Magnetic Resonance in Medicine 42:1072–1081 (1999)

On Standardizing the MR Image Intensity Scale

László G. Nyúl and Jayaram K. Udupa*

The lack of a standard image intensity scale in MRI causes many difficulties in image display and analysis. A two-step postprocessing method is proposed for standardizing the intensity scale in such a way that for the same MR protocol and body region, similar intensities will have similar tissue meaning. In the first step, the parameters of the standardizing transformation are "learned" from a set of images. In the second step, for each MR study these parameters are used to map their histogram into the standardized histogram. The method was tested quantitatively on 90 whole-brain studies of multiple sclerosis patients for several protocols and qualitatively for several other protocols and body regions. Measurements using mean squared difference showed that the standardized image intensities have statistically significantly ($P < 0.01$) more consistent range and meaning than the originals. Fixed gray level windows can be established for the standardized images and used for display without the need of per case adjustment. Preliminary results also indicate that the method facilitates improving the degree of automation of image segmentation. Magn Reson Med 42:1072–1081, 1999. © **1999 Wiley-Liss, Inc.**

Key words: image normalization; magnetic resonance imaging; image display; image processing

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 available with or without the use of contrast agents. These protocols allow the setting up of different contrasts among the different tissues within the same organ system. Unfortunately, one of the major difficulties with MRI techniques has been that intensities do not have a fixed meaning, not even within the same protocol, for the same body region, for images obtained on the same scanner, for the same patient. This implies that MR images cannot be displayed at preset windows; one has to often adjust the window settings per case. The lack of a meaning for intensities also poses problems in image segmentation (1,2) and quantification (3–5), which are the main motivating factors for this work.

Most visualization and analysis methods have parameters. Setting values for the parameters for these methods becomes more difficult without the same protocol-specific intensity meaning. What we need is that, for protocols that are the same or "close" to each other, the resulting images should be "close" (6). A standardizer can be incorporated at two stages of the image acquisition/processing flow. It can be built into the imaging device in order to produce images with a standard scale at the time of acquisition, or it can be used as a postprocessing step in a later phase. Attempts have been made to calibrate MR signal characteristics at the time of acquisition using phantoms (7,8). Although it is feasible to do such a calibration of all patient scans, it is somewhat cumbersome. Moreover, such a technique is not applicable to image data that have already been acquired without the required calibration phantoms. Postprocessing techniques that are applied to the image data that do not have any special acquisition requirements are, therefore, more attractive. There is a natural tendency to think that a simple scaling of the minimum to maximum intensity range of the given image to a fixed standard range may solve this problem. This usually does not help in achieving a similarity of intensities, as demonstrated in Nyúl and Udupa (9).

A postprocessing technique to automatically adjust the contrast and brightness of MR images (i.e., "windowing") for image display has been presented in Wendt (10). 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 meaning after the windowing transformation. There does not seem to have been any serious attempt to address this latter problem in the past.

The method described in this article offers a simple way of transforming the images nonlinearly so that there is a significant gain in similarity of the resulting images. It is based on transforming the intensity histogram of each given volume image into a "standard" histogram. This is achieved in two steps—a training step that is executed only once for a given protocol and body region and a transformation step that is executed for each given volume image. In the training step, certain landmarks of a standard histogram (for the body region and protocol under consideration) are estimated from a given set of volume images. In the transformation step, the actual intensity transformation from the intensity scale of the input volume image to the standard scale is computed by mapping the landmarks determined from the histogram of the given volume image to those of the standard histogram. The method is easy to implement and rapid in execution. The actual transformation itself can be stored as a lookup table in the image header. The transformed volume images permit predetermined display window settings and also facilitate quantitative image analysis.

THEORY AND ALGORITHMS

The method consists of two steps. In the first ("training") step, a set of volume images of the same body region and protocol corresponding to a population of patients is given as input. The parameters (landmarks) of a "standard" histogram are estimated from these image data. This step needs to be executed only once for a given protocol and

Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania.

Grant sponsor: NIH; Grant number: NS 37172; Grant sponsor: Department of Army; Grant number: DAMD 179717271.

*Correspondence to: Jayaram K. Udupa, Medical Image Processing Group, Department of Radiology, University of Pennsylvania, 423 Guardian Drive, 4th Floor Blockley Hall, Philadelphia, PA 19104-6021. E-mail: jay@mipg.upenn.edu

Received 18 May 1999; revised 20 July 1999; accepted 22 July 1999.

body region. In the second (“transformation”) step, any given volume image acquired as per the protocol and for the body region utilized in the training step is transformed so that its histogram parameters match those of the standard histogram. In this fashion, every patient volume image histogram is deformed to match it with the standard histogram. This step is image-dependent and needs to be executed for each given volume image. This step usually results in a nonlinear intensity transformation for the given image (since the two segments are mapped independently). However, the relationship between tissue intensities is maintained and intensity comparisons can be made using the standardized images.

Notation

We denote the set of MRI protocols by $\mathscr{P}$ and the set of body regions by $\mathscr{D}$. We represent a volume image by a pair $\mathscr{I} = (V, g)$ where V is a 3-dimensional array of volume elements (voxels) covering a body region of the particular patient for whom image data $\mathscr{I}$ are acquired, and g is a function (called intensity function) that assigns an integer intensity value for each $v \in V$. (We will often refer to a volume image simply as an image for short. This is not to be confused with the 2D slice images. All processing operations described in this article are carried out in 3D.) 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 $\mathscr{I}_{PD}$ the set of all images that can possibly be generated as per protocol $P \in \mathscr{P}$ for body region $D \in \mathscr{D}$.

The histogram of any image $\mathscr{I}$ is a pair $\mathscr{H} = (G, h)$ where G is the set of all possible intensity values (gray values) in $\mathscr{I}$ (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 gray values in $\mathscr{I}$, respectively.

The tails of the histogram often cause problems. Usually the high intensity tail corresponds to artifacts and outlier intensities, and causes considerable inter- and intrapatient/scanner variations. With this in mind, let pc_1 and pc_2 denote the minimum and maximum percentile values, respectively, that are used to select a range of intensity of interest (IOI).

Algorithms

Based on over 20 body region/protocol combinations, we have observed two types of histograms: unimodal and bimodal. We have also observed that the histograms of volume images of the same body region and protocol are always of the same type. Figure 1 illustrates them schematically. In case of bimodal histograms, we can usually use the 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 (identified by, for example, the point at which the histogram slope becomes −1). The locations of the histogram-specific parameters in these two cases are illustrated in Fig. 1. Since most of the protocols we studied produce bimodal histograms, we will concentrate on this

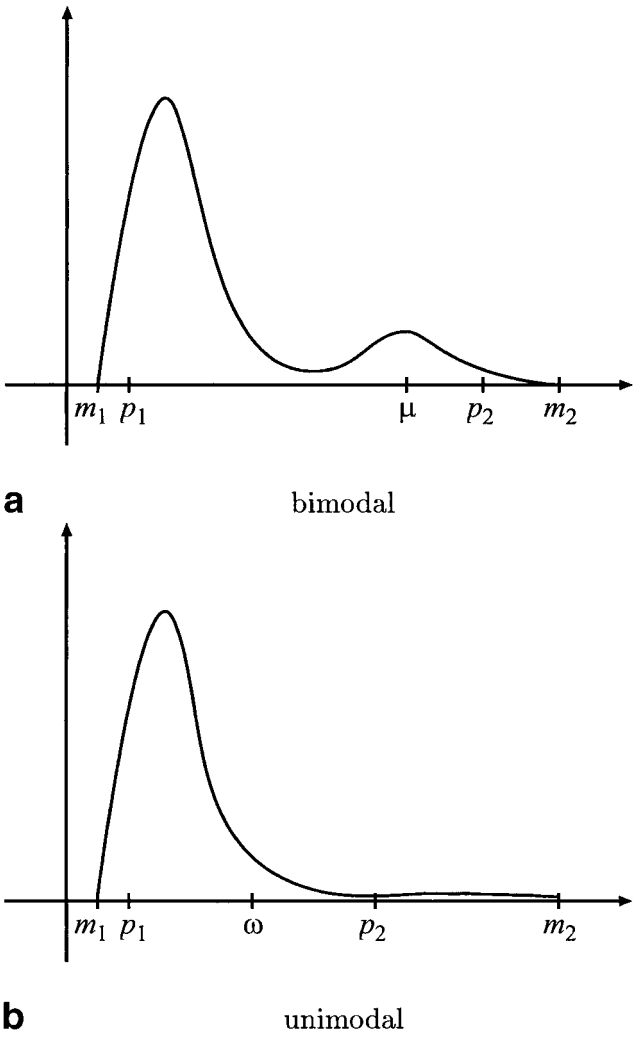

FIG. 1. Location of the histogram-specific parameters (landmarks). m_1 and m_2 are the minimum and maximum intensities in the image, p_1 and p_2 are minimum and maximum percentile intensities, μ is the second mode of the histogram (in the bimodal case) and ω is the shoulder of the “background hump” (in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

case in this article. Similar methods can be devised for the unimodal case.

Our overall approach is as follows. Let the minimum and the maximum intensities on the standard scale (corresponding to the standard histogram) for the IOI be s_1 and s_2, respectively. In the training step, the landmarks p_{1j}, p_{2j}, and μ_j obtained from the histogram of each image $\mathscr{I}_j$ of a subset of images of $\mathscr{I}_{PD}$ are mapped to the standard scale by mapping the intensities from $[p_{1j}, p_{2j}]$ to $[s_1, s_2]$ linearly. The formula for mapping $x \in [p_{1j}, p_{2j}]$ to x' is the following:

$$x' = s_1 + \frac{x - p_{1j}}{p_{2j} - p_{1j}}(s_2 - s_1). \qquad [1]$$

This mapping is utilized to determine the map μ'_j of μ_j on $[s_1, s_2]$, and subsequently the rounded mean μ_s of the μ'_js is computed (see Algorithm 1, below). In the transformation step, for any given image $\mathscr{I}_i = (V_i, g_i) \in \mathscr{I}_{PD}$, the actual second mode μ_i obtained (as described below) from the histogram of $\mathscr{I}_i$ is matched to μ_s by doing two separate

linear mappings (see Algorithm 2): the first from $[p_{1i}, \mu_i]$ to $[s_1, \mu_s]$ and the second from $[\mu_i, p_{2i}]$ to $[\mu_s, s_2]$. Figure 2 shows the plot of the mapping function. The lower and the upper ends of the standard scale are subsequently extended to s'_{1i} and s'_{2i}, respectively, by mapping $[m_{1i}, p_{1i}]$ to $[s'_{1i}, s_1]$ and $[p_{2i}, m_{2i}]$ to $[s_2, s'_{2i}]$, as illustrated in Fig. 2. We call this mapping from the intensities $[m_{1i}, m_{2i}]$ of $\mathcal{I}_i$ to $[s'_{1i}, s'_{2i}]$ of the standard scale the standardizer of $\mathcal{I}_i$ and denote it by $\tau_{\mathcal{I}_i}$. The expression for $\tau_{\mathcal{I}_i}$ (from Fig. 2) is as follows:

$$\tau_{\mathcal{I}_i}(x) \begin{cases} \left\lfloor \mu_s + (x - \mu_i)\dfrac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rfloor, & \text{if } m_{1i} \le x \le \mu_i, \\ \left\lceil \mu_s + (x - \mu_i)\dfrac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil, & \text{if } \mu_i \le x \le m_{2i}, \end{cases} \quad [2]$$

where $\lfloor \cdot \rfloor$ denotes the ''ceiling'' or ''floor'' operator. (These operators convert any real number y to the closest integer Y such that $Y \ge y$ or $Y \le y$, respectively.) Note that $s'_{1i} = \tau_{\mathcal{I}_i}(m_{1i})$, and $s'_{2i} = \tau_{\mathcal{I}_i}(m_{2i})$. The image $\mathcal{I}_{si} = (V_i, g_{si})$ resulting from the standardizing mapping of $\mathcal{I}_i$ is given by, for all $v \in V_i$, $g_{si}(v) = \tau_{\mathcal{I}_i}(g_i(v))$. 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 $\mathcal{I}_i$. In other words, the range $[s'_{1i}, s'_{2i}]$ may vary from image. However, $[s_1, s_2]$ is independent of $\mathcal{I}_i$, and this is the interval within which a uniformity of intensity meaning is achieved.

In order to find the second mode of the histogram, we eliminate the first mode (corresponding to the background) by thresholding. The threshold is chosen to be the overall mean intensity in the whole image. Several hundred studies have been successfully processed with this background removal process in our ongoing projects for different protocol and body region combinations.

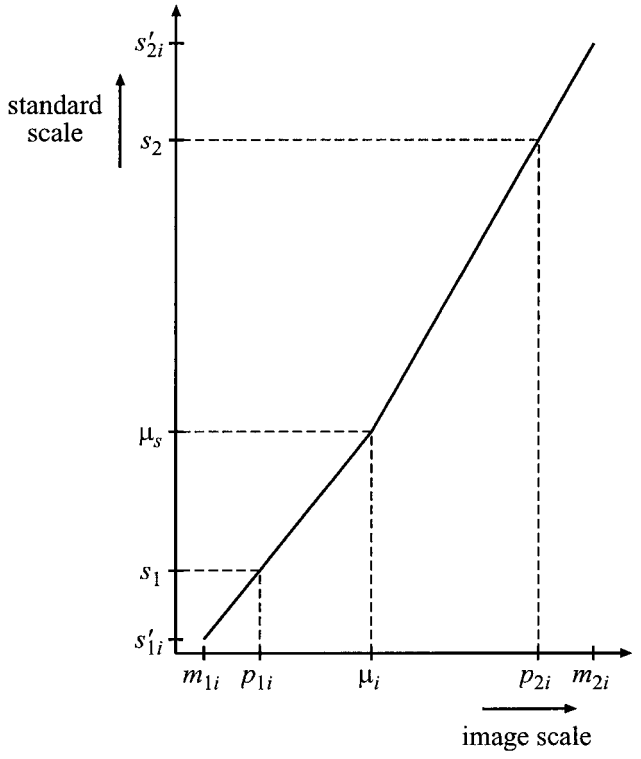

FIG. 2. The intensity mapping function for the transformation phase.

Algorithm 1. *Training*

Input: A set of images $\mathcal{I}_j (j = 1, 2, \ldots, N)$ that is a subset of $\mathcal{I}_{PD}$, the histogram parameters pc_1, pc_2, and s_1, s_2.
Output: μ_s.

begin
1. **for** $j = 1$ to N **do**
2. compute the histogram $\mathcal{H}_j$ of $\mathcal{I}_j$;
3. determine intensity values p_{1j} and p_{2j} corresponding to pc_1 and pc_2 and the mode μ_j on $\mathcal{H}_j$;
4. map $[p_{1j}, p_{2j}]$ of $\mathcal{H}_j$ onto $[s_1, s_2]$;
5. find the new mapped landmark location μ'_j;
6. **endfor**
7. calculate the rounded mean μ_s of μ'_j over $j = 1, 2, \ldots, N$;

end

Algorithm 2. *Transformation*

Input: An image $\mathcal{I}_i \in \mathcal{I}_{PD}$, pc_1, pc_2, s_1, s_2, μ_s.
Output: The transformed image $\mathcal{I}_{si}$ and/or a lookup table (LUT) that stores the standardizer $\tau_{\mathcal{I}_i}$.

begin
1. compute the histogram $\mathcal{H}_i = (G_i, h_i)$ of $\mathcal{I}_i$;
2. determine intensity values p_{1i} and p_{2i} corresponding to pc_1 and pc_2 and the second mode μ_i on $\mathcal{H}_i$;
3. map sections of the scale of $\mathcal{H}_i$ linearly according to Fig. 2 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 $v \in \mathcal{I}_i$ according to $\tau_{\mathcal{I}_i}$ to get $\mathcal{I}_{si}$ and/or output the mapping $\tau_{\mathcal{I}_i}$ in a LUT;

end

Choosing the Standardization Parameters

Although once the training step is completed the corresponding transformation step is fully determined, there are several possibilities to tailor the standardization to the specific needs of an application. The scale parameters may depend on the actual application. For example, s_1 should not be 0 if the values that are below pc_1 percentile need to be distinguished from ''nothing'' (i.e., value 0). Further, s_2 should be large enough not to compress the parts of the histograms (i.e., not to merge neighboring intensities after the transformation). When $pc_1 > 0$ and/or $pc_2 < 100$, the values in $[m_{1i}, p_{1i}]$ and $[p_{2i}, m_{2i}]$ must be mapped to $[s'_{1i}, s_1]$ and $[s_2, s'_{2i}]$, respectively, where s'_{1i} and s'_{2i} are determined by applying the mapping in the two linear sections of Fig. 2.

We will need the following definitions for stating certain theorems that guarantee the correct behavior of the standardizer. Let $\mu'_{\min} = \min_{\mathcal{I}_i \in \mathcal{I}_{PD}} \{\mu'_i\}$, and $\mu'_{\max} = \max_{\mathcal{I}_i \in \mathcal{I}_{PD}} \{\mu'_i\}$. 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

$$\mu_l - p_{1l} = \min_{\mathcal{I}_i \in \mathcal{I}_{PD}} \{(\mu_i - p_{1i})\}, \quad [3]$$

$$\mu_L - p_{1L} = \max_{\mathscr{I}_i \in \mathscr{I}_{PD}} \{(\mu_i - p_{1i})\}, \quad [4]$$

$$p_{2r} - \mu_r = \min_{\mathscr{I}_i \in \mathscr{I}_{PD}} \{(p_{2i} - \mu_i)\}, \quad [5]$$

$$p_{2R} - \mu_R = \max_{\mathscr{I}_i \in \mathscr{I}_{PD}} \{(p_{2i} - \mu_i)\}. \quad [6]$$

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

We state below three theorems which are crucial to guarantee the correct behavior of the standardizer $\tau_{\mathscr{I}_i}$ for any given image $\mathscr{I}_i$. Their complete proofs are given in Nyúl and Udupa (9). Theorem 1 states the conditions under which it is guaranteed that no two distinct intensities in $\mathscr{I}_i$ are merged into a single intensity in the standardized image $\mathscr{I}_{si}$. 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 $\tau_{\mathscr{I}i}$.

Theorem 1. *For any protocol $P \in \mathscr{P}$, any body region $D \in \mathscr{D}$, any image $\mathscr{I}_i \in \mathscr{I}_{PD}$, and for any pc_1 and pc_2, such that $p_{1i} < \mu_i < p_{2i}$, the standardizer $\tau_{\mathscr{I}_i}$ of $\mathscr{I}_i$ 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 a guidance for selecting the values of s_1 and s_2 that cause no intensity loss.

Theorem 2. *For any protocol $P \in \mathscr{P}$, any body region $D \in \mathscr{D}$, any image $\mathscr{I}_i \in \mathscr{I}_{PD}$, and for any pc_1 and pc_2, such that $p_{1i} < \mu_i < p_{2i}$, the standardizer $\tau_{\mathscr{I}_i}$ of $\mathscr{I}_i$ is a one-to-one mapping if $s_2 - s_1 \geq (\mu_L - p_{1L} + p_{2R} - \mu_R) \cdot max\ (\mu_L - p_{1L})/(\mu_l - p_{1l})$, $(p_{2R} - \mu_R)/(p_{2r} - \mu_r)$.*

Note that Theorems 1 and 2 state conditions that require observing all images in $\mathscr{I}_{PD}$. In practice, since this is impossible to do, we estimate the right side of "≥" 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.

In testing 100 proton density (Pd) volume images of the brain of different patients, acquired as per a fixed fast-spin-echo protocol on two GE 1.5T scanners, we found that $\mu_L - p_{1L} = 1139$, $p_{2R} - \mu_R = 960$, $\mu_l - p_{1l} = 693$, and $p_{2r} - \mu_r = 502$ for $pc_1 = 0$ and $pc_2 = 99.8$. By the condition in Theorem 2, this implies that, with $s_1 = 1$, if we set s_2 to at least 4016, we can make sure that the mapping is lossless (though probably in many cases a smaller value would be sufficient). In all our experiments (and current use), we have set $s_2 = 4095$. This choice has been adequate in all protocols we currently use and is highly unlikely to lead to a lossy mapping at least for any proton density brain images acquired as per the above protocol.

We arrived at the choice of $pc_2 = 99.8$ by the following experiments. We randomly selected a few Pd brain volume images acquired as per the above protocol. With $pc_1 = 0$ fixed, we computed p_{2i} for several values of pc_2 (in steps of 0.05, between 99.0 and 100.0) for each test image $\mathscr{I}_i$. We chose the value of pc_2 to be the largest value at which the variation in p_{2i} with respect to changes in pc_2 reached an acceptably small value. Figure 3 shows the plot of the standard deviation of p_{2i} for the images in the test set versus pc_2 (Fig. 3a) and the derivative of that function (Fig. 3b). The value $pc_2 = 99.8$ corresponds to the shoulder in the graph on the right where the function starts increasing rapidly. This implies that beyond this pc_2 value, the variations in the intensities are not systematic but random.

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

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

That is, the actual order of brightness of tissue regions in $\mathscr{I}_i$ is maintained in the image $\mathscr{I}_{si}$ output by Algorithm 2, although their relative contrast may change. In a Pd image

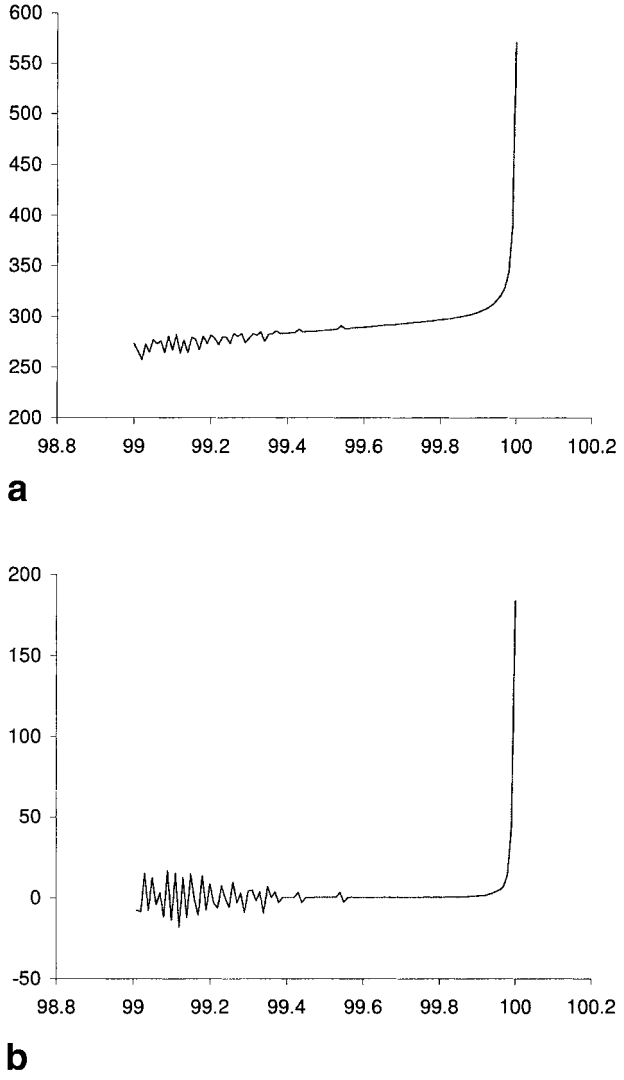

FIG. 3. A plot of the standard deviation of p_{2i} for the images in the test set vs. pc_2 (**a**), and its first derivative (**b**). $pc_2 = 99.8$ is approximately the largest value at which the derivative is still small (0.5).

of a brain, for example, the known brightness relationship gray matter > white matter > CSF is maintained in $\mathcal{Y}_{si}$.

EVALUATION

Our hypothesis is that, for any given protocol $P \in \mathcal{P}$, and for any body region $D \in \mathcal{D}$, the standardized images $\mathcal{Y}_{si}$ have more consistent tissue meaning for image intensities than the images $\mathcal{Y}_i$ before standardization. That is, after standardization voxels having the same intensity value are more likely to contain the same kind of tissue. For testing this hypothesis, for each protocol P and body region D, we need to consider the following variations in image data: (i) intrapatient (time-to-time) variation, (ii) interpatient variation, (iii) variations among different machines of the same brand, and (iv) variations among machines of different brands. Verifying this hypothesis rigorously taking all these factors into account is indeed a formidable task. Instead, we will rigorously test the hypothesis for several protocols and for only one body region, namely the brain, and for the factors indicated in only (i) and (ii) above. We will also provide qualitative evidences for the validity of the hypothesis through display examples before and after standardization for factors indicated in (i), (ii), and (iii) above for several P and D.

Qualitative Comparison

We conducted qualitative comparisons for the following MRI protocols and body regions: fast spin-echo (FSE) Pd, FSE T_2, spin-echo (SE) Pd, SE T_2, T_1 with Gadolinium enhancement (T_1E), and an SPGR sequence, all for the brain; a T_1-weighted gradient-echo sequence for the foot. The training set consisted of 10 studies (by a study we mean a volume image) of different patients in all cases. There were 30 studies each of FSE Pd, FSE T_2, and T_1E, and 10 studies each of SE Pd, SE T_2, T_1 GRE, and SPGR were transformed using the corresponding "trained" parameters. Images of all protocols had bimodal histograms except SPGR, which had unimodal histograms. In the latter case, a different landmark set was used in the standardization algorithm (see Fig. 1b), but the results are not discussed here because of space limitations.

The parameters for training were chosen as follows. The low and high percentiles were set to $pc_1 = 0$ and $pc_2 = 99.8$. Experiments were also done with other high percentiles (i.e., 100, 99.5, and 99.0). For the standard scale, we chose $s_1 = 1$, $s_2 = 4095$. Further, in all cases we made sure that the standardization mapping was lossless. In all cases, we derive the standardizer based on the whole volume image and not on the individual slices.

Histograms

Histograms of 10 FSE Pd studies at different stages of the transformation are displayed in Fig. 4. The low intensity part of the histograms that corresponds to the background voxels has been removed from the display in order to show the intensity of interest on a better scale. Original histograms of the volume images are plotted in Fig. 4a. Figure 4b displays the histograms after a simple linear mapping of $[p_{1i}, p_{2i}]$ to $[s_1, s_2]$, with $pc_1 = 0$ and $pc_2 = 99.8$. Histograms after the final standardizing mapping are plotted in Fig. 4c. A visual comparison shows that the histograms are more similar in shape and location after standardization than before. This implies that the actual intensity values and their distribution in the images are more similar after standardization. Note also that the simple linear mapping illustrated in the graph in Fig. 4b that accounts for outlier

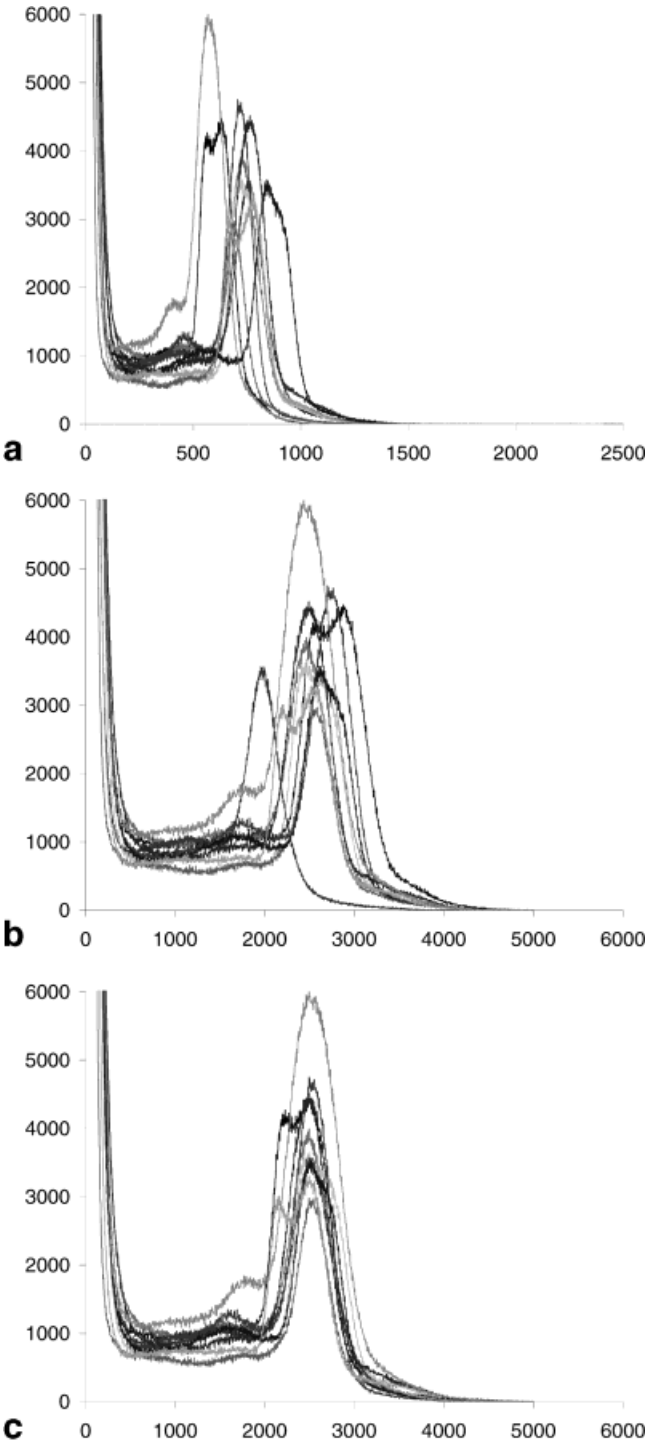

FIG. 4. Histograms at different stages of the standardization process for 10 different FSE Pd studies. Original histograms (**a**), histograms after intensity scaling from $[p_{1i}, p_{2i}]$ to $[s_1, s_2]$, with $s_1 = 1$, $s_2 = 4095$ (**b**), and after final standardization (**c**).

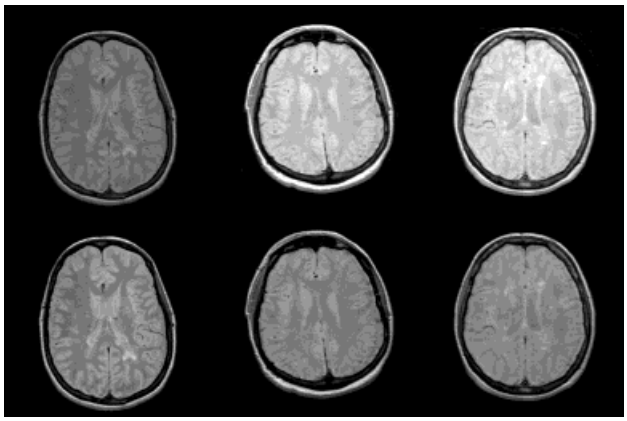

FIG. 5. Original slices from three studies acquired as per the same FSE Pd protocol before standardization displayed at a fixed window that was actually set up for the first image (first row), and after standardization displayed at a fixed ''standard'' window (second row).

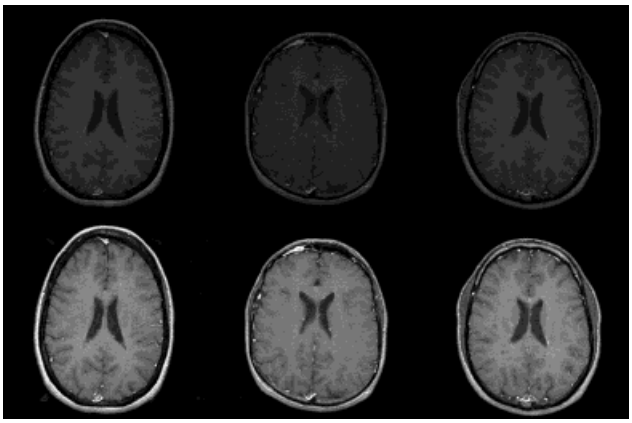

FIG. 6. Original slices from three studies acquired as per the same T_1 protocol with Gadolinium enhancement before standardization displayed at default windows (first row), and after standardization displayed at a fixed ''standard'' window (second row).

intensities is also not enough in making the histograms cluster.

Display at Fixed Window

Images in the first row of Fig. 5 show a slice from each of three different patient FSE Pd studies, all acquired using the same protocol. They are all displayed at a fixed gray level window that was actually set up for the first image. This window does not seem to be appropriate for the other two datasets because they have quite different intensity ranges. In the second row of Fig. 5, the same slices are displayed, after standardization (always on the whole volume image) using $pc_2 = 99.8$, at a fixed ''standard'' brain window that we devised after examining a few standardized images. In Table 1, we list for several different protocols such ''standard'' window settings (level and width) that we have devised in this fashion. The structures are well portrayed and the brightness and contrast are more similar than that of the originals. A similar behavior was observed on FSE T_2, SE Pd, and SE T_2 images.

Figure 6 is analogous to Fig. 5. It shows images before and after standardization for a T_1 protocol with Gadolinium enhancement. All images shown are acquired using the same protocol. The first row shows the original data displayed using the default window setting (i.e., window level set to the middle of the full range, and window width set to the full range of intensities of the study). The second row shows the same slices after standardization displayed using the ''standard'' brain window settings from Table 1.

Table 1
''Standard'' Windows for the Brain Images for Different Protocols Arrived at From the Standardized Images

	Level	Width
FSE Pd	2510	3444
FSE T_2	2306	4523
T_1E	2048	3500
SE Pd	2812	4095
SE T_2	2327	4095

Similar tables can be constructed for other body regions, protocols, and for specific tissue regions.

Figure 7 illustrates the improvements resulting from standardization for a different body region, namely the foot. The imaging protocol in this case was a T_1-weighted gradient-echo sequence and was the same for the three studies.

The following example is included to demonstrate that the standardizer still works when substantial deviations exist in the test images compared to the images used for training. Figure 8 shows a dataset of the brain of a patient with a large tumor. The first row of Fig. 8 shows a slice of a FSE Pd and a FSE T_2 dataset before standardization and the second row shows the same slices after standardization. In this case, the standardizer was arrived at from the training data utilized for the multiple sclerosis (MS) patient FSE Pd (Fig. 5) and FSE T_2 images with a similar (but not identical) protocol.

In Figure 9, we illustrate that the method is tolerant to small variations in protocol settings as well as variations that may exist among machines of the same brand in different hospitals. The figure shows three different patient studies obtained from a GE 1.5T Signa scanner at the University of Colorado Health Sciences Center (data courtesy of Dr. Jack Simon) using an SE Pd protocol. Images

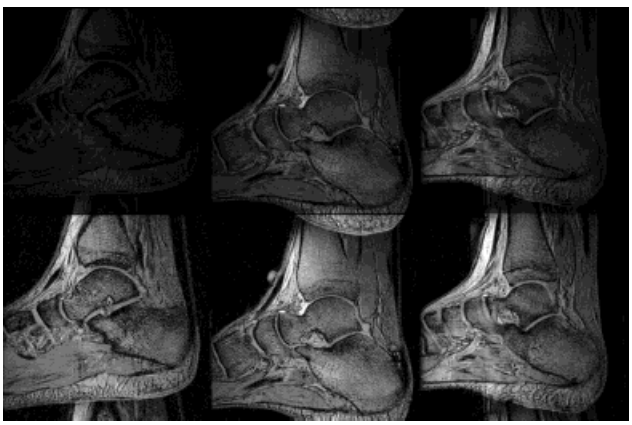

FIG. 7. Original slices of three foot studies before (first row), and after standardization (second row). The imaging protocol for all three datasets was a T_1-weighted gradient-echo sequence with identical parameters.

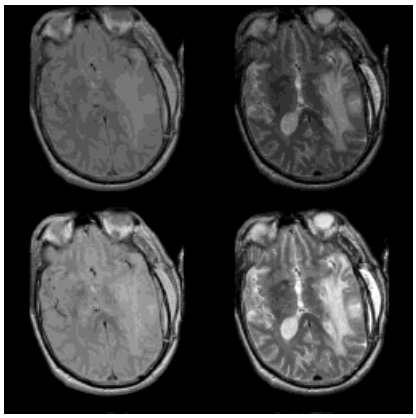

FIG. 8. Original slices, before standardization, of FSE Pd (left) and T_2 (right) datasets of a patient's brain with a large tumor displayed at default windows (first row), and after standardization, displayed at a standard window setting (second row) listed in Table 1 for the FSE Pd and T_2 case.

before standardization (displayed using the default window settings) are shown in the first row. The second row shows the same slices after standardization displayed by using the ''standard'' window setting from Table 1. In this case, training was done utilizing the SE Pd image data acquired on a scanner of the same brand in our hospital.

Cross and Mixed Training

The example illustrated in Fig. 10 shows that the method is robust against small variations in protocol settings. The first row of Fig. 10 shows three SE Pd studies before standardization (displayed with the default window setting). The second row shows the same slices after standardization with the parameters that were derived from a training set of an identical SE Pd protocol. These images are displayed with the ''standard'' window setting for SE Pd listed in Table 1. The third row shows the same slices after standardization with the parameters that were derived from a training set of images acquired as per an FSE Pd protocol. These images are displayed with the ''standard'' window setting for FSE Pd listed in Table 1. The main difference between the two mappings is a shift of the second mode. This is silently corrected for display when the proper window setting is applied. After both transformations and by using the corresponding window settings, the images look similar and have good brightness and contrast.

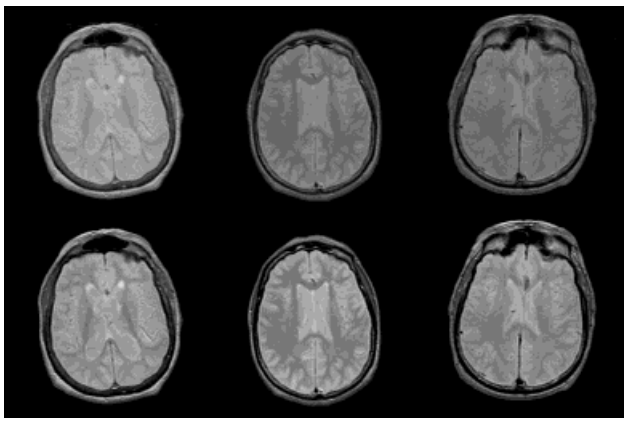

FIG. 9. Three SE Pd studies acquired as per the same protocol before standardization at default windows (first row), and after standardization displayed at a standard window (second row). The training data were acquired as per a similar protocol (with slightly different parameters) from a scanner of the same brand at a different hospital.

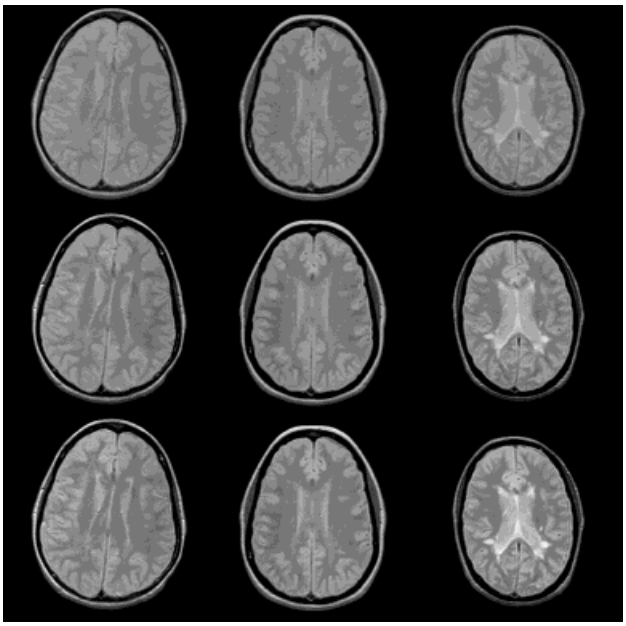

FIG. 10. Original slices from three studies acquired as per the same SE Pd protocol before standardization displayed at default windows (first row), the same slices displayed at a fixed ''standard'' window after standardization with the parameters that were derived from a training set of an identical SE Pd protocol (second row), and after standardization with the parameters that were derived from a training set of images acquired as per an FSE Pd protocol displayed at the ''standard'' window for FSE Pd (third row).

The training based on a set of images of a given protocol P and body region D makes the standardizer tuned tightly to the images in $\mathscr{Y}_{PD}$. This is due to the fact that the variation among histograms of images in $\mathscr{Y}_{PD}$ is less than that of images taken from $\mathscr{Y}_{PD}$ and $\mathscr{Y}_{P'D'}$ where $P \neq P'$ and/or $D \neq D'$. However, images of different protocols and body regions can be mixed for training and still create a standardizer that fosters consistency of tissue meaning of intensities. This is illustrated in Fig. 11, where images from four different combinations of protocol and body region were utilized in the training step and in devising the standardizer. All images (FSE Pd, T_2, T_1E brain, and GRE foot) were collected into a pool and were not distinguished during the training. That is, a single parameter configuration (i.e., $pc_1 = 0$, $pc_2 = 99.8$, $s_1 = 1$, and $s_2 = 4095$) were applied to all images in the training pool. The figure quite elegantly demonstrates that the standardized images show a better consistency of brightness and contrast at fixed window displays (bottom row) than the original images displayed at a default window for each image (top row). This observation invites the question whether it is possible to devise a whole-body standardizer which will obviate the need for protocol- and body-specific standardizers. We will not pursue this investigation here.

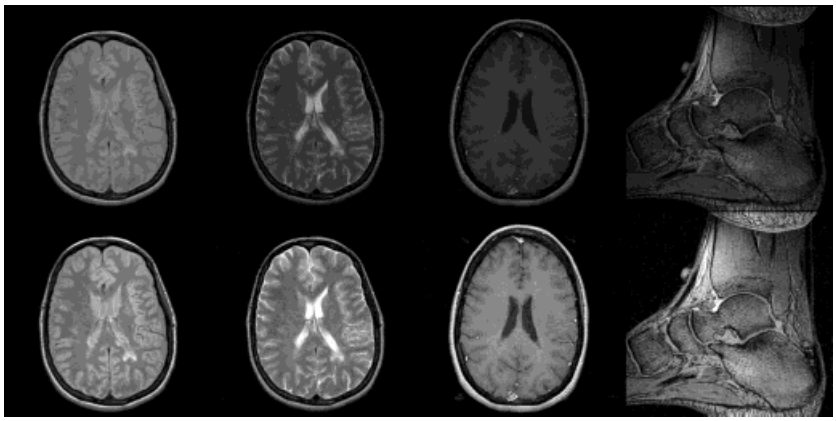

FIG. 11. Illustration of the effect of mixed training. Ten brain studies from each of FSE Pd (first column), FSE T_2 (second column), and T_1E (third column), and 10 T_1-weighted gradient-echo studies of the foot (fourth column) were used in training. The first row shows the display at a study-specific default window of one slice of a study from each of these protocols before standardization. The second row shows the same slices after standardization displayed at one fixed window.

Quantitative Comparison

In order to assess the effectiveness of the standardization method objectively, we conducted two types of quantitative tests on datasets of brain obtained from three protocols: FSE Pd, FSE T_2, and T_1E. The first test is for assessing intrapatient variation before and after standardization. In the second test, we used an algorithm (5) to segment different tissue regions and compared intensity statistics in different segmented tissue regions for assessing interpatient variation before and after standardization.

Intrapatient Variation

We used the same training datasets and parameter configurations as for qualitative comparison. The test method for all three protocols was the same. Two studies (volume images) acquired at different time instances were selected randomly for each of 15 MS patients from our database. The time distance between the two scans of the same patient varied between 1 and 6 years. For each patient, we registered the first scan to the second via a rigid transformation based on intensity value correlations. Because these patients had MS, the lesions were segmented (4,5) and removed for the purpose of comparison for minimizing variations in the two images of the same patient that may have been caused by the disease. The similarity of a pair of these registered, lesion-removed images was measured by the mean squared intensity difference between the two images normalized by the range of the image intensities of the older image in the pair. This similarity measure, denoted NMSD, was computed for every pair of volume images before and after intensity standardization for each of the three protocols.

Table 2 shows that the mean and the standard deviation of the NMSD over all studies after standardization are smaller than that before standardization. The values of NMSD for the 15 pairs of studies were compared using a paired *t*-test under the null hypothesis that there is no difference in NMSD before and after standardization. The hypothesis was rejected at a significance level of $P < 0.01$, indicating that the change in NMSD is statistically significant.

Intensities in those parts of the image that have experienced no changes due to changes in the patient over time are made considerably more repeatable after standardization than before. Further, the lower standard deviation of NMSD achieved after standardization indicates that the uniformity of intensity meaning resulting from standardization is less dependent on the patient.

Interpatient Variation

For this comparison, we randomly selected 12 FSE Pd and 12 FSE T_2 volume images from our database. For each patient only one time instance is considered. All images were then segmented into white mater (WM), gray matter (GM), cerebrospinal fluid (CSF), and MS lesion (LS) regions (5). For each of these regions in each image $\mathcal{I}_i$ 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 the standardized image $\mathcal{I}_{si}$ of $\mathcal{I}_i$ wherein normalization was done by dividing by $s_2 - s_1$. The standard deviations of the NMI values over the 12 volume images before and after standardization for both protocols and for WM, GM, and CSF tissue regions are shown in Table 3, together with the corresponding 95% confidence intervals (assuming a normal distribution for all NMI values) and percent coefficient of variation (CV) of NMI. The results indicate that for every tissue region in both protocols tested, the standard deviation of NMI values is reduced by a factor of 2 to 3 after standardization and %CV is reduced considerably. This implies that a substantially improved uniformity of tissue meaning for intensities is obtained across patient studies after standardization.

CONCLUDING REMARKS

We have described a method for standardizing MR image intensity scales in this article. Our goal was to devise a method that post-hoc makes an intensity transformation of images that are routinely acquired, without requiring spe-

Table 2
Mean and Standard Deviation of Normalized Mean Squared Differences (NMSD) Before and After Standardization for 15 Pairs of Studies and for Three Different Protocols

	FSE Pd		FSE T_2		T_1E	
	Mean	SD	Mean	SD	Mean	SD
Before	0.0099	0.0094	0.0093	0.0085	0.0025	0.0018
After	0.0039	0.0055	0.0036	0.0050	0.0020	0.0018
P-value	0.0094		0.0036		0.0450	

Each pair represents the studies obtained for the same patient at two time instances. The *P*-values of paired *t*-tests are also shown.

Table 3
Standard Deviation, the 95% Confidence Interval (CI) of the Normalized Mean Intensity (NMI) Values and % Coefficient of Variation (CV) of Different Tissues in FSE Pd and T_2 Images

	FSE Pd				FSE T_2			
	SD	CI low	CI high	%CV	SD	CI low	CI high	%CV
WM								
Before	0.1047	0.6572	0.7757	14.61	0.0580	0.3580	0.4236	14.83
After	0.0113	0.5523	0.5651	2.03	0.0091	0.3456	0.3559	2.59
GM								
Before	0.1144	0.7189	0.8483	14.59	0.0643	0.4124	0.4852	14.33
After	0.0067	0.6230	0.6306	1.07	0.0119	0.4502	0.4636	2.60
CSF								
Before	0.0884	0.5323	0.6323	15.18	0.0791	0.4564	0.5458	15.78
After	0.0172	0.5309	0.5504	3.18	0.0348	0.6596	0.6989	5.12

cialized acquisition protocols and calibration phantoms. The standardizing method aims at achieving consistency of tissue meaning of intensities by devising a transformation that is specific to a given MRI protocol and body region. The basic idea is to deform the histogram of a given volume image so that it matches a ''standard'' histogram for that protocol-body-region group. The parameters of the standard histogram are learned in a training step. We have provided theoretical guidelines, and a practical demonstration of how to utilize them, for selecting the values of the parameters of the method and proved that lossless intensity transformation and order is guaranteed, if choices are made as per guidelines.

The choice of the actual landmarks is an important factor. Other landmarks (e.g., median instead of the mode of the second hump) can be used that are more suitable, or more sophisticated curve fitting can be done to make the histograms more similar. We mention some possible changes to the basic standardization method that can be incorporated to make it better fit the type of input data and the actual application: 1) use more histogram landmarks, such as quartiles and deciles, 2) use polynomial functions to ''stretch'' the histogram segments, and 3) use spline-fitting techniques instead of segment-by-segment linear ''stretching.''

We have presented two types of studies to assess the degree of uniformity of intensity meaning achieved after standardization. In qualitative studies, we have shown through image display examples for several protocols and body regions that the consistency of the brightness level and contrast of images is considerably improved after standardization. This permits standardizing and fixing ''windows'' by protocol, body region, and tissue regions. This will minimize or eliminate the human interaction required in the per-case manual window adjustments that are currently required in visualizing MR images on physician viewing stations. This may also help in filming MR studies with a uniform appearance of the images. In quantitative studies, we have assessed the scanner-dependent intra- and interpatient intensity variations and demonstrated that these are substantially reduced after standardization. In a subjective examination of the image displays from over 100 studies, we have not come across any case where the method seemed to have failed. A formal observer study is currently under way to assess this component.

The method is simple, fast, easy to implement, and completely automatic. In a Picture Archiving and Communication System, it can be incorporated 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. It can even be built into the MR scanner to automatically produce images with the standard scale.

Our preliminary studies indicate that image analysis and tissue segmentation methods—our main motivation to undertake this research—are considerably improved in terms of their constancy of parameter settings and their degree of automation. With standardization, numerical meaning is achieved and, hence, numerical diagnosis and study of diseases may become possible. For example, for MS studies we may be able to specify an interval $[t_1, t_2]$ of the standardized Pd values which correspond to normal WM tissue, another interval $[t_3, t_4]$ corresponding to WM that is normal appearing but that is actually affected by the disease. Our preliminary results indicate that when the values of the imaged parameter (T_2, Pd, etc.) do not intrinsically overlap for the different tissues, then such a numerical characterization may be possible. We use the standardization method as the first step in all MR image visualization and analysis tasks in which we are currently engaged.

Figures 10 and 11 indicate a possible simplification of our approach. Instead of devising the standardization transform by protocol and body region, it may be possible to devise them based only on the type of the histogram (Fig. 1). That is, the protocol- and body-region-specific training and transformation will then be obviated, hopefully achieving consistency of intensity meaning independent of the protocol and body region. If we match the shoulder of the single mode in Fig. 1b with the second mode in Fig. 1a for images with different types of histograms, it may even be possible to devise standardization schemes that are independent of the histogram type.

ACKNOWLEDGMENTS

The authors thank Drs. Robert Grossman, Bruce Hirsch, and David Hackney for the MRI data sets utilized in this research.

REFERENCES

1. Bezdek JC, Hall LO, Clarke LP. Review of MR image segmentation techniques using pattern-recognition. Med Phys 1993;20:1033–1048.
2. Clarke LP, Velthuizen RP, Phuphanich S, Schellenberg JD, Arrington JA, Silbiger M. MRI: stability of three supervised segmentation techniques. Magn Reson Imaging 1993;11:95–106.
3. Kikinis R, Shenton ME, Gerig G, Martin J, Anderson M, Metcalf D, Guttmann CRG, McCarley RW, Lorensen W, Cline H, Jolesz FA. Routine quantitative analysis of brain and cerebrospinal-fluid spaces with MR imaging. J Magn Reson Imaging 1992;2:619–629.
4. Samarasekera S, Udupa JK, Miki Y, Grossman RI. A new computer-assisted method for the quantification of enhancing lesions in multiple sclerosis. J Comput Assist Tomogr 1997;21:145–151.
5. Udupa JK, Wei L, Samarasekera S, Miki Y, van Buchem MA, Grossman RI. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. IEEE Trans Med Imaging 1997;16:598–609.
6. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. Radiology 1998;209P:581–582.
7. Edelstein WA, Bottomley PA, Pfeifer LM. A signal-to-noise calibration procedure for NMR imaging systems. Med Phys 1984;11:180–185.
8. Yamamoto T, Nambu T, Miyasaka K, Morita Y. Accurate and practical calibration of MR signal intensities by the new transmission amplitude method: application of the numerical diagnosis to MRI. Radiology 1998;209P:582.
9. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. Technical Report MIPG.250, Medical Image Processing Group, Department of Radiology, University of Pennsylvania, January 1999.
10. Wendt RE. Automatic adjustment of contrast and brightness of magnetic resonance images. J Digit Imaging 1994;7:95–97.