions. The tissue segmentations were subsequently corrected by an expert (neuroradiologist) where needed. (These datasets have been used in our earlier evaluation studies for segmentation and filtering.) 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 gray matter and CSF. 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 the reliable estimation of the figures of merit. To get voxels in the interior of the WM region, we use an erosion operation on the segmented binary scenes so that a layer of two voxels form the boundary is removed.

For the WM region in each scene so obtained, we computed the normalized mean intensity (NMI) in each scene before and after standardization by diving the mean intensity in the region by p_2-p_1. This was repeated for each set of standardization scenes, wherein normalization was done by dividing the mean in the WM region by s_2-s_1. The standard deviations of the mean of the NMI values within the WM region before and after different standardization transforms for the different sets of patient studies were computed and are listed in Table III. The results indicate that the intensities on the standard scale have more consistent tissue meaning than those for the original scale for all datasets. Further, while the results indicate no significant difference in σ_{NMI} values between the existing and generalized scale-based

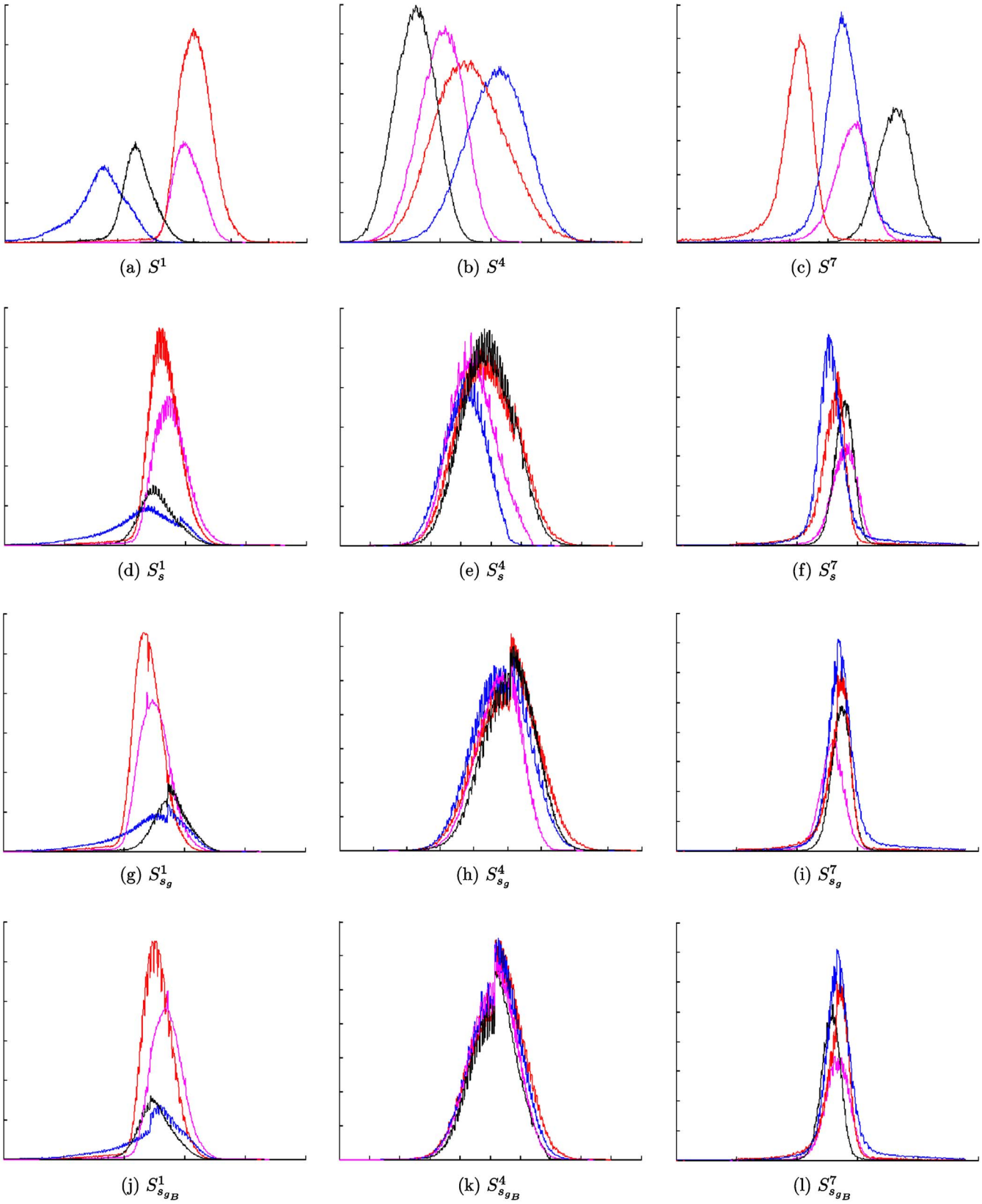

FIG. 2. Histograms of 4 different PD, T2, and T1E scenes, selected from datasets S^1, S^4, and S^7 corresponding to normal, MS, and brain tumor patient studies, before and after standardization by using the existing and new methods. For the normal patient studies (S^1), the histograms corresponding to the scenes produced by the existing method (S_s^1), the g-scale ($S_{s_g}^1$) and g_B-scale ($S_{s_{g_B}}^1$) appear similar in shape and alignment [(d),(g), (j)]. For the MS studies (S^4), the histograms of the scenes in $S_{s_g}^4$ and $S_{s_{g_B}}^4$ seem more closely aligned with one another than the histograms in S_s^4. Finally, for the tumor studies (S^7), the histograms of the scenes in $S_{s_g}^7$ and $S_{s_{g_B}}^7$ are clearly better aligned with one another than for the scenes in S_s^7.

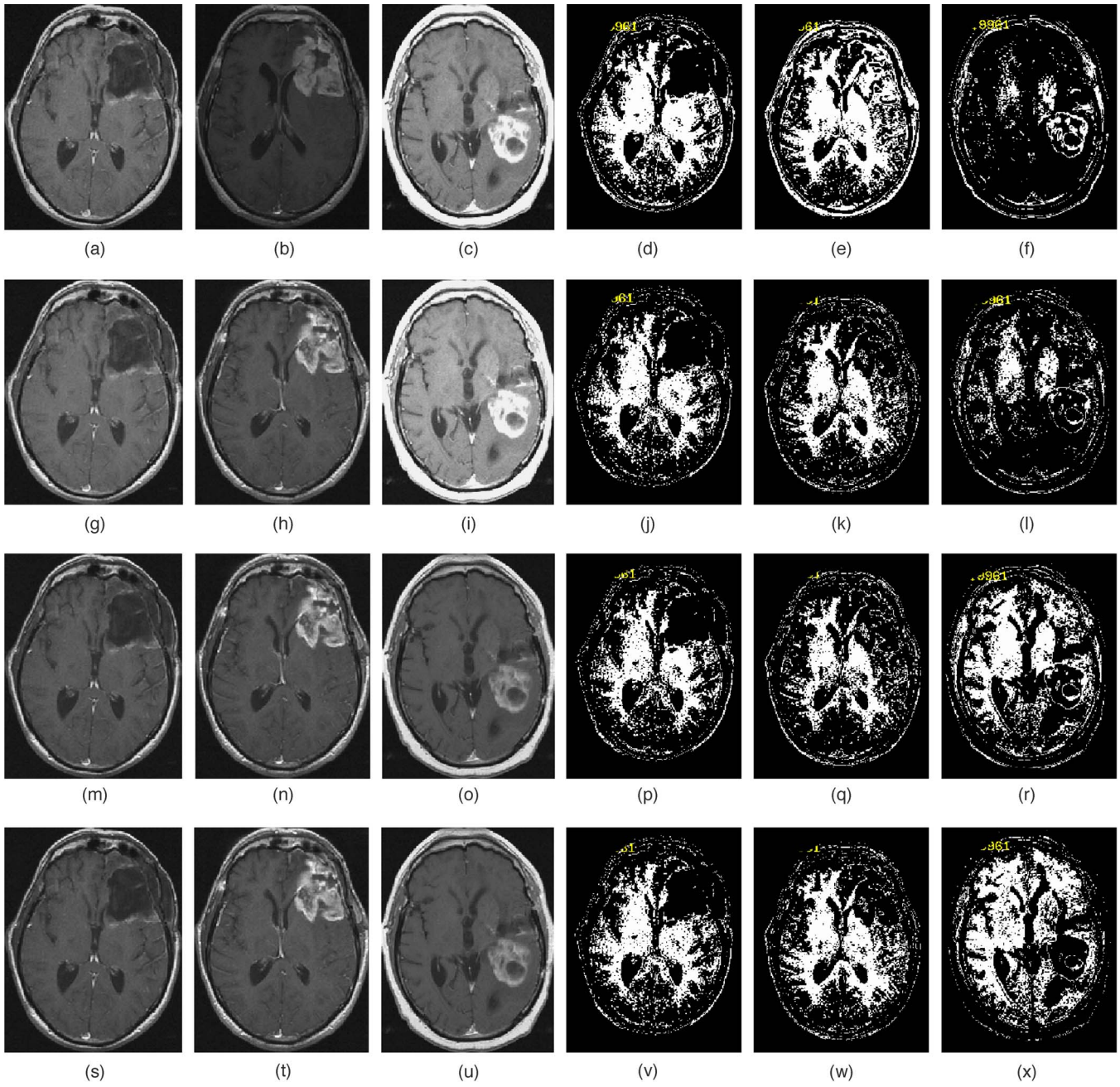

FIG. 3. Slices displayed at fixed window settings from three scenes from sets (a)–(c) S^7, (g)–(i) S^7_s, (m)–(o) $S^7_{s_g}$, and (s)–(u) $S^7_{s_{g_B}}$. In each row, slices from binary scenes resulting from fixed thresholding of gray scenes from the same row are displayed in the right half. Note that the WM segmentation achieved by the same fixed threshold interval in studies S^7 and S^7_s is not as good as the corresponding segmentations obtained in $S^7_{s_g}$ and $S^7_{s_{g_B}}$.

methods for the normal and MS data sets, a significant differences exists between them for the tumor studies. This implies that a substantially improved uniformity of tissue meaning for intensities is obtained for the g- and g_B-scale methods compared to the existing method for severely diseased or abnormal cases.

The σ_{NMI} values for the seven sets of studies were compared for each pair of conditions/methods, by using a paired t test under the null hypothesis that there is no difference in σ_{NMI} values between conditions/methods ($p \leq 0.05$); see Table IV. A statistically significant difference in σ_{NMI} values was observed for all sets after standardization compared to before (Table IV). Further, while no statistically significant difference was found between the existing histogram-based and g-scale-based methods, a significant difference was found between the existing and g_B-scale-based methods. The difference in σ_{NMI} values between the g- and g_B-scale-based methods was close to being statistically significant.

IV. CONCLUSIONS

We have described some of the problems with the original MRI scale standardization methods reported in Refs. 2 and 4,

TABLE III. σ_{NMI} values for the original scenes in the sets $S^1, S^2, S^3, S^4, S^5, S^6$, and S^7 and for the corresponding standardized scenes obtained by using the existing, g- and g_B-scale-based standardization methods.

Set	S^1	S^2	S^3	S^4	S^5	S^6	S^7
σ_{NMI}	0.0417	0.0851	0.0298	0.0196	0.0259	0.0536	0.0328
Set	S_s^1	S_s^2	S_s^3	S_s^4	S_s^5	S_s^6	S_s^7
σ_{NMI}	0.0181	0.0215	0.0164	0.0097	0.0132	0.0192	0.0161
Set	$S_{s_g}^1$	$S_{s_g}^2$	$S_{s_g}^3$	$S_{s_g}^4$	$S_{s_g}^5$	$S_{s_g}^6$	$S_{s_g}^7$
σ_{NMI}	0.0197	0.0160	0.0187	0.0085	0.0112	0.0119	0.0074
Set	$S_{s_{g_B}}^1$	$S_{s_{g_B}}^2$	$S_{s_{g_B}}^3$	$S_{s_{g_B}}^4$	$S_{s_{g_B}}^5$	$S_{s_{g_B}}^6$	$S_{s_{g_B}}^7$
σ_{NMI}	0.0128	0.0125	0.0195	0.0051	0.0087	0.0065	0.0107

and introduced two new scale-based methods that can help to overcome these problems. We have shown that landmarks derived from the largest g- and g_B-scale regions are more robust compared to landmarks derived from image intensity histograms, especially in the case of diseased or abnormal patient studies. While the scale-based methods require significantly more computational time than the histogram-based method, we note that most of this additional time was on account of training, which is done offline and only once for a given protocol and body region. The average times required for transforming the intensities for the g- and g_B-scale methods on a single dataset are 28 and 30 s, respectively, which, while being significantly longer than that for the histogram-landmark-based method (0.2 s), would be acceptable in a clinical scenario. Further, the higher levels of accuracy required in most quantitative image analysis applications would offset the additional computational expense of the scale-based standardization methods. Table V summarizes the performance of the three methods on the clinical studies considered in our quantitative evaluation. While marginally significant difference in performance was observed between the g- and g_B-scale methods, the overall g_B-scale outperformed the g-scale. Given that there is no significant difference in the efficiency of the two methods, the g_B-scale appears to be the standardization method of choice.

An assumption of the generalized scale-based standardization methods is that the largest scale region represents the same dominant normal tissue region in all studies pertaining to the same body region and imaging protocol. We have verified the validity of this assumption on hundreds of clinical and phantom datasets that we have evaluated in the context of our experiments with the generalized scale and its application.[1] However, in extreme circumstances, since the validity of the assumption underlying scale-based methods cannot be guaranteed, an interactive method may always be needed to select the scale region(s) corresponding to the same normal tissue as a backup and fully foolproof standardization strategy.

TABLE IV. p values for period t tests for comparing σ_{NMI} values for different pairs of conditions/methods.

Orig/hist	Orig/g-scale	Orig/g_B-scale	Hist/g-scale	Hist/g_B-scale	g-scale/g_B-scale
0.0069	0.0064	0.0055	0.0736	0.0158	0.0512

TABLE V. Best standardization methods for the seven clinical data sets $S^1 - S^7$.

	Normal		MS		Tumor		
Datasets	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Best results	g_B-scale	g-scale	hist	g_B-scale	g_B-scale	g_B-scale	g-scale

We believe that the robustness of the generalized scale-based methods compared to the existing methods is important for practical applications. By using the new standardized images in display, standard windows for the different tissues (not only for the main object itself) can be either automatically applied or manually selected (from a short list of available window settings), hence saving human interaction time on the per-case manual adjustments. Since the new methods work in the same way as the existing methods,[2,4] they are easy to implement, rapid in execution, and completely automatic like the original, and can be easily incorporated in a picture archiving and communication system as a DICOM value of interest lookup table. Hence the images can be automatically transformed or accompanied by the correct lookup table when they are downloaded to the viewing station.

Some possible future avenues are as follows. Other scale-based landmarks (such as corresponding to L_1, L_2, L_3, and L_4) can be used to perhaps further improve performance. Additional landmarks may provide the new methods improved anchorage capabilities, especially for scenes with bi- or multimodal histograms. Additional improvements may also be made by the use of polynomial functions to stretch the histogram segments, and the use of spline fitting techniques instead of segment-by-segment linear stretching. Another avenue may be the use of multiple tissue regions. In our current implementation in 3DVIEWNIX,[19] an interactive method permits multiple tissue regions (www.mipg.upenn.edu). However, obtaining landmark information from multiple tissue regions by using scale-based methods is more complicated mainly due to the difficulty involved in ascertaining to what tissue a given scale region corresponds. We believe that, from the perspective of quality of standardization, it is better to derive multiple landmarks from multiple tissue regions rather than from a single scale region.

ACKNOWLEDGMENT

The research reported here is supported by DHHS Grant No. NS37172.

[a]Address for correspondence: Jayaram K. Udupa, Medical Image Processing Group, 423 Guardian Drive, Blockley Hall, 4th Floor, Philadelphia, Pennsylvania 19104-6021. Phone: (215) 662-6781; fax: (215) 898-9145; electronic mail: jay@mipg.upenn.edu

[1]A. Madabhushi, J. Udupa, and A. Souza, "Generalized scale: Theory, algorithms, properties, and application to image inhomogeneity correction," Comput. Vis. Image Underst. **101**, 100–121 (2006).

[2]L. G. Nyul and J. K. Udupa, "On standardizing the MR image intensity

scale," Magn. Reson. Med. **42**, 1072–1081 (1999).
[3]R. E. Wendt, "Automatic adjustment of contrast and brightness of magnetic resonance images," J. Digit Imaging **x**, 95–97 (1994).
[4]L. G. Nyul, J. K. Udupa, and X. Zhang, "New Variants of a method of MRI Standardization," IEEE Trans. Med. Imaging **19**, 143–150 (2000).
[5]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. **12**, 715–721 (2000).
[6]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. Imaging **20**, 58–69 (2001).
[7]M. Styner, C. Brechbuhler, G. Szekely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI," IEEE Trans. Med. Imaging **19**, 153–165 (2000).
[8]T. Lindeberg, *Scale-Space Theory in Computer Vision* (Kluwer Academic, New York, 1993).
[9]P. Burt, "Fast filter transform for image processing," Comput. Vis. Graph. Image Process. **21**, 368–382 (1982).
[10]P. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **COM-31,4**, 532–540 (1983).
[11]J. L. Crowley and R. M. Stern, "Fast computation of the difference of low-pass transform," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 212–222 (1984).
[12]P. K. Saha, J. K. Udupa, and D. Odhner, "Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation, "Comput. Vis. Image Underst. **77**, 145–174 (2000).
[13]P. K. Saha, "Tensor scale: A local morphometric parameter with applications to computer vision and image processing," Comput. Vis. Image Underst. **99**, 384–413 (2005).
[14]S. M. Pizer, D. Eberly, and D. S. Fritsch, "Zoom-invariant vision of figural shape: the mathematics of core," Comput. Vis. Image Underst. **69**, 55–71 (1998).
[15]M. Tabb and N. Ahuja, "Multiscale image segmentation by integrated edge and region detection," IEEE Trans. Med. Imaging **6**, 642–655 (1997).
[16]P. K. Saha and J. K. Udupa, "Scale-based image filtering preserving boundary sharpness and fine structures," IEEE Trans. Med. Imaging **20**, 1140–1156 (2001).
[17]A. Madabhushi and J. K. Udupa, "Interplay of inhomogeneity correction and intensity standardization in MR image analysis,"IEEE Trans. Med. Imaging **24**, 561–576 (2005).
[18]J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," CVGIP: Graph. Models Image Process. **58**, 246–261 (1996).
[19]J. K. Udupa *et al.*, "3DVIEWNIX: an open, transportable, multidimensional, multi-modality, multi-parametric imaging software system," Proc. SPIE **2164**, 58–73 (1994).