A New Method for MRI Intensity Standardization with Application to Lesion Detection in the Brain

Florian Jäger[1], Yu Deuerling-Zheng[1], Bernd Frericks[2], Frank Wacker[2], Joachim Hornegger[1]

[1]Lehrstuhl für Mustererkennung, Universität Erlangen,
Martensstraße 3, 91058 Erlangen, Germany;
[2]Klinik für Radiologie und Nuklearmedizin, Charité CBF,
Hindenburgdamm 30, 12200 Berlin, Germany
Email: {jaeger,hornegger}@informatik.uni-erlangen.de
{bernd.frericks, frank.wacker}@charite.de

Abstract

A major problem of segmentation of magnetic resonance imaging is that intensities are not standardized like in computed tomography. In this article we will present a new method for MRI intensity standardization by aligning histograms of higher dimensions. So the correction process is independent from spatial coherences and prior segmentations of the reference and newly acquired images. Although the approach is not limited to a specific application or area of the body, it is utilized for fast classification of brain tissue. Therefore, reference statistics are computed once using a hidden Markov random field approach. The intensity corrected images are then classified using these learned statistics. In order to evaluate the presented methods the mean differences of intra-patient time series are chosen as reference. Furthermore, results of fast brain tissue classification are presented.

1 Introduction

For magnetic resonance imaging no intensity standard, like Hounsfield units in computed tomography, is available due to magnetic field inhomogeneities in both B_0 and RF excitation fields, etc. The disturbances can be characterized in two different ways: first, intensities of the same tissue class differ throughout a single volume. In order to deal with that problem a variety of algorithms for bias field correction were developed in the last decade. Especially statistical and clustering approaches yield good correction results [1, 2, 3]. However, these methods do not solve the problem of assigning a specific meaning to observed image intensities. Hence a certain measured intensity of the same or different patient cannot be associated to a specific anatomical meaning. For segmentation, a missing standard has the disadvantage that for every new suspect an individual training of the used (statistical) model has to be performed. For this reason the clinical applicability of many algorithms is low due to runtime restrictions. Furthermore, visualization systems cannot use standard presets (e.g. transfer functions) to visualize certain organs or tissue classes. The settings have to be adjusted for every single scan. These are the reasons why a second class of approaches dealing with inter-scan intensity standardization was developed by several authors.

In [4, 5] a 1-d histogram matching approach was presented. First, they detected some landmarks (percentiles, modes, ...) on the template and the reference histogram, matched them and finally interpolated linearly between the detected locations.

Pierre Hellier presented in [6] a correction method that estimates a mixture of Gaussians that approximates the histogram [7] first. Then he computes a polynomial correction function that aligns the mean intensities of the tissues.

A multiplicative correction field is estimated in [8] that adapts the intensity statistics of an acquired MR volume to a previously created model. This is achieved by minimizing the Kullback-Leibler divergence between the model and the template intensity distribution.

In [9] a method including spatial information between the reference and the template image is presented. In order to match the images a non-linear registration algorithm was used. On the aligned images a scalar multiplicative correction weight is computed.

How intensity standardization and bias correction influence each other is evaluated in [10]. The authors conclude that both steps are necessary but the correction of inhomogeneities has to be done beforehand.

For detection and quantification of lesions in MRI brain images usually T1- and T2-weighted as well as FLAIR (Fluid Light Attenuation Inversion Recovery) datasets are acquired. T1-weighted images have a higher spatial resolution but a lower tissue contrast than T2-weighted. FLAIR images have the advantage compared to T2-weighted images that it is easier to discriminate between edema and cerebral spinal fluid (CSF). State-of-the-art algorithms for intensity correction use a single sequence for standardization and drop information of others. For many applications this is sufficient, because in many regions of the body a gray value in one image is associated to exactly one intensity in another sequence (e.g. the brain). In general, however, this is not the case. The algorithm presented in this article utilizes all acquired sequences for intensity correction. With that, it is possible to separately correct tissue classes, that have the same intensity in one image but can be distinguished using more datasets. The presented method is based on a non-rigid registration of higher dimensional probability functions of two or more volumes (template density) with a reference density. From the registration result a non-parametric correction function is gained to standardize the intensities of the template volumes. With that the statistics are adapted to a previously calculated standard. Furthermore, the introduced approach does not rely on any assumptions about the shape of the probability distributions used. Thus the method is completely independent of any application, region of interest (brain, thorax, pelvis, ...), scanning protocol (T1-, T2-weighted, ...) or modality (MRI, CT, SPECT, ...) accordingly. However, as most of the state–of–the–art algorithms are tested with application to brain tissue classification, we will focus on this subject in the following as well. But one should keep in mind that the brain intensity statistics are rather simple and thus standardization is fairly trivial compared to other body regions as the dependence for brain tissue intensities is usually only monofunctional but for other regions the functional dependency is of a higher degree.

2 Segmentation of the Brain

2.1 Markov Random Field Theory

Let $\mathcal{S} = \{1, 2, \cdots, N\}$ be a discrete set with N sites. A *label* is an event that may happen to a site, which may be both continuous or discrete.

In our case, two discrete label sets are defined: one for the intensity values denoted as $\mathcal{Y}$ and one for the class memberships denoted as $\mathcal{X}$. Note that a label in $\mathcal{Y}$ assumes its value in the range $[1, 2^d]$ and a label in $\mathcal{X}$ assumes its value in the range $[1, K]$, with d being the pixel depth and K the number of classes.

Next the neighborhood system and cliques are introduced, which builds the basis for the contextual constraints. Most authors define a neighborhood for $\mathcal{S}$ as

$$\mathcal{N} = \{\mathcal{N}_i, i \in \mathcal{S}\}$$

where $\mathcal{N}_i$ is the set of site i's neighbors (e.g. [11]). The neighborhood system has the following properties:

1. a site is not a neighbor of itself: $i \notin \mathcal{N}_i$
2. the neighboring relationship is mutual:
   $i \in \mathcal{N}_j \Leftrightarrow j \in \mathcal{N}_i$

A clique for $(\mathcal{S}, \mathcal{N})$ is defined as a subset of sites in $\mathcal{S}$, in which all the pairs of distinct sites are neighbors, except for single-site cliques. Let $\mathcal{C}_n$ denote the collection of cliques of size n. In the context of this article only pairwise interaction between pixels is considered. Every clique therefore consists of two sites.

Furthermore let $X = \{X_1, X_2, \cdots, X_N\}$ be a set of random variables defined on the set $\mathcal{S}$, in which each random variable X_i takes a value $x_i \in \mathcal{X}$. Ξ is the set of all possible configurations of X. X is said to be a Markov random field on $\mathcal{S}$ with respect to a neighborhood system $\mathcal{N}$ if and only if the following two conditions are satisfied [11]:

1. positivity: $P(\boldsymbol{x}) > 0, \forall \boldsymbol{x} \in \Xi$
2. Markovianity: $P(x_i|x_{\mathcal{S}-\{i\}}) = P(x_i|x_{\mathcal{N}_i})$
   where $x_{\mathcal{S}-\{i\}}$ denotes the set of labels at the sites $\mathcal{S} - \{i\}$ and $x_{\mathcal{N}_i}$ stands for the sets of i's neighbors.

Markovianity depicts the local characteristics of the random field. This yields that a label at a site i depends only on its neighboring pixels. In other words, a site i has direct interaction with its neighbors. Markovianity can always be satisfied as a sufficiently large neighborhood $\mathcal{N}_i$ can be selected. The largest neighborhood consists of all other sites.

In this article the label set $\mathcal{X}$ for the class memberships satisfies the conditions of a MRF because (1) each pixel belongs to exactly one tissue class so that the positivity holds and (2) neighboring pixels are more likely to have the same class membership so that the Markovianity holds as well. The latter feature is more important because it favors locally clustered pixels more than single pixels. This can for example exclude sporadic located pixels in noisy images and results in a more robust and more homogenous segmentation.

A set of random variables X is said to be a Gibbs random field (GRF) on $\mathcal{S}$ with respect to $\mathcal{N}$ if and only if its configurations obey a Gibbs distribution [11]. A Gibbs distribution is defined as

$$P(x) = \frac{1}{Z} e^{-\frac{1}{T} U(x)} \qquad (1)$$

where

$$Z = \sum_{x \in F} e^{-\frac{1}{T} U(x)} \qquad (2)$$

is a normalizing constant called the partition function, T is a constant called the temperature that controls the sharpness of the distribution. $U(x)$ is the energy function

$$U(x) = \sum_{c \in \mathcal{C}} V_c(x) \qquad (3)$$

with $V_c(x)$ being the potential of the clique c and $\mathcal{C}$ denotes the set of all possible cliques using the neighborhood system $\mathcal{N}$.

According to the Hammersley-Clifford theorem a MRF is equivalent to a GRF. The theorem states that *X is called MRF on $\mathcal{S}$ with respect to $\mathcal{N}$ if and only if X is a GRF on $\mathcal{S}$ with respect to $\mathcal{N}$.* The practical value of the theorem is that it provides a simple way to specify the joint probability $P(\mathbf{x})$, $\mathbf{x} \in \Xi$ by specifying the clique potential functions $V_c(x)$ [11].

In our case only cliques of size two are considered. Thus the energy function can be written as

$$U(x) = \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} V_2(x_i, x_{i'}) \qquad (4)$$

and the conditional probability as

$$P(x_i | x_{\mathcal{N}_i}) = \frac{e^{-\sum_{i' \in \mathcal{N}_i} V_2(x_i, x_{i'})}}{\sum_{x_j \in \mathcal{X}} e^{-\sum_{i' \in \mathcal{N}_j} V_2(x_j, x_{i'})}}. \qquad (5)$$

2.2 FGM model

Finite mixture (FM) models are adopted in the domain of brain tissue segmentation by various authors [7, 12, 1], where most of them assume a Gaussian distribution for the tissue classes (Finite Gaussian Mixture (FGM) model). More specifically, the entire brain is to be classified into 3 classes: white matter (WM), gray matter (GM) and CSF. Assume the observable gray value is generated by a random process whose distribution is a mixture of these classes which can be characterized by the parameters of its normal distribution. The segmentation is accomplished by assigning each voxel, represented by its intensity, to the class with maximum posterior probability.

Suppose each image consists of K classes (tissue types). Let $\boldsymbol{y_i} = (y_{i,1}, y_{i,2}, \ldots, y_{i,L})$ denote the intensity vector of the voxel at site i, L being the number of MRI sequences. The FGM model is of the form

$$p(\boldsymbol{y_i}; \boldsymbol{\Theta}) = \sum_{k=1}^{K} w_k p(\boldsymbol{y_i}|k; \theta_k), \qquad (6)$$

$$w_k \geq 0, \sum_{k=1}^{K} w_k = 1$$

where $\boldsymbol{\Theta} = \{w_k; \theta_k | k = 1, ..., K\}$ and $\theta_k = (\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$, $\boldsymbol{\mu_k}$ is the L-component mean vector and $\boldsymbol{\Sigma_k}$ is the $L \times L$ covariance matrix. w_k is the locally independent prior probability ($w_k = p(k)$). $p(\boldsymbol{y_i}|k; \theta_k)$ is the class-conditional distribution density

$$p(\boldsymbol{y_i}|k; \theta_k) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma_k}|}} \cdot$$
$$\exp(-\frac{1}{2}(\boldsymbol{y_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_k})) \qquad (7)$$

with $|\boldsymbol{\Sigma_k}|$ and $\boldsymbol{\Sigma}_k^{-1}$ being the determinant and the inverse of $\boldsymbol{\Sigma_k}$, respectively.

The estimation of the unknown parameter vector $\boldsymbol{\Theta}$ can be done using the Expectation Maximization (EM) algorithm [7].

2.3 HMRF model

The HMRF model used in this article is defined using two random fields [13]. The first field is called observable random field $Y = \{Y_i, i \in \mathcal{S}\}$ with the events $Y_i \in \mathcal{Y}$ and $\mathcal{Y}$ being the set of all observable gray value vectors. Given a configuration $\boldsymbol{x} \in \Xi$,

it follows a known conditional probability function $p(\boldsymbol{y_i}|x_i;\boldsymbol{\Theta})$. The second is called hidden Markov random field $X = \{X_i, i \in \mathcal{S}\}$. It is characterized by a not observable labeling corresponding to unknown tissue classes. Furthermore, local characteristics of the HMRF can be defined using the neighborhood of a given site i:

$$P(\boldsymbol{y_i}, x_i|x_{\mathcal{N}_i};\boldsymbol{\Theta}) = P(\boldsymbol{y_i}|x_i;\boldsymbol{\Theta})P(x_i|x_{\mathcal{N}_i}). \tag{8}$$

Thus, the marginal probability distribution of $\boldsymbol{y_i}$ depends on the parameter set $\boldsymbol{\Theta}$ and the local neighborhood $x_{\mathcal{N}_i}$. It can be computed as

$$
\begin{aligned}
p(\boldsymbol{y_i}|x_{\mathcal{N}_i};\boldsymbol{\Theta}) &= \sum_{k=1}^{K} p(\boldsymbol{y_i}, k|x_{\mathcal{N}_i};\boldsymbol{\Theta}) \\
&= \sum_{k=1}^{K} p(\boldsymbol{y_i}|k;\theta_k)p(k|x_{\mathcal{N}_i})
\end{aligned} \tag{9}
$$

which is the definition of the hidden Markov random field model. Thus the posterior probability depends not only on the parameter set $\boldsymbol{\Theta}$, but also on the neighborhood $x_{\mathcal{N}_i}$:

$$
\begin{aligned}
p(k|\boldsymbol{y_i}, x_{\mathcal{N}_i};\boldsymbol{\Theta}) &= \frac{p(k|x_{\mathcal{N}_i})p(\boldsymbol{y_i}|k;\theta_k)}{p(\boldsymbol{y_i}|x_{\mathcal{N}_i};\boldsymbol{\Theta})} \\
&= \frac{p(k|x_{\mathcal{N}_i})p(\boldsymbol{y_i}|k;\theta_k)}{\sum_{j=1}^{K} p(j|x_{\mathcal{N}_i})p(\boldsymbol{y_i}|j;\theta_j)}.
\end{aligned} \tag{10}
$$

In the maximization step of the EM-algorithm, the mean of each class-conditional distribution and its covariance matrix are updated as followed:

$$\boldsymbol{\mu}_{\boldsymbol{k}}^{(t+1)} = \frac{\sum_{i=1}^{N} p(k|\boldsymbol{y_i}, x_{\mathcal{N}_i};\boldsymbol{\Theta}^{(t)})\boldsymbol{y_i}}{\sum_{i=1}^{N} p(k|\boldsymbol{y_i}, x_{\mathcal{N}_i};\boldsymbol{\Theta}^{(t)})} \quad \text{and} \tag{11}$$

$$\boldsymbol{\Sigma}_{k}^{(t+1)} = \frac{\sum_{i=1}^{N} p(k|\boldsymbol{y_i}, x_{\mathcal{N}_i};\boldsymbol{\Theta}^{(t)})\boldsymbol{\Sigma}_{k,i}^{(t+1)}}{\sum_{i=1}^{N} p(k|\boldsymbol{y_i}, x_{\mathcal{N}_i};\boldsymbol{\Theta}^{(t)})} \tag{12}$$

where

$$\boldsymbol{\Sigma}_{k,i}^{(t+1)} = (\boldsymbol{y_i} - \boldsymbol{\mu}_{k}^{(t+1)})(\boldsymbol{y_i} - \boldsymbol{\mu}_{k}^{(t+1)})^T. \tag{13}$$

Finally we have to define the locally dependent prior probability $p(x_i|x_{\mathcal{N}_i})$. In this article we use a second-order neighborhood system. The energy function $U(X)$ is constructed similarly to [13], however, the potentials $V(x_i)$ differ slightly:

$$U(x) = \sum_{i=1}^{N} V(x_i), \tag{14}$$

$$
\begin{aligned}
V(x_i) = &\sum_{r \in c_i^1} [1 - \delta(x_i - x_r)] + \\
&\sum_{r \in c_i^2} [1 - \delta(x_i - x_r)]/\sqrt{2} + \\
&\sum_{r \in c_i^3} [1 - \delta(x_i - x_r)]\triangle x/\triangle z, \tag{15}
\end{aligned}
$$

where

$$\delta(x) = \left\{ \begin{array}{ll} 1, & x = 0; \\ 0, & x \neq 0, \end{array} \right.$$

c_i^1 represents the four nearest- and c_i^2 the four nearest diagonal- neighbors. The neighborhood c_i^3 includes the two nearest neighbors in z-direction. $\triangle x$ and $\triangle z$ are the pixel sizes in x/y direction and z direction accordingly. The locally dependent prior probability is given by

$$p(x_i|x_{\mathcal{N}_i}) = \frac{e^{-V(x_i)}}{\sum_{x_j \in \mathcal{X}} e^{-V(x_j)}}. \tag{16}$$

Thus the contribution of a neighbor is penalized with its Euclidian distance to the central voxel. As MRI images are usually acquired with a square pixel size, the slice thickness is often much larger than the pixel size (otherwise the acquisition time would take too much time) which results in non-isotropic voxels. This feature must be incorporated into the energy function because a neighbor voxel in the next slice has clearly much smaller contribution than a neighbor voxel in the same slice. Therefore an anisotropic 3D-neighborhood system is utilized with a smaller weight over the slices.

3 Intensity Standardization

The goal of the intensity standardization approach is to find a mapping between the intensities of a set of images $U = (U_1, U_2, \ldots, U_n)$, where n is the number of images and a reference set of images $R = (R_1, R_2, \ldots, R_n)$ so that an arbitrary intensity vector $\boldsymbol{i} \in \mathcal{I}^n$ describes the same tissue class in both sets. The main idea of the contribution is that this can be approximated by the minimization of the distance between the joint pdfs of both image sequences. The required pdfs have a dimensionality of n, corresponding to the number of images. The domain is $]-\infty, +\infty[^n$. In practice, however, it can be scaled to $[0, 1]^n$ due to limited gray values observed. The pdfs of both tuples will never

be equal (at least for real datasets) as the volume of equal tissue classes differs for inter- as well as for intra-patient measurements (e.g. partial volume effects, positioning of the patient, ...). Thus the search for a gray value mapping is equivalent to finding the deformation between the pdfs so that they are closest with respect to a given distance measure. If the joint pdfs are treated as images, this task is called image registration or image fusion.

Image registration can be summarized as the problem of finding a deformation between a reference image A and a template image B so that the deformed template image B_φ is similar regarding a certain distance measure $\mathcal{D}$. Here the reference image A corresponds to the pdf p_R representing the multi-dimensional density function of R and respectively $B = p_U$ the density of U.

The used distance measure depends on the application. The most common ones are the sum of squared differences (SSD) for mono-modal applications, the normalized cross correlation (NCC) and mutual information (MI) for multi-modal problems [14, 15]. Because the function values of the pdfs have equal meaning the usage of SSD is sufficient in this article. However, the minimization of the proposed distance measures yields an ill-posed optimization problem. For this reason, further regularization terms have to be added to smooth the objective function. These so-called smoothers restrict the deformation of the template image in general. Mostly either elastic, fluid, curvature or diffusion approaches are utilized. Here the deformation field shall not change very fast, thus we chose a curvature based regularizer.

As we have to register density functions variational registration methods are chosen. These have the advantages that the images to be aligned are treated as functions $f : \mathbb{R}^n \mapsto \mathbb{R}$. The method used is based on an approach introduced by J. Modersitzki [16]. The minimization problem to be solved can be formulated as

$$\mathcal{J}[p_R, p_U; \boldsymbol{u}] = \mathcal{D}[p_R, p_U; \boldsymbol{u}] + \alpha\mathcal{S}[\boldsymbol{u}], \quad (17)$$

where the function $\boldsymbol{u} : \mathbb{R}^n \mapsto \mathbb{R}^n$ corresponds to the deformation field and n is the dimensionality of the pdfs. Furthermore, $\mathcal{D}$ is the distance measure and $\mathcal{S}$ represents the smoother. The factor α defines the influence of the regularizer on the objective function. The deformed density function p_{U_φ} can be computed as $p_{U_\varphi}(\boldsymbol{i}) = p_U \circ \varphi(\boldsymbol{i})$ with

$$\varphi(\boldsymbol{i}) = \boldsymbol{i} - \boldsymbol{u}(\boldsymbol{i}). \quad (18)$$

As mentioned before we use a similarity measure based on SSD to compute the distance between the pdfs. In our context this can be formulated as

$$\mathcal{D}^{SSD}[p_R, p_U; \boldsymbol{u}] = \frac{1}{2}\int_\Omega (p_{U_\varphi}(\boldsymbol{i}) - p_R(\boldsymbol{i}))^2 \, \mathrm{d}\boldsymbol{i}. \quad (19)$$

It calculates the distance between the functions related to their function values at a position $\boldsymbol{i}$, with $\Omega = [0, 1]^n$ representing the image domain.

The curvature based regularizer used in this article can be formulated as

$$\mathcal{S}^{curv}[\boldsymbol{u}] := \frac{1}{2}\sum_{l=1}^{n}\int_\Omega (\triangle u_l)^2 d\boldsymbol{x}, \quad (20)$$

with $\triangle$ being the Laplacian operator. In order to find the minimum of the objective function $\mathcal{J}$, a variational problem of first order has to be solved. Therefore, the Gâteaux derivative has to be applied to $\mathcal{J}$. Thus the variational gradient for the proposed distance measure yields [16]

$$f^{SSD}(\boldsymbol{i}, \boldsymbol{u}(\boldsymbol{i})) = d\mathcal{D}^{SSD}[p_R, p_U; \boldsymbol{u}]$$
$$= (p_R(\boldsymbol{i}) - p_{U_\varphi}(\boldsymbol{i}))\nabla p_{U_\varphi}(\boldsymbol{i}), \quad (21)$$

where the operator $d\mathcal{D}$ is the Gâteaux derivative.

The derivative $d\mathcal{S}^{curv}$ with respect to Neumann boundary conditions results in

$$\mathcal{A}^{curv}[\boldsymbol{u}] = d\mathcal{S}^{curv}[\boldsymbol{u}] = \triangle^2\boldsymbol{u}. \quad (22)$$

Using the introduced similarity measures (eqn 21) and the proposed smoother (eqn 22) the solution to the variational optimization problem can be found by solving the Euler Lagrange equation

$$\mathcal{A}^{curv}[\boldsymbol{u}] - f^{SSD}(\boldsymbol{i}, \boldsymbol{u}(\boldsymbol{i})) = 0, \quad \text{for all} \quad \boldsymbol{i} \in \Omega. \quad (23)$$

The result of the optimization problem is the deformed density function p_{U_φ} and much more important the deformation function $\boldsymbol{u} : \mathbb{R}^n \mapsto \mathbb{R}^n$. In the case of the registration of multi-dimensional pdfs it describes how to transform the gray values of one set of images U such that the intensity distribution matches a reference distribution best regarding $\mathcal{D}^{SSD}$ and the restriction $\mathcal{S}^{curv}$. Hence the intensity standardization can be done by

$$\boldsymbol{i}_{corr} = \boldsymbol{i}_U + \boldsymbol{u}(\boldsymbol{i}_U). \quad (24)$$
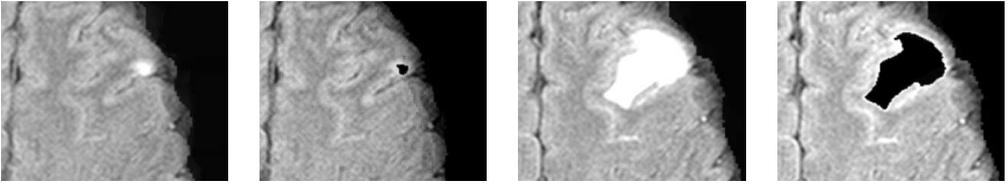
Figure 1: Images from left to right: The first two images show the used reference FLAIR image including the lesion and with the lesion extracted. The third and forth image show the template FLAIR images with and without lesion. All images are from the same patient and approximately same slice.

4 Experiments and Results

4.1 Experimental Setup

For the evaluation basically T1- and T2/FLAIR images were used. The T2-weighted FLAIR datasets were acquired on a Siemens Symphony 1.5 T scanner with 408x512x19, 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 208x256x19 with 0.86 mm^2 and 7.2 mm slice thickness and TE = 14 and TR = 510. All images used were real patient data including evolving lesions.

All algorithms used were implemented in C++ and integrated in the medical software package In-Space3D.

4.2 Results

In order to evaluate the intensity standardization the mean distance of the reference and the template volumes of one patient was chosen. Hence a good standardization result has a much smaller bias between the reference and the corrected than the unprocessed images. Furthermore, the variance of both distances is given by:

$$\mu = \frac{1}{N} \sum_i (x_i - y_i) \qquad \text{and} \qquad (25)$$

$$\sigma^2 = \frac{1}{N} \sum_i (x_i - y_i - \mu)^2, \qquad (26)$$

with N being the number of used voxels, x_i being a voxel in the template and y_i a corresponding voxel in the reference volume. However, the evaluation method has the drawback, that real patient data with evolving structures was used and thus the anatomy of the brain slightly changed. The lesions were removed by a segmentation step beforehand

(just for evaluation, not for the intensity standardization step!). Only those voxels that are classified as healthy brain tissue in both volumes are considered. In Figure 1 slices including and excluding lesions are shown. Here it is obvious, that even after extraction, the surrounding tissue changed. Furthermore, a rigid registration using normalized mutual information [17] as distance measure and interpolation had to be utilized for evaluation, so that the reference and template volumes match each other. The mean μ using the corrected images yielded 0.33 and $\sigma^2 = 180$ without the standardization the mean difference was 4.56 and $\sigma^2 = 183$ for the FLAIR images. In both cases the number of used voxels N was about 10^6. The difference for the T1 images resulted in $\mu = -0.96$ and $\sigma^2 = 134$ for the corrected and $\mu = -2.51$ and $\sigma^2 = 147$ for the uncorrected images with $N \approx 3 \cdot 10^5$.

The motivation of the introduced method is to adjust the image statistics so that a learned model can be utilized for classification. The performance difference of training and standardization are very significant (about one min for standardization and classification; about 30-45 min for training and classification). In Figure 2 the classification results of two different volume pairs (FLAIR and T1) of different patients are shown. The standardization for both sets was done using the same reference volume and statistics. It is apparent that the classification yielded better results using learned statistics and intensity classification. Especially in the lower row, without previous standardization everything is classified as lesion (white). In the upper row the small lesion on the righthand side is still feasible. In Figure 3 the marginals of the joint pdf before and after the standardization are shown. The marginals of the joint pdfs correspond to the pdfs of the single volumes. The standardization and classification was tested on seven image pairs.
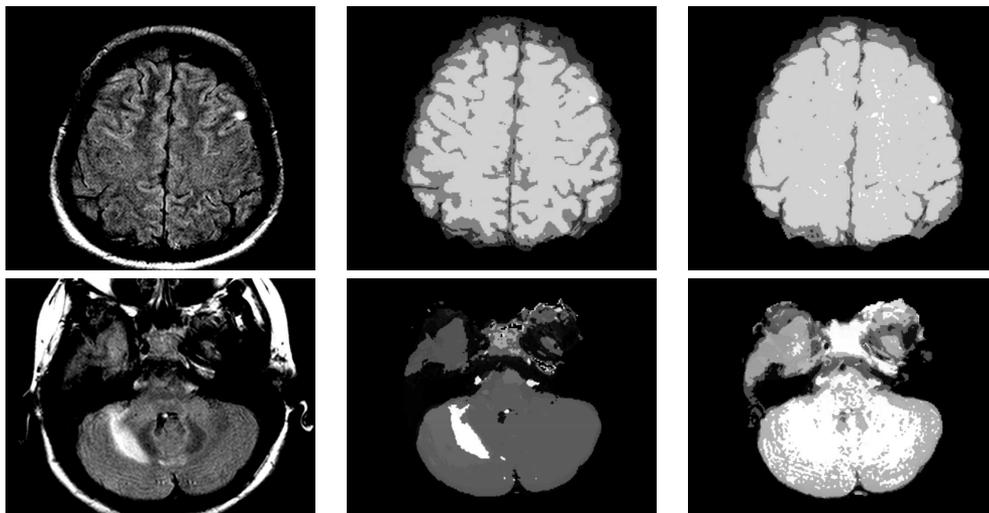
Figure 2: Images from left to right: In the first column FLAIR images of two newly acquired volumes is shown. Middle column: classification result using the learned statistics and intensity standardization. Right column: Classification result of the newly acquired images using the learned statistics without standardization.

5 Summary

Intensity standardization of MRI images is an important task as it highly influences the performance and quality of segmentation algorithms. In this article we presented a new method relying on the nonrigid registration of joint probability densities. Thus the introduced approach is independent of the application, protocol, region of interest and even of the acquisition modality. Furthermore, we showed that the corrected data offers a great benefit for segmentation algorithms, as it can adapt the image statistics to reference ones. As it uses information of all sequences equally more complex intensity distributions can be dealt with. This was shown with application to fast classification of brain tissue using a statistical approach called hidden Markov random field.

6 Conclusion

The presented standardization method is a reliable way to adjust image statistics of multiple series of MRI images. However, the results have to be verified in a broader range and evaluated for different body regions as well. After standardization statistical classification methods can be used more eas-ily in other regions of interest with more complex statistics as well.

The method is independent according to the acquisition protocol and body region, thus it is useful for other modalities, too. In future there is a SPECT normalization planned.

Acknowledgment

References

[1] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of mri data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–442, August 1996.

[2] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based bias field correction of MR images of the brain," *IEEE Transactions on Medical Imag-*
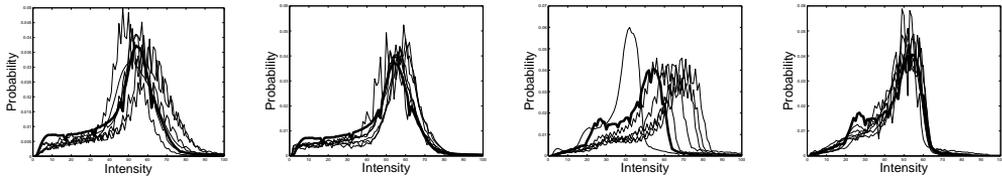
Figure 3: Images from left to right: All images show marginals of the joint probability density. The first images shows the densities of the FLAIR images before correction, the second after standardization. The third image shows the marginals of the T1-images before and the fourth after correction. The bold plot is the reference.

ing, vol. 18, no. 10, pp. 885–896, October 1999.

[3] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.

[4] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, February 2000.

[5] Y. Ge, J. K. Udupa, L. G. Nyúl, L. Wei, and R. I. Grossman, "Numerical tissue characterization in MS via standardization of the MR image intensity scale," in *Journal of Magnetic Resonance Imaging*, vol. 12, 2000, pp. 715–721.

[6] P. Hellier, "Consistent intensity correction of MR images," *International Conference on Image Processing (ICIP 2003)*, vol. 1, pp. 1109–12, September 2003.

[7] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897–908, October 1999.

[8] N. Weisenfeld and S. Warfield, "Normalization of joint image-intensity statistics in MRI using the kullback-leibler divergence," in *IEEE International Symposium on Biomedical Imaging*, vol. 1, 2004, pp. 101–104.

[9] M. Schmidt, "A method for standardizing mr intensities between slices and volumes," University of Alberta, Tech. Rep. TR05-14, 2005.

[10] A. Madabhushi and J. K. Udupa, "Interplay between intensity standardization and inho-

mogeneity correction in MR image processing," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 561–576, May 2005.

[11] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, ser. Computer Science Workbench, T. L. Kunii, Ed. Tokyo: Springer, 2001.

[12] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 677–688, August 2001.

[13] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, January 2001.

[14] G. Hermosillo, C. Chefd'hotel, and O. Faugeras, "Variational methods for multimodal image matching," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 329–343, 2002.

[15] F. Jäger, J. Han, J. Hornegger, and T. Kuwert, "A variational approach to spatially dependent non-rigid registration," in *Proc. SPIE*, J. M. Reinhardt and J. P. W. Pluim, Eds., vol. 6144, San Diego, February 2006, pp. 860–869.

[16] J. Modersitzki, *Numerical Methods for Image Registration*. Oxford University Press, Oxford New York, 2004.

[17] D. A. Hahn, J. Hornegger, W. Bautz, T. Kuwert, and W. Römer, "Unbiased rigid registration using transfer functions," in *Proc. of SPIE on Medical Imaging*, vol. 5747, 2005, pp. 151–162.