Intensity normalization for improved MR images analysis

Florent Lalys[1], Sushmita Datta[1], Léorah Freeman[1], Stacey S. Cofield[2], Gary R. Cutter[2], Fred D. Lublin[2], Jerry S. Wolinsky[1], and Ponnada A. Narayana[1]
[1]University of Texas Health Science Center at Houston, Houston, Texas, United States, [2]University of Alabama at Birmingham, Birmingham, Alabama, United States

Targeted audience: Computer and image processing scientists

Purpose: Most MRI segmentation techniques are intensity based. Since the image intensity varies from patient-to-patient, scanner-to-scanner, and across imaging sessions, intensity normalization (IN) is a critical step in image processing, and particularly for consistent segmentation. The IN technique described by Nyul et al [1] is commonly used because of its simplicity, effectiveness, accuracy and computation time [2]. The strategy underlying this algorithm is to deform in a piece-wise linear manner based on pre-defined landmarks in the image histograms and match these landmarks to a reference histogram determined through training. This method has been recently validated in the context of multiple sclerosis (MS) segmentation on a relatively small sampling of patients [3]. However, estimation of the optimal parameters for IN is challenging, and not evaluated critically on large cohorts. Further, there are no studies to assess a specific reference histogram or a specific set of histogram landmarks. The purpose of this work was to significantly improve IN by identifying an optimal set of parameters and verify them on a large cohort of MS patients.

Methods Experiments were carried out using MRI data from 1008 MS subjects enrolled in the NIH-sponsored CombiRx multi-center clinical trial [4]. The protocol included the acquisition of multi-channel images (T1w pre- and post-contrast, T2w, PDw and FLAIR) with patients scanned baseline, 6, 12, 24, and 36 months on study. A total of 27500 images were used. Image histogram landmarks defined in different ways were tested: median, decile, vingitiles, percentile and tissue-intensity formulations. Then, different reference scales were investigated: multi-subject multi-channel (all subjects with all the five channels), multi-subject single-channel, single-subject multi-channel, single-subject single-channel and single time-point scan (chosen as baseline scan). For evaluation, pairwise comparison between each time point scan and its corresponding baseline scan was performed. As a quantitative measure of IN quality, the normalized mean square difference (NMSD) was considered on the whole brain. The Kullback-Leibler (KL) histogram divergence was also considered for assessing the within tissue homogeneity based on a prior validated segmentation of cerebrospinal fluid, gray matter, white matter and lesion compartments of each subject's brain. Analyses were carried out for each MR channel separately, and multiple comparisons were used for validating statistical significance.

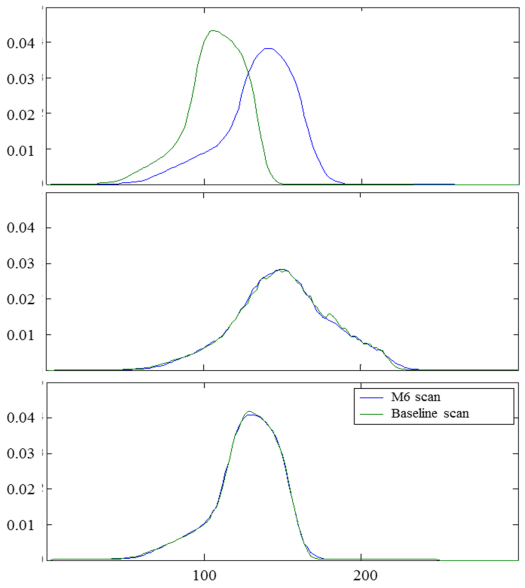

Fig. 1. Profiles of two intensity histograms (baseline scan and 6 months scan) with native images (above), intensity normalized images using a multi-subject multi-channel reference scale (middle) and a single-subject mono-channel reference scale (below). Decile formulation was used here.

Results: For each channel, patient scans after IN were found to have greater homogeneity (Fig. 1; mean KL = 0.039) and to be closer (Fig. 2; mean NMSD = 4.1) when using a single-subject single-channel reference scale along with the percentile formulation ($p<0.001$). At the other extreme, multi-subject multi-channel reference scans led to higher KL (mean = 0.048) and NMSD (mean = 7.5), independent of the type of formulation used ($p<0.0001$). An alternative compromise used a multi-subject single-channel scale with the vingitile formulation (mean NMSD = 5.9, mean KL = 0.044), which leads to low within tissue divergence without losing tissue types.

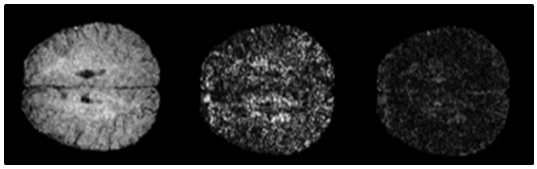

Fig. 2. Difference maps between baseline and 6 month scans with: native images (left), intensity normalized images using a multi-subject multi-channel reference scale (middle) and a single-subject mono-channel reference scale (right). Decile formulation was used here.

Discussion & conclusion: Here we propose a set of IN parameters that are superior to the ones most commonly used. The superior performance of our approach was evaluated on a large cohort of MS patients. To ensure homogeneity of tissue type intensities for image subtraction purposes, a single-subject single-channel reference scale should be used with the percentile formulation. For other applications (e.g. image segmentation) that need higher degrees of automation with a common reference scale, a multi-subject single-channel reference scale should be used along with the vingitile formulation. These findings contradict the parameters used by the majority of published studies.

References

1. Nyul et al. New variants of a method of MRI scale standardization. IEEE trans. Med Imaging. 2000; 19(2), 143-150
2. Bergeest and Jäger. A comparison of five methods for signal intensity standardization in MRI. Informatik aktuell. 2008; 36-40
3. Shah et al. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. MeDIA. 2011; 15(2):267-82
4. Lublin et al. Randomized study combining interferon and glatiramer acetate in multiple sclerosis. 2013; Ann Neurol 73, 327-340.