Inter-Station Intensity Standardization for Whole-Body MR Data

Oleh Dzyubachyk,[1]* Marius Staring,[1] Monique Reijnierse,[1] Boudewijn P. F. Lelieveldt,[1,2] and Rob J. van der Geest[1]

Purpose: To develop and validate a method for performing inter-station intensity standardization in multispectral whole-body MR data.

Methods: Different approaches for mapping the intensity of each acquired image stack into the reference intensity space were developed and validated. The registration strategies included: "direct" registration to the reference station (Strategy 1), "progressive" registration to the neighboring stations without (Strategy 2), and with (Strategy 3) using information from the overlap regions of the neighboring stations. For Strategy 3, two regularized modifications were proposed and validated. All methods were tested on two multispectral whole-body MR data sets: a multiple myeloma patients data set (48 subjects) and a whole-body MR angiography data set (33 subjects).

Results: For both data sets, all strategies showed significant improvement of intensity homogeneity with respect to vast majority of the validation measures ($P < 0.005$). Strategy 1 exhibited the best performance, closely followed by Strategy 2. Strategy 3 and its modifications were performing worse, in majority of the cases significantly ($P < 0.05$).

Conclusions: We propose several strategies for performing inter-station intensity standardization in multispectral whole-body MR data. All the strategies were successfully applied to two types of whole-body MR data, and the "direct" registration strategy was concluded to perform the best. **Magn Reson Med 77:422–433, 2017. © 2016 The Authors Magnetic Resonance in Medicine published by Wiley Periodicals, Inc. on behalf of International Society for Magnetic Resonance in Medicine**

Key words: whole-body MRI; multispectral MRI; multistation acquisition; intensity standardization

INTRODUCTION

Whole-body MR is gaining increasing interest as a noninvasive method for diagnosing systemic diseases, e.g., cancer or diseases of the circulatory system (1,2). Due to the limited size of the magnet coil, whole-body acquisition is typically performed in several stations (3). Here, by *station* we mean a region of coverage for a single acquisition (4). In addition, most of the modern whole-body scanning protocols provide multiple complementary contrast channels. Image intensity of such an acquired set of multispectral 3D image stacks suffers from two types of intensity variations (5): intensity inhomogeneity within each station (*class I*; bias) and inter-scan signal intensity variation (*class II*).

While many image postprocessing methods were developed for correction of the first type of inhomogeneity (6–8), the literature on correction of the inter-scan intensity variation is rather sparse. Most of the published algorithms in this category, typically referred to as *intensity standardization* methods, were developed for brain images, although several authors show that their methods can also be applied for MR data of other anatomical regions (knee MR, etc.). In particular, Nyúl and Udupa (9) present a parametric intensity standardization method where the parameters are learned during the training stage. Weisenfeld and Warfield (10) developed a standardization method based on minimizing the Kullback–Leibler divergence between two histograms that can also be used for multispectral data. Schmidt (11) presented a method that matches intensity of each image to that of the target image using a smooth multiplicative field and can perform both inter-slice and inter-volume intensity correction. Madabhushi and Udupa (12) investigated the influence of prior bias correction on the intensity standardization and concluded that it provides a small positive effect. Bergeest and Jäger (13) compared five different intensity standardization methods and concluded superiority of more advanced approaches. Jäger and Hornegger (5) developed a method that can be used for standardizing image intensities between two different whole-body scans. Iglesias et al. (14) and Jog et al. (15) presented approaches that are considerably different from all the other methods as they perform standardization based on MR physics acquisition equations.

Commercial products from three largest MRI hardware vendors: CLEAR (Philips), PURE (GE Healthcare), and Prescan Normalize (Siemens) implement the same idea of performing B_1 field correction by acquiring a prescan (16). Such correction, although not being a primary target, typically also results in improved inter-volume intensity homogeneity. Another method that is based on

[1]Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands.

[2]Intelligent Systems Department, Delft University of Technology, Delft, The Netherlands.

Grant sponsor: Dutch Technology Foundation STW (Stichting Technische Wetenschappen); Grant number: 10894; Grant sponsor: Netherlands Organization for Scientific Research (NWO); Grant number: VENI 639.021.124 (to M.S.).

*Correspondence to: Oleh Dzyubachyk, Division of Image Processing, Department of Radiology, Leiden University Medical Center, Postbus 9600, Leiden, The Netherlands. E-mail: o.dzyubachyk@lumc.nl

Received 21 July 2015; revised 20 November 2015; accepted 28 November 2015

DOI 10.1002/mrm.26098
Published online 1 February 2016 in Wiley Online Library (wileyonlinelibrary.com).

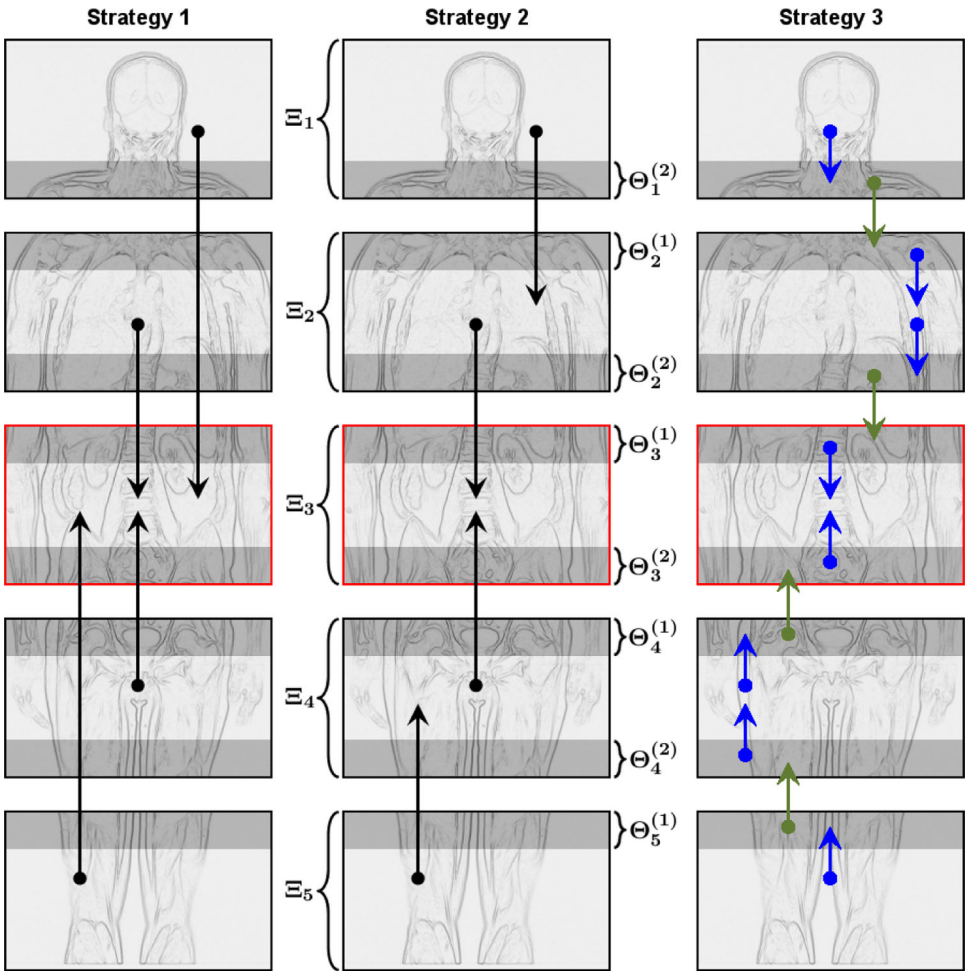

FIG. 1. Schematic illustration of different registration strategies for inter-station intensity standardization. The reference station $\Xi_{\text{ref}=3}$ is marked in red. Black arrows represent registration between $\mathcal{H}(\Xi_{s_1})$ and $\mathcal{H}(\Xi_{s_2})$; blue ones — between $\mathcal{H}(\Xi_s)$ and $\mathcal{H}(\Theta_s^{(1;2)})$; and green ones — between $\mathcal{H}(\Theta_s^{(1)})$ and $\mathcal{H}(\Theta_{s\pm1}^{(2)})$. Here Ξ_s, $\Theta_s^{(1)}$, and $\Theta_s^{(2)}$, respectively, denote the entire station and its overlap region with the previous and the next station, and $\mathcal{H}(:)$ are the histograms of these regions. Note, that the strategies $\mathcal{S}_{3a}$ and $\mathcal{S}_{3b}$ impose constraints on the registrations represented by the blue arrows.

estimation of real parameter maps during acquisition was developed by Warntjes et al. (17).

However, all the aforementioned methods concentrate on standardizing the intensities between different scans corresponding to the same anatomical location. At the same time, there is a clear need for standardizing intensity between different stations in the setting of whole-body MR imaging (3,18,19). This type of intensity standardization is of significantly higher complexity than that between different images of the same anatomical region because of large variation of intensity distributions of different stations throughout the entire body. Robinson et al. (18) developed such an inter-station intensity inhomogeneity correction method, but only for a very specific case, as their approach is based on the typical histogram appearance of a T_1w image. Thus, it can hardly be extended to different data types. Recently, Romu et al. (20) developed a method for correcting intensity inhomogeneity of whole-body two-point Dixons MR volumes and used it in several subsequent studies (21). The presented approach improves, in particular, inter-station intensity homogeneity due to its multiscale nature.

In this article, we present a registration-based method for standardization of image intensities between different stations of a whole-body MR acquisition protocol. Within our approach, we develop several registration strategies and validate their performance on two different types of multispectral whole-body MR data. In addition, we investigated the impact of prior bias correction on the quality of the subsequent intensity standardization step. Our method is general in the sense that it does not make assumptions about the type of the particular whole-body data set (number and type of available contrast channels, etc.). To our knowledge, except for the aforementioned works of Robinson et al. (18), Romu et al. (20), and our earlier conference publication (22), this problem has not yet been addressed in the literature.

METHODS

Let N_s and N_c denote the total number of acquired stations and the number of contrast channels, respectively. In the following, without loss of generality, we assume $N_c = 2$. By $\mathbf{I}_s = (I_{s,1}, I_{s,2})$ we denote the image intensity within each station Ξ_s for $s = \overline{1, N_s}$. Finally, let $\Theta_s^{(1)}$ and $\Theta_s^{(2)}$, respectively, denote the overlap regions of Ξ_s with the previous and the next station, and $\Omega = \cup_{s=1}^{N} \Xi_s$ be the entire volume. Figure 1 illustrates the introduced notations on a sample whole-body volume. Note, that here we work with data stacks in the coordinate system associated with the acquired images. Hence, although $\Theta_s^{(1)}$ and $\Theta_{s-1}^{(2)}$ represent the same region in the world coordinates, they are two different regions in the image coordinates.

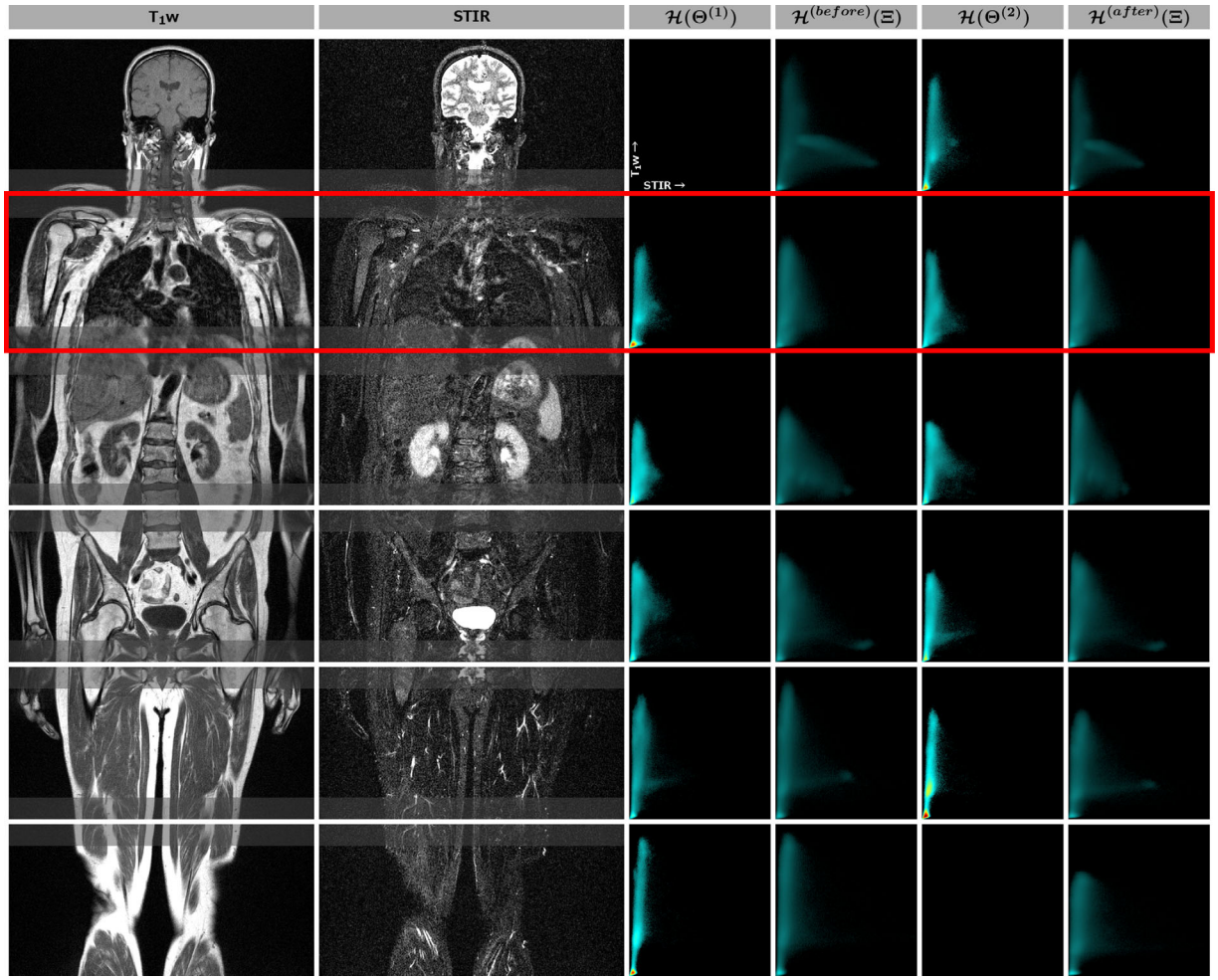

FIG. 2. Joint intensity histograms corresponding to different stations for one multiple myeloma patient. For each station, illustrated by a single slice, corresponding histograms of the entire station before and after standardization and the histograms of the overlap regions (shaded) are shown. The reference station is marked with a red border. Here $\mathcal{H}(\Xi)$, $\mathcal{H}(\Theta^{(1)})$, and $\mathcal{H}(\Theta^{(2)})$ are, respectively, the histograms of the entire station and its overlap region with the previous and the next station.

Data Description

For validation of the developed methodology, we apply it to two different types of whole-body MR data: a multiple myeloma patients data set and a whole-body MR angiography (WB-MRA) data set.

Data Set 1: Multiple Myeloma Patients

Forty-eight whole-body scans of 16 multiple myeloma subjects were acquired on a commercial human whole-body 1.5T MR system (Philips Intera, Philips Medical Systems, Best, The Netherlands) at the Leiden University Medical Center (Leiden, The Netherlands) using the standard quadrature body coil. For retrospective anonymized studies from routine patient care in Dutch University Medical Centers, institutional review board approval is not required. Each subject was scanned between one and five times, with an approximate 6 months interval between consecutive follow-up scans. Each whole-body volume was acquired in 6 stations in an interleaved manner: first the T_1w image stack, then the STIR or the T_2-STIR, after which the scanner bed was moved to the next station. The size of the overlap region between two neighboring stations was equal to $5.38 \pm 0.05\%$ of their total size. Typical scan parameters were: coronal slice orientation, pulse repetition time 520–755 ms (T_1w) and 2290–3054 ms (STIR; T_2-STIR), echo time 17.5 ms (T_1w) and 64 ms (STIR; T_2-STIR), field-of-view 530×530 mm^2, data matrix 512×512, in-plane resolution 1.03×1.03 mm^2, 5 mm slice thickness, 42–56 slices, 0.5 mm inter-slice gap, 45–60 minutes total acquisition time. A typical example of a whole-body volume of a multiple myeloma patient is shown in Figure 2.

Data Set 2: Whole-Body MRA

Thirty-three subjects were selected from a population-based cohort (23) and scanned using the standard quadrature body coil on a 1.5T scanner (Philips Gyroscan Intera, Philips Medical Systems, Best, The Netherlands) at the Uppsala University Hospital (Uppsala, Sweden). The study was approved by the Ethics Committee of the University of Uppsala and the participants gave informed consent. Imaging was performed using a WB-MRA protocol with the subject placed in supine

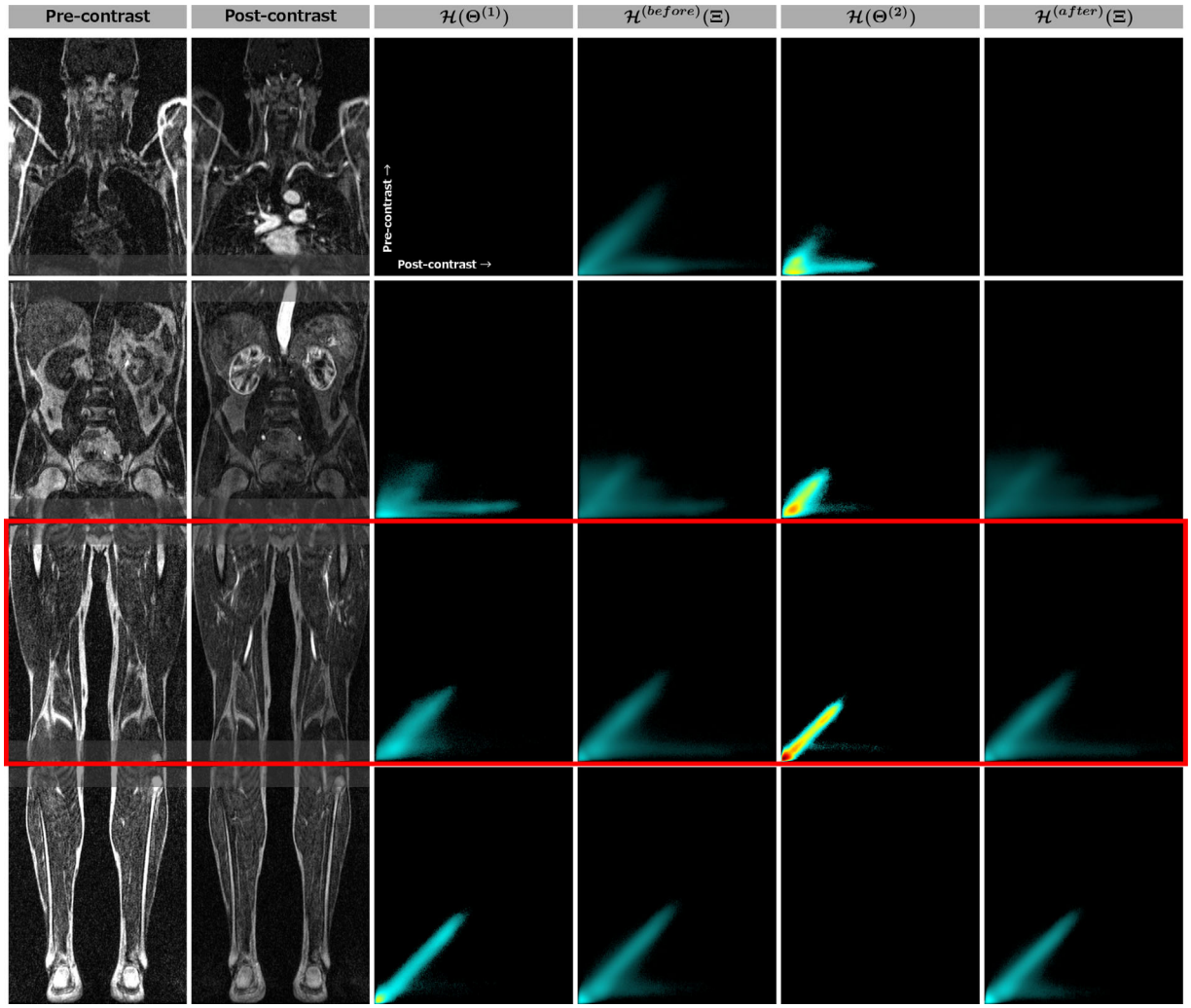

FIG. 3. Joint intensity histograms corresponding to different stations for one whole-body MRA subject. For each station, illustrated by a single slice, corresponding histograms of the entire station before and after standardization and the histograms of the overlap regions (shaded) are shown. The reference station is marked with a red border. Note, that the first station was excluded from further analysis due to severe fold-over artefacts. Here $\mathcal{H}(\Xi)$, $\mathcal{H}(\Theta^{(1)})$, and $\mathcal{H}(\Theta^{(2)})$ are, respectively, the histograms of the entire station and its overlap region with the previous and the next station.

position, feet-first, on the scanner bed with an extension. The arms were positioned above the patient's head.

A 3D RF-spoiled T_1w gradient echo acquisition was performed at 4 stations, beginning with the fourth station (ankle), before contrast injection. Thereafter, 40-mL of gadodiamide (Omniscan; GE Healthcare, Oslo, Norway) was injected intravenously with an automated injector (MR Spectris; Medrad, Pittsburgh, PA) at a rate of 0.6-mL/s in 67 s and flushed with 20-mL of a saline solution. The stations were scanned in reversed order during the contrast administration, starting with the first station (supra-aortic arteries and the thoracic aorta). The average total scan time was 87 s (17 s per station), including instructions for breath-holding and table movement (4 s for each of the three movements).

The sequence parameters were: coronal slice orientation, TR/TE/flip angle 2.5 ms/0.94 ms/30°; bandwidth 781.3 Hz/pixel; matrix size 256×256; number of slices = 60; slices thickness = 4 mm; 80% scan percentage. The acquired voxel size was $1.76 \times 1.76 \times 4.0$ mm^3, which was reconstructed by zero-filling to $0.88 \times 0.88 \times 2.0$ mm^3. Overlap between consecutive stations in the feet-head direction was $4.88 \pm 1.91\%$ of their total size. The first station (head region) was suffering from severe fold-over artefacts due to the positioning of the arms. Thus, in the following, we excluded this station from further analysis for sake of purity in the experiments. In this work, we treated the pre- and post-contrast images as two contrast channels of the multispectral volume. A typical example of a WB-MRA volume is shown in Figure 3.

Registration of Joint Intensity Histograms

For performing intensity standardization, we use an approach based on registering the joint intensity histogram of the current station to that of the reference one with subsequent intensity warping (5). All the processing and analysis is performed on the logarithm-transformed histograms, scaled in such a way that they become

probability density functions. In the following, $\mathcal{H}^*$ and $\mathcal{H}$, respectively, denote the intensity histogram and its logarithm-transformed counterpart. The histograms are calculated using all the occurring intensity values on each contrast channel (9). Similar to the aforementioned work, we truncate high intensity values using a threshold. The latter is chosen as the last point whose histogram value (on the logarithmic scale) was higher than 10% of the histogram maximum. In the following, by $\mathcal{H}(\Omega)$, $\mathcal{H}(\Xi_s)$, $\mathcal{H}(\Theta_s^{(1)})$, and $\mathcal{H}(\Theta_s^{(2)})$ we denote the histograms of the entire volume, of a single station, and of its overlap regions with the previous and the next station, respectively.

For aligning two histograms, an intensity-based image registration framework was used. However, instead of applying the framework to the imaging data, registration is performed on the histograms, denoted as $\mathcal{H}_{\text{fix}}$ and $\mathcal{H}_{\text{mov}}$. Image registration is formulated as an optimization problem, where a cost function is minimized with respect to the transformation parameters $\boldsymbol{\mu}$. Normalized correlation NC was used as the cost function to drive the registration. In other words, we solve

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \ \text{NC}(\boldsymbol{T}_{\boldsymbol{\mu}}; \mathcal{H}_{\text{fix}}, \mathcal{H}_{\text{mov}}), \qquad [1]$$

where $\boldsymbol{T}_{\boldsymbol{\mu}} = \boldsymbol{T}(\mathcal{H}_{\text{fix}}; \mathcal{H}_{\text{mov}})$ is the coordinate transformation parameterized by $\boldsymbol{\mu}$, and $\hat{\boldsymbol{\mu}}$ is the optimal transformation parameter setting. Let $\mathbf{I}_{\text{fix}}$ and $\mathbf{I}_{\text{mov}}$ be the multispectral images corresponding to $\mathcal{H}_{\text{fix}}$ and $\mathcal{H}_{\text{mov}}$, respectively. The calculated transformation $\boldsymbol{T}_{\hat{\boldsymbol{\mu}}}$ is applied to map $\mathbf{I}_{\text{mov}}$ into the intensity space of $\mathbf{I}_{\text{fix}}$:

$$\mathbf{I}_{\text{mov}}^{\text{corr}} = \mathbf{I}_{\text{fix}}\big(\boldsymbol{T}_{\hat{\boldsymbol{\mu}}}(\mathbf{I}_{\text{mov}})\big). \qquad [2]$$

For alignment, an affine registration using a multiresolution approach for both histograms and transformations with a Gaussian image pyramid is performed using two resolutions. Assuming affine transformation eliminates the need to perform additional regularization or to use optimization constraints.

An adaptive stochastic gradient descent optimizer (24) was used for solving Eq. [1]. All registrations were performed with the open source registration software `elastix` (25). The parameters were optimized per test data set, and were kept fixed for all registrations between different types of histograms belonging to the same data set. In practice, the `MaximumStepLength` was the only parameter that needed to be optimized, and for both data sets the same value `MaximumStepLength` = 5 was found to be optimal. A sample parameter file that was used for both data sets can be downloaded from `http://elastix.bigr.nl/wiki/index.php/Parameter_file_database`.

Registration Strategies

In this section, we suggest several strategies for interstation intensity standardization. An overview of the described registration strategies is given in Figure 1. For each strategy, the joint intensity histogram of each of the stations is registered to that of the reference station Ξ_{ref}. The latter is chosen separately for each of the two validation data sets as the one whose histogram has the minimum average distance to the histograms of the remaining stations. In this way, the second (chest and shoulders) and the third station (upper legs) were, respectively, chosen for Data Sets 1 and 2. Let $\mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_s))$ denote the final transformation between $\mathcal{H}(\Xi_s)$ and $\mathcal{H}(\Xi_{\text{ref}})$. Several registrations strategies can be applied to register $\mathcal{H}(\Xi_s)$ to $\mathcal{H}(\Xi_{\text{ref}})$.

- **"Strategy 1"** ($\boldsymbol{\mathcal{S}_1}$). All the histograms $\mathcal{H}(\boldsymbol{\Xi}_s)$ are directly registered to the histogram of the reference station $\mathcal{H}(\boldsymbol{\Xi}_{\text{ref}})$; see Figure 1a:

$$\mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_s)) = \boldsymbol{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_s)). \qquad [3]$$

- **"Strategy 2"** ($\boldsymbol{\mathcal{S}_2}$). All the histograms $\mathcal{H}(\boldsymbol{\Xi}_s)$ are progressively registered to the histogram of the next station, in both directions from the reference station $\mathcal{H}(\boldsymbol{\Xi}_{\text{ref}})$; see Figure 1b. The final transformation on each station is obtained by composing the corresponding transforms:

$$\begin{aligned}\mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_s)) = \mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_{s\pm1})) \\ \circ \boldsymbol{T}(\mathcal{H}(\Xi_{s\pm1}), \mathcal{H}(\Xi_s)).\end{aligned} \qquad [4]$$

- **"Strategy 3"** ($\boldsymbol{\mathcal{S}_3}$). Each histogram $\mathcal{H}(\boldsymbol{\Xi}_s)$ is registered to $\mathcal{H}(\boldsymbol{\Theta}_s^{(1)})$ and/or $\mathcal{H}(\boldsymbol{\Theta}_s^{(2)})$, and also $\mathcal{H}(\boldsymbol{\Theta}_s^{(1)})$ and $\mathcal{H}(\boldsymbol{\Theta}_{s-1}^{(2)})$ are coregistered; see Figure 1c. Similar to the previous case, the final transformation for each station is obtained by transform composition:

$$\mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_s)) = \begin{cases} \mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_{s+1})) \circ \boldsymbol{T}(\mathcal{H}(\Xi_{s+1}), \mathcal{H}(\Theta_{s+1}^{(1)})) \circ \\ \quad \boldsymbol{T}(\mathcal{H}(\Theta_{s+1}^{(1)}), \mathcal{H}(\Theta_s^{(2)})) \circ \boldsymbol{T}(\mathcal{H}(\Theta_s^{(2)}), \mathcal{H}(\Xi_s)), & s < \text{ref}, \\ \mathcal{T}(\mathcal{H}(\Xi_{\text{ref}}), \mathcal{H}(\Xi_{s-1})) \circ \boldsymbol{T}(\mathcal{H}(\Xi_{s-1}), \mathcal{H}(\Theta_{s-1}^{(2)})) \circ \\ \quad \boldsymbol{T}(\mathcal{H}(\Theta_{s-1}^{(2)}), \mathcal{H}(\Theta_s^{(1)})) \circ \boldsymbol{T}(\mathcal{H}(\Theta_s^{(1)}, \mathcal{H}(\Xi_s)), & s > \text{ref}. \end{cases} \qquad [5]$$

- **"Strategy 3a"** ($\boldsymbol{\mathcal{S}_{3a}}$). Similar to $\mathcal{S}_3$, but in this case the registrations between $\mathcal{H}(\boldsymbol{\Xi}_s)$ and $\mathcal{H}(\boldsymbol{\Theta}_s^{(1;2)})$ are regularized for all the stations:

$$\hat{\boldsymbol{\mu}}_1 = \arg\min_{\boldsymbol{\mu}} \sum_{s=2}^{N_s} \omega_s \, \text{NC}(\boldsymbol{T}_{\boldsymbol{\mu}}; \mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(1)})), \qquad [6]$$

$$\hat{\boldsymbol{\mu}}_2 = \arg\min_{\boldsymbol{\mu}} \sum_{s=1}^{N_s-1} \omega_s \, \text{NC}(\boldsymbol{T}_{\boldsymbol{\mu}}; \mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(2)})), \qquad [7]$$

with equal weights: $\omega_{\{s=\overline{1,N_s-1}\}} = 1/(N_s - 1)$.

- **"Strategy 3b"** ($\boldsymbol{\mathcal{S}_{3b}}$). Similar to $\boldsymbol{\mathcal{S}_{3b}}$, but the weights are linearly decreasing with respect to the current

station s_*: $\boldsymbol{\omega}_{\{s=\overline{1,N_s-1}\}}=1-|s-s_*|/(N_s-1)$. The weights are normalized in such a way that $\sum_{s=1}^{N_s-1}\boldsymbol{\omega}_s=1$.

Strategy $\mathcal{S}_1$ is the most straightforward one, but suffers from low similarity between the histograms as they correspond to different anatomical locations and thus can have significantly varying amount of different tissues. Strategy $\mathcal{S}_2$ can potentially overcome this issue as neighboring histograms are more alike. However, it also has a drawback that the final transform is obtained as a composition of registrations. Thus, it is sensitive to misregistrations at early stages as they get propagated. $\mathcal{S}_3$ performs registration taking into account the overlap regions. Ideally, if the images would not be affected by a bias field, matching the corresponding histograms of two overlapping stations would potentially provide a very accurate intensity standardization. However, presence of bias and difference in tissue composition between the entire station and the regions where it overlaps with its neighbors also require coregistering $\mathcal{H}(\Xi_s)$ and $\mathcal{H}(\boldsymbol{\Theta}_s^{(1;2)})$. Two modifications of this strategy are built on an assumption that, if the tissue composition of Ξ_s and $\boldsymbol{\Theta}^{(1;2)}$ were the same, the difference between the histograms would be explained merely by the bias due to hardware imperfection. Thus, the transform is computed jointly from all the stations, with a possibility to set different weights for each particular station.

Bias Correction

Bias correction was performed as a preprocessing step using the N4 method of Tustison et al. (26). The results with prior bias correction are marked with a superscript "(+)" sign, e.g. "$\mathcal{S}_1^{(+)}$". Success of the bias correction step was assessed by calculating the entropy of the joint intensity histogram. More precisely, we calculate the average histogram entropy of all stations, and then take the ratio of the aforementioned measure after and before bias correction. Values lower than unity indicate lower bias in the corrected data, and vice versa. For both our test data sets, the corresponding ratios were equal to 0.98 ± 0.00.

Distance Between Histograms for Validation

Inspired by Jäger and Hornegger (5), we use the Jeffrey divergence

$$d_J\left[\mathcal{H}_{(1)},\mathcal{H}_{(2)}\right]=\sum_{i=1}^{L}\left(h_i^{(1)}\log\frac{2h_i^{(1)}}{h_i^{(1)}+h_i^{(2)}}+h_i^{(2)}\log\frac{2h_i^{(2)}}{h_i^{(1)}+h_i^{(2)}}\right) \quad [8]$$

as a similarity measure between two histograms, where L is the number of elements in each joint intensity histogram. Note, that this measure is a metric, and, in particular, satisfies the triangle inequality (27).

Validation Measures

Five following measures were used for validating the inter-station intensity homogeneity and its improvement:

1. Average distance between the histograms of each station and that of the entire volume:

$$D_{\text{vol}}=\overline{d_J}[\mathcal{H}(\Xi_s),\mathcal{H}(\Omega)]=\frac{1}{N_s}\sum_{s=1}^{N_s}d_J[\mathcal{H}(\Xi_s),\mathcal{H}(\Omega)]. \quad [9]$$

2. Average distance between the histograms of each station and that of the reference station:

$$D_{\text{ref}}=\overline{d_J}[\mathcal{H}(\Xi_s),\mathcal{H}(\Xi_{\text{ref}})]=\frac{1}{N_s-1}\sum_{s\neq\text{ref}}d_J[\mathcal{H}(\Xi_s),\mathcal{H}(\Xi_{\text{ref}})]. \quad [10]$$

3. Ratio of the per-station histogram distance to the entire volume, averaged for all stations:

$$R_{\text{vol}}=\prod_{s=1}^{N_s}\left(\frac{d_J^{(\text{after})}[\mathcal{H}(\Xi_s),\mathcal{H}(\Omega)]}{d_J^{(\text{before})}[\mathcal{H}(\Xi_s),\mathcal{H}(\Omega)]}\right)^{\frac{1}{N_s}}. \quad [11]$$

4. Ratio of the per-station histogram distance to the reference station, averaged for all stations except for the reference:

$$R_{\text{ref}}=\prod_{s\neq\text{ref}}\left(\frac{d_J^{(\text{after})}[\mathcal{H}(\Xi_s),\mathcal{H}(\Xi_{\text{ref}})]}{d_J^{(\text{before})}[\mathcal{H}(\Xi_s),\mathcal{H}(\Xi_{\text{ref}})]}\right)^{\frac{1}{N_s-1}}. \quad [12]$$

5. Ratio of the histogram entropy $E(\mathcal{H}^*(\boldsymbol{\Omega}))$ for the entire volume $\boldsymbol{\Omega}$:

$$R_{\text{ent}}=\frac{E^{(\text{after})}(\mathcal{H}^*(\Omega))}{E^{(\text{before})}(\mathcal{H}^*(\Omega))}. \quad [13]$$

The first four of the defined measures are distance-based, and the remaining one is entropy-based. Note, that the distance-based measures used for validation are completely different in nature from the normalized correlation distance metric used for performing the registration. This allows objective validation of the proposed method.

More similar histograms have lower values of D_{vol} and D_{ref}. For all R_{vol}, R_{ref}, and R_{ent}, values below unity indicate improvement of the inter-station homogeneity and vice versa. Note, that the last three measures are meaningless for the bias-corrected data, because in this case the values before the intensity standardization are based on the histogram of the corrected images, which might influence similarity between histograms. Hence, we do not calculate and report them, as well as the D_{vol} and D_{ref} for the unstandardized data for this case. Also note that, since Data Set 2 has three stations with the reference station in the middle, $\mathcal{S}_1$ and $\mathcal{S}_2$ are identical for this case. Thus, for this case we refer to them by using a combined label "$\mathcal{S}_{(1,2)}$".

Visual Quality Valitation

To better assess the performance of our method, we conducted an additional experiment in which an experienced radiologist (M.R.) visually ranked each of the stitched volumes. For each validation data set, the

Table 1
Average distance between histograms of different regions before and after intensity standardization using different registration strategies with and without bias correction.

	Raw	S_1	S_2	S_3	S_{3a}	S_{3b}
Data Set 1						
D_{vol}	19.39 ± 1.34*	**12.34 ± 1.16**	12.38 ± 1.18	13.68 ± 1.77*	14.70 ± 1.54*	14.21 ± 1.41*
D_{ref}	17.89 ± 2.05*	**10.26 ± 1.18**	10.71 ± 1.35	12.72 ± 2.06*	13.46 ± 1.70*	12.93 ± 1.61*
R_{vol}		**64.22 ± 3.78**	64.48 ± 4.04	68.92 ± 6.15*	70.13 ± 4.75*	68.70 ± 4.25*
R_{ref}		**59.02 ± 4.17**	61.10 ± 4.19	71.88 ± 9.27*	77.00 ± 6.56*	73.88 ± 5.56*
R_{ent}		89.70 ± 5.08*	87.86 ± 5.96*	**83.88 ± 6.56**	93.48 ± 5.74*	92.62 ± 5.50*
		$S_1^{(+)}$	$S_2^{(+)}$	$S_3^{(+)}$	$S_{3a}^{(+)}$	$S_{3b}^{(+)}$
D_{vol}		11.94 ± 1.14	**11.93 ± 1.15**	13.48 ± 1.53	14.15 ± 1.42	13.94 ± 1.36
D_{ref}		**9.98 ± 1.19**	10.51 ± 1.51	12.65 ± 1.85	13.26 ± 1.63	12.96 ± 1.62
	Raw	$S_{(1,2)}$	S_3	S_{3a}	S_{3b}	
Data Set 2						
D_{vol}	12.95 ± 2.39*	**10.40 ± 2.35**	11.08 ± 2.72	10.97 ± 2.48	10.74 ± 2.48	
D_{ref}	16.68 ± 2.79*	**12.34 ± 2.30**	13.11 ± 2.38	14.68 ± 2.40*	13.83 ± 2.24*	
R_{vol}		**86.82 ± 7.13**	87.48 ± 7.57	92.94 ± 6.38*	91.17 ± 6.00*	
R_{ref}		**73.62 ± 12.28**	78.59 ± 12.37	88.32 ± 10.29*	83.17 ± 11.19*	
R_{ent}		103.73 ± 4.39*	104.50 ± 4.59*	**99.70 ± 6.59**	100.74 ± 5.44	
		$S_{(1,2)}^{(+)}$	$S_3^{(+)}$	$S_{3a}^{(+)}$	$S_{3b}^{(+)}$	
D_{vol}		11.20 ± 2.43	11.80 ± 2.92	11.58 ± 2.59	11.47 ± 2.62	
D_{ref}		13.25 ± 2.36	14.00 ± 2.45	15.16 ± 2.34	14.53 ± 2.31	

The best performing strategy for each case is highlighted in bold. An asterisk denotes statistical difference with 5% confidence interval, with respect to: the best performing strategy, or the worst performing strategy (for the raw data), or the corresponding results without the bias correction. All the values are scaled by a factor of 100 for presentation. Table columns represents different inter-station intensity standardization strategies described in the "Registration Strategies" section and table rows represent different validation measures defined in the "Validation Measures" section.

volumes were assessed per-channel and consisted of the most representative slice and the maximum intensity projection. Each volume pair, before and after applying our inter-station intensity standardization algorithm (the $\mathcal{S}_1$ strategy was used in this case), was placed next to each other (left and right) in random order. The observer, blinded to the order in which the volumes were placed, was instructed to rate the intensity homogeneity of the volume pairs using one of the three options: (*i*) left is better; (*ii*) right is better; (*iii*) no difference in image quality between left and right.

RESULTS

In this section, we report results of inter-station intensity standardization on both validation data sets. In addition, we investigate the influence of prior bias correction on the quality of intensity standardization. Statistical significance was calculated by applying the two-sample Kolmogorov–Smirnov test (28). Difference in performance between strategies was assessed by comparing each of them to the best performing strategy for each particular case. For all strategies, the effect of prior bias correction

Table 2
Average Distance Between Histograms of Different Regions Before and After Registration

	Data Set 1			Data Set 2		
Histogram distance	Before	After	Ratio	Before	After	Ratio
D_{ref}	0.18 ±0.02	0.13 ±0.02	0.72 ±0.06	0.17 ±0.03	0.12 ±0.02	0.75 ±0.12
$\overline{d_J}[\mathcal{H}(\Xi_s), \mathcal{H}(\Xi_{s+1})]$	0.16 ±0.01	0.10 ±0.01	0.61 ±0.05			
$\overline{d_J}\left[\mathcal{H}(\Theta_s^{(2)}), \mathcal{H}(\Theta_{s+1}^{(1)})\right]$	0.21 ±0.02	0.10 ±0.01	0.46 ±0.06	0.27 ±0.04	0.09 ±0.01	0.32 ±0.05
$\overline{d_J}\left[\mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(1)})\right]$, $\mathcal{S}_3$	0.35 ±0.03	0.14 ±0.01	0.41 ±0.03	0.47 ±0.05	0.18 ±0.03	0.39 ±0.06
$\mathcal{S}_{3a}$		0.16 ±0.02	0.46 ±0.04		0.19 ±0.03	0.41 ±0.06
$\mathcal{S}_{3b}$		0.16 ±0.02	0.45 ±0.04		0.19 ±0.03	0.40 ±0.06
$\overline{d_J}\left[\mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(2)})\right]$, $\mathcal{S}_3$	0.28 ±0.03	0.12 ±0.01	0.41 ±0.03	0.15 ±0.04	0.10 ±0.02	0.68 ±0.08
$\mathcal{S}_{3a}$		0.13 ±0.02	0.47 ±0.04		0.10 ±0.03	0.71 ±0.07
$\mathcal{S}_{3b}$		0.13 ±0.02	0.46 ±0.04		0.10 ±0.02	0.69 ±0.07

Here D_{ref} is the average distance between the histograms of each station and that of the reference station; $\overline{d_J}$ is the average value of all pairwise histogram distances defined in the "Distance Between Histograms for Validation" section; $\mathcal{H}(\Xi_s)$, $\mathcal{H}(\Theta_s^{(1)})$, and $\mathcal{H}(\Theta_s^{(2)})$ are, respectively, the histograms of the entire station and its overlap region with the previous and the next station. Inter-station intensity standardization strategy $\mathcal{S}_3$ and its modifications $\mathcal{S}_{3a}$ and $\mathcal{S}_{3b}$ are defined in the "Registration Strategies" section.

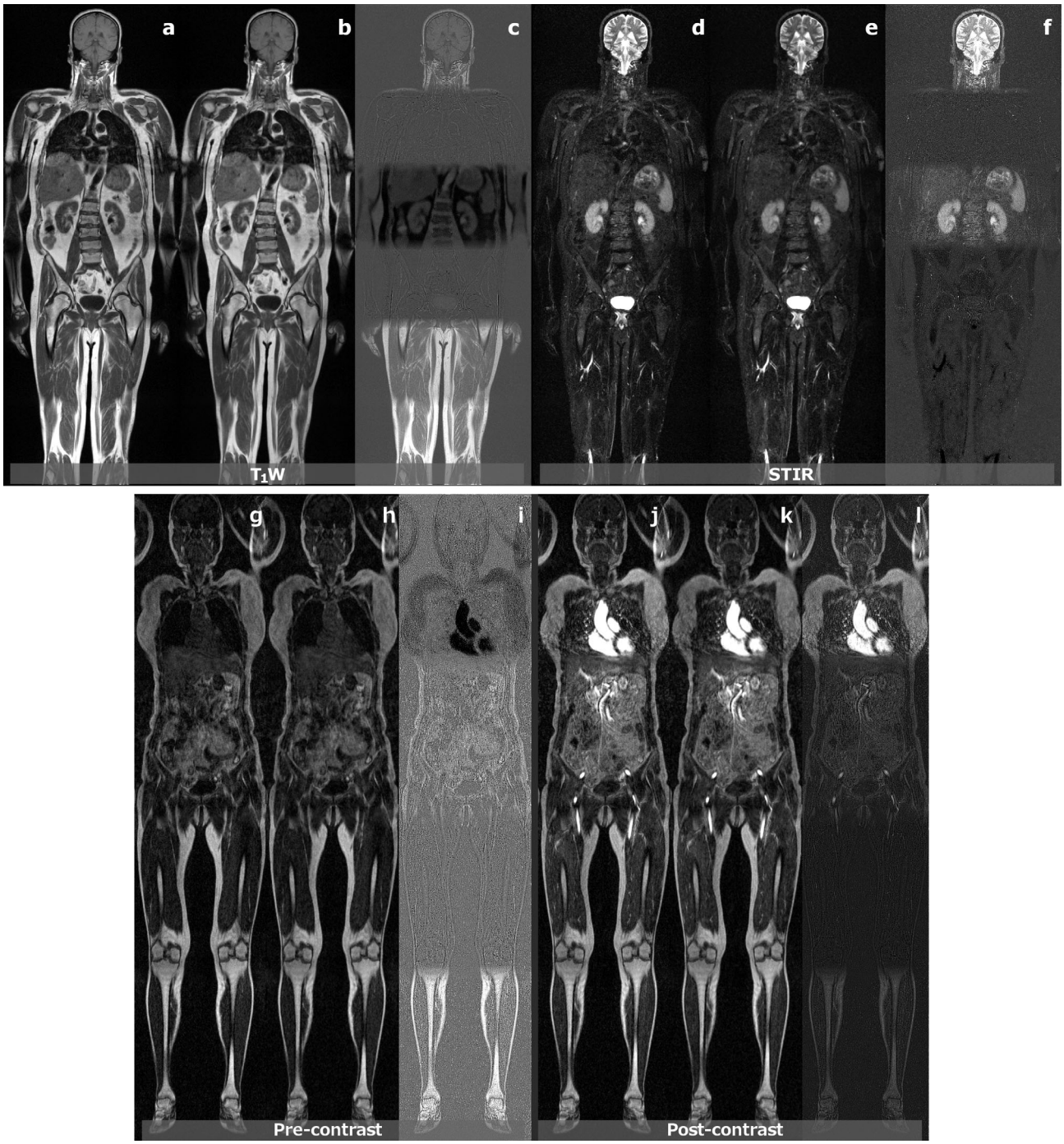

FIG. 4. Complete reconstructed volumes of the multispectral whole-body MR data before (a,d,g,j) and after the intensity standardization (b,e,h,k), and the corresponding difference images (c,f,i,l). One slice from each 3D image volume is shown. Images (a–f) correspond to Data Set 1, and images (g–l) — to Data Set 2. Intensity of all images was enhanced for visualization purposes. Note, that even though for Data Set 2 the first station was excluded from analysis, we still show it for completeness of the figure.

was assessed by comparing each particular measure with and without bias correction.

Different Strategies, Data Sets, with and without Bias Correction

Results of inter-station intensity standardization are reported in Table 1. Average distance between histograms of different regions before and after intensity standardization is reported in Table 2. Effect of the inter-station intensity standardization on reconstructed whole-body MR volumes is shown in Figure 4 and Supporting Movie S1. Figure 5 and Supporting Figures S1–S6 illustrate performance of different registration strategies in terms of R_{vol}, R_{ref}, and R_{ent} for each particular multispectral volume from both test data sets. Finally, Figure 6 shows the effect

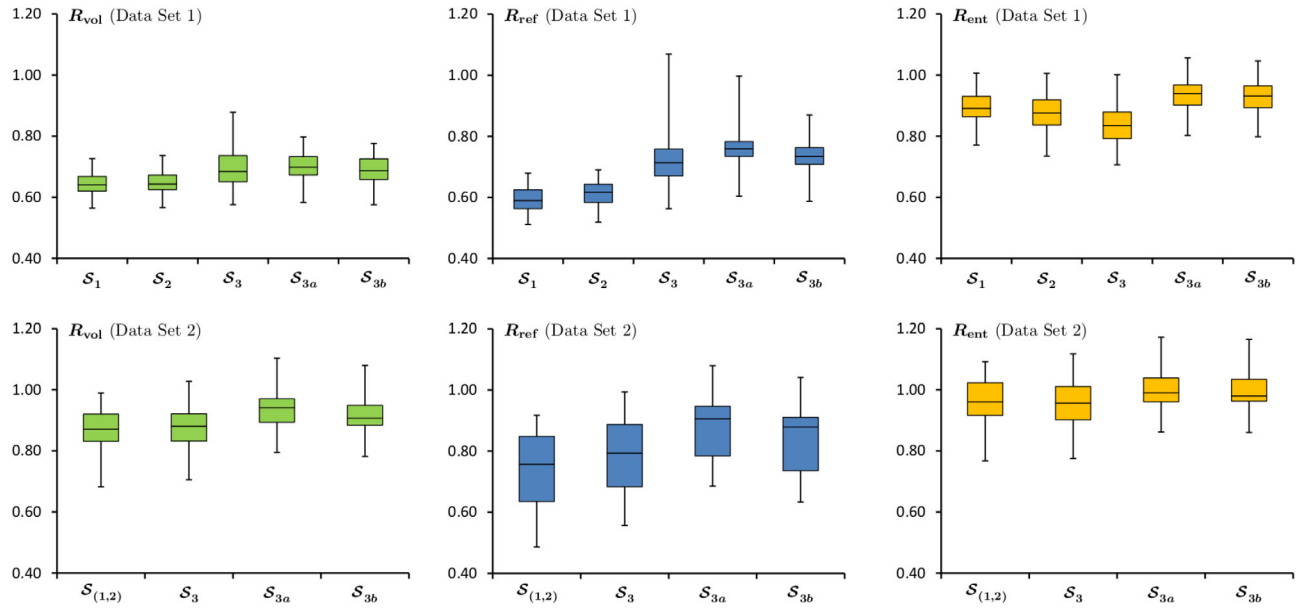

FIG. 5. Inter-station intensity standardization in terms of the R_{vol}, R_{ref}, and R_{ent} measures for both data sets and different registration strategies. R_{vol} and R_{ref} are respectively defined as the average ratio (after-/before standardization) of the per-station histogram distance to the entire volume or to the reference station, and R_{ent} is the ratio (after-/before standardization) of the histogram entropy for the entire volume. Values below unity indicate improvement in intensity homogeneity.

of prior bias correction on the resulting quality of intensity standardization in terms of D_{vol} and D_{ref} for different registration strategies and both test data sets.

Visual Quality Valitation Results

Our expert was only able to detect relevant visual changes in the T_1w contrast channel of Data Set 1. The corresponding results were very consistent with our quantitative results: our radiologist preferred the corrected volume in 43/48 cases, and no visual quality difference was concluded in the 5 remaining cases. The visual quality difference on the STIR contrast channel was detected in 9/48 cases, from which in 3 cases the corrected and in 6 cases the raw image was preferred. For Data Set 2, our expert was able to see relevant quality improvement in 2 post-contrast volumes only, one corrected and one raw.

Implementation Details and Execution Times

Our method was implemented in MATLAB R2012b (The MathWorks, Inc.). All experiments were executed on a 3.60 GHz Intel(R) Xeon(R) computer with 32 GB RAM. Average computational time for Data Set 1 was in the range of 5 s and 15 s for registering one pair of histograms with and without regularization, and correspondingly 1 min for $\mathcal{S}_1$ and $\mathcal{S}_2$ and 4.5 min for $\mathcal{S}_3$ and its modifications for the entire standardization procedure. Corresponding values for Data Set 2 were: 5 s, 12 s, 30 s, and 80 s.

DISCUSSION AND CONCLUSIONS

Two important conclusions can be drawn from the results on divergence-based validation measures reported in Table 1. The first important observation is that all the measures indicate improvement of inter-station intensity homogeneity, which is, in particular, confirmed by values of R_{vol} and R_{ref} being less than unity. The second observation is that the "direct" registration strategies $\mathcal{S}_1$ and $\mathcal{S}_2$ perform better than the ones using the information from the overlapping regions. For both validation data sets, $\mathcal{S}_1$ exhibits the best performance with respect to all four divergence-based validation measures. Strategy $\mathcal{S}_2$ performs insignificantly worse in all cases (note, that this conclusion is only applicable to Data Set 1). Performance of $\mathcal{S}_3$ and its modifications is worse than that of the other two strategies, in majority of the cases significantly. Among the latter approaches, $\mathcal{S}_3$ performs best on both validation data sets with respect to three out of four measures. Strategy $\mathcal{S}_{3b}$ that uses the linear regularization of the deformation field exhibits the best performance in the remaining two cases.

The results of the entropy-based validation measure reported in Table 1 are somewhat less conclusive. While this measure indicates definite improvement of inter-station intensity homogeneity for Data Set 1, for Data Set 2 only the best performing method shows a small but consistent improvement. Moreover, contrary to the divergence-based validation measures, R_{ent} indicates that the "progressive with overlaps" strategies perform best: $\mathcal{S}_3$ for Data Set 1, significantly in all cases, and $\mathcal{S}_{3a}$ for Data Set 2, significantly in two out of three cases. We hypothesize that such performance with respect to the entropy-based validation measure is strongly related to the structure of the intensity histogram. Namely, for the histograms with more pronounced peaks our entropy-based measure reflects alignment of the corresponding peaks between the histograms of each station. Whereas for the cases when the peaks are not well defined this

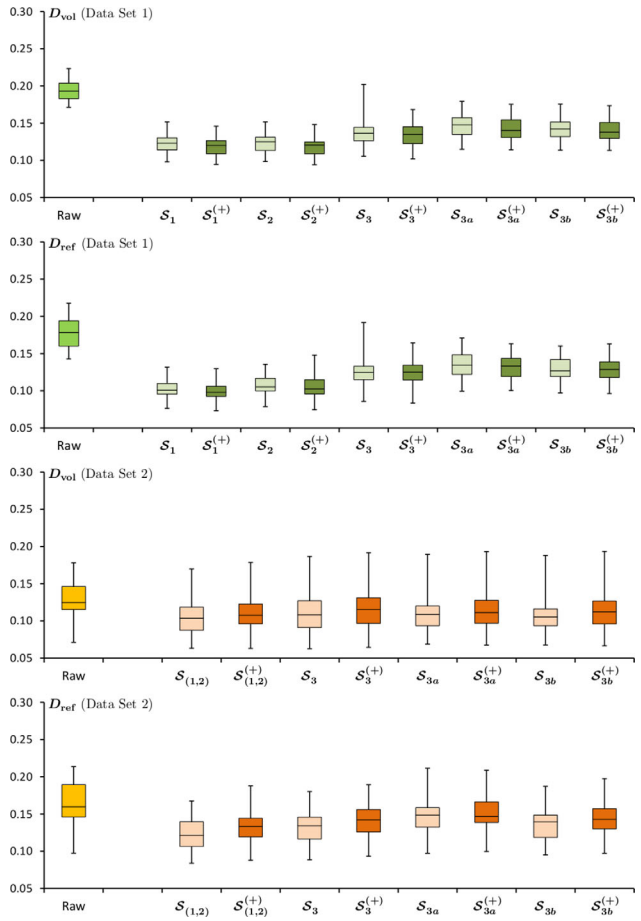

FIG. 6. Distribution of D_{vol} and D_{ref} values for different registration strategies on both test data sets, with and without prior bias correction. D_{vol} and D_{ref} are the average distance between the histograms of each station and that of the entire volume and the reference station, respectively.

measure becomes sensitive to sharpening or blurring of each particular peak resulting from the performed intensity mapping.

Our structured qualitative reading experiment convincingly confirmed improved volume homogeneity on the T_1w contrast channel of Data Set 1. At the same time, our expert found it difficult to detect improvement in interstack intensity homogeneity on the STIR contrast channel of Data Set 1 and on both contrast channels of Data Set 2. The given explanation for this was that all those volumes, in comparison with the T_1w data, (*i*) were already much more homogeneous, as shown in Figure 4 and Supporting Movie S1, and (*ii*) contained relatively a lot less information; see Figures 2–4 and Supporting Movie S1. This difficulty is also related to the tissue composition of T_1w images, whose intensity histogram typically has three very pronounced peaks corresponding to air, muscle, and fat. Large image areas containing one primary tissue are easy to use as a reference for visual scoring. In contrast, intensity histograms of the other contrast channels are much less structured, which makes visual assessment of quality improvement very challenging. The results of the presented visual scoring experiment are consistent with the quantitative results reported in the "Different Strategies, Data Sets, With and Without Bias Correction" section in the sense that the latter confirm larger image quality improvement on Data Set 1 in comparison with Data Set 2.

Results reported in Table 2 indicate that in all the cases registration improves the similarity between pairs of histograms. For Data Set 1, the most similar initial histograms are the ones corresponding to the neighboring stations. Similarity between the histogram of each station and that of the reference station is somewhat lower. For Data Set 2, $\overline{d_J}[\mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(2)})]$ is found to be the smallest average distance between initial histogram pairs. However, the relative improvement as result of the registration for these cases is much smaller than that of the registrations between the histograms of the overlapping reasons. Pairs of histograms of the corresponding overlapping regions, $\mathcal{H}(\Theta_s^{(2)})$ and $\mathcal{H}(\Theta_{s+1}^{(1)})$ exhibit the highest degree of similarity after registration, on both test data sets. In both cases, $\overline{d_J}[\mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(2)})]$ is lower than $\overline{d_J}[\mathcal{H}(\Xi_s), \mathcal{H}(\Theta_s^{(1)})]$, before as well as after registration.

Another important conclusion from Table 1, also illustrated in Figure 6, is that the difference between the inter-station intensity standardization with and without bias correction was insignificant in all cases. Moreover, in all but one case prior bias correction has improved the intensity standardization on Data Set 1, whereas for Data Set 2 the results without the bias correction were better in all cases. This observation is in line with what was reported by Madabhushi and Udupa (12).

Performing all the processing on the intensity histogram level makes it independent from the slice orientation. Thus, our method can be directly extended to data sets acquired with sagittal or transversal orientation of the slices. Performing joint intensity standardization on all contrast channels has two important consequences. First, in contrast to the marginal intensity histograms where some peaks can be overshadowed by stronger neighbors, joint intensity histograms are much more informative and easy to register. Second, in this way we preserve the relation between the contrast channels, which is important for the clinical usability of the data. A possible drawback of such joint processing is that, in case the contrast channels contain significantly different amount of information, the registration might be drawn towards the "more dominant" channel.

We are not aware about any specific limitations of our approach, as long as the histogram pairs have sufficient similarity for the registration method to be able to align them. We also want to point out that here we use a simple affine transformation model. Other data types might require more advanced registration procedures. In particular, different approaches might be applied for registering intensity histograms of different stations. Developing targeted approaches for registering histograms of each station to that of the reference one would, in particular, greatly improve the final results on Data Set 2. However, in this work we decided to use the same registration approach for all the histograms for the sake of purity of experiments.

Also we want to emphasize here that in this work the validation was performed on two data sets that were acquired on Philips hardware using the same type of coil

(body coil). Although our method was designed in a way to be generic enough with respect to the input data, its application to data sets acquired on scanners of different vendors and/or with different coil types (e.g., surface- or organ-oriented coils) requires additional validation and possible tailoring of the method.

The presented validation of our method is based on general commonly used computer vision measures, such as entropy and divergence. While these measures indicate quality improvement as result of performed intensity standardization, the ultimate validation of success or failure of any image processing method, and our method in particular, should be considered in the context of intended application of the data. Hence, in our future work, we are planning to incorporate the developed methodology into a larger framework for reconstruction of whole-body volumes from multispectral MR data, in which the images will also be corrected for the intensity inhomogeneity within each stack (bias). Next, the entire developed framework will be applied to enable objective assessment of progression or regression of cancerous lesions in multiple myeloma patients.

In conclusion, in this work we have presented a generic approach for inter-station intensity standardization within a whole-body MR volume. Our approach can be applied to any type of whole-body MR data, in particular, multispectral, and it does not make any assumptions about the data. We have developed several registration strategies and showed that the "direct" registration approaches are superior in comparison with approaches that employ information from the overlap regions of the neighboring stations. This was confirmed by applying our algorithm to two large multispectral whole-body MR data sets of very different nature. Results of the performed validation study confirm efficiency and generality of our inter-station intensity standardization approach.

ACKNOWLEDGMENTS

The authors thank Joel Kullberg, Lars Johansson, and Håkan Ahlström from Uppsala University Hospital (Uppsala, Sweden) for providing the whole-body MRA data set.

REFERENCES

1. Raab MS, Podar K, Breitkreutz I, Richardson PG, Anderson KC. Multiple myeloma. Lancet 2009;374:324–339.
2. Ruehm SG, Goyen M, Barkhausen J, Kröger K, Bosk S, Ladd ME, Debatin JF. Rapid magnetic resonance angiography for detection of atherosclerosis. Lancet 2001;357:1086–1091.
3. Börnert P, Aldefeld B. Principles of whole-body continuously-moving-table MRI. J Magn Reson Imaging 2008;28:1–12.
4. Brant W, de Lange E. Essentials of Body MRI. Oxford: Oxford University Press; 2012. p 416.
5. Jäger F, Hornegger J. Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. IEEE Trans Med Imag 2009;28:137–150.
6. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87–97.
7. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based bias field correction of MR images of the brain. IEEE Trans Med Imaging 1999;18:885–896.
8. Wells W, Grimson W, Kikinis R, Jolesz F. Adaptive segmentation of MRI data. IEEE Trans Med Imaging 1996;15:429–442.
9. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. Magn Reson Med 1999;42:1072–1081.
10. Weisenfeld N and Warfield S. Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, USA, 2004. pp 101–104.
11. Schmidt M. A method for standardizing MR intensities between slices and volumes. Technical Report TR05-14, University of Alberta, Edmonton, AB, Canada. 2005.
12. Madabhushi A, Udupa J. Interplay between intensity standardization and inhomogeneity correction in MR image processing. IEEE Trans Med Imag 2005;24:561–576.
13. Bergeest JP, Jäger F. A comparison of five methods for signal intensity standardization in MRI. In: Proceedings of the Bildverarbeitung für die Medizin, Berlin, Germany, 2008. pp 36–40.
14. Iglesias J, Dinov I, Singh J, Tong G, Tu Z. Synthetic MRI signal standardization: Application to multi-atlas analysis. In: Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention, Beijing, China, 2010. pp 81–88.
15. Jog A, Roy S, Carass A, Prince JL. Pulse sequence based multi-acquisition MR intensity normalization. In: Proceedings of the SPIE: Medical Imaging, Orlando, FL, USA, 2013. p 86692H.
16. Bernstein MA, Huston J, Ward HA. Imaging artifacts at 3.0T. J Magn Reson Imaging 2006;24:735–746.
17. Warntjes J, Dahlqvist Leinhard O, West J, Lundberg P. Rapid magnetic resonance quantification on the brain: optimization for clinical usage. Magn Reson Med 2008;60:320–329.
18. Robinson K, Ghita O, Whelan PF. Intensity non-uniformity correction in multi-section whole body MRI. In: Proceedings of the SPIE, Dublin, Ireland, 2005. p 164–174.
19. Tizon X, Lin Q, Hansen T, Borgefors G, Johansson L, Ahlström H, Frimmel H. Identification of the main arterial branches by whole-body contrast-enhanced MRA in elderly subjects using limited user interaction and fast marching. J Magn Reson Imaging 2007;25:806–814.
20. Romu T, Borga M, Dahlqvist Leinhard O. MANA–multi scale adaptive normalized averaging. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, 2011. pp 361–364.
21. Andersson T, Romu T, Karlsson A, Norén B, Forsgren MF, Smedby O, Kechagias S, Almer S, Lundberg P, Borga M, Dahlqvist Leinhard O. Consistent intensity inhomogeneity correction in water-fat MRI. Magn Reson Med 2015;42:468–476.
22. Dzyubachyk O, van der Geest RJ, Staring M, Börnert P, Reijnierse M, Bloem JL, Lelieveldt BPF. Joint intensity inhomogeneity correction for whole-body MR data. In: Proceedings of the 16th International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 2013. pp 106–113.
23. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly the prospective investigation of the vasculature in Uppsala seniors (PIVUS) study. Arterioscler Thromb Vasc Biol 2005;25:2368–2375.
24. Klein S, Pluim JPW, Staring M, Viergever MA. Adaptive stochastic gradient descent optimisation for image registration. Int J Comput Vision 2009;81:227–239.
25. Klein S, Staring M, Murphy K, Viergever M, Pluim J. `elastix`: a toolbox for intensity-based medical image registration. IEEE Trans Med Imag 2010;29:196–205.
26. Tustison N, Avants B, Cook P, Zheng Y, Egan A, Yushkevich P, Gee J. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29:1310–1320.
27. Endres DM, Schindelin JE. A new metric for probability distributions. IEEE Trans Inform Theory 2003;49:1858–1860.
28. Eadie W. Statistical methods in experimental physics. Amsterdam: North-Holland Pub. Co.; 1971. p 296.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Inter-station intensity standardization for Data Set 1 in terms of the Rvol measure for different registration strategies. Rvol is defined as the average ratio (after-/before standardization) of the per-station histogram distance to the entire volume. Values below unity indicate improvement in intensity homogeneity. The solid line indicates the mean value and the

dashed lines indicate one standard deviation interval around the mean. Markers of the same color mean that the data points correspond to the same subject.

Figure S2. Inter-station intensity standardization for Data Set 1 in terms of the Rref measure for different registration strategies. Rref is defined as the average ratio (after-/before standardization) of the per-station histogram distance to the reference station. Values below unity indicate improvement in intensity homogeneity. The solid line indicates the mean value and the dashed lines indicate one standard deviation interval around the mean. Markers of the same color mean that the data points correspond to the same subject.

Figure S3. Inter-station intensity standardization for Data Set 1 in terms of the Rent measure for different registration strategies. Rent is defined as the ratio (after-/before standardization) of the histogram entropy for the entire volume. Values below unity indicate improvement in intensity homogeneity. The solid line indicates the mean value and the dashed lines indicate one standard deviation interval around the mean. Markers of the same color mean that the data points correspond to the same subject.

Figure S4. Inter-station intensity standardization for Data Set 2 in terms of the Rvol measure for different registration strategies. Rvol is defined as the average ratio (after-/before standardization) of the per-station histogram distance to the entire volume. Values below unity indicate improvement in intensity homogeneity. The solid line indicates the mean value and the dashed lines indicate one standard deviation interval around the mean.

Figure S5. Inter-station intensity standardization for Data Set 2 in terms of the Rref measure for different registration strategies. Rref is defined as the average ratio (after-/before standardization) of the per-station histogram distance to the reference station. Values below unity indicate improvement in intensity homogeneity. The solid line indicates the mean value and the dashed lines indicate one standard deviation interval around the mean.

Figure S6. Inter-station intensity standardization for Data Set 2 in terms of the Rent measure for different registration strategies. Rent is defined as the ratio (after-/before standardization) of the histogram entropy for the entire volume. Values below unity indicate improvement in intensity homogeneity. The solid line indicates the mean value and the dashed lines indicate one standard deviation interval around the mean.

Movie S1. Complete reconstructed volume of the multi-spectral whole-body MR data (Data Set 1) before (a,d) and after the intensity standardization (b,e), and the corresponding difference images (c,f). Intensity of all images was enhanced for visualization purposes.