University College London

Information Processing in Medical Imaging - Coursework Report

---

Author:
Răzvan Valentin Marinescu
razvan.marinescu.14@ucl.ac.uk

EPSRC Centre for Doctoral Training in Medical Imaging
University College London

April 23, 2015

Task I - Full brain segmentation

Proposed pipeline

Segmentation propagation (`p1reg.py`) has been performed on the extra 5 AD and 5 controls, both baseline and followup images, using the tutorial instructions from the CMIC TIG website. This meant that we needed to propagate segmentations for 20 different images, and since we had 10 source templates available, this meant 200 propagations. A discussion about how I optimised this process and chose the parameters for the functions is given at the end of this section. Figure 1 shows the difference between the reference image and the floating (template) image, and we can clearly see that they are not aligned at all. Therefore, we first perform an affine transformation (`reg_aladin`) and plot the difference images in figure 2. After this linear tranformation, the brains align better, with the overall boundaries fitting well. Afterwards, we perform a non-linear registration using a free-form deformation (`reg_f3d`) and using the affine transformation as a starting position (see figure 3). Here, there is even more improvement in the registration of certain microstructures such as the ventricles and lobes, as well as in the overall brain boundary. We then use this tranformation to propagate the segmentation of the template image into the space of the reference image. This process is repeated for all 10 template images and then a simple majority voting is used to find the final segmentation. Figure 4 shows that the propagated segmentation matces well with the underlying T1 scan. We then run this pipeline for all the AD and control subjects, both baseline and followup scans.

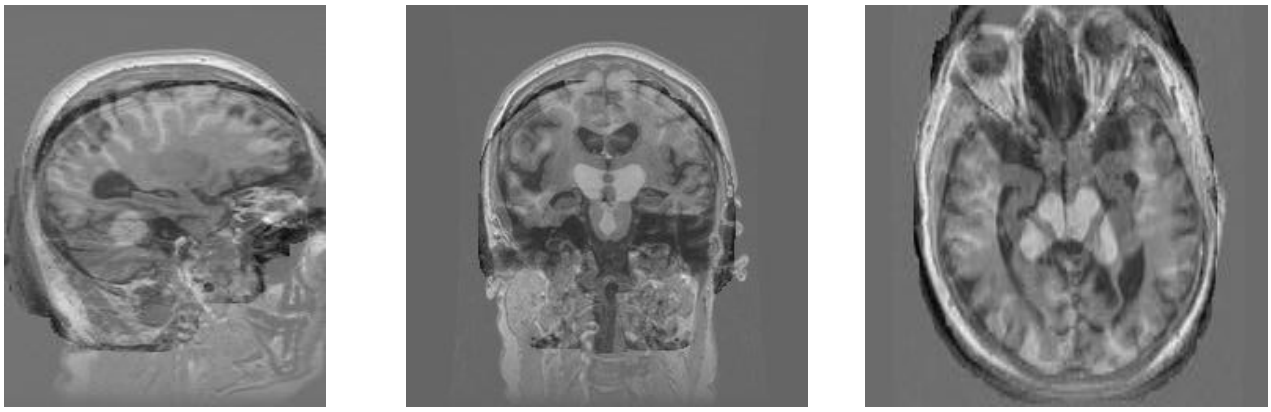

Figure 1: Difference image after no transformation

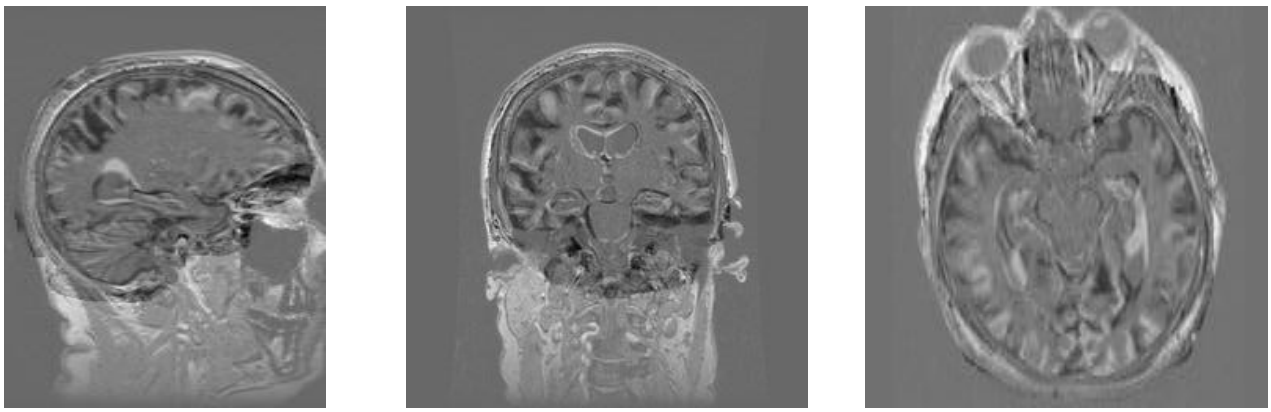

Figure 2: Difference image after affine transformation

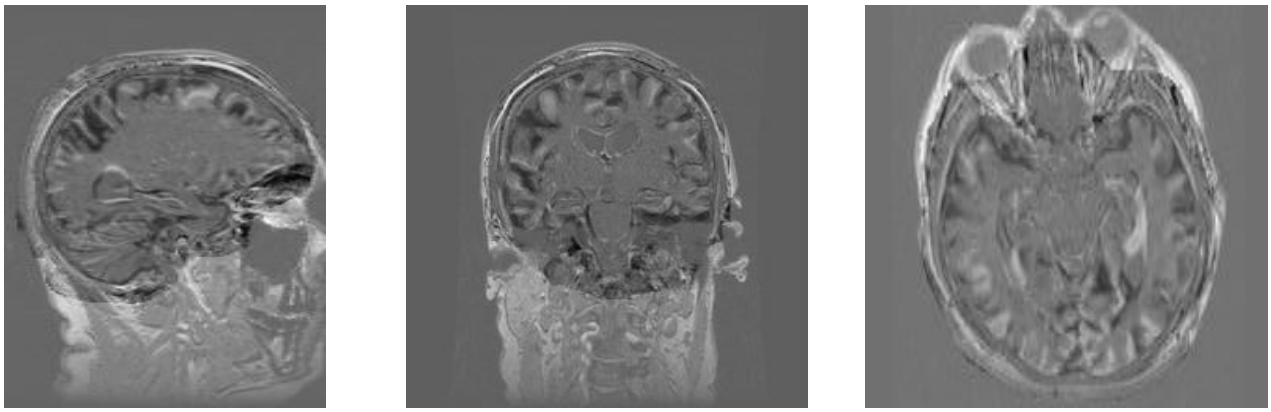

Figure 3: Difference image after affine and nonlinear transformation using the free-form deformation algorithm

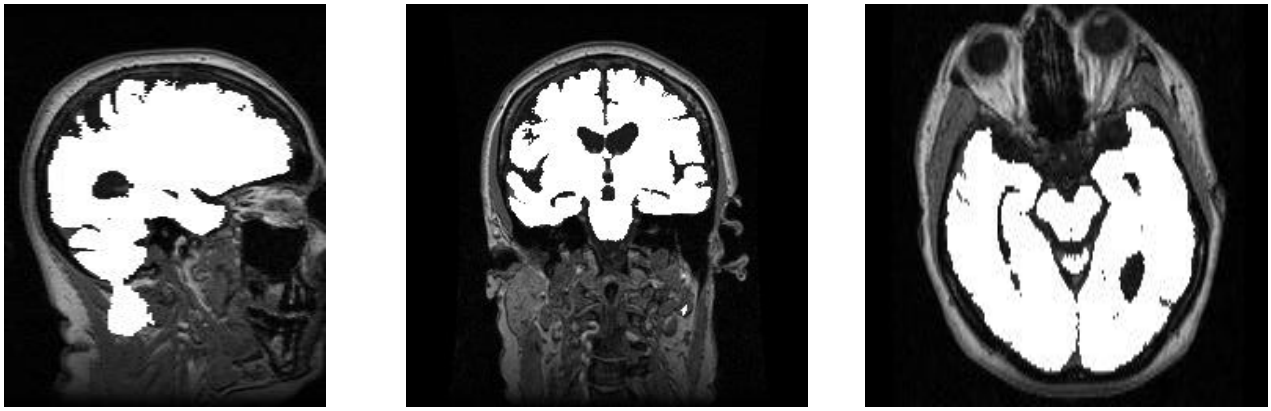

Figure 4: T1 image and corresponding segmentation after label fusion with a simple majority voting

Parameters

As each propagation took around 1.5 minutes on an Intel(R) Xeon(R) CPU @ 3.60GHz, a careful balance between computation time and accuracy had to be obtained. In order to find out which set of parameters to choose, we computed the dice score on the template database using a leave one out approach. More precisely, for each template image, we propagated the segmentations from the 9 other templates, performed label fusion and computed the dice score. This pipeline was then repeated for the other template images and for other sets of parameters. For the nonlinear deformation `reg_f3d` function, we tried 4 values for the number of levels: [2,3,4,5] and 4 values for the number of iterations: [50, 150, 200, 300]. In figure 5 we plotted the dice score for all these combinations, while in figure 6 we calculated wall-clock time taken to perform the computation on the PC. We notice that the dice score and computation times are not much affected by the number of levels used (notice the small range of the Y-axis in the dice score plot). However, as the number of iterations increases, we see slight improvements in the dice scores but also increases in the computation time. Finally, we decided to use 4 levels and 200 iterations, in order to keep a balance between computation time and dice scores.

Task II - Atrophy measurement

In order to compute the BSI, we first co-registered the baseline and followup images to midspace and also propagated the segmentations to midspace. Afterwards, the BSI has been computed using the method described in Leung et al. [1]. First, the baseline and followup images have been aligned using the 9DOF affine registration. Then, the union and intersection regions of the images is computed. The union is dilated once, while the intersection is eroded once using a structure that looks like a 3x3x3 sphere. The brain boundary shift region is then given by the XOR of the dilated union and eroded intersection. Figure 7 shows the boundary region for the first AD patient. The intensity of both images is then normalised by

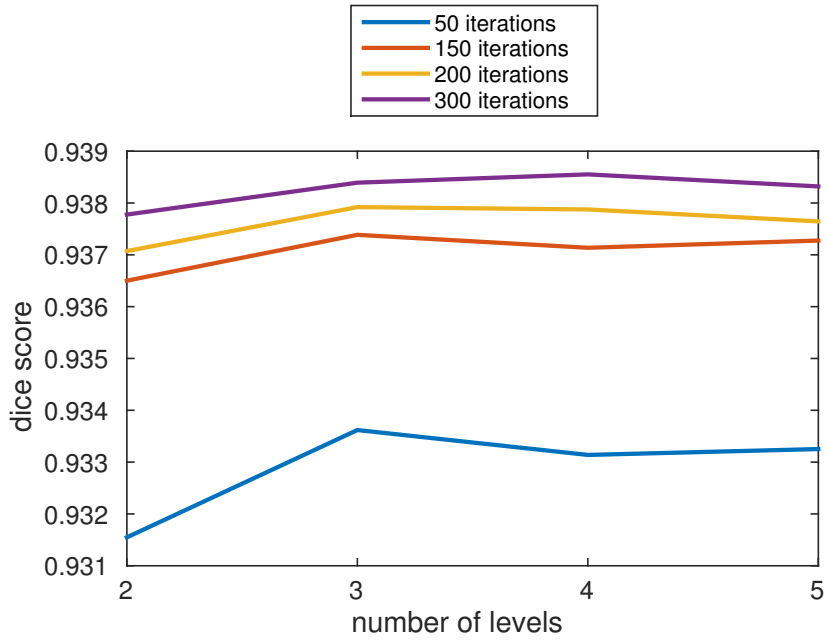

Figure 5: Dice scores for different number of levels and iterations.

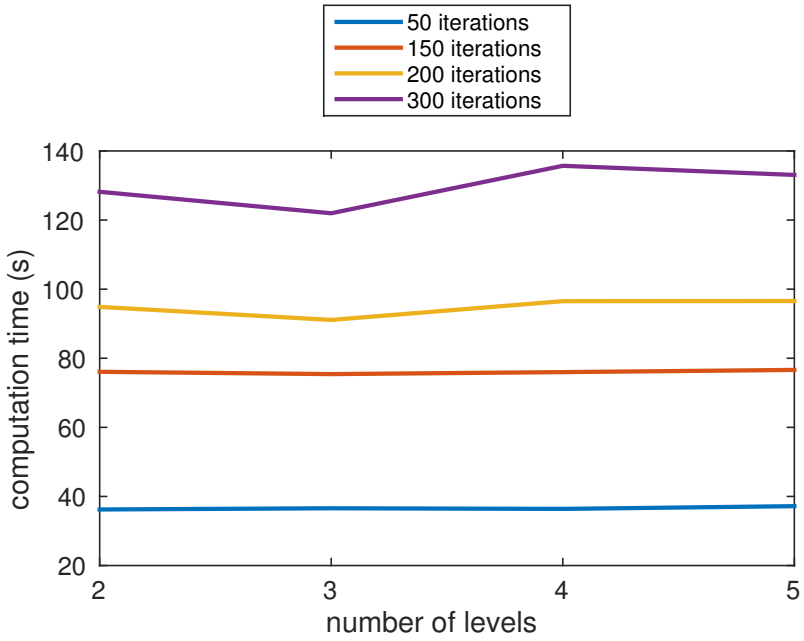

Figure 6: Computation time for different number of levels and iterations.

dividing by the mean intensity inside the intersect region. Finally, the BSI is computed using a manually chosen intensity window of [0.45,0.65] recommended by Freeborough and Fox, 1997. [2]. Apart from the BSI, we have also computed the segmentation volume difference.

Figures 8 and 9 show the BSI and segmentation volume difference respectively, for all age-matched AD and control subject pairs. We can clearly notice that the BSI is much better at separating the age-matched AD and controls, which is an indication that BSI can better measure atrophy in the brain. Only for AD-control pair 15 do we get that the BSI is lower in the AD subject than in the age-matched control, suggesting this pair might be an outlier. On the other hand, the segmentation volume difference cannot easily discriminate between AD and controls (figure 9).

Figure 7: BSI boundary region for an AD patient

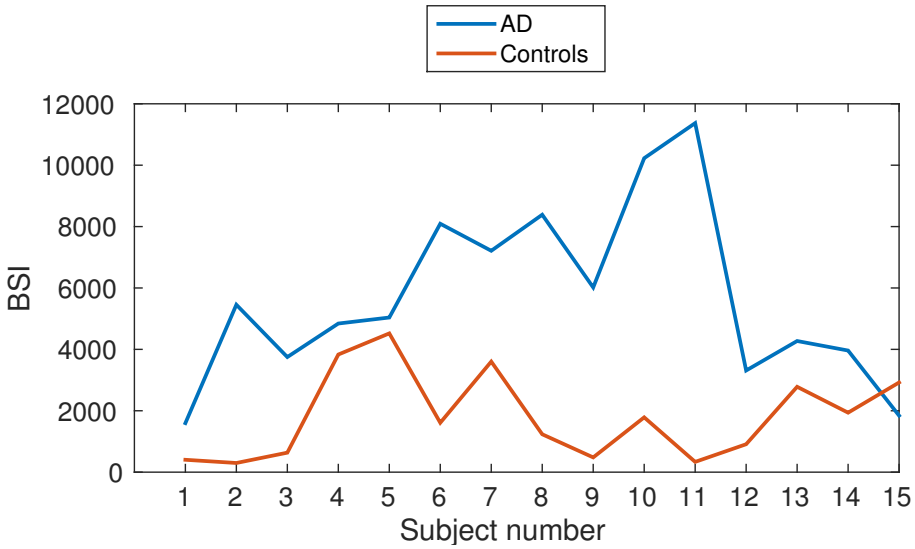

Figure 8: BSI measurements for all AD and control subjects. Each AD patient is paired with an age-matched control, typically a spouse or carer. [3] There is visibly more atrophy (as measured by BSI) for AD subjects compared to control subjects. The only exception is subject 15, which might be an outlier.

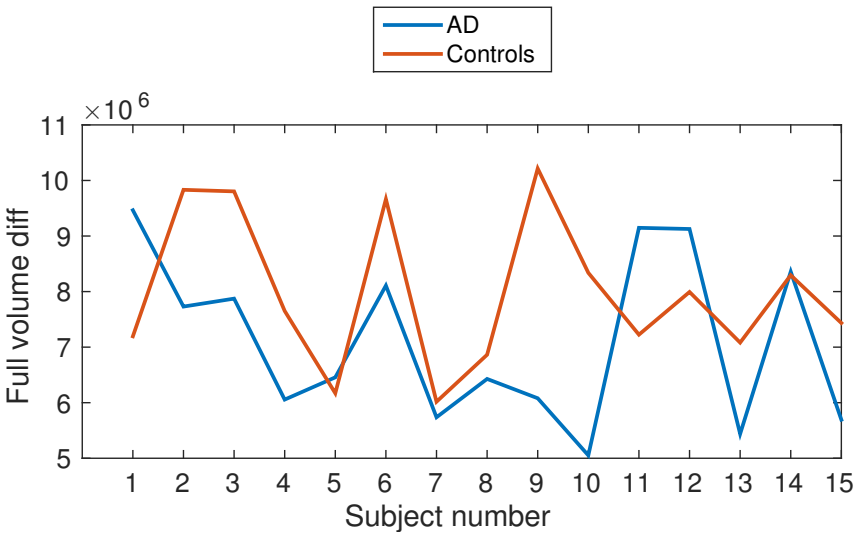

Figure 9: Full volume difference for all AD and control subjects. There is considerable more overlap between AD and control scores compared with the BSI.

Task III - Statistical Analysis

T-tests

A two-sample t-test has been performed for AD and control groups using the BSI, in order to see if there is a statistically significant difference in atrophy between AD and control groups (`p3.m`). Since each AD patient had an age-matched control (normally a spouse or carer) [3], this allows us to also perform a paired t-test, checking for a pairwise difference in atrophy. A similar analysis has been made using the full-volume segmentation difference. Results are presented in the following table:

Metric	paired-sample t-test		two-sample t-test	
	H_0	p-value	H_0	p-value
BSI	rejected	5.23e-04	rejected	6.45e-05
Seg. volume diff.	not rejected	0.0844	not rejected	0.1062

For the BSI metric, both t-tests rejected the null hypothesis H_0, while for the full-volume metric, H_0 could not be rejected. We therefore conclude that there is a significant difference of atrophy between AD and control groups, as measured by the BSI. On the other hand, using the full-volume segmentation difference, we conclude that there is no difference between the two groups. This suggests that the BSI is better at discriminating between AD and control subjects than the full-volume segmentation difference.

Sample size analysis

MATLAB function `samplesizepwr` has been used to calculate the sample size required to detect a 25% atrophy rate with an 80% power (file `p3.m`). For the BSI, minimum sample sizes of 34 and 77 are required to detect an atrophy reduction relative to AD and normal ageing respectively. Similarly, for the full segmentation volume difference, minimum sample sizes of 8 and 6 are required to detect an atrophy reduction relative to AD and normal ageing, respectively. The reason for the lower sample size required using the full volume atrophy is because the standard deviation of the groups was much lower relative to the mean.

Future improvements

In order to improve the statistical power of this technique, there are several things one could do:

- **Data acquisition**: Obtain higher-resolution MRI data using 7T (or higher) scanners. In our MIRIAD dataset [3], a 1.5T scanner was used.

- **Registration**: A closer look at the registration suggests that there is a lot of room for improvement. We can improve it by calculating the transformation using more iterations at the free-form deformation step. However, other more recent methods such as those based on progressive principal component registration by Melbourne et al. [4] might obtain better results.

- **Segmentation**: This can be improved by using more templates, or by implementing smarter fusion methods such as weighted majority voting or probabilistic methods such as Non-local-STAPLE or STEPS. One other limitation of this method is that the templates might produce similar labeling errors, but this can be overcome with the method proposed by Wang et al, 2013 [5]. One could also combine the segmentation propagation method with other methods based on level sets [6], fuzzy c-means [7], Gaussian mixture models using EM, Markov random fields, Self organising maps [8] or Learning vector quantization [9].

- **Atrophy estimation**: Perform a bias field correction before BSI computation and do a more robust intensity normalisation and automatic parameter selection based on the intrinsic tissue contrast of the MR images, as described in Leung et al, 2010 [1].

Demographic information could also be included in order to improve the statistical analysis. For example, one could compensate for age in the atrophy measurement using the BSI by fitting a linear or polynomial model over the age & BSI data. The two-sample t-test should show improvements after this change, while the paired t-test shouldn't be affected too much, as the AD and control pairs are already age-matched. If gender information is also available, one could use that to perform a t-test on male and female groups in order to see if the atrophy rate is statistically different. Other demographic information could be used in a similar manner. If multiple types of demographic data is used, one could perform *Canonical Correlation Analysis* in order to find out how much each of these demographic factors are correlated with increased atrophy rate and full volume difference.

Bibliography

[1] Kelvin K Leung, Matthew J Clarkson, Jonathan W Bartlett, Shona Clegg, Clifford R Jack, Michael W Weiner, Nick C Fox, Sébastien Ourselin, Alzheimer's Disease Neuroimaging Initiative, et al. Robust atrophy rate measurement in alzheimer's disease using multi-site serial mri: tissue-specific intensity normalization and parameter selection. *Neuroimage*, 50(2):516–523, 2010.

[2] Peter A Freeborough and Nick C Fox. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat mri. *Medical Imaging, IEEE Transactions on*, 16(5):623–629, 1997.

[3] Ian B Malone, David Cash, Gerard R Ridgway, David G MacManus, Sebastien Ourselin, Nick C Fox, and Jonathan M Schott. Miriadpublic release of a multiple time point alzheimer's mr imaging dataset. *NeuroImage*, 70:33–36, 2013.

[4] A Melbourne, D Atkinson, MJ White, D Collins, M Leach, and D Hawkes. Registration of dynamic contrast-enhanced mri using a progressive principal component registration (ppcr). *Physics in medicine and biology*, 52(17):5147, 2007.

[5] Hongzhi Wang, Jung Wook Suh, Sandhitsu R Das, John B Pluta, Caryne Craige, and Paul A Yushkevich. Multi-atlas segmentation with joint label fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):611–623, 2013.

[6] Daniel Cremers, Mikael Rousson, and Rachid Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International journal of computer vision*, 72(2):195–215, 2007.

[7] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2):191–203, 1984.

[8] Suchendra M Bhandarkar, Jean Koh, and Minsoo Suk. Multiscale image segmentation using a hierarchical self-organizing map. *Neurocomputing*, 14(3):241–272, 1997.

[9] Nicolaos B Karayiannis and Pin-I Pai. Segmentation of magnetic resonance images using fuzzy algorithms for learning vector quantization. *Medical Imaging, IEEE Transactions on*, 18(2):172–180, 1999.