A Comparison of Five Methods
for Signal Intensity Standardization in MRI

Jan-Philip Bergeest, Florian Jäger

Lehrstuhl für Mustererkennung, Friedrich-Alexander-Universität Erlangen-Nürnberg
jan.p.bergeest@informatik.stud.uni-erlangen.de

Abstract. A major problem in magnetic resonance imaging (MRI) is the lack of a pulse sequence dependent standardized intensity scale like the Hounsfield units in computed tomography. This affects the post processing of the acquired images as, in general, segmentation and registration methods depend on the observed image intensities. Different approaches dealing with this problem were proposed recently. In this article we will describe and compare five state-of-the-art standardization methods regarding speed, applicability and accuracy. As a quality measure the mean distance and the Kullback-Leibler divergence are considered. For the experiments 28 MRI head volume images, acquired during clinical routine, were used.

1   Introduction

MRI has the drawback of having intensity variations due to magnetic field inhomogeneities and scanner-related intensity artifacts. Thus, it is not possible to create a pulse sequence dependent, general intensity scale. Algorithms dealing with the problem of the correction of signal intensity inhomogeneities usually focus on intra-volume signal intensity distortions and do not correct the inherently non-standardness. Hence, intensities obtained from the same or different patients have no meaning related to a specific anatomical tissue even if they are acquired on the same MRI scanner. This implies a significant affect on the accuracy and precision of following image processing, analysis, segmentation and registration methods relying on intensity similarity. Furthermore, standard presets cannot be used to display MR images or to visualize certain tissue classes and/or pathologies. These settings have to be adjusted for every single case.

First we will give a brief overview of the intensity standardization method described in Florian Jäger et al. [1] that basically performs a registration of joint histograms. Further, we will discuss 1-d histogram based approaches presented in Nyúl et al. [2] that matches landmarks on histograms, in [3] a method is introduced that is based on histogram estimation using Gaussian mixtures, as well as an algorithm adapted from dynamic histogram warping [4]. In addition we will have a brief look at the algorithm proposed by Weisenfeld and Warfield [5], in which a multiplicative correction field is estimated. Finally the experimental setup is described and the evaluation results are presented and discussed.

2 Methods

2.1 MRI Intensity Standardization by Non-rigid Registration of Joint Histograms

The idea proposed in [1] is that a normalization can be achieved by finding a deformation of the joint histograms of two sets of images with respect to a certain distance measure. Each of these histograms is at least two dimensional and contains the intensity information of two or more MRI sequences (e.g. T1 or T2 weighted images). If the probability density functions are considered as images, the normalization can be treated as a registration problem. The resulting non-linear correction function is used to adjust the image intensities of the MRI image series. In [6] this approach is extended to be applicable to a whole body MRI scan. Therefore, the volume is split up into K sub volumes which are corrected separately. In order to include the influence of small local structures the other $K - 1$ sub volumes are used as regularizer.

2.2 Landmark-based MRI Scale Standardization

The basic idea of the method presented in [2] is to find a mapping that deforms the template intensity histogram of an input image so that it matches a reference histogram based on landmarks. In the first step the landmarks of the mean histogram are detected which has to be done only once for each given protocol and body region (training). Second, the real image-dependent normalization is applied to a given volume image, of the same body region and protocol as used in the training step, by matching its detected landmarks to the standard positions (transformation). Hence each segment, between two landmarks, is mapped linearly and independently. The resulting transformation is nonlinear.

2.3 MRI Intensity Correction using Mixture Mapping

A normalization that is only applicable to images of the head region is described in [3]. This is done by matching intensities of head specific anatomical tissue classes. Therefore five main classes are used: background, cerebrospinal fluid (CSF), gray matter (GM), white matter (WM) and a mixture of fat and muscle. In the first step the histograms from a source and a target data set, i.e. two 3D images, are approximated by a Gaussian mixture. Each tissue class k is modeled by a Gaussian probability density function which has a mean μ_k that is approximated using the Expectation-Maximization (EM) algorithm. In the second step a polynomial correction function f^p of order p is used to interpolate the correction of the intensities smoothly: $f^p(x) = \sum_{i=0}^{p} \Theta_i x^i$. By minimizing the cost function $\sum_{k=1}^{n} (f^p(\mu_k) - v_k)^2$ the coefficients Θ_i are obtained, where μ_k and v_k are the means of the reference and template image respectively.

2.4 Dynamic Programming Applied to Intensity MRI Correction

Dynamic programming can be used to find an optimal alignment between two sequences of data of different length with respect to a monotonic and separable cost function. Furthermore, this function is constrained to hold the principle of optimality. In [4], one histogram is mapped to another by a one-to-one, one-to-many and many-to-one mapping. Let f_i^t denote the frequency of occurrence in the i-th bin of the template histogram, respectively f_j^r of the j-th bin in the reference histogram. A plausible cost function $c(i,j)$ for two intensities is the difference $\left| f_i^t - f_j^r \right|$. In order to find an appropriate mapping a cost matrix C is defined where $C(i,j)$ corresponds to a optimal sub path and $C(i_{max}, j_{max})$ contains the minimal distance of the histograms. The cheapest path and the corresponding mapping can be computed efficiently using dynamic programming.

2.5 Intensity Normalization by Minimizing the Kullback-Leibler Divergence

A multiplicative correction field is estimated in [5] to match a template histogram to a reference model density. The observed image O is composed of a multiplicative pixelwise intensity corruption field F, additional acquisition noise n and the correct image I in the following way: $O = FI + n$. After neglecting n for having only little influence on the problem of intensity normalization and solving the equation for I the uncorrupted image is obtained as $I \approx F^{-1}O$. The parameter field F^{-1} has to be chosen in a way that the Kullback-Leibler divergence between source and target sets is minimized. The Simultaneous Perturbation Stochastic Approximation (SPSA) is used to generate the gradient estimate.

3 Results

The images used for evaluation were T1- and T2/FLAIR images. The T2-weighted FLAIR datasets were acquired on a Siemens Avanto 1.5 T scanner with $408 \times 512 \times 19$, pixel size of 0.43 mm^2 and 7.2 mm slice thickness and TE = 143 and TR = 9000. The T1-weighted images had a resolution of $208 \times 256 \times 19$ with 0.86 mm^2 and 7.2 mm slice thickness and TE = 14 and TR = 510. In total 28 volumes from eleven different patients were used. Since the methods need to be applicable for medical diagnosis in a real clinical environment, all images used were chosen to be real patient data including evolving lesions.

For the evaluation, T1-weighted volumes and the appropriate FLAIR data sets were evaluated separately. A single reference image was chosen for each patient. Then the follow-up studies were standardized to the intensities of the corresponding reference images. As a measure for the quality of the standardization the absolute mean (Equation 1) and the average Kullback-Leibler (KL) divergence (Equation 2) were computed. In order to minimize the influence of changing anatomy on the evaluation, lesions were segmented and removed beforehand. However, the normalization was applied to the original images since

Table 1. The measurement before standardization is given in parenthesis and the methods are numbered in the same order as described above. μ_{kl} is the average KL divergence and μ_m is the overall mean distance

Method	T1		FLAIR	
	μ_m (32.62)	μ_{kl} (0.46)	μ_m (53.22)	μ_{kl} (0.61)
2.1	6.82	0.12	7.47	0.16
2.2	10.41	0.14	10.30	0.27
2.3	14,69	0.16	12,33	0.56
2.4	7.65	0.07	5.99	0.30
2.5	8.25	0.07	16.79	0.15

the pathological distortions should be considered in the evaluation of the robustness of the algorithm. As quality measure the average absolute distance between all reference and template volumes, is given by:

$$\mu_m = \frac{1}{N} \sum_i |x_i - y_i| \tag{1}$$

with N being the number of compared images, x_i being the mean intensity of the reference and as y_i the corresponding template mean is used. As a measure of the difference between two probability densities, the overall average KL divergence between the reference and template images is computed as follows:

$$\mu_{kl} = \frac{1}{N} \sum_i \sum_j p_{ij} \log(p_{ij}/q_{ij}) \tag{2}$$

where p_{ij} is the probability of occurrence of intensity j in the template image i and q_{ij} is the probability of occurrence of intensity j in the reference image i.

In Figure 3 an example for the standardization is given. All images were binarized and displayed at the same threshold value. It can be seen that the intensities vary globally and become more similar regarding similar tissue classes. In Table 1 a quantitative overview of the computed distance measures is given. As mentioned above these values were computed after removal of the lesions but without any additional inhomogeneity correction. Thus, the difference between
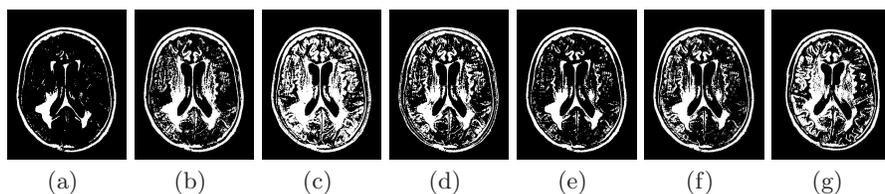

(a)   (b)   (c)   (d)   (e)   (f)   (g)

Fig. 1. Binary FLAIR image slices of two data sets from the same patient where (a) is standardized on (g) the reference image. (b) corresponds to the corrected image gained by the algorithm presented in 2.1, respectively (c) to 2.2, (d) to 2.3, (e) to 2.4 and (f) to 2.5

Table 2. General overview: + is good, o is average and − is below average

Method	Speed	Applicability	Accuracy
2.1	o ($\approx$ 1 minute)	+	+
2.2	+ ($\approx$ 2 seconds)	o	+
2.3	o ($\approx$ 5 minutes)	−	0
2.4	+ ($\approx$ 2 seconds)	o	+
2.5	− ($\approx$ 30 minutes)	o	+

the volumes will never become zero. A short overview of the methods for comparison can be found in Table 2, where applicability means that the algorithm can be easily adapted to different regions of the body. Accuracy refers to robustness and improvement of the image quality. The runtime measurements were performed on a Intel Core2 CPU T5500 with 1.66GHz and 2 gigabyte RAM.

4 Discussion

In this article we have compared five MRI signal intensity standardization methods in terms of speed, applicability and accuracy. More sophisticated methods that make use of all image sequences or do a pixelwise correction estimation are of course slower but lead to better results (approaches 2.1 and 2.5), whereas approach 2.3 leads to worse results. Algorithm 2.2 and 2.4 are fast and in general provide good results but 2.2 is vulnerable to distortions in the histograms due to the fact that it is piecewise linear and relies on a learned histogram shape. All methods can be potentially applied to every body area excepting 2.3 which uses head specific tissue classes.

References

1. Jäger F, Deuerling-Zheng Y, Frericks B, et al. A new method for MRI intensity standardization with application to lesion detection in the brain. In: Kobbelt L, et al, editors. Vision Modeling and Visualization. Köln: Aka GmBH; 2006. p. 269–76.
2. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging. 2000;19(2):143–50.
3. Hellier P. Consistent intensity correction of MR images. Proc ICIP. 2003;1:1109–12.
4. Cox IJ, Roy S, Hingorani SL. Dynamic histogram warping of image pairs for constant image brightness. Proc ICIP. 1995; p. 2366–9.
5. Weisenfeld N, Warfield S. Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence. Proc IEEE Int Symp Biomed Imaging. 2004;1:101–4.
6. Jäger F, Nyul L, Frericks B, et al. Whole body MRI intensity standardization. Proc BVM. 2007; p. 459–63.