HIPPOCAMPAL ATROPHY RATE USING AN EXPECTATION MAXIMIZATION CLASSIFIER WITH A DISEASE-SPECIFIC PRIOR

*Jyrki Lötjönen[a], Robin Wolz[b], Juha Koikkalainen[a], Valeria Manna[a], Christian Ledig[b], Lennart Thurfjell[c], Roger Lundqvist[c], Gunhild Waldemar[d], Hilkka Soininen[e], Daniel Rueckert[b], and for the Alzheimer's Disease Neuroimaging Initiative**

[a]Knowledge Intensive Services, VTT Technical Research Centre of Finland, Tampere, Finland
[b]Department of Computing, Imperial College London, London, UK
[c]Medical Diagnostics R&D, GE Healthcare, Uppsala, Sweden
[d]Department of Neurology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[e]Department of Neurology, Kuopio University Hospital, University of Eastern Finland, Kuopio, Finland

ABSTRACT

Hippocampal atrophy is a well-known characteristic associated with Alzheimer's disease. In this work, we propose a 4D Expectation Maximization framework for measuring the atrophy rate of the hippocampus from serial magnetic resonance images. One novelty of the framework is a disease-specific prior that regularizes the segmentation near the borders of the hippocampus. Regions where the hippocampus tends to get larger in the follow-up images than in the baseline are penalized. Using the ADNI cohort, we obtained classification accuracies of 83 % for healthy control and Alzheimer's disease patient groups and 60 % for stable and progressive MCI groups using the baseline and 12-month follow-up images.

Index Terms—expectation maximization classifier, segmentation, atrophy rate, Alzheimer's disease

1. INTRODUCTION

Early diagnostics of Alzheimer's disease (AD) plays a central role in developing solutions for this devastating disease. The current diagnostic guidelines emphasize the importance of biomarkers [1]. In addition to proven memory impairment, at least one supportive finding is needed either from magnetic resonance imaging (MRI), positron emission tomography (PET), cerebrospinal fluid (CSF) biomarkers or autosomal dominant mutation. For MRI, the recommended assessment criteria include volume loss of hippocampi, entorhinal cortex and/or amygdala. Several automated methods have been proposed for estimating atrophy from brain MRI images [2-4].

In [5], we presented a multi-atlas segmentation method for the hippocampus utilizing the expectation maximization (EM) classifier as a post-processing step. The atrophy rate of the hippocampus was measured in our previous work based on simultaneous 4D segmentation of all sequential [4]. In [4], a spatial prior was defined from multi-atlas registrations and the segmentation of all time-points was performed simultaneously using graph-cuts. In this work, we extend our approaches reported in [4] and [5] by proposing a 4D EM framework with a dynamic spatially varying Markov-Random Field (MRF) prior.

Although hippocampal neurogenesis is currently under active investigations [6], the size of the hippocampus is known to decrease even in healthy aging as well as in AD. Typical annualized atrophy rates reported for the hippocampus are 4.7 % for AD patients and 1.4 % for healthy age-matched controls [7]. Therefore, we propose an MRF term that penalizes segmentation label configurations against this constraint, i.e., situations where the size of the hippocampus tends to grow in follow-up images. This constraint is designed to allow a more robust segmentation with respect to image noise and interpolation artefacts, and is based on the assumption that in the context of AD, no increase of hippocampal volume over time can be expected.

2. METHODS

2.1. Intensity normalization and alignment of baseline and follow-up images

Intensity normalization is performed for skull-stripped images. The brain tissue was extracted using an in-house skull-stripping tool (non-published).

White-matter (WM), gray-matter (GM) and CSF were segmented from skull-stripped images using the EM classification method performing simultaneously intensity inhomogeneity correction [8]. The mean intensities of WM, GM and CSF were computed both from the baseline and follow-up images and the intensities of the follow-up images were mapped to the intensities of baseline images using a semi-linear transformation.

The baseline and follow-up images were then aligned using first a 9-parameter affine transformation and then a non-rigid transformation [5]. More detailed intensity normalization was

*This work is partially funded under the 7th Framework Programme by the European Commission http://cordis.europa.eu/ist/) and Tekes -Finnish Funding Agency for Technology and Innovation (www.tekes.fi).
Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

performed using the method proposed in [9]. As the images were non-rigidly registered, the kernel size of the median filter was only 2 instead of 5 as used in [9]. The correction fields computed using [9] were finally transformed to the follow-up images using the spatial transformations defined earlier and the intensities of follow-up images were corrected to be consistent with the baseline images.

2.2. Expectation maximization classification

The labelings $\{f^T,..,f^{T+N}\}$ of the baseline and follow-up images $\{I^T,..,I^{T+N}\}$ where N is the number of follow-up images, minimize the energy function:

$$\{f^T,..,f^{T+N}\} = arg\min_{\{f^T,..,f^{T+N}\}}(\sum_t E(f^t)) + w\gamma E_{MRF}(f^T,..,f^{T+N}) \quad (1)$$

where

$$E(f^t) = E_{intensity}(f^t) + \alpha E_{atlas}(f^t) + \beta E_{MRF}(f^t). \quad (2)$$

The parameter $w=w(f^T,..,f^{T+N})$ is a weight factor defined as a dynamic spatially changing 4D MRF prior (Section 2.5). $E_{intensity}$ measures the likelihood (Section 2.3). E_{atlas} is a spatial prior (Section 2.4) and EMRF is a 3D MRF prior (Section 2.5). The parameters α, β and γ are user defined weight factors.

In practice, the energy for each label f^t is computed separately for each time-point *t* using Eq. 2, and thereafter the label configuration $f^T,..,f^{T+N}$ minimizing the energy in Eq. 1 is defined. The initialization of the label configurations is obtained from multi-atlas segmentation, i.e., each voxel has the label having the highest likelihood in the probabilistic atlases (Section 2.4).

The user defined weight factors (α, β and γ) used in this work were 1.0, 0.025 and 0.1, respectively. We did not perform a thorough sensitivity analysis for the weighting values and the combination used may not be optimal. However, our preliminary analysis showed that the approach is not very sensitive to the values chosen. We observed in previous work that a spatial prior from multi-atlas segmentation plays an important role in measuring atrophy. The corresponding finding was made also in this work reflected by the relatively strong weighting of the probabilistic atlas.

2.3. Intensity term

The intensity of each structure *k* (WM, GM, CSF and hippocampus in this work) was assumed to have a Gaussian density function, described by mean μ and standard deviation σ:

$$E_{intensity} = -\sum_{p\in I} \ln p(I_p|\, f_p = k),$$

where

$$p(I_p|f_p = k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(I_p-\mu_k)^2}{2\sigma_k^2}}.$$

The parameters μ and σ were estimated from the target volume by weighting each voxel with the likelihood that it belongs to the class *k*.

2.4. Spatial prior for baseline and follow-up images

The spatial prior for the baseline image was computed using a multi-atlas approach, as described in [5]. A separate template space was used for atlas-selection. The non-rigid transformation ($T_{BA,i}$) from the baseline image to the selected atlas A_i was initialized by con-catenating the transformations from the baseline to the template (T_{BT}) and from the pre-computed transformation from the template to each atlas ($T_{TA,i}$) as presented in Fig. 1. The label map f^i of the atlas A_i is propagated to the baseline coordinates by the inverse transformation of $T_{BA,i}$. The probability that a certain baseline voxel *p* belongs to the class *k* is

$$p_k(f_p) = \frac{1}{N}\sum_{i=1,..,N}\begin{cases}1, & if\ f_p^i = k\\ 0, & if\ f_p^i \neq k\end{cases}.$$

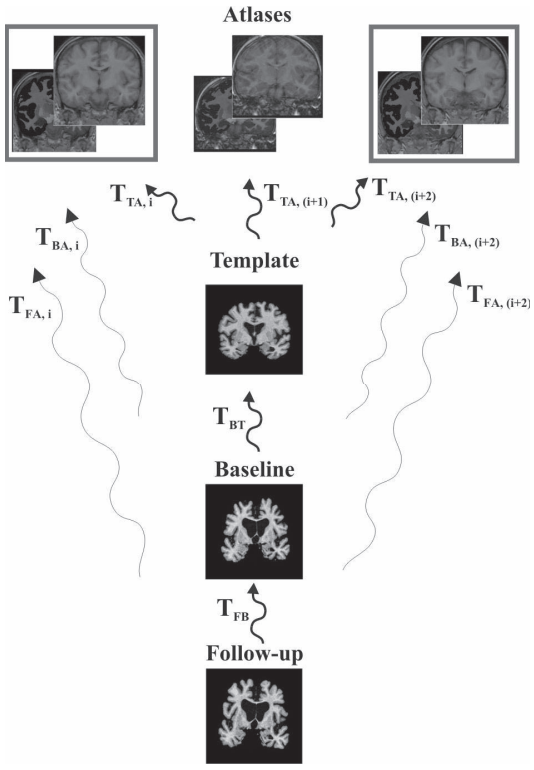


Fig. 1. Transformations defined for computing probabilistic atlases of different tissues for baseline and follow-up images.

The probabilistic atlases for the follow-up images are computed as described above except that the transformation $T_{FA,i}$ is initialized by con-catenating the transformations T_{FB} and $T_{BA,i}$. Therefore, only a few iterations are needed for computing the final transformation. As registrations with real data contain always errors, the need for only a few iterations leads to smaller discrepancies between the probabilistic atlases of the baseline and follow-up images than constructing the probabilistic atlases independently. The spatial prior is computed as follows:

$$E_{atlas} = -\sum_{p\in I} \ln p(f_p = k).$$

2.5. Regularization prior for 3D and 4D data

The MRF-based regularization prior was defined for keeping structures smooth. The formulation presented in [10] was used:

$$E_{MRF} = \sum_{p \in I} \sum_{q \in N_p} \frac{a_{pq}}{d(p,q)},$$

where N_p is the 6-neighborhood around voxel p, $d(p,q)$ is the distance between centers of voxels p and q, and

$$a_{pq} = \begin{cases} -2 & if\ f_p = f_q \\ 1 & otherwise \end{cases}.$$



Fig. 2. Spatial weighting factors for regularizing segmentation of the hippocampus: $w_{outside}$ = white, w_{border} = black and $w_{otherwise}$ = gray.

The same prior was used for regularizing segmentations between images from different time points. However, this term was weighted with a factor varying spatially and dynamically during the EM iterations.

A hard segmentation of the hippocampus was defined during each iteration of the EM algorithm. Voxels outside, on the border and inside of the hippocampus were detected using morphological erosion and dilation operations with 6-neighborhood in 3D (Fig. 2). The weighting factors for a pair of images $\{I^{T+i}, I^{T+i+1}\}$ are defined as follows:

$$w(f^{T+i}, f^{T+i+1}) = \begin{cases} \varphi & if\ outside\ HC\ at\ f^{T+i} \\ 0 & if\ border\ of\ HC\ at\ f^{T+i} \\ 1 & otherwise \end{cases}$$

where $\varphi > 1$ is a user-defined weighting factor and HC stands for the hippocampus. This weighting factor penalizes different labels in a layer outside the hippocampus (value $\varphi = 10$ used in this work) and eliminates the MRF constraint on the border where the atrophy is expected to occur, and use a constant otherwise (defined by γ in Eq. 2).

If we assume that atrophy was distributed uniformly on the hippocampal surface, the surface would shrink about 0.1 mm for AD patients and 0.03 mm for healthy controls based on the typical size of the hippocampus and its atrophy rates. The magnitude of annual atrophy is clearly smaller than a typical voxel size in MRI images. Atrophy can be therefore expected to appear mostly in a one-voxel layer, i.e., on the voxels in the border of the hippocampus in image I^{T+i} when the image pair $\{I^{T+i}, I^{T+i+1}\}$ is compared. If image resolution was clearly higher or atrophy of several years was measured, the border area should be made wider for corresponding the range of shrinking in millimeters.

Fig. 3 summarizes the algorithm proposed.

3. DATA AND RESULTS

3.1. Subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). ADNI contains data from about 800 persons: 200 AD patients, 400 MCI (mild-cognitive impairment) and 200 cognitively normal (CN). In this work, T1-weighted 1.5T MRI images of the baseline and 12-month follow-up measurements were analysed from 476 subjects. The MCI subjects were divided into two groups: stable MCIs (SMCI) did not convert during the ADNI follow-up time while progressive MCIs (PMCI) converted during the follow-up. An overview on the demographics of these subjects is given in Table 1.

- Skull-strip MRI images
- Remove intensity inhomogeneities
- Register images affinely and non-rigidly
- Normalize intensities
- Generate probabilistic atlases
- Perform EM classification by iterating:
  - M-phase:
    - Estimate the mean and variance of each tissue class by weighting each voxel with the likelihood of the tissue.
  - E-phase:
    - Compute energies for each tissue class of each time point (Eq. 2)
    - Compute spatially varying MRF weights using labels from the previous iteration
    - Compute labels for all time points simultaneously by minimizing energy (Eq. 1)
    - Compute likelihoods for each tissue class

Fig. 3. Algorithm for estimating atrophy-rate of the hippocampus.

Table 1. Number (female fraction), average age and average result of the Mini Mental State Examination (MMSE) for the study population.

	N (female)	Age	MMSE
CN	152 (75)	76.2 ± 4.9	29.2 ± 0.9
SMCI	109 (37)	75.2 ± 6.8	27.4 ± 1.7
PMCI	112 (41)	74.7 ± 6.8	26.7 ± 1.7
AD	103 (49)	75.5 ± 7.2	23.2 ± 2.0

3.2. Evaluation method

In this work, the classification accuracy for separating healthy controls and AD cases as well as the sample size needed in a clinical trial to detect significant atrophy rate changes were studied. For classification, we used the simplest possible linear classifier: $C = \beta_0 + \beta_1 F_1$,where β_i is the weight of the feature (independent variable) F_1 (atrophy rate) and C represents a dependent variable used to predict the class, e.g. class 'AD' if $C \geq 0$ and class 'healthy' if $C < 0$. We defined the weights β_i using

extensive search. Separate training and test sets were chosen to avoid over-learning. We used 2/3 of cases in the training set to build the classifier and 1/3 of cases to test the classification rate. We repeated the selection 1000 times and computed different statistical measures (mean, standard deviation, 95%-confidence interval) for the results.

The statistically significant differences between groups were studied using the t-test. The difference was considered statistically significant if $p < 0.05$.

The number of required study subjects is a critical factor in clinical trials [11]. Here, we tested the number of subjects needed in a hypothetical two-arm study to detect a 25% change in atrophy rate with a power of 80% and 5% significance using the proposed method.

3.3. Results

The classification results are presented in Table 2. For comparison, we show the results for our previous method [4] using the subjects summarised in Table 1. The atrophy rates obtained were 6.6±2.2% for AD patients, 5.7±1.4% for PMCI subjects, 4.5±1.8% for SMCI subjects, and 3.6±1.4 % for healthy controls.

Table 2. Classification accuracies and 95%-confidence intervals for CN-AD and SMCI-PMCI groups. Statistical significant differences in the test set are shown by '*'.

	Training set [%]	Test set [%]	CI-95 [%]
CN-AD	84.3±1.6	83.4±3.4	76-90
SMCI-PMCI	64.4±2.1	59.7±4.7	49-68
CN-AD[4]	82.0±1.6	79.8±3.5*	73-87
SMCI-PMCI[4]	65.1±1.9	59.2±4.4*	51-67

As proposed in [11], it is essential to correct atrophy rates for healthy ageing in order to produce a reliable estimate for the sample sizes needed in clinical trials. Correcting for healthy ageing in the hypothetical study described above, leads to sample sizes of N=210 per arm for MCI subjects and N=68 for AD subjects. This is compared to N=254 and N=55 respectively with our previous method [4].

4. DISCUSSION

We proposed a method for segmenting the hippocampus from sequential MRI images. The method was implemented in the Expectation Maximization framework although an implementation in other frameworks is also straightforward. A disease-specific spatially and dynamically varying Markov Random Field prior was proposed. First, it is known that the size of the hippocampus should decrease even due to normal aging. In our approach, an increase in the volume was penalized. Second, the regularization between different time points should be suppressed in regions where atrophy appears in images.

Our preliminary results suggest good performance of the approach. A thorough analysis with larger data sets and different cohorts is needed to verify the results. We compared our method also to our previous work [4]. Detailed comparison is, however, difficult as the proposed and previous methods are using different software components and slightly different methods, e.g., we used graph-cuts in our previous work.

In this work, we applied the varying MRF prior only to the hippocampus but it could be used also for the gray-matter class. This extension would potentially improve the consistency of the atrophy rate measurement when computed for the whole brain.

5. REFERENCES

[1] B. Dubois, H. Feldman, C. Jacova, et al. "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria," *Lancet Neurol*, vol. 8, pp. 734-746, 2007.

[2] J. Morra, Z. Tu, L. Apostolova, et al. "Automated mapping of hippocampal atrophy in 1-year repeat MRI data form 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls," *NeuroImage*, vol. 45, pp.3-15, 2009.

[3] K. Leung, M. Clarkson, J. Bartlett et al., "Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection," *NeuroImage*, vol. 50: pp.516-523, 2010.

[4] R. Wolz, R. Heckemann, P. Aljabar, et al., "Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI," *NeuroImage*, vol. 52, pp.109-118, 2010.

[5] J. Lötjönen, R. Wolz, J. Koikkalainen, et al., "Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease," *NeuroImage*, vol. 56, pp. 185-196, 2011.

[6] J. Snyder and H. Cameron. "Could adult hippocampal neurogenesis be relevant for human behavior," Behavioural Brain Research, in press, 2011.

[7] J. Barnes, J. Bartlett, L. van de Pol et al., "A meta-analysis of hippocampal atrophy rates in Alzheimer's disease," *NeuroImage*, vol. 30, pp. 1711-1723, 2009.

[8] K. van Leemput, F. Maes, D. Vandermeulen, P. Suetens. "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag*., vol. 18, pp. 897-908, 1999.

[9] E. Lewis and N. Fox. "Correction of differential intensity inhomogeneity in longitudinal MR images," *NeuroImage*, vol. 23, pp. 75-83, 2004.

[10] J. Tohka, A. Zijdenbos, and A. Evans. "Fast and robust parameter estimation for statistical partial volume models in brain MRI," *NeuroImage*, vol. 23, pp 84-97, 2004.

[11] N. C. Fox, G. R. Ridgway, J. M. Schott. "Algorithms, atrophy and Alzheimer's disease: Cautionary tales for clinical trials" NeuroImage, vol 57, pp 15-18, 2011.