New Methods of MR Image Intensity Standardization Via Generalized Scale

Anant Madabhushi[a] and Jayaram K. Udupa[a]

[a]Medical Image Processing Group, Department of Radiology
University of Pennsylvania, Philadelphia, PA 19104

ABSTRACT

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 the image intensity characteristics may occur. In an attempt to overcome this problem, in this paper, we present two new intensity standardization methods based on the concept of generalized scale. In[1] we introduced the concept of generalized scale (g-scale) to overcome the shape, topological, and anisotropic constraints imposed by other local morphometric scale models.[3,4] Roughly speaking, the g-scale of a voxel in a scene was defined as the largest set of voxels connected to the voxel that satisfy some homogeneity criterion. We subsequently formulated a variant of the generalized scale notion, referred to as generalized ball scale (g_B-scale), which, in addition to having the advantages of g-scale, also has superior noise resistance properties. These scale concepts are utilized in this paper to accurately determine principal tissue regions within MR images, and 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 3D MR images corresponding to four different protocols and to normal, Multiple Sclerosis (MS), and brain tumor patient studies. The generalized scale-based methods were found to be better than the existing methods, with a significant improvement observed for severely diseased and abnormal patient studies.

1. INTRODUCTION

Magnetic Resonance (MR) Imaging (MRI) is a non-invasive 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,[5–7] 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. A major difficulty in MR image analysis 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. 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 yet entered routine radiological practice.) More importantly, 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 which 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.[8] 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

Correspondence to Jayaram K. Udupa: E-mail: {anantm, jay}@mipg.upenn.edu, Telephone: (215) 662-6780, Fax: (215) 898-9145

Medical Imaging 2005: Image Processing, edited by J. Michael Fitzpatrick, Joseph M. Reinhardt, Proc. of SPIE Vol. 5747 (SPIE, Bellingham, WA, 2005)
1605-7422/05/$15 · doi: 10.1117/12.595925

that address the problem of standardization of image intensities explicitly are.[9–11] 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. 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.[12] describe a methodology to perform three-dimensional multi-modal brain image 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,[13] 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,[9] Nyul and Udupa presented a method which transforms images non-linearly 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[9–11] 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,[9] the mode on the histogram was used as the landmark for transforming the scene intensities. In later work,[10] it was felt that the mode was not a suitably robust landmark, since it could correspond to one particular tissue in one image and to another tissue in another image. Consequently a variant of the original standardization procedure was introduced in[10] which replaced the mode with the median and other quartile locations on the histogram. These quartile-based standardization methods were shown to be more robust compared to the original mode-based method.

In cases of diseases, and 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 methods that uses a locally adaptive concept of 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 found useful in almost all image processing and analysis tasks including segmentation, filtering, interpolation, registration, visualization, and quantitative analysis. Broadly speaking, scale concepts can be divided into three categories: (1) multi-scale 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.[14] 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 multi-scale representation of the image.[15] Recently, there has been considerable interest in developing *locally adaptive scales*,[3,4] 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,[1] we proposed a generalized scale (abbreviated as *g-scale* from hereon) model which is adaptive like other local morphometric models, and which possesses the global spirit of multi-scale representations. We also subsequently developed a variant of the generalized scale notion that we refer to as the generalized ball scale (abbreviated as g_B*-scale* from hereon), which in addition to having the advantages of the g-scale model also has superior noise resistance properties.[2] To overcome the sensitivity of the existing standardization methods to the landmarks on the histogram, we present two new intensity standardization methods in this paper based on the g- and g_B-scale models. These new methods exploit the ability of 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 non-linear mapping of intensities. We contend that this makes the scale-based methods more robust compared to the existing methods, especially for the cases of patient studies with abnormalities.

The rest of this paper is organized as follows. In Section 2, we describe the various steps involved in the new intensity standardization methodologies. In Section 3, we present the results of qualitatively and quantitatively

comparing the new methods with the existing methods. Finally, we present our concluding remarks in Section 4.

2. METHODOLOGY

2.1. Notation

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 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 $D \in \mathcal{D}$.

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 $u \in C$. 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.

A local scale model, called *ball scale*[3] 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 pre-defined 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 *b*-scale, we proposed a generalized scale model .[1,2] 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 subset of C containing c such that all voxels in $G(c)$ satisfied a pre-determined homogeneity criterion. Subsequently we defined a variant of the generalized scale, called *generalized ball-scale* or *g_B-scale*.[2] For any voxel c in a scene $\mathcal{C}$, 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 *g*-scale is estimated by the addition of individual voxels into a *g*-scale set based on the homogeneity criterion, g_B-scale is determined by the inclusion of voxels which 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 *g*-scale. The *g*-scale corresponds essentially to a fuzzy connected component (based on homogeneity) of $\mathcal{C}$, and hence, it is computed via dynamic programming. 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 *g*- and g_B-scales associated with any voxel c in a scene are denoted by $G(c)$ and $G_B(c)$, respectively. The set of all *g*- and g_B-scales associated with $\mathcal{C}$ are denoted by $\mathcal{G}(\mathcal{C})$ and $\mathcal{G}_\mathcal{B}(\mathcal{C})$, 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. Since this paper pertains to an application of *g*- and g_B-scales, we will not delve into their theoretical and algorithmic aspects in this paper; see[2] for details. We will use sub-scripts s, s_g and s_{g_B} to denote the scenes and the sets of scenes resulting from applying the histogram landmark-based ,[9] *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.

2.2. Motivation for generalized scale-based standardization

As previously mentioned, the original paper on intensity standardization[9] used the mode of the histogram as a landmark. This was subsequently replaced by the median and other quartile locations on the scene histogram,[10] which were shown to be more robust than the original mode-based method.

In the cases of diseases, and when a disease is so pervasive that normal tissue 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. This is demonstrated in Figure 1 which shows slices from two T1-weighted MRI scenes of two different patients taken after gadolinium administration, one normal

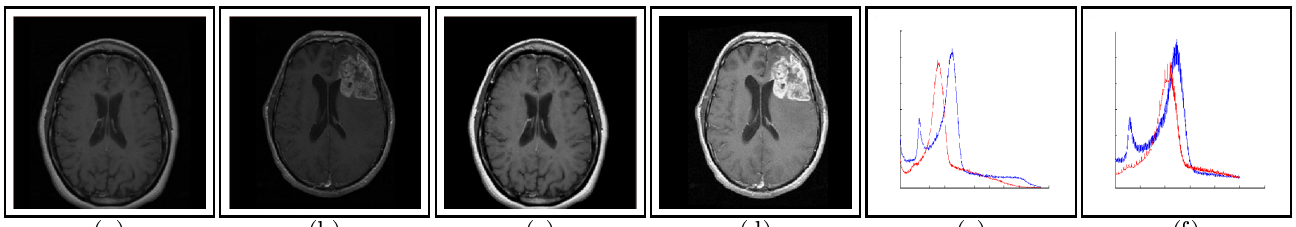

Figure 1. Slices of (a) a normal T1-weighted scene of the brain, (b) a T1-weighted scene of the brain with a tumor, both taken after gadolinium administration, and (c), (d) the scenes in (a), (b) after intensity standardization by using the histogram landmark-based method in,[10] and (f), (g) the intensity histograms of the scenes in (a) and (b), and (c), (d), respectively.

and the other with a brain tumor. The intensity histogram of the foreground (body region only) MRI scene for the normal study is unimodal, while the histogram of the tumor scene contains two peaks. (All image processing operations are carried out in three dimensions unless mentioned otherwise.) Figures 1(a)-(f) show the same scenes and their histograms before and after standardization by using the median method of.[10] The residual non-standardness after standardization can be readily seen in Figure 1(f). A standardization method that relies on landmarks deliberately estimated from the same tissue regions in the scene will be more stable than a method that relies on gross landmarks derived from the histogram, especially for the case of abnormal patient studies with severe disease. Figure 2 shows the tumor scene example of Figure 1(b) and the largest g- and g_B-scale regions obtained for this scene (Figures 2(b), (c)). The estimated largest g- and g_B-scale regions lie almost entirely within the dominant tissue region (white matter (WM) in this case). Hence a method that uses landmarks estimated from the dominant tissue region estimated from the scene either automatically (scale-based methods) or interactively (by using human help either to paint a region of interest or to select a scale region) will be more robust to pathological variations. Note in Figures 2(b) and (c) that the g- and g_B-scale regions are not meant to represent accurate segmentations of the largest tissue region (in this case WM). The idea is to use generalized scale to accurately estimate samples of the same (largest) tissue region.

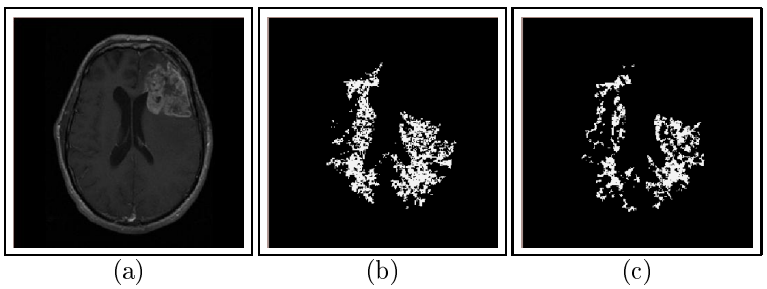

Figure 2. Slices of a (a) T1-weighted (with gadolinium) MR scene of the brain containing a tumor, and the corresponding largest (b) g- and (c) g_B-scale regions.

2.3. Landmarks via scale regions

The basic idea of the standardization methods described in[9,10] is to identify a set of landmarks on the gray scale of the scenes via its 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[9,10] is in the manner in which the landmarks are identified. Subsequently, the mapping is done in exactly the same way as in.[9,10]

As described in,[9,10] 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 intra-patient/scanner variations. With this in mind, let pc_1 and pc_2 denote the minimum and maximum cut-off 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,[10] 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[9,10] the mode, median, deciles and quartiles were all used in the following landmark configurations for the histogram based standardization method:

$$L_1 = \{pc_1, \mu, pc_2\}, L_2 = \{pc_1, \mu_{50}, pc_2\}, L_3 = \{pc_1, \mu_{25}, \mu_{50}, \mu_{75}, pc_2\}, \text{and } L_4 = \{pc_1, \mu_{10}, \mu_{20}, \mu_{30}, ..., \mu_{90}, pc_2\}, \tag{2.1}$$

where μ_p for $p \in \{10, 20, 25, 30, .., 75, ..., 90\}$ represents the intensity value corresponding to the p^{th} 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,[10] 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 $L_{2g} = \{pc_1, \mu_{50g}, pc_2\}$ and $L_{2g_B} = \{pc_1, \mu_{50g_B}, pc_2\}$ be the configurations similar to L_2 but used in the g- and g_B-scale methods where μ_{50g} and μ_{50g_B} denote the median value determined from the respective methods.

2.4. Generalized scale based standardization

The method comprises of two separate steps: training, transformation. In the first step (training) of the generalized scale-based methods, 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 learnt 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 learnt 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[9,10] 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,[1,2] and the largest scale region is determined. The median intensity μ_{50g}/μ_{50g_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_{50g}/μ_{50g_B} is transformed to μ'_{50g}/μ'_{50g_B} on the standard scale for each scene $\mathcal{C} \in T^{PD}$.

(v) The rounded median intensity $\mu_{s_g}/\mu_{s_{g_B}}$ on the standard scale is computed from the average of μ'_{50g}/μ'_{50g_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_{50g}/μ_{50g_B} of the scale region is then computed.

(ii) A piecewise linear mapping is then determined as described in[9,10] 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_{50g}/μ_{50g_B} with $\mu_{s_g}/\mu_{s_{g_B}}$. Figure 3 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 Figure 3. 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 $\mathcal{C}$

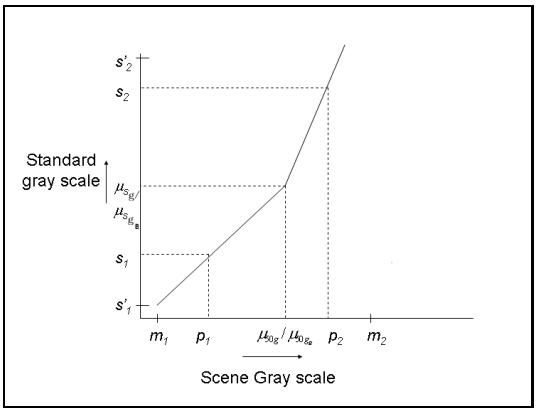

Figure 3. g-scale-based standardization mapping with the various parameters indicated.

and denote it by $\psi_{\mathcal{C}_g}/\psi_{\mathcal{C}_{g_B}}$. The expression for $\psi_{\mathcal{C}_g}$ for the g-scale-based method (from Figure 3) is as follows,

$$\psi_{\mathcal{C}_g}(x) = \begin{cases} \lceil \mu_{s_g} + (x - \mu_{50g}) \frac{s_1 - \mu_{s_g}}{p_1 - \mu_{50g}} \rceil, & \text{if } m_1 \leq x \leq \mu_{50g} \\ \lceil \mu_{s_g} + (x - \mu_{50g}) \frac{s_2 - \mu_{s_{g_B}}}{p_2 - \mu_{50g}} \rceil, & \text{if } \mu_{50g} \leq x \leq m_2, \end{cases} \tag{2.3}$$

where $\lceil \cdot \rceil$ denotes the *ceiling* operation. (This operator converts any real number y to the closest integer Y such that $Y \geq y$. Instead, a *floor* operator ($Y \leq y$) may also be used.) Note that $s_1' = \psi_{\mathcal{C}_g}(m_1)$, and $s_2' = \psi_{\mathcal{C}_g}(m_2)$. 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 $c \in C$, $f_{s_g}(c) = \psi_{\mathcal{C}_g}(f(c))$ wherein μ_{50g} is used in (2.3). $\mathcal{C}_{s_{g_B}}$ is similarly defined by replacing μ_{50g} by μ_{50g_B} and μ_{s_g} by $\mu_{s_{g_B}}$ in Equation (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.

2.5. Algorithms

We now present the algorithms for the training and transformation steps. Both the algorithms are straightforward and require no special data structures or optimization in implementation.

Algorithm 1: Training
Input: A set of scenes $\mathcal{C}_j$, j=1,2,....,N, that is a subset of $\mathcal{S}^{PD}$ for some fixed protocol $P \in \mathcal{P}$, and for a given body region $D \in \mathcal{D}$, the histogram parameters pc_1 and pc_2, and s_1, s_2.
Output: $\mu_{s_g}/\mu_{s_{g_B}}$.
begin
1. *for* $j = 1$ to N *do*

2. determine histogram $\mathcal{H}_j$ of $\mathcal{C}_j$ and intensities p_{1j}, p_{2j} in $\mathcal{C}_j$ corresponding to percentiles pc_1 and pc_2;
3. determine median intensities μ_{50gj}/μ_{50g_Bj}, corresponding to the largest g-/g_B-scale region in $\mathcal{C}_j$;
4. map $[p_{1j}, p_{2j}]$ of $\mathcal{H}_j$ onto $[s_1, s_2]$ linearly;
5. find the new mapped landmark locations $\mu'_{50gj}/\mu'_{50g_Bj}$;
6. *endfor*;
7. calculate the rounded mean $\mu_{s_g}/\mu_{s_{g_B}}$ over all $\mu'_{50gj}/\mu'_{50g_Bj}$ for $j = 1, 2, .., N$;

end

Algorithm 2: Transformation
Input: A scene $\mathcal{C} \in \mathcal{S}^{PD}$, pc_1, pc_2, s_1, s_2, $\mu_{s_g}/\mu_{s_{g_B}}$.
Output: The transformed scene $\mathcal{C}_{s_g}/\mathcal{C}_{s_{g_B}}$.
begin

1. determine histogram $\mathcal{H}=(\mathbf{G}, h)$ of $\mathcal{C}$ and intensities p_1 and p_2 in $\mathcal{C}$ corresponding to pc_1 and pc_2;
2. determine the median intensity μ'_{50g}/μ'_{50g_B} corresponding to the largest g-/g_B-scale region in $\mathcal{C}$.
3. map sections of the gray scale $\mathbf{G}$ of $\mathcal{H}$ linearly according to Figure 3 to the standard scale $\mathbf{G_s}$;
4. map intensity value of every voxel $c \in \mathcal{C}$ according to mapping function $\psi_{\mathcal{C}_g}/\psi_{\mathcal{C}_{g_B}}$ to get $\mathcal{C}_{s_g}/\mathcal{C}_{s_{g_B}}$;

end

2.6. Choosing the Standardization Parameters

Although, once the training step is done, the corresponding transformation step is fully determined, there are several possibilities to tailor the standardizer to the specific needs of an application. For example, s_1 should not be zero if the values below p_1 need to be distinguished from nothing (i.e., value zero).

In,[9,10] a theory was delineated to show that the choice of landmark configurations guaranteed the correct behavior (i.e., preserving the intensity relations between voxels) of the standardizer for any scene $\mathcal{C}$. The difference between the landmark configuration we use (L_{2g}/L_{2g_B}) and that employed in[9] (L_2) is that, while we use the median intensity of the largest scale region, in,[9] the intensity at the median location of the scene histogram of the foreground was used. Hence the theory in[9,10] readily generalizes to our methods. Consequently we limit our discussion to only reproducing the main essence of the three theorems stated in.[9,10]

Theorems 1 and 2 of [9,10] state the conditions under which it is guaranteed that no two distinct intensities in $\mathcal{C}$ are merged into a single intensity in the standardized scene $\mathcal{C}_s/\mathcal{C}_{s_g}/\mathcal{C}_{s_{g_B}}$. Thus, if standardization is done respecting these conditions, then there is no loss of information, and the original scene can be obtained by inverting the standardizer. Theorem 1 gives a guide to selecting the proper values of s_1 and s_2 that cause no intensity loss. Theorem 2 requires that s_2 - s_1 should be greater than a certain number derived from $\mathcal{C}_s/\mathcal{C}_{s_g}/\mathcal{C}_{s_{g_B}}$. Based on observing several scenes in $\mathcal{S}^{PD}$, we set s_2-s_1 to a number sufficiently larger than the number derived from the scenes we observed in $\mathcal{S}^{PD}$. As per the recommendations outlined in,[9,10] we set s_1=0 and s_2=4095. Further, the lower and upper percentiles were set to pc_1=0, pc_2=99.8. In,[9] it was found that beyond pc_2=99.8, the variations in the intensities are not systematic but random. Table 1 shows the different parameter settings that were used for the three methods (histogram landmark-based,[9,10] g-scale-based, and g_B-scale-based methods).

Theorem 3 states that, once the conditions of Theorem 1 have been satisfied, the order of the input intensities is maintained in the output scenes. That is, the actual order of brightness of tissue regions in $\mathcal{C}$ is maintained in the scene $\mathcal{C}_s/\mathcal{C}_{s_g}/\mathcal{C}_{s_{g_B}}$, although their relative contrast may change. In a PD-weighted scene of the brain, for example, the known brightness relationship gray matter > white matter > CSF is maintained in the transformed scene.

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	$\mu_{50_{g_B}}$

Table 1. Parameter configurations used for the different standardization methods.

3. RESULTS AND EVALUATION

We evaluated the three methods listed in Table 1 both qualitatively and quantitatively by using a set of 67 clinical MR image data sets corresponding to four different protocols and acquired from normal subjects, subjects with MS, and from brain tumor patients. Table 2 gives a description of these data sets. For all sets, the body region D considered was the head.

For qualitative evaluation, we considered 1) plotting the standardized histograms, 2) displaying the scenes before and after standardization at fixed gray level window settings, and 3) displaying the binary versions of standardized scenes obtained after thresholding at fixed levels. 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. These studies are described in greater detail below.

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=22cm^2	256×256× $\mathcal{N}$ 40≤ $\mathcal{N}$ ≤ 44	0.86×0.86 ×3
S^2	11	T2	Normal	TR/TE_{eff}=2500/90, FOV=22cm^2	256×256× $\mathcal{N}$ 40≤ $\mathcal{N}$ ≤ 44	0.86×0.86 ×3
S^3	11	PD	MS	TR/TE_{eff}=2500/18, FOV=22cm^2	256×256× $\mathcal{N}$ 45≤ $\mathcal{N}$ ≤ 60	0.86×0.86 ×3
S^4	11	T2	MS	TR/TE_{eff}=2500/90, FOV=22cm^2	256×256× $\mathcal{N}$ 45≤ $\mathcal{N}$ ≤ 60	0.86×0.86×3 ×3
S^5	11	T1E	MS	TR/TE_{eff}=600/27, FOV=22cm^2	256×256× $\mathcal{N}$ 45≤ $\mathcal{N}$ ≤ 60	0.86×0.86 ×3
S^6	6	T1	Tumor	TR/TE_{eff}=600/27, FOV=22cm^2	256×256× $\mathcal{N}$ 28≤ $\mathcal{N}$ ≤32	0.86×0.86 ×5
S^7	6	T1E	Tumor	TR/TE_{eff}=600/27, FOV=22cm^2	256×256× $\mathcal{N}$ 28≤ $\mathcal{N}$ ≤32	0.86×0.86 ×5

Table 2. Description of the data sets used in evaluation.

3.1. Qualitative

We conducted qualitative comparisons for the following MRI protocols: PD-, T2-, T1-, and T1-weighted with gadolinium enhancement (T1E). Our hypothesis was that the performance of the new scale-based standardization methods would be comparable to that of the existing methods[9,10] on normal data sets and on data sets 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.

3.1.1. Histograms

Histograms of PD, T2, and T1E-weighted scenes corresponding in turn to normal, MS, and brain tumor patients before and after standardization by using the existing and new methods are shown in Figure 4. To avoid clutter we have shown only 4 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^1_s), the g-scale ($S^1_{s_g}$) and g_B-scale ($S^1_{s_{g_B}}$) appear similar in shape and alignment (Figures 4(d), (g), (j)). For the MS studies (S^4), the histograms of the scenes in $S^4_{s_g}$ and $S^4_{s_{g_B}}$ seem more closely aligned with one another than the histograms in S^4_s. Further, the scenes in $S^4_{s_{g_B}}$ appear to have less residual non-standardness than the scenes in $S^4_{s_g}$. Finally, for the tumor studies (S^7), the histograms of the scenes in $S^7_{s_g}$ and $S^7_{s_{g_B}}$ are clearly better aligned with one another than the scenes in S^7_s.

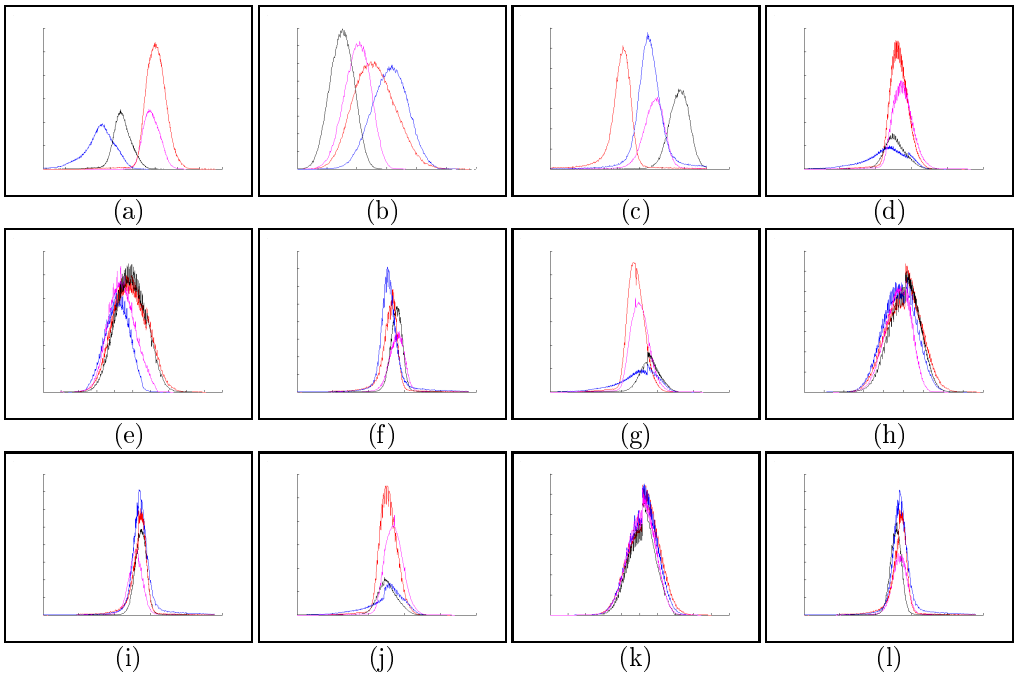

Figure 4. Histograms of 4 different PD-, T2-, and T1E-weighted scenes corresponding to normal, MS, and brain tumor patient studies: (a) S^1, (b) S^4, (c) S^7, (d) S^1_s, (e) S^4_s, (f) S^7_s, (g) $S^1_{s_g}$, (h) $S^4_{s_g}$, (i) $S^7_{s_g}$, (j) $S^1_{s_{g_B}}$, (k) $S^4_{s_{g_B}}$, and (i) $S^7_{s_{g_B}}$.

3.1.2. Display at a fixed window setting

Images in each row of Figure 5 show a slice from each of three different T1E-weighted brain tumor MR studies. The rows from top to bottom correspond respectively to the scenes from sets S^7, S^7_s, $S^7_{s_g}$, and $S^7_{s_{g_B}}$. For each method, the scenes are displayed at a fixed window setting arrived at interactively for the first image in the row. Figure 5 demonstrates that the structures are better portrayed and the contrast is more similar for the standardized scenes in S^7_s, $S^7_{s_g}$, and $S^7_{s_{g_B}}$ (Rows 2, 3, 4) than for the original data sets in S^7 (Row 1). Since the intensity ranges are quite different for the three tumor studies (Figures 5(a)-(c)), the display for the histogram-based standardization method is inadequate for the third study (Figure 5(f)). Note also the striking similarity of the displays among the g- and g_B-scale-based standardization methods (Figures 5(g), (h), (i) and 5(j), (k), (l)). No visual difference was observed among the scenes standardized by the three methods for the normal and the MS studies.

3.1.3. Binary Scenes at Fixed Thresholds

Figure 6 demonstrates in another way the fact that the intensity values have no consistent meaning in the original scenes but they do have after standardization. Images are displayed in binary form by using fixed

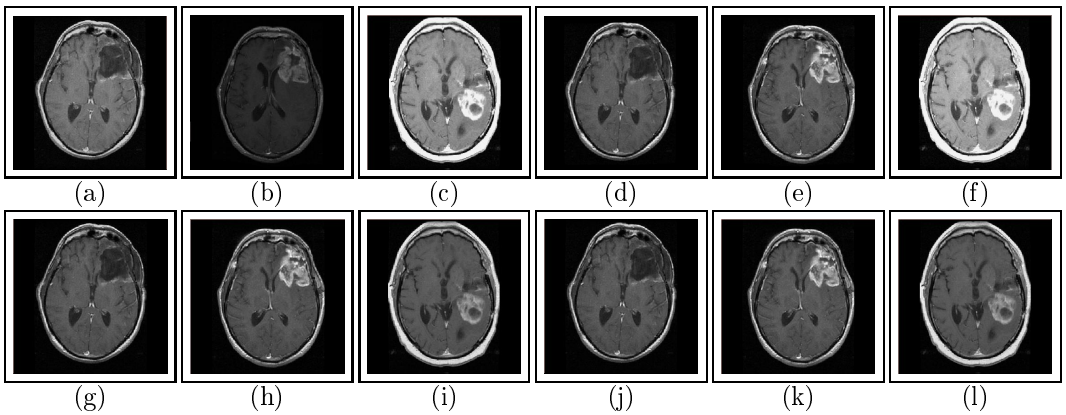

Figure 5. Slices displayed at fixed window settings from three scenes from sets (a)-(c) S^7, (d)-(f) S^7_s, (g)-(i) $S^7_{s_g}$, and (j)-(l) $S^7_{s_{g_B}}$.

thresholds to segment approximately the WM region of the brain. The displayed scenes correspond exactly to the scenes in Figure 5. For 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 used for the remaining two studies. In Row 1, it is well demonstrated that the same fixed threshold interval does not segment 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 (Figures 6(d), (e)), but misses out on most of the WM in the third study (Figure 6(f)). The segmentation for the g- and g_B-scale-based methods is clearly more consistent and accurate compared to the histogram-based method (compare Figure 6(f) with Figures 6(i), (l)). 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.

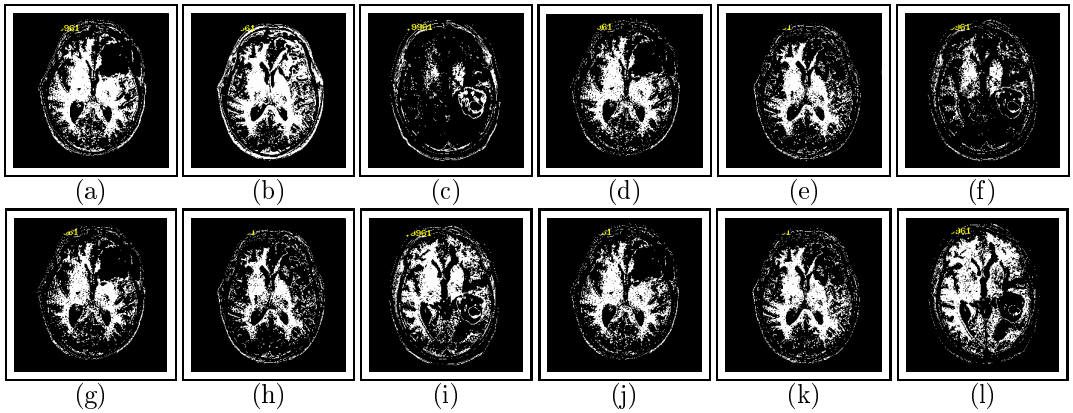

Figure 6. Slices from binary scenes resulting from fixed thresholding of scenes from the sets (a)-(c) S^7, (d)-(f) S^7_s, (g)-(i) $S^7_{s_g}$, and (j)-(l) $S^7_{s_{g_B}}$.

3.2. Quantitative

In order to assess the effectiveness of the standardization methods objectively, we computed the WM intensity statistics over the population of 67 data sets. The tissue segmentations obtained via the method of fuzzy connectedness,[3] were subsequently corrected by an expert (neuroradiologist) where needed. 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. For the WM region in each scene so obtained, we computed the normalized mean intensity (NMI) in each scene before and after standardization by dividing the mean intensity in the region by p_2-p_1. This was repeated for each set of standardized 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 (σ_{NMI}) before and after different standardization transforms for the different sets of patient studies were computed and are listed in Table 3. The results indicate that the intensities on the standard scale have more consistent tissue meaning than those for the original scale for all data sets. Further, while the results indicate no significant difference in σ_{NMI} values between the existing and generalized scale-based methods for the normal and MS data sets, a significant difference 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.

Set	S^1	S^2	S^3	S^4	S^5	S^6	S^7
σ_{NMI}	0.041	0.085	0.029	0.019	0.025	0.053	0.032
Set	S^1_s	S^2_s	S^3_s	S^4_s	S^5_s	S^6_s	S^7_s
σ_{NMI}	0.018	0.021	0.016	0.009	0.013	0.019	0.016
Set	$S^1_{s_g}$	$S^2_{s_g}$	$S^3_{s_g}$	$S^4_{s_g}$	$S^5_{s_g}$	$S^6_{s_g}$	$S^7_{s_g}$
σ_{NMI}	0.019	0.016	0.018	0.008	0.011	0.011	0.007
Set	$S^1_{s_{g_B}}$	$S^2_{s_{g_B}}$	$S^3_{s_{g_B}}$	$S^4_{s_{g_B}}$	$S^5_{s_{g_B}}$	$S^6_{s_{g_B}}$	$S^7_{s_{g_B}}$
σ_{NMI}	0.0128	0.0125	0.0195	0.0051	0.0087	0.0065	0.0107

Table 3. 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.

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 4. A statistically significant difference in σ_{NMI} values was observed for all sets after standardization compared to before (Table 4). 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.

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 4. p-values for paired t-tests for comparing σ_{NMI} values for different pairs of conditions/methods.

4. CONCLUSIONS

We have described some of the problems with the original MRI scale standardization methods reported in,[9,10] 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 off-line and only once for a given protocol and body region. The average time required for transforming the intensities for the g- and g_B-scale methods on a single data set (size: 256×256×60) are typically 27.89s and 29.95s, respectively, which, while being significantly longer than that for the histogram-landmark-based method (0.22s), would be acceptable in a clinical scenario (All operations were carried out on a 2.3 GHz, Pentium IV PC.). Further, the higher levels of accuracy required in most quantitative image analysis applications would off-set the additional computational expense of the scale-based standardization methods. Table 5 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, overall g_B-scale outperformed g-scale. Given that there is no significant difference in the efficiency of the two methods, g_B-scale appears to be the standardization method of choice.

	Normal		MS			Tumor	
Data sets	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

Table 5. Best methods for the 7 data sets S_1-S_7.

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 data sets that we have evaluated in the context of our experiments with generalized scale and its application.[1,2] However, in extreme circumstances, since the validity of the assumption underlying scale-based methods cannot be guaranteed, an interactive method may always be needed as a backup and fully fool-proof standardization strategy.

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,[9,10] 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.

REFERENCES

1. A. Madabhushi, J. Udupa, A. Souza, "Generalized Scale: Theory, Algorithms, and Application to Inhomogeneity Correction", *Proc. of SPIE: Med. Imag.*, vol. 2, pp. 765-76, 2004.
2. A. Madabhushi, J. Udupa, A. Souza, "Generalized Scale: Theory, Algorithms, Properties, and Application to Image Inhomogeneity Correction", Submitted to *CVIU*.
3. P. Saha, J. K. Udupa, and D. Odhner "Scale-based Fuzzy Connected Image Segmentation: Theory, Algorithms, and Validation", *CVIU*, vol. 77, pp. 145-174, 2000.
4. P. K. Saha, "Novel Theory and Methods for Tensor Scale: A Local Morphometric Parameter", *Proc. of SPIE: Med. Imag.*, vol. 2, pp. 743-53, 2003.
5. J. C. Bezdek, L. O. Hall, and L. P. Clarke, "Review of MR image segmentation techniques using pattern recognition", *Med. Physics*, vol. 20, no. 4, p. 1033-1048, 1993.
6. D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation", *Annual Review of Biomedical Eng.*, vol. 2, pp. 315-338, 2000.
7. J. Udupa, L. Wei, S. Samarasekera, Y. Miki, M. V. Buchem, and R. Grossman, "Multiple sclerosis lesion quantification using fuzzy-connectedness principles", *IEEE Trans. on Med. Imag.*, vol. 16, pp. 598-609, 1997.
8. R. E. Wendt, "Automatic adjustment of contrast and brightness of magnetic resonance images", *Journal of Dig. Imag.*, pp. 95-97, 1994.
9. L. G. Nyul, J. K. Udupa, "On standardizing the MR Image Intensity Scale", *Magnetic Reson. in Medicine*, vol. 42, pp. 1072-1081, 1999.
10. L. G. Nyul, J. K. Udupa and X. Zhang, "New Variants of a method of MRI Standardization", *IEEE Trans. on Med. Imag.*, vol. 19, pp. 143-150, 2000.
11. Y. Ge, J. Udupa, L. G. Nyul, L. Wei and R. Grossman, "Numerical Tissue Characterization in MS via Standardization of the MR Image Intensity Scale", *Jrnl. of Mag. Reson. Imag.*, vol. 12, pp. 715-721, 2000.
12. A. Guimond, A. Roche, N. Ayache, J. Meunier, "Three-Dimensional Multi-modal Brain Warping Using the Demons Algorithm and Adaptive Intensity Corrections", *IEEE Trans. on Med. Imag.*, vol. 20, pp. 58-69, 2001.
13. M. Styner, C. Brechbuhler, G. Szekely, G. Gerig, "Parametric Estimate of Intensity Inhomogeneities Applied to MRI", *IEEE Trans. on Med. Imag.*, vol. 19, pp. 153-165, 2000.
14. Tony Lindeberg, "Scale-Space Theory in Computer Vision", *Kluwer Academic Publishers*, 1993.
15. P. Burt, "Fast Filter Transform for Image Processing", *Computer Graphics and Image Processing*, pp. 20-51, 1981.
16. P. Saha and J. Udupa, "Scale-based image filtering preserving boundary sharpness and fine structures", *IEEE Trans. on Med. Imag.*, vol. 20, pp. 1140-1156, 2001.