Medical Image Analysis 38 (2017) 50–64

Contents lists available at ScienceDirect

Medical Image Analysis

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

Longitudinal segmentation of age-related white matter hyperintensities

Carole H. Sudre[a,b,*], M. Jorge Cardoso[a,b], Sebastien Ourselin[a,b], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Translational Imaging Group, Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK

[b] Dementia Research Centre, UCL Institute of Neurology, University College London, London, UK

ARTICLE INFO

Article history:
Received 1 August 2016
Revised 13 February 2017
Accepted 15 February 2017
Available online 24 February 2017

Keywords:
White matter hyperintensities
Segmentation
Longitudinal

ABSTRACT

Although white matter hyperintensities evolve in the course of ageing, few solutions exist to consider the lesion segmentation problem longitudinally. Based on an existing automatic lesion segmentation algorithm, a longitudinal extension is proposed. For evaluation purposes, a longitudinal lesion simulator is created allowing for the comparison between the longitudinal and the cross-sectional version in various situations of lesion load progression. Finally, applied to clinical data, the proposed framework demonstrates an increased robustness compared to available cross-sectional methods and findings are aligned with previously reported clinical patterns.

1. Introduction

White matter hyperintensities (WMH), also known as leukoaraiosis, as observed in FLuid Attenuated Inversion Recovery (FLAIR), T2-weighted (T2) and proton density weighted (PD) magnetic resonance (MR) images are widely observed in the ageing population. The abnormal signal, explained by a change in the fat/water ratio, reflects a damage to the white matter. Hypotheses related to deleterious changes in the blood supply and in the blood brain barrier (Wardlaw et al., 2013) have been put forward to explain the occurrence of such damage, and cardiovascular risk factors such as hypertension have been shown to be associated to the WMH burden (Abraham et al., 2015; Vuorinen et al., 2011). Furthermore, such lesions have been linked with cognitive impairment, in particular with respect to processing speed and executive function (Prins and Scheltens, 2015; Wakefield et al., 2010).

To further assess potential causality effects between lesion burden and clinical outcome, new emphasis has been given to longitudinal studies of lesion load and cognitive assessment (Schmidt et al., 2005). In normal ageing, increase in the lesion volume with time was observed with a higher rate of change correlated with more severe baseline lesion volume (Pantoni and The LADIS Study group, 2011). For a normal population, progression in leukoaraiosis has been related to motor decline (Silbert et al., 2008), and cognitive disabilities (Schmidt et al., 2005) as well as memory impairment (Gunning-Dixon and Raz, 2000). Additionally, lesion burden at baseline has been associated with faster cognitive decline in Alzheimer's disease (AD), mild cognitive impairment (MCI) and normal populations (Carmichael et al., 2010).

The evaluation of WMH progression, however, remains difficult. In many cases, visual rating scales are used to assess the increase in severity of the lesion burden (Gouw et al., 2008). Most of them have however been developed for cross-sectional studies and are difficult to utilise in longitudinal cases due to the lack of sensitivity to change (Schmidt et al., 2005). Specific progressive rating scales have been proposed to alleviate this drawback (Prins et al., 2004), but volumetric measurements appear to allow for more accurate group differentiation (Pantoni and The LADIS Study group, 2011). Even when using semiautomatic segmentation methods for volume assessment (Schmidt et al., 2005) instead of performing the segmentation manually, the process remains time-consuming and the strategy of looking at images back-to-back can introduce bias

* Corresponding author at: Translational Imaging Group, Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, 8th Floor Malet Place Engineering Building, 2 Malet Place, WC1E 7JE, London UK

E-mail address: carole.sudre.12@ucl.ac.uk (C.H. Sudre).

[1] Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

http://dx.doi.org/10.1016/j.media.2017.02.007

(Schmidt et al., 2005). Therefore, longitudinal, robust automatic lesion segmentation solutions are greatly needed.

Even though imaging time points can be considered independently when automatically measuring the volume of WMH in longitudinal studies (Carmichael et al., 2010), it has been shown that considering the time points separately within subject introduced an additional source of variability in the results (Elliott et al., 2013). Accounting for the structural similarities between time points, or relating the information from one time point to others may increase the robustness of the method.

The problem of longitudinal lesion assessment is of great interest in other fields of neuroimaging such as multiple sclerosis (MS), and various methods have been designed to assess longitudinal lesion change. This issue is especially sensitive in MS, in which the lesion load progression is non-monotonic. Methods relying on the analysis of the differences between registered serial images, as in Rey et al. (2002), may be hindered by other volumetric changes occurring between the time points. In studies with long-term follow-up, in which the drop-off rate can be high (*e.g.* in age-related studies), being able to handle different numbers of time points is an additional challenge.

In the context of age-related WMH, the progressive nature of the damage can be taken as an argument to consider consecutive image pairs as in Bosc et al. (2003). However, noise and artefacts, prevalent in aging or in the demented population, may affect methods based on direct comparison; other solutions based on image averaging and model building may be advantageous. For instance, the use of average images to guide the processing of longitudinal data has been promoted in Reuter et al. (2012).

The solution developed in this work first consists in creating a longitudinal intra-subject average (Section 2.1), followed by the estimation of an appropriate joint Gaussian mixture model (GMM) (Section 2.2) that will finally be used to constrain the lesion segmentation at each time point (Section 2.3). The main assumption of this work is that all time points can be diffeomorphically mapped to a subject-specific mean appearance.

To assess the relevance of the proposed technique for the study of WMH progression, a longitudinal lesion simulator was developed (Section 3.1) so as to test the method with various longitudinal patterns and lesion loads. A surrogate clinical validation was performed using data from the Alzheimer's disease Neuroimaging Initiative (ADNI) to test whether documented cross-sectional as well as longitudinal findings reported in the literature could be reproduced.

2. Method

In the following the subscript τ denotes a specific time point and GW the groupwise average appearance model. Prior to the construction of the average, an expectation maximisation (EM) algorithm with outlier detection and bias field correction is performed on each individual time point. The intensities $\mathbf{Y}_\tau$ are the resulting log-transformed, normalised and bias field corrected intensities of the skull-stripped images. With N the number of voxels and D the number of modalities, image intensities are vectorised into $Y^{(d)} = \{y_{d1}, \cdots, y_{dn}, \cdots y_{dN}\}$ with y_{dn} the intensity at voxel n of modality d, so that

$$\mathbf{Y} = \begin{pmatrix} Y^{(1)} \\ \vdots \\ Y^{(D)} \end{pmatrix}.$$

2.1. Longitudinal intra-subject average

In order to build the average appearance model, two main components linking the individual images to the average space are needed: a spatial transformation and an intensity transformation. An intensity matching between images is needed to account for changes in contrast, MR scanning variations and some artefacts. These transformations are obtained through an iterative process, proved to limit bias towards a specific time point. In order to avoid unrealistic spatial deformations, affine transformations roughly aligning the images are first applied before considering non-rigid transformations to obtain the final spatial transformations $T_{\tau \rightarrow \text{GW}}$. At each iteration, the intensities of the images spatially transformed to the GW space are mapped to the intensities of the current average image using a polynomial fit of degree 2 for each modality used. More formally, the intensity mapping and the resulting mapping coefficients $h_\tau^{(d)}$ for one modality d can be expressed as

$$\underset{h_\tau^{(d)}}{\text{argmin}} \parallel A\left(T_{\tau\rightarrow\text{GW}}(Y_\tau^{(d)})\right) \cdot h_\tau^{(d)} - Y_{\text{GW}}^{(d)} \parallel^2$$

where $A(T_{\tau\rightarrow\text{GW}}(Y_\tau^{(d)}))$ is the polynomial matrix transformation of $T_{\tau\rightarrow\text{GW}}(Y_\tau^{(d)})$ such that

$$A(Y) = \begin{pmatrix} 1 & y_1 & y_1^2 \\ \vdots & \vdots & \vdots \\ 1 & y_N & y_N^2 \end{pmatrix}.$$

The steps to create an average appearance model are:

Step 1 Register each of the individual time points to the current average image.
Step 2 Map the intensities of each resampled image to the current average image using a polynomial fit of degree 2.
Step 3 Average all resampled and intensity transformed images to create the new current average image.
Step 4 Go back to step 1.

With this set up the loop is performed five times: the first iteration consists in the estimation of a rigid transformation followed by two affine transformations before allowing for a non-rigid registration at the last two iterations.

2.2. Model selection

After creating the average appearance model, patient-specific tissue priors and brain mask are obtained using the GIF (Geodesic Information Flow) pipeline developed in Cardoso et al. (2015). In this method, label-fusion is used to generate subject specific tissue priors ($\mathbf{A}$) by propagating pre-segmented templates and fusing them locally according to Cardoso et al. (2015). Using the priors and brain mask as inputs, BaMoS (Sudre et al., 2015) is used to model the data according to a three-level Gaussian mixture. The first level segments inliers from outliers observations while the anatomical tissue information is introduced at the second level so that each of the inlier and outlier tissue classes is modelled by a Gaussian mixture at the third level of the model. The final distribution model is then expressed as

$$f(\mathbf{Y}|\,\Xi_{\mathbf{K}}) = \prod_{n=1}^{N} \sum_{l\in I,O} \sum_{j=1}^{J} \sum_{k=1}^{K_{l_j}+1} \pi_{nl_{j_k}} \mathcal{M}\left(\mathbf{y}_n|\theta_{l_{j_k}}\right)$$

where $\pi_{nl_{j_k}}$ are the spatially varying weights in the mixture obtained by multiplying the class mixing proportions, and the inlier and tissue at the previous levels, l refers to the segmentation between inliers (I) and outliers (O), j to the anatomical classes and k to the individual Gaussian components. The notation $\mathbf{K}$ is used to encompass the model complexity (number of components for each tissue class K_{l_j}), while $\boldsymbol{\Theta}$ gathers the model parameters of each individual component (mixture weight $w_{l_{j_k}}$, mean $\boldsymbol{\mu}_{l_{j_k}}$ and