Correspondence

New Variants of a Method of MRI Scale Standardization

László G. Nyúl, Jayaram K. Udupa*, and Xuan Zhang

***Abstract*—One of the major drawbacks of magnetic resonance imaging (MRI) has been the lack of a standard and quantifiable interpretation of image intensities. Unlike in other modalities, such as X-ray computerized tomography, MR images taken for the same patient on the same scanner at different times may appear different from each other due to a variety of scanner-dependent variations and, therefore, the absolute intensity values do not have a fixed meaning. We have devised a two-step method wherein all images (independent of patients and the specific brand of the MR scanner used) can be transformed in such a way that for the same protocol and body region, in the transformed images similar intensities will have similar tissue meaning. Standardized images can be displayed with fixed windows without the need of per-case adjustment. More importantly, extraction of quantitative information about healthy organs or about abnormalities can be considerably simplified. This paper introduces and compares new variants of this standardizing method that can help to overcome some of the problems with the original method.**

***Index Terms*—Image display, image normalization, image processing, magnetic resonance imaging (MRI).**

I. Introduction

Magnetic resonance imaging (MRI) has revolutionized radiological imaging of the internal structures of the human body. It has the advantage of being noninvasive with no known health hazards. A variety of MRI protocols are currently available, with and without the use of contrast agents, such as T1, T1 with a contrast agent, T2 and proton density (Pd) with spin-echo (SE) or fast spin-echo (FSE) sequences, magnetization transfer (MT), FLAIR, SPGR, and GRASS. These protocols allow the setting up of different contrasts among the different tissues within the same organ system.

Ironically, this richness of acquisition schemes comes with a major difficulty. The image intensities in MRI do not have a fixed meaning, not even within the same protocol for the same body region obtained on the same scanner for the same patient. This implies that MR images cannot be displayed at preset windows; one may have to adjust the window settings per case. The lack of a meaning for intensities also poses problems in image segmentation [1], [2] and quantification [3], [4].

Most visualization and analysis methods have parameters. The exceptions are perhaps manual methods wherein the human knowledge can be thought of as representing the parameters. However, the results of segmentation by two physicians are likely to differ because of the differences in their training. Setting values for the parameters for the nonmanual methods becomes more difficult without the same protocol-specific intensity meaning. What we need is that the resulting images be close when the protocols are the same or close to each other.

Manuscript received July 20, 1999; revised October 28, 1999. This work was supported in part by the NIH under Grant NS 37172 and in part by the Department of Army under Grant DAMD 179717271. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Vannier. *Asterisk indicates corresponding author.*

L. G. Nyúl, *J. K. Udupa, and X. Zhang are with the Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104-6021 USA (e-mail: jay@mipg.upenn.edu).

Publisher Item Identifier S 0278-0062(00)02301-6.

Attempts have been made in the past to calibrate MR signal characteristics at the time of acquisition, using phantoms [5], [6]. Postprocessing techniques that are applied to the image data that do not have any special acquisition requirements are clearly more attractive. Such methods would not only make acquiring images simpler, but also make it possible to process the already acquired data. There does not seem to have been any serious attempt to address this problem in the past.

The method described in [7]–[9] offers a simple way of transforming the images so that there is a significant gain in similarity of the resulting images. It is a two-step process consisting of a training step (executed only once for each protocol and body region) and a transformation step (executed on each given image). This new transformation results in a standard scale for each protocol and body region. Intensities in the transformed images have consistent tissue meanings and standard window settings can be determined for different tissues just as in CT.

The main idea underlying the methods is to deform the image histograms so that they match a mean histogram determined through training. The actual matching is based on certain landmarks identified on the histograms. In the original method [9], we utilized a low- and a high-percentile point together with the mode corresponding to the foreground in the image as the landmarks. Although, for display, this achieves a remarkably consistent window level and width setting, for image segmentation the mode-based matching is often too sensitive to the actual location of the mode, which is often quite variable. In this paper we utilize other landmarks, the quartiles and deciles, in particular the median corresponding to the foreground in the image, in addition to the low and high percentiles, as described in Sections II and III to overcome the difficulties with the mode. We conduct quantitative comparisons of intensities for the methods utilizing different configurations of landmarks, as described in Section IV. We state our concluding remarks in Section V.

II. The Standardizing Method

In this section, we will describe our standardizing method. The framework will be set up in a general way so that the earlier method [9] can also be described in the same framework.

A. Notation

We denote the set of MRI protocols by $\mathcal{P} = \{P_1, P_2, \cdots, P_n\}$ and the set of body regions by $\mathcal{D} = \{D_1, D_2, \cdots, D_m\}$. We represent a volume image (or an image for short) by a pair $\mathcal{V} = (V, g)$ where V is a three-dimensional (3-D) array of volume elements (voxels) covering a body region of the particular patient for whom image data $\mathcal{V}$ are acquired, and g is a function, called the intensity function, that assigns an integer intensity value for each $v \in V$. We assume that $g(v) \geq 0$ for all $v \in V$ and $g(v) = 0$ if and only if there are no measured (and computed) data for voxel v. We denote by $\mathcal{V}_{PD}$ the set of all images that can possibly be generated as per protocol $P \in \mathcal{P}$ for body region $D \in \mathcal{D}$.

The histogram of any image $\mathcal{V}$ is a pair $\mathcal{H} = (G, h)$ where G is the set of all possible intensity values (gray values) in $\mathcal{V}$ (i.e., the range of g) and h is a function whose domain is G and whose value for each $x \in G$ is the number of voxels $v \in V$ for which $g(v) = x$. Let $m_1 = \min\{g(v) | v \in V \text{ and } g(v) > 0\}$ and $m_2 = \max\{g(v) | v \in V \text{ and } g(v) > 0\}$, the minimum and maximum intensity values in $\mathcal{V}$, respectively.

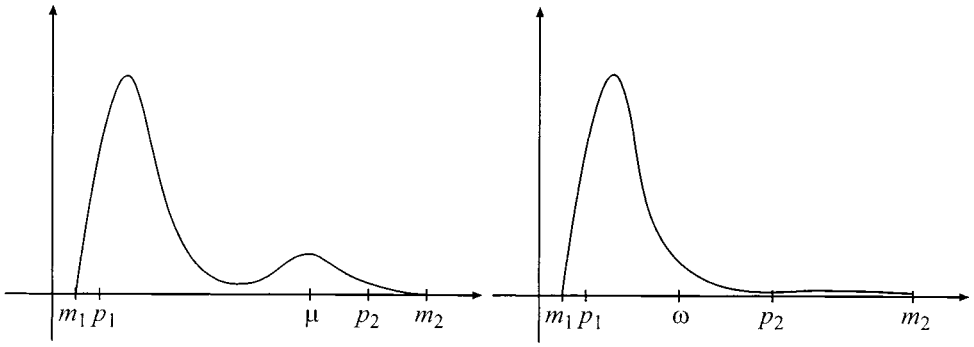


Fig. 1. Location of the histogram-specific parameters (landmarks). m_1 and m_2 are the minimum and maximum intensities in the image, p_1 and p_2 are the minimum and maximum percentile intensities, μ is the second mode of the histogram (in the bimodal case), and ω is the shoulder of the background hump (in the unimodal case).

B. Landmarks on the Histogram

It is desirable to cut off the tails of the histogram of the image 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 percentile values, respectively, that are used to select a range of intensity of interest (IOI). Let the actual intensity values corresponding to pc_1 and pc_2 in the histogram $\mathcal{H}$ be p_1 and p_2. We will come back later to how pc_1 and pc_2 are determined.

Based on examining over 20 body region/protocol combinations, we have observed mainly two types of histograms among MR images: unimodal and bimodal (see Fig. 1). In the case of bimodal histograms, we have used in the past the second mode (μ) that corresponds to the main foreground object in the image as a histogram landmark. With unimodal histograms the mode usually corresponds to the background, so we need to select some other landmark. This may be, for example, the shoulder (ω) of the hump of the background intensities. The locations of the histogram-specific parameters in these two cases are illustrated in Fig. 1 schematically. Since all of the protocols we studied in this paper and in our ongoing applications produce bimodal histograms, we will confine our description only to such protocols. With some additional work, the methods can be extended to the unimodal case as well as to other (multimodal) cases.

We will consider the following landmark configurations:

$$\begin{aligned} L_1 &= \{pc_1, pc_2, \mu\} \\ L_2 &= \{pc_1, pc_2, \mu_{50}\} \\ L_3 &= \{pc_1, pc_2, \mu_{25}, \mu_{50}, \mu_{75}\} \\ L_4 &= \{pc_1, pc_2, \mu_{10}, \mu_{20}, \cdots, \mu_{90}\} \end{aligned} \tag{1}$$

where μ_p for $p \in \{10, 20, \cdots, 90, 25, 75\}$ represents the pth percentile in the histogram associated with the foreground part of a given image and μ represents its second mode. For any image $\mathcal{V}_i$ we will denote by p_{1i} and p_{2i} the image intensities corresponding to the percentiles pc_1 and pc_2, respectively, of the histogram of $\mathcal{V}_i$. We will use the notation $\mu_{1i}, \mu_{2i}, \cdots, \mu_{li}$ to denote the image intensities in $\mathcal{V}_i$ corresponding to the landmarks other than pc_1 and pc_2 that are derived from $\mathcal{V}_i$ from among those defined in (1). For example, for the set L_3, $\mu_{1i}, \mu_{2i}, \mu_{3i}$ represent the intensities corresponding to the 25th, 50th, and 75th percentiles of the histogram of the foreground of $\mathcal{V}_i$, respectively.

C. Overview of the Standardizing Method

The method consists of two steps. In the first (training) step, a set of images of the same body region and protocol corresponding to a population of patients is given as input. The parameters of a histogram transformation are learned from these image data and a few additional input parameters are determined. These additional parameters may vary and are used to fine tune the transformation to the protocol, the body region, and possibly the application that will work on the transformed data. This step needs to be executed only once for a given protocol and for a given body region. In the second (transformation) step, the images are transformed using the parameters learned in the first step. This transformation is image dependent and needs to be done for each given image.

Our overall approach is as follows. Let the minimum and the maximum intensities on the standard scale for the IOI be s_1 and s_2, respectively. In the training step, the landmarks ($p_{1j}, p_{2j}, \mu_{1j}, \mu_{2j}, \cdots, \mu_{lj}$) obtained (as described later) from each image $\mathcal{V}_j$ of a set of images are mapped to the standard scale by mapping the intensities from $[p_{1j}, p_{2j}]$ onto $[s_1, s_2]$ linearly (see Fig. 2). Then, for each k, $1 \le k \le l$, the mean μ_{ks} of the mapped μ_{kj}s over the images $\mathcal{V}_j$ is computed. In the transformation step, for any given image $\mathcal{V}_i$, the actual landmark locations μ_{ki} obtained from its histogram are matched to μ_{ks} by doing several separate linear mappings: the first from $[p_{1i}, \mu_{1i}]$ to $[s_1, \mu_{1s}]$; the second from $[\mu_{1i}, \mu_{2i}]$ to $[\mu_{1s}, \mu_{2s}]$, $\cdots$; and the last from $[\mu_{li}, p_{2i}]$ to $[\mu_{ls}, s_2]$, as illustrated in Fig. 3. In the past, we have used the configuration denoted L_1, with the corresponding transformation that results for a given image illustrated in Fig. 3(a). Fig. 3(b) demonstrates the general case of l landmarks in addition to p_{1i} and p_{2i}. In Fig. 3, s'_{1i} and s'_{2i} represent the actual extreme values on the standard scale. They are determined by the linear mappings of the extreme segments: $[p_{1i}, \mu_{1i}]$ to $[s_1, \mu_{1s}]$ and $[\mu_{li}, p_{2i}]$ to $[\mu_{ls}, s_2]$. We call the overall mapping from the intensities $[m_{1i}, m_{2i}]$ of $\mathcal{V}_i$ to $[s'_{1i}, s'_{2i}]$ of the standard scale the standardizer of $\mathcal{V}_i$ and denote it by $\tau_{\mathcal{V}_i}$.

D. Algorithms

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

Algorithm 1. Training

Input: A set of images $\mathcal{V}_j$ $(j = 1, 2, \cdots, N)$ that is a subset of $\mathcal{V}_{PD}$ for some fixed $P \in \mathcal{P}$ and $D \in \mathcal{D}$, the histogram parameters pc_1, pc_2, and s_1, s_2, and $L \in \{L_1, L_2, L_3, L_4\}$.

Output: $\{\mu_{ks} | 1 \le k \le l\}$.

begin

1. **for** $j = 1$ to N **do**
2. compute the histogram $\mathcal{H}_j$ of $\mathcal{V}_j$;

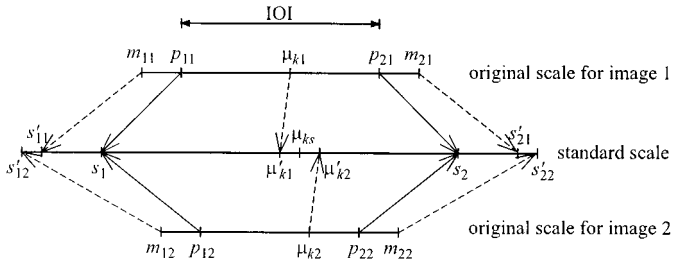


Fig. 2. Finding the parameters of the standard histogram. For illustration, only two input images are shown. For $j = 1, 2$, m_{1j} and m_{2j} are the minimum and maximum intensities in the image $\mathcal{V}_j$, p_{1j} and p_{2j} are the minimum and maximum percentile intensities, μ_{kj} is one of the landmarks of the histogram, μ'_{kj} is the mapped value of μ_{kj}, and μ_{ks} is the mean of the μ'_{kj}s: the actual parameter we are looking for on the standard scale.

3. determine intensity values p_{1j} and p_{2j} corresponding to pc_1 and pc_2 and the landmark locations μ_{1j}, μ_{2j}, $\cdots$, μ_{lj} on $\mathcal{H}_j$ (see comments below);
4. map $[p_{1j}, p_{2j}]$ of $\mathcal{H}_j$ onto $[s_1, s_2]$ linearly;
5. find the new mapped landmark locations μ'_{1j}, μ'_{2j}, $\cdots$, μ'_{lj};
6. **endfor;**
7. calculate the rounded means μ_{1s}, μ_{2s}, $\cdots$, μ_{ls} of μ'_{1j}, μ'_{2j}, $\cdots$, μ'_{lj}, respectively, over $j = 1, 2, \cdots, N$;

end

Algorithm 2. Transformation

Input: An image $\mathcal{V}_i \in \mathcal{V}_{PD}$, pc_1, pc_2, s_1, s_2, μ_{1s}, μ_{2s}, $\cdots$, μ_{ls}.

Output: The transformed image $\mathcal{V}_{si}$ and/or a lookup table (LUT) that stores the standardizer $\tau_{\mathcal{V}_i}$.

begin

1. compute the histogram $\mathcal{H}_i = (G_i, h_i)$ of $\mathcal{V}_i$;
2. determine intensity values p_{1i} and p_{2i} corresponding to pc_1 and pc_2 and the landmark locations μ_{1i}, μ_{2i}, $\cdots$, μ_{li} on $\mathcal{H}_i$ (see comments below);
3. map sections of the scale of $\mathcal{H}_i$ linearly according to Fig. 3 to the standard scale G_s of the standard histogram $\mathcal{H}_s = (G_s, h_s)$;
4. map the intensity value of every voxel $v \in \mathcal{V}_i$ according to $\tau_{\mathcal{V}_i}$ to get $\mathcal{V}_{si}$ and/or output the mapping $\tau_{\mathcal{V}_i}$ in a LUT;

end

We point out that the free ends characterized by the values of s'_{1i} and s'_{2i} of the standard scale depend on the given image $\mathcal{V}_i$. In other words, the range $[s'_{1i}, s'_{2i}]$ may vary from image to image. However, $[s_1, s_2]$ is independent of $\mathcal{V}_i$ and this is the interval within which a uniformity of intensity meaning is achieved.

In order to find $\mu_{1i}, \mu_{2i}, \cdots, \mu_{li}$ for a given image $\mathcal{V}_i$, we need to identify roughly the foreground part of the image. In an image of a head of a patient, for example, this part corresponds roughly to the set of all voxels that fall in the head. One of several simple methods can be utilized for this purpose. We employ thresholding. The threshold is chosen to be the overall mean intensity in the whole image. The method of background removal described in [4] can also be used. Several hundred studies have been processed automatically and successfully using this method in our ongoing projects for different protocol and body region combinations.

III. The Variants

A. 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).

We have delineated a theory and established a set of theorems in [9] for the configuration denoted L_1 that are crucial to guarantee the correct behavior (i.e., preserving the intensity relations between voxels) of the standardizer $\tau_{\mathcal{V}_i}$ for any given image $\mathcal{V}_i$. The theory readily generalizes to configurations L_2, L_3, and L_4.

We will need the following definitions for stating the theorems. Let $\mu'_{\min} = \min_{\mathcal{V}_i \in \mathcal{V}_{PD}} \{\mu'_i\}$, and $\mu'_{\max} = \max_{\mathcal{V}_i \in \mathcal{V}_{PD}} \{\mu'_i\}$. Note that $\mu'_{\min}$ and $\mu'_{\max}$ are dependent on s_1 and s_2. Let l, L, r, and R be indexes such that

$$\begin{aligned} \mu_l - p_{1l} &= \min_{\mathcal{V}_i \in \mathcal{V}_{PD}} \{(\mu_i - p_{1i})\} \\ \mu_L - p_{1L} &= \max_{\mathcal{V}_i \in \mathcal{V}_{PD}} \{(\mu_i - p_{1i})\} \\ p_{2r} - \mu_r &= \min_{\mathcal{V}_i \in \mathcal{V}_{PD}} \{(p_{2i} - \mu_i)\} \\ p_{2R} - \mu_R &= \max_{\mathcal{V}_i \in \mathcal{V}_{PD}} \{(p_{2i} - \mu_i)\}. \end{aligned} \tag{2}$$

We will assume throughout that pc_1 and pc_2 are such that, for any $\mathcal{V}_i \in \mathcal{V}_{PD}$, $p_{1i} < \mu_i < p_{2i}$.

Theorem 1 states the conditions under which it is guaranteed that no two distinct intensities in $\mathcal{V}_i$ are merged into a single intensity in the standardized image $\mathcal{V}_{si}$. Thus, if standardizing is done respecting these conditions, then there is no loss of information and the original image can be obtained by inverting the standardizer $\tau_{\mathcal{V}_i}$.

Theorem 1: For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{PD}$, and for any pc_1 and pc_2 such that $p_{1i} < \mu_i < p_{2i}$, the standardizer $\tau_{\mathcal{V}_i}$ of $\mathcal{V}_i$ is a one-to-one mapping if $\mu'_{\min} - s_1 \geq \mu_L - p_{1L}$ and $s_2 - \mu'_{\max} \geq p_{2R} - \mu_R$.

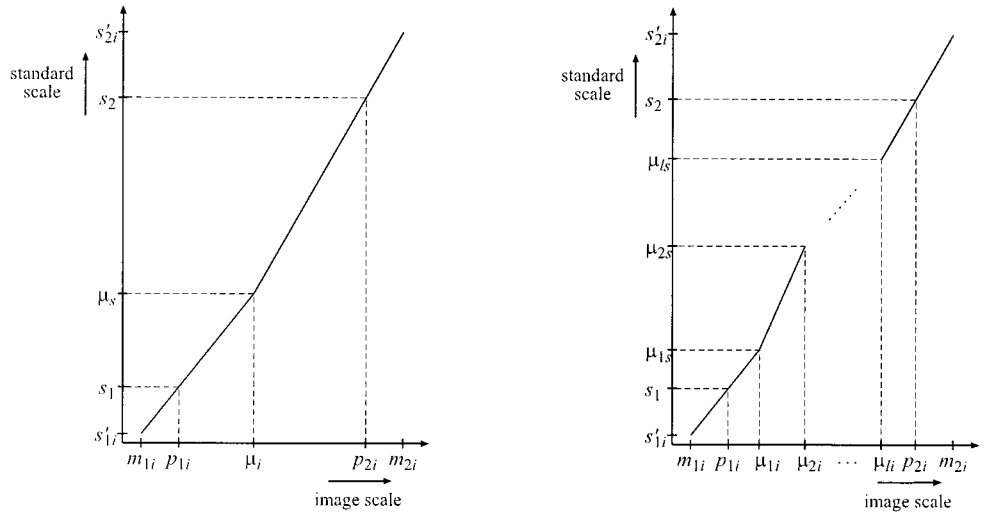
The following theorem gives guidance for selecting the values of s_1 and s_2 that cause no intensity loss.

Theorem 2: For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{PD}$, and for any pc_1 and pc_2 such that $p_{1i} < \mu_i < p_{2i}$, the standardizer $\tau_{\mathcal{V}_i}$ of $\mathcal{V}_i$ is a one-to-one mapping if

$$s_2 - s_1 \geq (\mu_L - p_{1L} + p_{2R} - \mu_R) \cdot \max\left\{\frac{\mu_L - p_{1L}}{\mu_l - p_{1l}}, \frac{p_{2R} - \mu_R}{p_{2r} - \mu_r}\right\}.$$

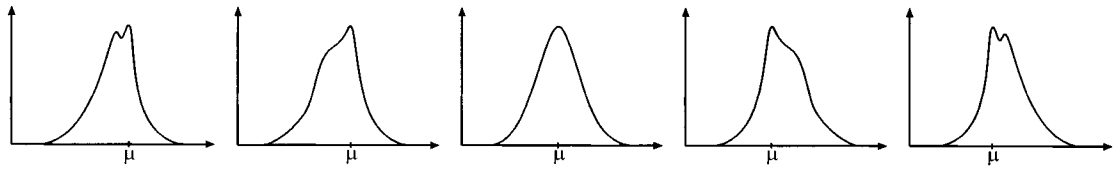
Note that Theorems 1 and 2 state conditions that require observing all images in $\mathcal{V}_{PD}$. In practice, since this is impossible to do, we estimate the right side of $\geq$ in the expression in Theorem 2 by examining a sufficient number of volume images and set $s_2 - s_1$ to a number sufficiently greater than this estimated entity. Our implemented software gives a warning message should an image be encountered for which this condition is violated. Even in such cases of violation, the software can be used in such way that, when a violation is detected, $s_2 - s_1$ is automatically updated so that this condition is indeed satisfied.

Theorem 3 states that once the conditions of Theorem 1 are satisfied, the order of the input intensities is maintained in the output images.



(a) (b)

Fig. 3. The intensity mapping function for the transformation phase. (a) For three landmarks (for configuration L_1). (b) For the general case with l landmarks.



(a) (b) (c) (d) (e)

Fig. 4. Schematic representation of the shapes of brain MRI histograms. For clarity, only the histogram corresponding to the foreground (brain) is shown.

Theorem 3: For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{PD}$, any standardizer $\tau_{\mathcal{V}_i}$ of $\mathcal{V}_i$ that satisfies the conditions in Theorem 1 and for any intensities x_1 and x_2 of $\mathcal{V}_i$, $\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2)$ if and only if $x_1 < x_2$.

That is, the actual order of brightness of tissue regions in $\mathcal{V}_i$ is maintained in the image $\mathcal{V}_{si}$ output by Algorithm 2, although their relative contrast may change. In a Pd image of a brain, for example, the known brightness relationship gray matter (GM) > white matter (WM) > CSF is maintained in $\mathcal{V}_{si}$.

B. Choosing the Landmark Configuration

The choice of the actual landmark configuration is also an important factor. The mode-based method described previously [9] (configuration L_1) works wellfor several MR protocols and several body regions for several applications but there are cases (and applications) wherein this simple method is not appropriate. For illustration, we will consider the shape of the GM, WM, and CSF distributions in FSE Pd brain images. Their relative locations vary among studies and even among studies of the same patient. Fig. 4 shows some histogram shapes, all of which were found in the histograms of multiple sclerosis (MS) patient brain studies. We recall here that in FSE Pd images, GM regions are brighter than WM regions.

Mainly the WM and GM distributions combine to give the histogram hump illustrated in Fig. 4 while the CSF distribution falls in the lower tail of this hump. The weakness of the mode-based method is that sometimes the mode [the intensity corresponding to the peak, denoted μ_i in Fig. 3(a)] corresponds to GM intensity [Fig. 4(a) and (b)], while in other cases it corresponds to WM intensity [Fig. 4(d) and (e)] or may also correspond to intensities that lie between real GM and WM intensities [Fig. 4(c)]. Therefore, when we match the mode to a fixed location on the standard scale, we may match GM in some cases and WM in the others. Because of this switching behavior, the mode-based method is often not appropriate if the application is image segmentation, where we need more accurate meaning on the standard scale even for relatively small ranges, although for display window setting this behavior does not cause many problems. In order to eliminate the switching behavior, one approach is to choose the median of the histogram of the foreground as a landmark to match. By studying approximately 100 histograms and their landmarks for images of several protocol/body region combinations, we observed that this landmark remains consistent, even in cases where the histogram has two similar peaks [Fig. 4(a) and (e)] and an asymmetric shape [Fig. 4(b) and (d)].

We may also use more histogram landmarks, such as quartiles and deciles, to better define the standard histogram. In these percentile-based methods, in addition to the cut-off percentiles p_{1i}, p_{2i}, we select several percentile markers μ_{1i}, μ_{2i}, $\cdots$, μ_{li} within the IOI. In the case of quartiles, $l = 3$ and μ_{1i}, μ_{2i}, μ_{3i} correspond to the 25th, 50th, and 75th percentiles within the IOI, respectively. In the case of deciles, $l = 9$ and μ_{1i}, μ_{2i}, $\cdots$, μ_{9i} correspond to the 10th, 20th, $\cdots$, 90th percentiles, respectively.

C. Choosing the Mapping Function

There are other possible changes to the basic standardizing method that can be used to make it better fit the type of input data and the actual application: 1) using histogram landmarks defined in different ways (i.e., not by percentiles); 2) using polynomial functions to stretch the histogram segments; and 3) use of spline fitting techniques instead of segment-by-segment linear stretching. When these variants are

TABLE I
PARAMETER CONFIGURATIONS USED IN THE DIFFERENT COMPARISONS

method	pc_1	pc_2	s_1	s_2	landmarks
0	—	—	—	—	none
1	0	99.8	1	4095	mode
2	0	99.8	1	4095	median
3	0	99.8	1	4095	quartiles
4	0	99.8	1	4095	deciles

used, changes are required in Step 3 of Algorithm 2, wherein the segment-wise linear mapping should be replaced by the actual mapping used.

IV. EVALUATION

For an evaluation of the method, for each protocol and body region, we need to consider the following variations in image data: 1) intra-patient (time-to-time) variation; 2) inter-patient variation; 3) variations among different machines of the same brand; 4) variations among machines of different brands. An evaluation that takes into account all these factors is indeed very formidable. The following sections describe the methods of evaluation that we used to examine how different kinds of variations are affected by the different variants of the standardizing method. Although we will address mainly the factors indicated in 1) and 2) rigorously, we also performed visual comparisons to check how different variants of the method handle the factors indicated in 3). We will consider several protocols but will confine ourselves to only one body region, namely the head.

Table I shows the different parameter settings that were used for the methods in carrying out the comparisons. We refer to the methods by numbers. Zero identifies the identity transformation (i.e., no transformation at all). First, display examples are illustrated with images before and after the different transformations for several P. Then quantitative studies are described with illustrating tables. For all tests, segmentwise linear mapping was used. The training was done by using ten different patient studies of the particular protocol and body region.

A. Qualitative Comparison

We conducted qualitative comparisons for the following MRI protocols: FSE Pd, FSE T2, spin-echo (SE) Pd, SE T2, T1 with Gadolinium enhancement (T1E), and SPGR. 30 studies each of FSE Pd, FSE T2 and T1E, and ten studies each of SE Pd, SE T2, and SPGR were transformed using the corresponding trained parameters. Within any of these protocols, the image acquisition parameters were identical for all patient studies. All studies used in obtaining the results reported here consisted of 45–60 contiguous 3-mm-thick axial slices of size 256 × 256 with 0.86 mm × 0.86 mm in-plane resolution. The voxel intensities were represented as 12-bit integers. No additional preprocessing was performed on any of these images. 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 volume histograms and not to individual slices, the slice orientation and the resolution has negligible effect on the transformation. Two types of visual comparisons were made: By displaying at fixed gray level window settings, and by displaying the binary images obtained at fixed threshold ranges.

1) Display at Fixed Windows: Images in each row of Fig. 5 show a slice from each of three different patient Pd studies. The rows from top to bottom correspond to Methods 0–4. For each method, the images are displayed at a fixed window setting. For Method 0, the window was set up correctly (visually) for the first image. For other methods, we devised standard window settings, as listed in Table II, by examining a few standardized studies. This figure demonstrates that the structures are well portrayed and the contrast is more similar for Methods 1–4 than for Method 0. Since the intensity ranges are quite different for the three studies, the displays for Method 0 exhibit inadequate window settings for the second and third study. Note also the striking similarity of the displays among Methods 1–4. Other protocols we tested exhibit a similar phenomenon and, hence, image displays are not included here for them (see [9] for a variety of display examples of this type comparing Methods 0 and 1 for several combinations of P and D. An observer

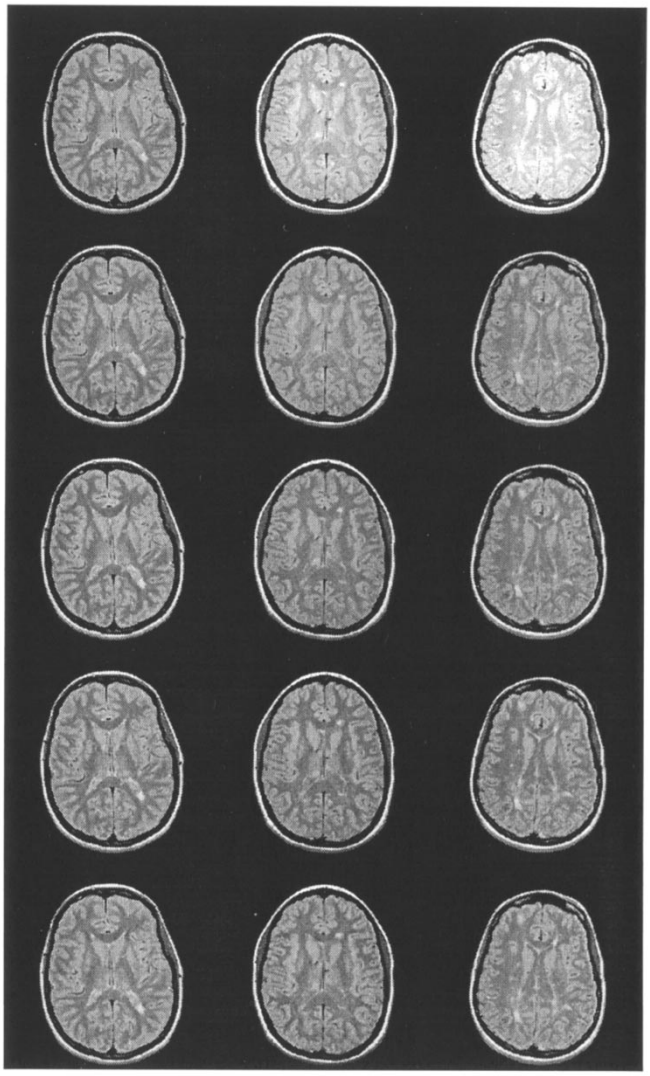


Fig. 5. Images displayed at fixed gray level windows. Rows 1–5 correspond to Methods 0–4. In each row, a slice from the Pd studies of three MS patients is displayed.

TABLE II
STANDARD WINDOWS FOR THE BRAIN FOR DIFFERENT PROTOCOLS AND METHODS. SIMILAR TABLES CAN BE CONSTRUCTED FOR OTHER PROTOCOLS AND BODY REGIONS

method	FSE Pd		FSE T2		T1E	
	level	width	level	width	level	width
1	2360	3444	2306	4523	2048	3500
2,3,4	2460	3444	2446	4523	2048	3500

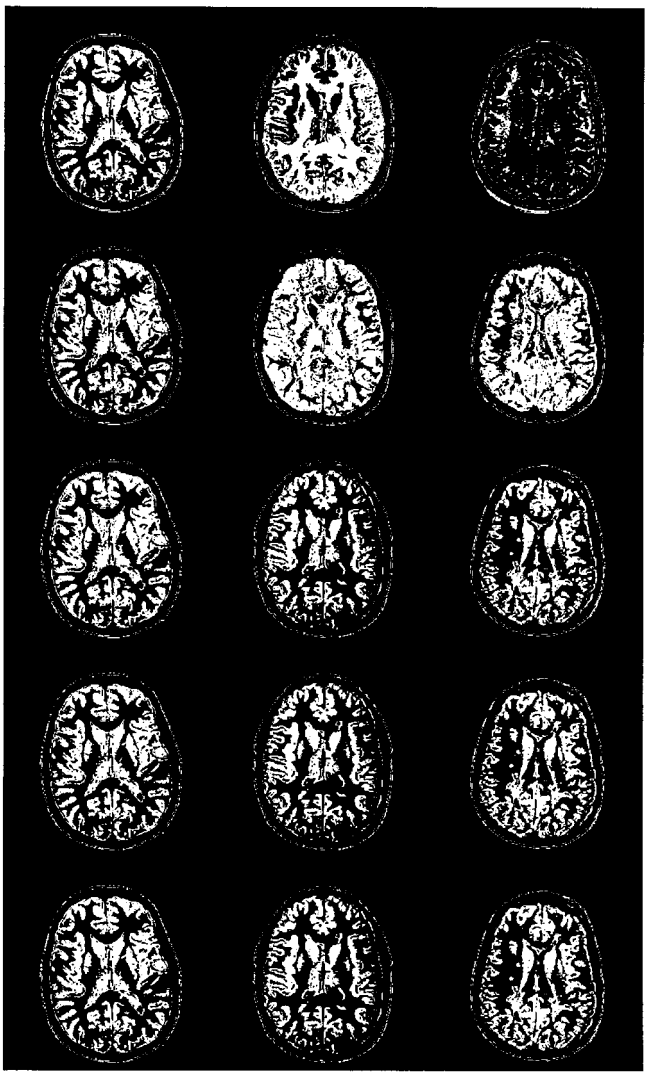


Fig. 6. Images in Fig. 5 displayed as binary using a fixed threshold interval for each row. The threshold interval for the first image in each row was determined interactively to give a good GM definition by visual inspection. Rows 1–5 correspond to Methods 0–4.

study to compare the display quality among methods is currently underway).

2) Binary Images at Fixed Thresholds: Fig. 6 demonstrates in another way that the intensity values have no consistent meaning in the original images but they do have after the transformation. Images are displayed as binary using fixed thresholds to segment approximately the GM region of the brain. These images correspond exactly to those in Fig. 5.

For each row, the threshold intervals were chosen to roughly segment the GM region in the first study by visual inspection. In Row 1 it is well demonstrated that the same thresholds do not segment the same tissue in different studies. In the second study the given interval corresponds to mainly WM, and in the third it is far below the brain tissue intensities. The switching problem of the mode method is well illustrated by the images in the second row. Although the transformed images displayed at a standard window show similar brightness and contrast characteristics, the fixed thresholds for small tissue intensity ranges are not quite applicable. The histograms of the three images in this case are like those shown in Fig. 4 (b), (d), and (e). The fixed interval thresholds segment GM in all cases, for Methods 2–4 more consistently than in rows 1 and 2. This implies that the meaning of intensities on the standard scale is more consistent after median-based and other percentile-based transformations.

TABLE III
MEAN AND STANDARD DEVIATION OF NORMALIZED MEAN SQUARED DIFFERENCES (NMSD) BEFORE (METHOD 0) AND AFTER DIFFERENT STANDARDIZING TRANSFORMS (METHODS 1–4) FOR 15 PAIRS OF STUDIES AND FOR THREE DIFFERENT PROTOCOLS. EACH PAIR REPRESENTS THE STUDIES OBTAINED FOR THE SAME PATIENT AT TWO TIME INSTANCES

method	FSE Pd		FSE T2		T1E	
	mean	sd	mean	sd	mean	sd
0	0.0099	0.0094	0.0093	0.0085	0.0025	0.0018
1	0.0039	0.0055	0.0036	0.0050	0.0020	0.0018
2	0.0039	0.0057	0.0035	0.0050	0.0020	0.0018
3	0.0040	0.0057	0.0033	0.0045	0.0020	0.0018
4	0.0039	0.0054	0.0028	0.0036	0.0019	0.0016

B. Quantitative Comparison

Two types of quantitative tests on data sets of brains obtained from three protocols FSE Pd, FSE T2, T1E were conducted: intra- and interpatient variations.

1) Intrapatient Variation: We used the same training data sets and parameter configurations as for the qualitative comparison described in the previous section. The test method for all three protocols was the same. Two scans acquired at different time instances were randomly selected for 15 patients from our database of MS patient studies. The time distance between the two scans of the same patient varied between 1 6 yr. For each patient we registered the first scan to the second via rigid transformation based on intensity value correlations. All original images had the same spatial resolution and the paired images covered the same volume of interest. The correlation-based registration method we used proved to result in registration of up to a voxel accuracy. Because these patients had MS and the images of the same patient were acquired 1–6 yr. apart, the number and volume of lesions changed as the disease evolved. The variation of voxel intensities within these lesion regions is due to the change of the tissue characteristics caused by the disease, rather than to the acquisition process. To eliminate these variations in the intensities and to compare only those voxels that supposedly contain the same tissue (mixture), in the two images of the same patient, the lesions were segmented [4] and removed for the purpose of the comparison. However, for the training process no segmentation was done and the whole volume histograms were used. The lesions in these MS patient studies occupy only a small fraction of the volume of the study and the lesion voxels have high intensity values that are mostly above the 99.8 percentile value, i.e., outside the IOI used by the method. This small volume of high-intensity lesion voxels does not significantly change the median and other landmarks.

The similarity of a pair of these registered lesion-removed images was measured by the mean squared intensity difference normalized to the original range of the images and denoted NMSD. This similarity measure was computed for every pair of images for each method. Table III shows that the mean value and the standard deviation of the NMSD after each transformation are smaller than that before transformation. The mean values of NMSD for the pairs of studies were compared using the paired t-test. The results displayed in Table IV show that the change in the means of NMSD due to the standardizer is statistically significant for all variants (first block). However, the differences in the improvements among the mode-based and other percentile-based variants are not statistically significant (second block). We encountered several patient studies [with histograms as illustrated in Fig. 4(a) and

TABLE IV
p-VALUES OF PAIRED t-TESTS APPLIED TO THE NMSD DATA

method	method	FSE Pd	FSE T2	T1E
0	1	0.0094	0.0036	0.0450
0	2	0.0101	0.0036	0.0374
0	3	0.0108	0.0031	0.0373
0	4	0.0102	0.0027	0.0241
1	2	0.1722	0.0040	0.5772
1	3	0.0154	0.0396	0.8138
1	4	0.7285	0.0795	0.1862
2	3	0.3415	0.0744	0.5628
2	4	0.6217	0.1007	0.2068
3	4	0.3450	0.1202	0.1925

TABLE V
THE STANDARD DEVIATION, THE 95% CONFIDENCE INTERVAL (CI), AND % COEFFICIENT OF VARIATION (CV) OF THE NORMALIZED MEAN INTENSITY (NMI) VALUES OF DIFFERENT TISSUES IN FSE Pd AND T2 IMAGES

tissue	method	FSE Pd				FSE T2			
		sd	CI low	CI high	% CV	sd	CI low	CI high	% CV
WM	0	0.1047	0.6572	0.7757	14.61	0.0580	0.3580	0.4236	14.83
	1	0.0113	0.5523	0.5651	2.03	0.0091	0.3456	0.3559	2.59
	2	0.0074	0.5728	0.5812	1.29	0.0090	0.3455	0.3557	2.56
	3	0.0111	0.5693	0.5819	1.93	0.0086	0.3407	0.3504	2.49
	4	0.0106	0.5695	0.5816	1.85	0.0099	0.3377	0.3489	2.88
GM	0	0.1144	0.7189	0.8483	14.59	0.0643	0.4124	0.4852	14.33
	1	0.0067	0.6230	0.6306	1.07	0.0119	0.4502	0.4636	2.60
	2	0.0065	0.6397	0.6470	1.00	0.0097	0.4440	0.4549	1.85
	3	0.0045	0.6449	0.6501	0.70	0.0137	0.4299	0.4453	3.13
	4	0.0052	0.6458	0.6518	0.81	0.0137	0.4268	0.4423	3.15
CSF	0	0.0884	0.5323	0.6323	15.18	0.0791	0.4564	0.5458	15.78
	1	0.0172	0.5309	0.5504	3.18	0.0348	0.6596	0.6989	5.12
	2	0.0185	0.5470	0.5679	3.33	0.0318	0.6572	0.6932	5.15
	3	0.0172	0.5481	0.5675	3.08	0.0296	0.6422	0.6757	4.49
	4	0.0185	0.5450	0.5660	3.33	0.0259	0.6141	0.6435	4.12

(e)] where the mode-based standardizer failed, yet the median-based standardizer succeeded. These exceptional cases are as important in practice as the nonexeceptionals. If the method breaks down on one of these cases, effort has to be made to specially treat the study in order to deal with the problem yet to retain the overall procedure applied to the normal cases as much as possible (e.g., being able to apply the same postprocessing steps with the same parameter settings to all studies). The median-based standardizer seems to handle those cases that were exceptional for the mode-based method, yielding more robust methods. The practical significance of this should not be overlooked yet this is difficult to capture within the statistical mold where the exceptional one usually gets lost among the many normals. Had we included a sufficient number of such exceptional cases in our analysis, we would have seen a statistically significant difference between Method 1 and the rest.

2) Interpatient Variation: For this comparison we randomly selected 12 FSE Pd and 12 FSE T2 data sets from our database. All images were previously segmented into WM, GM, and CSF regions [4]. The lesions were also segmented [4] and excluded from the tissue regions for comparison. (The reason for not including T1E in this evaluation is that the segmentation of WM, GM, and CSF in this protocol is significantly more difficult.) We calculated the intensity statistics over the population of images for each of these tissue regions separately. The standardizing parameters were the same as those for the other comparisons. For each of these regions in each image $\mathcal{V}_i$ in each of these protocols, we calculated the normalized mean intensity (NMI) by dividing the mean intensity in the region by $p_{2i} - p_{1i}$. This was repeated for each set of the transformed images wherein normalization was done by dividing the mean intensity in the region by $s_2 - s_1$. The standard deviations of the mean values of the NMI values before and after different standardizing transforms are shown in Table V, together with the corresponding 95% confidence intervals and the coefficients of variation. The table indicates that the intensities on the standard scale have more consistent tissue, meaning than those on the original scale and that the median-based standardizer outperforms the mode-based method in achieving similar tissue meaning of intensities. Although for CSF (that included the ventricles as well as the peripheral CSF, where the partial volume effect is significant and results in low intensities) the percent CV is slightly higher for the median-based method than that for the mode-based, the overall improvement (although probably not statistically significant) is seen in cases of WM and GM. More importantly, the exceptional cases (as mentioned previously) are handled correctly by the median- and percentile-based methods.

V. CONCLUDING REMARKS

We have described some of the problems with the original MRI scale standardization method reported in [9] and introduced new variants of the method that can help to overcome these problems. Median and other percentile values are more robust than the mode, therefore, using these as landmarks results in a more robust standardizer and, hence, more consistent meaning of intensities and better defined ranges for different tissues on the standard scale. It is possible to set better intensity of interest ranges while still being able to distinguish relevant information at the ends of the scale.

We have presented two types of studies to assess the effectiveness of the new variants. In qualitative studies we have shown through image display examples that the consistency of the brightness level and contrast of images is considerably improved after standardization. We have also shown that the tissues have much more consistent intensity ranges if the new variants of the standardizing method are used. In quantitative studies, we have demonstrated that the scanner dependent intra- and interpatient intensity variations are substantially reduced after standardizing using the new variants. Although the difference between the reduction yielded by the original standardizing method and by the new variants is not statistically significant, the new variants are more robust and this is important in practical applications.

Using the new standardized images in display, the standard windows for the different tissues (not only for the main object itself) can be either automatically applied or manually selected (e.g., from a short list of available windows), hence saving human interaction time on the per-case manual adjustments.

The new methods work in the same way as the original. They are easy to implement, rapid in execution, and completely automatic like the original, and can easily be incorporated into the original implementation framework. Like the original, they can also be incorporated in a picture archiving and communication system as a DICOM value of interest lookup table, so that images are automatically transformed or accompanied by the correct lookup table when they are downloaded to the viewing station.

Our preliminary studies indicate that image analysis and tissue segmentation methods are considerably improved in terms of their con-

stancy of parameter settings and their degree of automation upon scale standardization, especially using the median-based method.

ACKNOWLEDGMENT

The authors would like to thank R. I. Grossman for the MRI data sets utilized in this research.

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] L. P. Clarke, R. P. Velthuizen, S. Phuphanich, J. D. Schellenberg, J. A. Arrington, and M. Silbiger, "MRI Stability of three supervised segmentation techniques," *Magn. Reson. Imag.*, vol. 11, no. 1, pp. 95–106, 1993.

[3] R. Kikinis, M. E. Shenton, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C. R. Guttmann, R. W. McCarley, W. Lorensen, and H. Cline, "Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging," *J. Magn. Reson. Imag.*, vol. 2, no. 6, pp. 619–629, 1992.

[4] J. K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M. A. van Buchem, and R. I. Grossman, "Multiple sclerosis lesion quantification using fuzzy-connectedness principles," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 598–609, 1997.

[5] W. A. Edelstein, P. A. Bottomley, and L. M. Pfeifer, "A signal-to-noise calibration procedure for NMR imaging systems," *Med. Phys.*, vol. 11, no. 2, pp. 180–185, 1984.

[6] T. Yamamoto, T. Nambu, K. Miyasaka, and Y. Morita, "Accurate and practical calibration of MR signal intensities by the new transmission amplitude method: Application of the numerical diagnosis to MRI," *Radiology*, vol. 209P, p. 582, 1998.

[7] L. G. Nyúl and J. K. Udupa, "On standardizing the MR image intensity scale," *Radiology*, vol. 209P, pp. 581–582, 1998.

[8] ——, "An approach to standardize the MR image intensity scale," in *Proc. SPIE Medical Imaging*, vol. 3658, S. K. Mun and Y. Kim, Eds., San Diego, CA, 1999, pp. 595–603.

[9] ——, "On standardizing the MR image intensity scale," *Magn. Reson. Med.*, vol. 42, pp. 1072–1081, 1999.