Synthetic MRI Signal Standardization: Application to Multi-atlas Analysis

Juan Eugenio Iglesias[1], Ivo Dinov[2], Jaskaran Singh[2],
Gregory Tong[2], and Zhuowen Tu[2]

[1] Medical Imaging Informatics, University of California, Los Angeles
`jeiglesias@ucla.edu`
[2] Laboratory of Neuroimaging, University of California, Los Angeles
`kearny89@ucla.edu, gregory.tong@ucla.edu, ivo.dinov@loni.ucla.edu,`
`zhuowen.tu@loni.ucla.edu`

Abstract. From the image analysis perspective, a disadvantage of MRI is the lack of image intensity standardization. Differences in coil sensitivity, pulse sequence and acquisition parameters lead to very different mappings from tissue properties to image intensity levels. This presents challenges for image analysis techniques because the distribution of image intensities for different brain regions can change substantially from scan to scan. Though intensity correction can sometimes alleviate this problem, it fails in more difficult scenarios in which different types of tissue are mapped to similar gray levels in one scan but different intensities in another. Here, we propose using multi-spectral data to create synthetic MRI scans matched to the intensity distribution of a given dataset using a physical model of acquisition. If the multi-spectral data are manually annotated, the labels can be transferred to the synthetic scans to build a dataset-tailored gold standard. The approach was tested on a multi-atlas based hippocampus segmentation framework using a publicly available database, significantly improving the results obtained with other intensity correction methods.

1 Introduction

Magnetic resonance imaging (MRI) is the modality of choice for brain imaging due to its excellent contrast in soft tissue. MRI images are a function of three properties of the tissue: the spin-lattice relaxation time T_1, the spin-spin relaxation time T_2, and the proton density ρ [1]. The way in which these physical properties are mapped to image intensities depends heavily on the imaging pulse sequence. The acquisition is typically designed to enhance the contribution of one of the properties and minimize the impact of the other two, leading to the well-known T_1-weighted (T_1-w), T_2-w and ρ-w imaging. Acquisition of real T_1 and T_2 maps[2] is possible but not very extended yet. Automated analysis of T_1-w, T_2-w and ρ-w images is difficult because the statistical distribution of voxel intensities can change significantly from scan to scan. Intensity standardization methods can alleviate this problem by trying to match the histogram of one scan

to that of a target volume. However, they cannot completely solve the problem because the image intensity of a scan is a function of three physical properties of the tissue, and the relationship between gray levels in two volumes is in general not one-to-one. A mapping can be found if three or more channels are available for both the source and target images[4], but this is seldom the case.

Meanwhile, multi-atlas segmentation is becoming increasingly popular in brain image analysis[5]. The main idea is to register a number of manually-labeled volumes to a target scan, propagate the annotations with the resulting transforms, and fuse the propagated labels to generate a probabilistic map that can be thresholded at a certain level (typically 0.5) to yield the final segmentation. If the atlases and the target scan have been acquired using different protocols, it is possible to use mutual information (MI) as the metric for the registration and then use a simple fusion scheme in which the mode of the labels at each location is taken[5]. If the image intensities are matched, the segmentation results can be improved by estimating the local success of the registration as the gray level difference between each registered atlas and the target volume at each point, and then using these differences to give a higher weight to the more accurately registered atlases in the label fusion scheme.

In this study, a MRI standardization method based on synthetic MRI is presented. A number of scans for which three channels (T_1-w, T_2-w and ρ-w) are available are used to synthesize T_1-w volumes matched to a specific target dataset. Manual annotations of the hippocampus made on the original T_1-w scans can be made, propagated to the synthetic volumes and finally used to segment the hippocampus in the target dataset in a multi-atlas setup. The hippocampus was selected as the target of the segmentation because its morphometry is relevant in many diseases (Alzheimers, Parkinsons, schizophrenia...) and, as a consequence, there are publicly available datasets with manual annotations that can be used to validate the technique. The performance of the proposed method in the multi-atlas framework is compared with the cases in which other intensity standardization algorithms are used.

2   Materials and Methods

2.1   Data

Two datasets are used in this study; one for training and one for evaluation. The training dataset consists of 10 volumes acquired with a Siemens 1.5T scanner using two different pulse sequences: 3D MP-RAGE for the T_1-w scans ($1{\times}1{\times}1$ mm^3 resolution, TR/TE/TI = 1,900/4.38/1,100ms., flip angle α = 15°) and double-echo spin echo for the T_2-w and ρ-w (TR/TE1/TE2 = 3,300/17/100ms., 1.4 mm slices with $1{\times}1$ mm^3 resolution). The left and right hippocampus were manually delineated in the T_1-w scans by two trained neuroscientists. These scans play the role of atlases in this study.

Fifteen scans from the publicly available Hippocampus Segmentation Database (HSD, *www.radiologyresearch.org/HippocampusSegmentationDatabase*) were used with testing purposes. Ten of them belong to patients with temporal

lobe epilepsy and may have atrophic hippocampi, which makes the segmentation harder. These 15 scans were acquired with a 1.5T GE scanner using a inversion recovery spoiled gradient echo sequence (IR-SPGR) with TR/TE/TI = 7.6/1.7/500 ms, flip angle $\alpha = 20°$. Two mm coronal slices were acquired with a $0.78 \times 0.78\ mm^2$ in-plane resolution. The same two neuroscientists who annotated the three-channel dataset also delineated the hippocampus in a subset of coronal slices of 10 scans with inter-reader variability evaluation purposes.

2.2 Synthetic MRI Generation

Estimation of true T_1, T_2 and ρ: The first step in the generation of synthetic MRI images is to calculate the true physical properties of the tissue i.e. the true T_1, T_2 and ρ maps of the training dataset. In an ideal situation, a large number of volumes would be acquired with different parameters (typically TE/TR) and then fed to a least-squares or expectation maximization algorithm to compute a robust estimate of the physical values [6]. However, in this study we assume the typical clinical scenario in which the available volumes are T_2-w and ρ-w scans acquired with a double-echo sequence and a T_1-w scan acquired with a fast T_1 sequence such as SPGR or MP-RAGE. To minimize the impact of noise on our estimates, the volumes were first denoised with an implementation of the Perona-Malik filter[7]. Then, the scans were skull-stripped with the BET algorithm[8] and subsequently bias field corrected with the N3 method[9]. Finally, the T_1-w scans were registered to the T_2 and ρ-w scans. The software package ITK (*www.itk.org*) was used to maximize a MI metric over a rigid transform (i.e. rotation and translation), which suffices to align the scans because they are images from the same patient acquired a few minutes apart.

Now, assuming the pulse sequences for the training dataset (i.e. double-echo spin echo, MP-RAGE), the signal models for the T_2-w, ρ-w and T_1-w scans are[1]:

$$S_2(\boldsymbol{r}) = k_2 \cdot \rho(\boldsymbol{r}) \cdot (1 - e^{-\frac{TR_2}{T_1(\boldsymbol{r})}})e^{-\frac{TE_2}{T_2(\boldsymbol{r})}} \tag{1}$$

$$S_\rho(\boldsymbol{r}) = k_\rho \cdot \rho(\boldsymbol{r}) \cdot (1 - e^{-\frac{TR_\rho}{T_1(\boldsymbol{r})}})e^{-\frac{TE_\rho}{T_2(\boldsymbol{r})}} \tag{2}$$

$$S_1(\boldsymbol{r}) = k_1 \cdot \rho(\boldsymbol{r}) \cdot f(T_1(\boldsymbol{r})) \tag{3}$$

where $\boldsymbol{r}$ is the position vector, TR_1 and $TR_2 = TR_\rho$ are the repetition times and TE_1, TE_2 and TE_ρ are the echo times. The constants k_1, k_2 and k_ρ account for global gains due to coil sensitivity, digital signal stretching and other constant multiplicative factors. The function $\rho(\boldsymbol{r})$ is proportional to the magnetization, which is in turn proportional to the proton density, but also affected by other factors such as flow attenuation. The function $\rho(\boldsymbol{r})$ is arbitrarily rescaled so that its maximum is one. The function $f(T_1)$, which neglects T_2^* effects, has a complex expression which is not reproduced here; the reader is referred to [10] for the details. Nevertheless, $f(T_1)$ is strictly increasing and hence invertible, which will be useful in the proposed algorithm.

The system of equations can be solved if prior knowledge on the characteristic T_1 and T_2 of some type of tissue is used; otherwise k_1, k_2 and k_ρ cannot be

Table 1. Gauss-Siedel algorithm to estimate the true T_1, T_2 and ρ maps

1. Estimate $T_2(\boldsymbol{r})$ from equations 1 and 2 as $T_2(\boldsymbol{r}) = (TE_2 - TE_\rho)/\log \frac{k_2 S_\rho(\boldsymbol{r})}{k_\rho S_2(\boldsymbol{r})}$
where the ratio $\frac{k_2}{k_\rho}$ can be obtained as $\frac{k_2}{k_\rho} = \frac{\bar{S}_{2,WM}}{S_{\rho,WM}} \exp\left(\frac{TE_2 - TE_\rho}{T_{2,WM}}\right)$
2.Assuming $T_1(\boldsymbol{r}) = 0$, use equation 1 to initialize $\rho'(\boldsymbol{r}) \leftarrow S_2(\boldsymbol{r}) \exp\left(\frac{TE_2}{T_2(\boldsymbol{r})}\right)$
$k_2 \leftarrow \max_{\boldsymbol{r}} \rho'(\boldsymbol{r})$ and $\rho(\boldsymbol{r}) \leftarrow \rho'(\boldsymbol{r})/k_2$
3.Estimate k_1 using the current values of $\rho(\boldsymbol{r})$: $k_1 \leftarrow \bar{S}_{1,WM}/[f(T_{1,WM}) \cdot \bar{\rho}_{WM}]$
where $\bar{\rho}_{WM}$ is the median value of $\rho(\boldsymbol{r})$ in the white matter mask.
4.Update the real T_1 volume using equation 3: $T_1(\boldsymbol{r}) \leftarrow f^{-1}(S_1(\boldsymbol{r})/k_1\rho(\boldsymbol{r}))$
5.Update $\rho'(\boldsymbol{r}) \leftarrow S_2(\boldsymbol{r}) \exp[TE_2/T_2(\boldsymbol{r})]/[1 - \exp(-TR_2/T_1(\boldsymbol{r}))]$, as well as
$k_2 \leftarrow \max_{\boldsymbol{r}} \rho'(\boldsymbol{r})$ and $\rho(\boldsymbol{r}) \leftarrow \rho'(\boldsymbol{r})/k_2$
6.Go to step 3 until convergence; seven or eight iterations often suffice. Figure
1 shows the T_1, T_2 and ρ maps for an axial slice of one of the training cases.

Fig. 1. T_1, T_2 and ρ maps for an axial slice of one of the cases in the training dataset

estimated. We propose using the white matter, which can be segmented in the
T_1-w scan with the method from [11]. Even if the segmentation is not perfect, we
can use the median value (a robust estimate) of the images in the mask $\bar{S}_{1,WM}$,
$\bar{S}_{2,WM}$ and $\bar{S}_{\rho,WM}$ to match the predicted and theoretical values of T_1 and T_2
in the white matter: $T_{1,WM}$=785 ms (at 1.5T), $T_{2,WM}$=92 ms[12]. The physical
parameters are then estimated with the Gauss-Siedel method (Table 1).

Synthesizing dataset-matched volumes: Once the real T_1, T_2 and ρ maps
have been estimated, it is straightforward to compute how the volumes would
look like if they had been acquired with the T_1-w IR-SPGR pulse sequence of
the HSD dataset. The approximate signal equation for IR-SPGR is [1]:

$$S_{HSD} \propto \rho(\boldsymbol{r})[1 - 2e^{-\frac{TI}{T_1(\boldsymbol{r})}} + e^{-\frac{TR}{T_1(\boldsymbol{r})}}]$$

The intensities can be computed up to a multiplicative constant. The constant
is estimated by matching the median intensity of the brain (a very consistent
intensity landmark) in the synthesized and test volumes.

2.3   Multi-atlas Segmentation

The multi-atlas segmentation framework from [13] was used in this study. The atlases are registered to the (skull-stripped, bias field corrected) scan and the absolute difference image is calculated for each registered atlas. These difference images are blurred and inverted (previous addition of a small constant ϵ to prevent division by zero) to yield the weight images $\lambda_i(\boldsymbol{r})$, $i \in \{1, \ldots, N_{atlas}\}$, where $\boldsymbol{r}$ is again the position vector. Then, for each structure s to segment (i.e. the left and right hippocampus), the probability volume is given by:

$$P_s(\boldsymbol{r}) = \frac{\sum_{i=1}^{N_{atlas}} \lambda_i(\boldsymbol{r}) L_i(\boldsymbol{r})}{\sum_{j=1}^{N_{atlas}} \lambda_j(\boldsymbol{r})} \tag{4}$$

where $L_i(\boldsymbol{r})$ represents the propagated labels from atlas i (one if inside, zero if outside). The probability volume is blurred and thresholded at 0.5 to generate a binary map, out of which the largest connected component is extracted as the final segmentation. The performance depends on the quality of the weight images, which in turn depends on the image standardization.

3   Experiments and Results

3.1   Experiment Setup

Two experiments were designed in this study. In both of them, the first step was to register and match the intensities of the atlases to the test images. ITK was used to optimize a B-spline nonrigid transform using a MI metric between the T_1-w volumes. Then, the image intensity of the atlases was matched to that of the test image using our approach (which uses three channels) as well as three other methods (based solely on the T_1-w volume): 1) histogram stretching, disregarding the top and bottom 1% of the histogram; 2) histogram equalization; and 3) landmark-based matching[3] using eight landmarks (including the median white and gray matter intensities given by the method from [11]).

The first experiment quantizes the influence of the method on the mean joint histogram of the samples and the registered, intensity-corrected atlases. In the second experiment, the impact of the presented approach on a multi-atlas method to segment the hippocampus is assessed. The dependence of the performance on the intensity-matching strategy is analyzed (the registration was the same for all the methods) and compared to the inter-user variability in the manual annotations, which defines the boundary for the performance of the system. In both experiments, the width of the Gaussian kernels to smooth the difference image and the likelihood map was set to $\sigma=1$ mm, whereas the regularization parameter to invert the difference image was set to $\epsilon = 10^{-3}$.

3.2   Results

Joint histogram: The success of a intensity-matching strategy can be evaluated by building the mean joint histogram of the test images and registered,
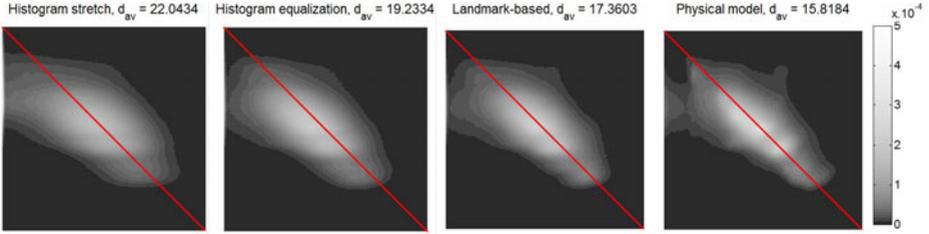
Fig. 2. Mean joint histogram of the images of the HSD dataset (rows) and the corresponding atlases (columns) after registration and intensity matching using different strategies. The color map follows a logarithmic scale. The weighted average distance from each bin to the diagonal is displayed above each joint histogram.

intensity corrected atlases. If the registration and matching were perfect, the histogram would be a diagonal matrix, but in practice the result is a cloud around the diagonal. The weighted average distance from each bin to the diagonal can thus be used as a performance metric. This measure is displayed, along with the histograms, in Figure 2. Our physical model outperforms the landmark-based approach, which in turn outperforms histogram equalization and stretching.

Multi-atlas segmentation: In order to evaluate the performance of the segmentation with an overlap measure (the Dice coefficient in this study), it is important to analyze the impact of the inter-reader variability on the measure first. For that purpose, the neuroscientists who annotated the scans in this study also delineated the hippocampus in a subset of slices of the test dataset (see section 2.1). The neuroscientists were not shown the ground truth annotations from the test dataset in order to prevent bias in their delineations. The average Dice coefficient was 0.783 ± 0.086 for the left hippocampus and 0.825 ± 0.033 for the right; these values are not significantly different according to a proportion equality test ($p \approx 0.45$). These values for the overlap measure are consistent with other values reported in the literature [14], and represent a boundary for the performance our algorithm can achieve.

Table 2 displays the Dice coefficients achieved by the multi-atlas scheme for the original fusion method from [5] (i.e. taking the mode of the labels) and for the intensity-based fusion from equation 4 when the different intensity matching methods are used. The algorithms outperform one another in the same order as in the previous experiment: our method provides again the highest performance. A paired t-test shows that our method is better than the landmark-based approach (which provides the second largest overlap) at a level of significance $\alpha = 0.05$. In absolute terms, the mean Dice coefficient provided by the proposed method (0.755) is in the range of results of other recent methods in the literature in which the training and test datasets have been acquired with different scanners and/or pulse sequences (for example [15]). The segmentations for the first two volumes of the HSD dataset are shown in figure 3.

Table 2. Dice coefficients for the multi-atlas segmentation of the left and right hippocampus. The last column is the p-value of a paired t-test comparing our method with the landmark-based approach.

Method	Mode[5]	Hist. stret.	Hist. equal.	Landmark-based	This study	p
Left	0.678±0.226	0.702±0.117	0.713±0.142	0.718±0.133	0.748±0.126	0.0306
Right	0.695±0.212	0.711±0.089	0.722±0.125	0.733±0.106	0.761±0.096	0.0562
Both	0.686±0.215	0.706±0.102	0.718±0.132	0.725±0.119	0.755±0.110	0.0403
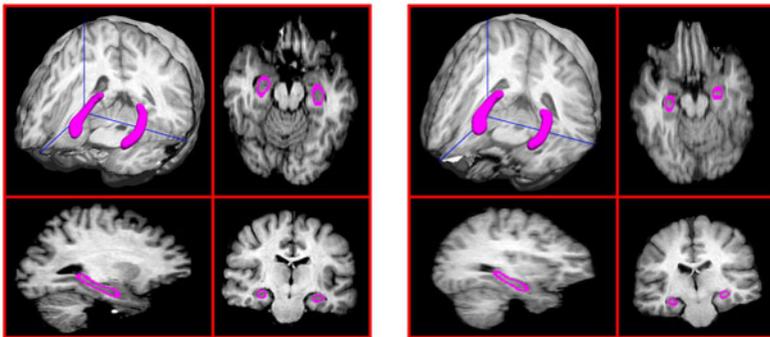
Fig. 3. Segmentation of the left and right hippocampus in the first two volumes of the test dataset: three orthogonal planes and 3D rendering

4   Discussion

A MRI image intensity standardization method based on synthetic MRI and its application to multi-atlas segmentation have been presented in this paper. Despite approximations such as neglecting T_2^* effects and despite the fact that only three channels are used to estimate the T_1, T_2 and ρ maps (unlike other synthetic MRI methods that use many more), the method provides a fairly accurate intensity mapping between different acquisition protocols. Though a multi-atlas framework was used to test the method, any other image analysis application could in principle benefit from the algorithm. The results for the hippocampus segmentation are reasonable given that only ten atlases were utilized, and they should improve if more atlases were available and a proper atlas selection process could be carried out. The performance is also limited by the fact that the annotations of our dataset and HSD are not very consistent; this inter-user variability gives an idea of how difficult automated hippocampus segmentation is.

A potential disadvantage of the proposed algorithm is that, whereas T_2 is relatively independent of the magnetic field strength of the scanner, T_1 presents a dependence on this parameter. This study is based on images acquired at 1.5T. As of today, most scans are acquired at 1.5T or 3T, which are the standard clinical field strengths. It would therefore be possible to annotate a set of atlases acquired at 3T to cover most of the clinical scans currently acquired in the World. Another possibility would be to modify the T_1 map of the atlases to

account for the field strength difference using a model of the relationship between them. Addressing this problem, testing the method on more scans from different datasets and utilizing more atlases in the segmentation remain as future work.

Acknowledgements

References

1. Bernstein, M., King, K., Zhou, X.: Handbook of MRI Pulse Sequences. Elsevier, Academic Press (2004)
2. Deoni, S., Rutt, B., Peters, T.: Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. Magnetic Resonance in Medicine 49(3), 515–526 (2003)
3. Nyul, L., Udupa, J., Zhang, X.: New variants of a method of MRI scale standardization. IEEE T. Med. Imaging 19(2), 143–150 (2000)
4. Jager, F., Hornegger, J.: Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. IEEE T. Med. Imaging 28(1), 137–150 (2009)
5. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33(1), 115–126 (2006)
6. Maitra, R., Riddles, J.J.: Synthetic magnetic resonance imaging revisited. IEEE T. Med. Imaging 29(3), 895–902 (2010)
7. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE T. Pattern Anal. 12(7), 629–639 (1990)
8. Smith, S.: Fast robust automated brain extraction. Hum. Brain Mapp. 17(3), 143–155 (2002)
9. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in mri. IEEE T.Med. Imaging 17, 87 (1998)
10. Deichmann, R., Good, C.D., Josephs, O., Ashburner, J., Turner, R.: Optimization of 3-D MP-RAGE for structural brain imaging. Neuroimage 12(1), 112–127 (2000)
11. Shattuck, D., Leahy, R.: Brainsuite: An automated cortical surface identification tool. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 50–61. Springer, Heidelberg (2000)
12. Nishimura, D.: Principles of Magnetic Resonance Imaging. Stanford U., Stanford (1996)
13. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in ct scans. IEEE T. Med. Imaging 28, 1000–1010 (2009)
14. Jeukens, C., Vlooswijk, M., Marian Majoie, H., de Krom, M., Aldenkamp, A., Hofman, P., Jansen, J., Backes, W.: Hippocampal MRI Volumetry at 3 Tesla: Reliability and Practical Guidance. Invest. Radiol. 44(9), 509 (2009)
15. Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE T. Med. Imaging 27, 495–508 (2008)