NeuroImage 90 (2014) 207–217

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Improved longitudinal gray and white matter atrophy assessment via application of a 4-dimensional hidden Markov random field model

Michael G. Dwyer *, Niels Bergsland, Robert Zivadinov

Buffalo Neuroimaging Analysis Center, Department of Neurology, School of Medicine and Biomedical Science, University at Buffalo, State University of New York, Buffalo, NY, USA

ARTICLE INFO

Article history:
Accepted 3 December 2013
Available online 12 December 2013

Keywords:
Gray matter
White matter
Tissue-specific
Atrophy
Longitudinal
Segmentation
Direct method
4-Dimensional HMRF
Random field

ABSTRACT

SIENA and similar techniques have demonstrated the utility of performing "direct" measurements as opposed to post-hoc comparison of cross-sectional data for the measurement of whole brain (WB) atrophy over time. However, gray matter (GM) and white matter (WM) atrophy are now widely recognized as important components of neurological disease progression, and are being actively evaluated as secondary endpoints in clinical trials. Direct measures of GM/WM change with advantages similar to SIENA have been lacking. We created a robust and easily-implemented method for direct longitudinal analysis of GM/WM atrophy, SIENAX multi-time-point (SIENAX-MTP). We built on the basic halfway-registration and mask composition components of SIENA to improve the raw output of FMRIB's FAST tissue segmentation tool. In addition, we created LFAST, a modified version of FAST incorporating a 4th dimension in its hidden Markov random field model in order to directly represent time. The method was validated by scan–rescan, simulation, comparison with SIENA, and two clinical effect size comparisons. All validation approaches demonstrated improved longitudinal precision with the proposed SIENAX-MTP method compared to SIENAX. For GM, simulation showed better correlation with experimental volume changes ($r = 0.992$ vs. 0.941), scan–rescan showed lower standard deviations (3.8% vs. 8.4%), correlation with SIENA was more robust ($r = 0.70$ vs. 0.53), and effect sizes were improved by up to 68%. Statistical power estimates indicated a potential drop of 55% in the number of subjects required to detect the same treatment effect with SIENAX-MTP vs. SIENAX. The proposed direct GM/WM method significantly improves on the standard SIENAX technique by trading a small amount of bias for a large reduction in variance, and may provide more precise data and additional statistical power in longitudinal studies.

Introduction

Brain atrophy measurement has become a key analysis in basic neuroimaging science, aging research, and research into pathologic conditions including multiple sclerosis (MS) (Bermel and Bakshi, 2006; Zivadinov and Bakshi, 2004), Alzheimer's disease (AD) (Sluimer et al., 2008), and Parkinson's disease (Burton et al., 2005). It is also becoming an important component of modern MS and AD clinical trials (Thal et al., 2006; Zivadinov et al., 2008). Readily available segmentation tools such as FMRIB's Structural Image Evaluation, using Normalisation, of Atrophy (SIENA) (Smith et al., 2001) allow atrophy measurements to be both reliable (Sormani et al., 2004) and highly standardized across studies and research centers (Jasperse et al., 2007).

A better understanding of the specific mechanisms of atrophy has led many researchers to focus on separate quantification of gray matter (GM) and white matter (WM) atrophy (Chard et al., 2004; Hulst and Geurts, 2011; Karas et al., 2003; Sanfilipo et al., 2006; Thompson et al., 2003; Zivadinov and Minagar, 2009; Zivadinov and Pirko, 2012). Although a variety of measurement approaches are available (Chen et al., 2004; Derakhshan et al., 2010; Nakamura and Fisher, 2009; Nakamura et al., 2011), one used by a number of groups is to perform independent tissue segmentations at both baseline and follow-up time points (potentially with some spatial normalization), and then calculate the changes via simple subtraction of the relevant total volumes (Healy et al., 2009; Horakova et al., 2009; Oreja-Guevara et al., 2005; Valsasina et al., 2005). Unfortunately, although this approach is straightforward, intuitive, and easily implemented, it is considerably less reproducible than a direct measurement like SIENA. In fact, even whole brain measures from SIENAX (the cross-sectional variant of SIENA) (Smith et al., 2002) are less reproducible than SIENA change measures (Cover et al., 2011), and GM/WM-specific measures are even more difficult. This reduction in precision can have serious consequences for the statistical power of planned studies, resulting either in the need for very large subject groups or in the inability to detect real changes (Anderson et al., 2007; Healy et al., 2009).

There are a number of potential reasons that reproducibility issues arise from this approach. First, and fundamentally, since two independent cross-sectional measures are used rather than a single direct measurement, there are two sources of measurement error. Without

* Corresponding author at: Buffalo Neuroimaging Analysis Center, Department of Neurology, School of Medicine and Biomedical Sciences, 100 High St., Buffalo, NY 14203, USA. Fax: +1 716 859 4005.
E-mail address: mgdwyer@bnac.net (M.G. Dwyer).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.12.004

a direct comparison, it is likely that the two segmentations will come to slightly different conclusions about the precise spatial and intensity distributions of the tissue classes involved. In particular, voxels of relatively ambiguous intensity (i.e., halfway between GM and WM) will often be classified differently at baseline and follow-up, despite a lack of change in actual tissue morphometry. Although these differences may ultimately cancel out in aggregate, they add to the overall variance of the measurement. This problem is also aggravated by the fact that the GM/WM border (as seen on conventional MRI) is generally not as clearly defined as the brain/cerebrospinal fluid (CSF) border, and the absolute intensity contrast between the two tissues is usually considerably lower than that between GM and CSF.

Second, scanner drift and differences in positioning can lead to minor geometric distortions in the acquired images that can change volumetric measurements (Freeborough, 1996). Even when subtle, these changes can dwarf small clinical changes that are the target of studies and clinical trials. Although their nonlinear nature can make them challenging to completely correct (Caramanos et al., 2010), they can be at least somewhat ameliorated by improved co-registration with full affine parameters.

Third, brain extraction can have a significant effect on measured tissue volumes (Battaglini et al., 2008; Keihaninejad et al., 2010; Leung et al., 2011; Popescu et al., 2012), so inconsistent brain extraction at baseline and follow-up can lead to tissue volume changes that do not reflect actual atrophy.

To address these issues, we developed a novel technique that augments FMRIB's Automated Segmentation Tool (FAST) algorithm with a 4-dimensional hidden Markov random field (HMRF) to ensure more consistent classification. In the current version of FAST, a 3-dimensional HMRF is used to impose local spatial constraints on the segmentation process. Essentially, this HMRF penalizes discrepancies in tissue classification for isolated voxels, but allows for contiguous areas of change (Zhang et al., 2001). So, for example, a shift in the WM/GM border when moving from slice to slice is not significantly penalized. By extending the basic spatial model to a full spatio-temporal (4-dimensional) model, we hoped to allow for the same sort of shift in border over the time dimension – atrophy/growth/shifting – instead of the slice direction dimension (or any other spatial dimension), while simultaneously penalizing small, localized discontinuities. Although this approach will introduce a small bias in the results toward no change and might also have limited benefit in the face of very large deformations, we hypothesized that these concerns would be outweighed by the potential reduction in noise-induced variance.

Additionally, we incorporated two of the key elements of SIENA into our tissue-specific analysis technique: skull-constrained halfway-space co-registration to address positioning and scanner drift issues, and uniform brain extraction to reduce extraction-related variance.

Materials and methods

Inclusion of a temporal component in the HMRF

The HMRF framework employed by FAST (Zhang et al., 2001) is designed to use spatial neighborhood information to elegantly mitigate the noise and homogeneity problems inherent in MRI-based tissue segmentation. Intuitively, when assigning tissue class labels to voxels, the algorithm attempts to minimize a global cost function that penalizes two separate elements: selection of class labels whose mean intensities do not match well with the labeled voxels (e.g., labeling a relatively bright voxel GM on a T1-weighted image), and creation of spatially isolated labels (e.g., a single GM voxel completely surrounded by WM). It is the tension and balance between these intensity-matching and spatial-homogeneity-preserving goals that allow for the quality of tissue segmentations achievable with FAST.

More rigorously, the class labeling posterior probability in standard FAST is updated iteratively according to

$$p^{(t)}\left(l|y_i\right) \propto \exp \beta w_{i,l} - \left(\log\left(\sqrt{2\pi\sigma_l^2}\right) + \frac{\left(y_i-\mu_l\right)^2}{2\sigma_l^2}\right)$$

where t is the iteration, l is the label of a specific class, i is the voxel index, β is the neighborhood weighting factor, w is a neighborhood weighting function, and the entire second term is the standard log Gaussian penalizing deviations of the voxel intensity from the proposed class mean. The proportionality rather than equivalence reflects the fact that FAST allows for the incorporation of spatial prior probabilities, and also for the fact that the absolute probabilities would need to be normalized. The key component in this context is the weighting function w, which expresses the overall a-priori probability of finding a three-dimensional class configuration where voxel i's class is l, given its already-classified neighbors. Internally, FAST calculates this term via a conversion to "MRF weights".

From a local perspective, the total label weight for a voxel i and label l is determined by iterating over its neighbors j in neighborhood N (where N is the neighborhood of the up to 26 voxels surrounding voxel i), and is calculated as:

$$w_{i,l} = \sum_{j \in N_i} \frac{1}{d(i,j)} \cdot p^{(t)}\left(l\middle|y_j\right)$$

where $d(i,j)$ is the distance between voxels i and j (diagonal neighbors are farther than horizontal neighbors, and the slice resolution is often lower than the in-plane resolution), and the right side is the current iteration's a posteriori probability of classification l for voxel j. Note that in implementation, half of the probabilities are calculated from the previous step due to linear scanning.

However, it is important to note that the mathematics behind the general HMRF model are not limited to the usual 3 spatial dimensions (Winkler, 2003), and in fact the implementation in FAST can be very naturally extended from 3 dimensions to 4 with minimal modification. As noted above, tissue atrophy or growth can be considered as a change in the border between tissues or between tissue and CSF when moving along the time dimension, and is analogous to the shifts that occur when moving from slice to slice. The main distinction is that for a two-point tissue change analysis, each voxel has only one temporal neighbor whereas there are usually 26 spatial neighbors. Recognizing this, we modified the above calculation to be:

$$w_{i,l} = \sum_{j \in N_i} \frac{1}{d(i,j)} \cdot p^{(t)}\left(l\middle|y_j\right) + z p^{(t-1)}\left(l\middle|y_i'\right)$$

where the second term is the prior iteration a posteriori probability of classification l for voxel i in the same physical position as i but at the other time-point, multiplied by a factor z. Thus, this calculation now includes a regularization term for maintaining constant voxel classification over time as well as for the usual agreement with spatially neighboring classifications. The z factor is a weighting coefficient to control the importance of temporal consistency in the model, and can be specified either as a constant or as a constant divided by the length of time between scans.

We implemented the above scheme, called LFAST (for longitudinal FAST), by modifying FAST to be multithreaded, and having the two segmentations proceed in parallel. We used a producer/consumer semaphore system to cause each thread of FAST to operate in symmetric lock-step, moving on to the next iteration only when the total cost function could be calculated by each side comparison. Furthermore, this was implemented in an unbiased way such that neither time point was processed "first" or influenced the other time point in a non-reciprocal manner.