

Using Manifold Learning for Atlas Selection in Multi-Atlas Segmentation

Albert K. Hoang Duc^{1*}, Marc Modat¹, Kelvin K. Leung², M. Jorge Cardoso¹, Josephine Barnes², Timor Kadir³, Sébastien Ourselin¹, for the Alzheimer's Disease Neuroimaging Initiative

1 Centre for Medical Image Computing, University College London, London, United Kingdom, **2** Dementia Research Centre, University College London, London, United Kingdom, **3** Mirada Medical, Oxford, United Kingdom

Abstract

Multi-atlas segmentation has been widely used to segment various anatomical structures. The success of this technique partly relies on the selection of atlases that are best mapped to a new target image after registration. Recently, manifold learning has been proposed as a method for atlas selection. Each manifold learning technique seeks to optimize a unique objective function. Therefore, different techniques produce different embeddings even when applied to the same data set. Previous studies used a single technique in their method and gave no reason for the choice of the manifold learning technique employed nor the theoretical grounds for the choice of the manifold parameters. In this study, we compare side-by-side the results given by 3 manifold learning techniques (Isomap, Laplacian Eigenmaps and Locally Linear Embedding) on the same data set. We assess the ability of those 3 different techniques to select the best atlases to combine in the framework of multi-atlas segmentation. First, a leave-one-out experiment is used to optimize our method on a set of 110 manually segmented atlases of hippocampi and find the manifold learning technique and associated manifold parameters that give the best segmentation accuracy. Then, the optimal parameters are used to automatically segment 30 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI). For our dataset, the selection of atlases with Locally Linear Embedding gives the best results. Our findings show that selection of atlases with manifold learning leads to segmentation accuracy close to or significantly higher than the state-of-the-art method and that accuracy can be increased by fine tuning the manifold learning process.

Citation: Hoang Duc AK, Modat M, Leung KK, Cardoso MJ, Barnes J, et al. (2013) Using Manifold Learning for Atlas Selection in Multi-Atlas Segmentation. *PLoS ONE* 8(8): e70059. doi:10.1371/journal.pone.0070059

Editor: Karl Herholz, University of Manchester, United Kingdom

Received: December 18, 2012; **Accepted:** June 15, 2013; **Published:** August 2, 2013

Copyright: © 2013 Hoang Duc et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was undertaken at UCL/UCLH which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research Trust Co-ordinating centre and has also received equipment funded by the Alzheimers Research Trust. AKHD is supported by an IMPACT studentship with a part of the funding from MIRADA MEDICAL. MJC is supported by a scholarship from the Fundação para a Ciência e a Tecnologia, Portugal (Scholarship number SFRH/BD/43894/2008). KKL acknowledges support from the MRC, ARUK and the NIHR. JB is supported by an Alzheimer's Research Trust (ART, UK) Research Fellowship partly supported by the Kirby Laing Foundation. NCF is funded by the Medical Research Council (UK). MM and SO are supported by CBRC number 168. SO is also supported by EPSRC/CRUK Comprehensive Cancer Imaging Centre of UCL and KCL (grant number C1519AO) and EPSRC EP/H046410/1. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorph Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Mesoscale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pzer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. The authors would like to thank the ADNI study subjects and investigators for their participation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: TK, one of the authors in this paper, is employed by a commercial company (Mirada Medical). AKHD, first author in this paper, is partly funded by a commercial company (Mirada Medical). There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: albert.hoang.duc.10@ucl.ac.uk

Introduction

Multi-atlas segmentation is an automated segmentation method that shows good robustness and accuracy in segmenting various anatomical structures [1–4]. In this framework, a segmentation of a target image is obtained through the propagation and fusion of multiple atlas images by mean of registration. As demonstrated by [5], propagation of atlases similar to the target image significantly improves the quality of the segmentation. As a result, it is crucial to

develop strategies for selecting the best atlases in the framework of multi-atlas segmentation in order to achieve optimal accuracy.

Several approaches for atlas selection have been proposed over the past few years [2,3,5–11]. For instance, in the multiple-atlas propagation and segmentation method (MAPS) [7], the most similar atlases are selected based on intensity similarity after rigid registration. In [12], manifold learning is used to select atlases which are located in the neighbourhood of the target on the manifold. This novel approach gives promising results. However,

Table 1. Subject demographics in control and probable AD subjects used for parameter optimization. Mean (SD) unless specified otherwise.

| | Control (n = 19) | AD (n = 36) |
|-----------------------|------------------|-------------|
| Age, years | 68.7 (7.0) | 69.6 (7.3) |
| Gender male (%) | 9 (47%) | 14(39%) |
| MMSE at baseline, /30 | 29.5 (0.7) | 19.4 (4.1) |

doi:10.1371/journal.pone.0070059.t001

some aspects in that study have not been investigated thoroughly such as the type of manifold learning or optimal manifold parameters. Therefore, our paper investigates further the usage of manifold learning for atlas selection in the framework of multi-atlas segmentation.

Manifold learning has been successfully used in multiple medical imaging applications including segmentation [13], registration [14,15], classification [16] and statistical population analysis [17,18]. The most popular manifold learning techniques used in medical imaging are Isomap [19], Locally Linear Embedding (LLE) [20] and Laplacian Eigenmaps (LEM) [21]. For instance, Laplacian Eigenmaps is used by [22] to reduce the computational complexity in multi-modal registration and by [23] for biomarker discovery in MR imaging. Isomap is used by [14] to tackle the problem of performing large deformation registration and by [24] to parametrize cardiac MRI images. [25] investigates the detection of seizures in EEG signals with Locally Linear Embedding.

Each manifold learning technique attempts to preserve a different geometrical property of the underlying manifold. Isomap is a global approach that attempts to preserve pairwise metrics. In contrast, LLE and LEM aim to preserve the local geometry of the data. Since each manifold learning technique is associated with a different objective function, it is legitimate to assume that, for a given data set, the associated embeddings are also different. Previous studies [14,22–25] only included a single technique in their design and manifold parameters appeared to have been chosen arbitrarily. To our knowledge, no studies have investigated (i) the effect of different manifold learning algorithms and (ii) the computation of optimal manifold parameters for a given application. This paper addresses these two points. In the context of atlas selection for multi-atlas segmentation, we investigate the appropriate choice of manifold learning technique and manifold parameters that result in optimal atlas selection and subsequently achieve optimal segmentation accuracy.

Table 2. Subject demographics in set of 30 labelled randomly selected subjects used for method validation.

| | Control (n = 10) | MCI (n = 10) | AD (n = 10) |
|-----------------|------------------|--------------|-------------|
| Age, years | 78.6 (5.4) | 75.3 (8.8) | 77.2 (6.8) |
| Gender male (%) | 6 (60%) | 7 (70%) | 7 (70%) |
| MMSE, /30 | 29.5 (0.7) | 27.4 (1.8) | 27.0 (2.7) |

Mean (SD) unless specified otherwise.

doi:10.1371/journal.pone.0070059.t002

Methods

1.1 Ethics Statement

This study was conducted in accordance with the ethical standards stated in the Declaration of Helsinki using publicly available imaging data.

1.2 Overview

This paper aims to qualitatively and quantitatively assess the selection of atlases to combine in the framework of multi-atlas segmentation using 3 different manifold learning techniques. We consider Isomap [19], Locally Linear Embedding (LLE) [20] and Laplacian Eigenmaps (LEM) [21] since those techniques are the most widely used in medical imaging.

Our method can be summarized in 3 steps. First, a low-dimensional manifold is learned from the space spanned by the set of atlases using the 3 different techniques (§ 1.3). The neighbourhood relationship on the manifold is derived from non-rigid transformations that align atlases to each other in the high-dimensional space (§ 1.4). Second, a new target image is embedded onto the previously computed manifold by means of the out-of-sample extension [26] (§ 1.5). Third, the target image is segmented using atlases that are within its vicinity on the manifold (§ 1.6).

For each manifold learning technique, we investigate the effects of (i) the number of dimensions of the resulting embedding, (ii) the number of neighbours used to build the k -nearest neighbour graph in the high-dimensional space, and (iii) the number of atlases used during the combination process.

An atlas data set composed of 110 manually segmented images of hippocampi from the MIRIAD public data set (www.ucl.ac.uk/drc/research/miriad) is used to optimize each manifold learning technique on a leave-one-out experiment (§ 2.1). Segmentation accuracy is then validated on an independent set of 30 manually segmented images from the Alzheimer's Disease Neuroimaging Initiative (ADNI, www.loni.ucla.edu/ADNI/) (§ 2.2). The MIRIAD data set is described in § 1.7. The ADNI data set is described in § 1.8.

1.3 Manifold Learning

Given a set of n atlases $A = (a_1, \dots, a_n) \in \mathbb{R}^D$, the goal is to identify atlases that are most similar to a target image $x \in \mathbb{R}^D$ using manifold learning. It has been suggested that the set of brain images has an intrinsic dimensionality meaning that points in data set A and image x are lying on or near a manifold with dimensionality d which is embedded in the D -dimensional space [17]. By using manifold learning, data set $A \in \mathbb{R}^D$ is transformed into a new dataset $Y = (y_1, \dots, y_n) \in \mathbb{R}^d$ with $d < D$, while preserving the non-linear geometry and neighbourhood information of the high-dimensional data in the low-dimensional space. The atlases that are nearest to x are identified on the low-dimensional manifold and used for segmentation.

Variation in brain images is best described by non-linear dimensionality reduction models compared to linear ones like Principal Component Analysis (PCA) or Multi-Dimensional Scaling (MDS) [17]. In our study, low-dimensional embeddings are computed with 3 different non-linear techniques: Isomap [19], Locally Linear Embedding (LLE) [20] and Laplacian Eigenmaps (LEM) [21]. The differences between those 3 techniques are emphasized by their unique objective functions. For Isomap, the objective function is:

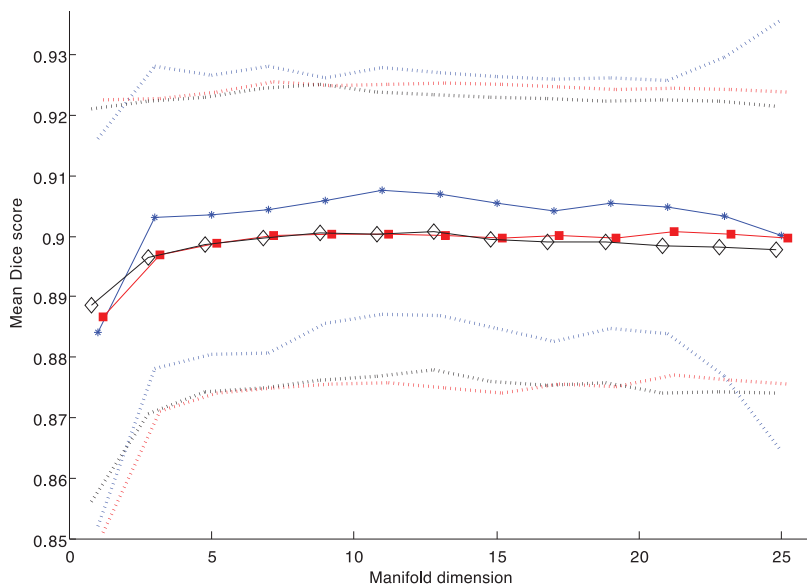
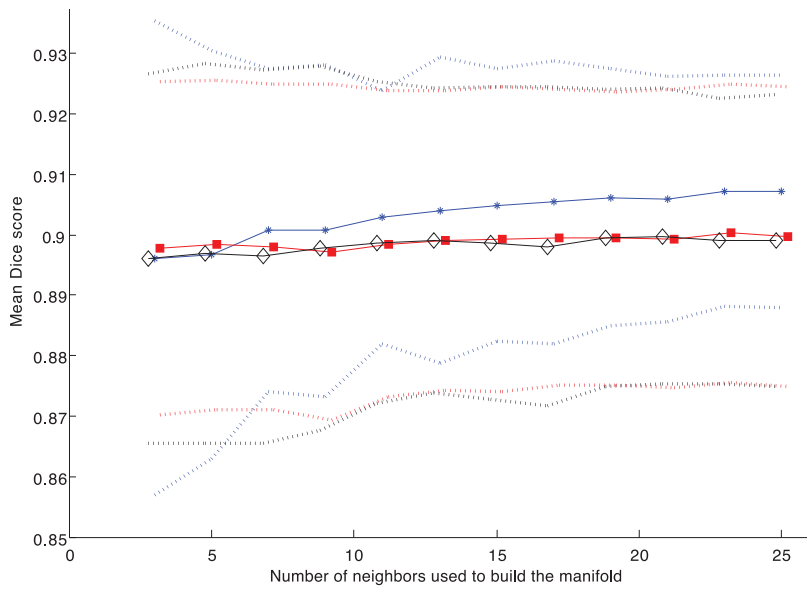
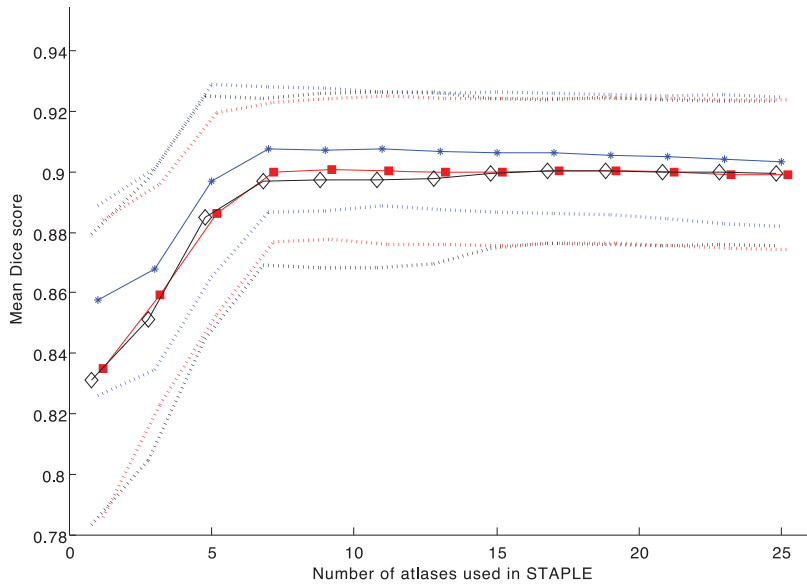


Figure 1. Mean Dice’s similarity index computed for $k_D \in [3, 25]$, $d \in [1, 25]$, $k_d \in [1, 25]$. Locally Linear Embedding is in blue, Isomap is in red and Laplacian Eigenmaps is in black. Solid lines represent the mean Dice’s similarity index, dotted lines represents the standard deviation. Mean Dice’s similarity index against: (a) the number of atlases fused in STAPLE (d and k_D fixed to best parameters), (b) the neighbourhood size k_D in computing the manifold (d and k_d fixed to best parameters), and (c) the manifold dimension d (k_D and k_d fixed to best parameters). doi:10.1371/journal.pone.0070059.g001

$$\phi(Y) = \sum_{i=1}^n \sum_{j=1}^n (g_{ij}^2 - \|y_i - y_j\|^2)$$

where g_{ij} represents the geodesic distance between a_i and a_j in the high-dimensional space. For LLE, the objective function is:

$$\phi(Y) = \sum_{i=1}^n \|y_i - \sum_{j \in N_k(i)} w_{ij} y_j\|^2$$

where $N_k(i)$ are the k -nearest neighbours of a_i and weight w_{ij} is the contribution of a_j in reconstructing a_i in the high-dimensional space. As demonstrated by [20], the optimal weights w_{ij} are obtained through minimization by solving a least-squares problem.

Finally, the objective function associated with LEM is:

$$\phi(Y) = \sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 e^{-\|a_i - a_j\|^2 / \sigma}$$

All 3 techniques require the construction of a connected graph in the high-dimensional space using the k -nearest neighbour algorithm. The number of neighbours used to build this connected graph is defined as k_D .

Unlike PCA, the embedding produced by these techniques is a function of a metric which determines the k_D -nearest neighbours in the high-dimensional space and subsequently the neighbouring images on the low-dimensional manifold. We use the metric presented in § 1.4 to find those k_D -nearest neighbours.

1.4 Distance between Pairs of Images

We derive the metric from the method presented by [27]. An atlas a and target image x are similar when the non-rigid transformation that aligns them produces a small deformation. Similarity is based on the displacement field $F_{x \rightarrow a}$ of the non-rigid transformation $T_{x \rightarrow a}$. In order to avoid the computational load of performing registrations between all atlases and every new unseen

target image, an average atlas M is built from the atlases in the data set using the iterative groupwise registration scheme described by [28]. This enables M to lie near the center of the space of all atlases. From the average atlas M , a displacement field $F_{M \rightarrow a}$ (resp. $F_{M \rightarrow x}$) is derived from the non-rigid transformation $T_{M \rightarrow a}$ (resp. $T_{M \rightarrow x}$) for each atlas a (resp. new target x). The similarity is then evaluated with:

$$s(x, a) = \sum_{l=1}^V \|F_{M \rightarrow a}(l) - F_{M \rightarrow x}(l)\|_2$$

where $\|\cdot\|_2$ is the L2 norm and V is the number of voxels in each atlas.

In this framework, the similarity between x and any atlases a can be evaluated by registering x to M . Since M lies near the center of the space of all atlases, the manifold resulting from the approximation of $F_{x \rightarrow a}$ with $F_{M \rightarrow a} - F_{M \rightarrow x}$ minimizes the error in estimating the neighbourhood relationship when compared to the manifold resulting from the direct computation of $F_{x \rightarrow a}$.

The non-rigid transformation T is performed using an efficient implementation [29] of the free-form deformation algorithm [30]. The transformation model is parameterized using a cubic B-Spline scheme and the transformation T is driven by the normalised mutual information.

1.5 Extending a Manifold with a New Target Image x

For Isomap, LLE and LEM, the out-of-sample extension is performed using the Nyström approximation [26]. Experiments on real high-dimensional data have demonstrated the accuracy of out-of-sample extension in positioning an out-of-sample point on a low-dimensional manifold [26]. The metric presented in § 1.4 is also used for extending the manifold.

Since the low-dimensional manifold is embedded in a Euclidean space, the L2 distance is used to determine the k_d -nearest neighbours of x on the manifold. Those k_d -nearest neighbours are subsequently used for label propagation.

Table 3. Mean Dice’s similarity indexes \overline{DS} (SD) obtained with manifold learning selection (LLE, ISO, LEM) and plain selection (BASE).

| | LLE | ISO | LEM | BASE |
|------------|-----------------------|-----------------------|------------------------|-----------------|
| | $d=11, k_D=23, k_d=7$ | $d=21, k_D=23, k_d=9$ | $d=13, k_D=21, k_d=19$ | $k_d=9$ |
| Mean DS | 0.9077 | 0.8995 | 0.8971 | 0.8756 |
| (SD) | (0.0211) | (0.0228) | (0.0245) | (0.0219) |
| p -value | LLE vs. | ISO vs. | LEM vs. | BASE vs. |
| | ISO, $p=0.0216$ | LLE, $p=0.0216$ | LLE, $p=0.0275$ | LLE, $p=0.0056$ |
| | LEM, $p=0.0275$ | LEM, $p=0.3250$ | ISO, $p=0.3250$ | ISO, $p=0.0137$ |
| | BASE, $p=0.0056$ | BASE, $p=0.0137$ | BASE, $p=0.0204$ | LEM, $p=0.0204$ |

p -values comparing each approach with each other are reported.

doi:10.1371/journal.pone.0070059.t003

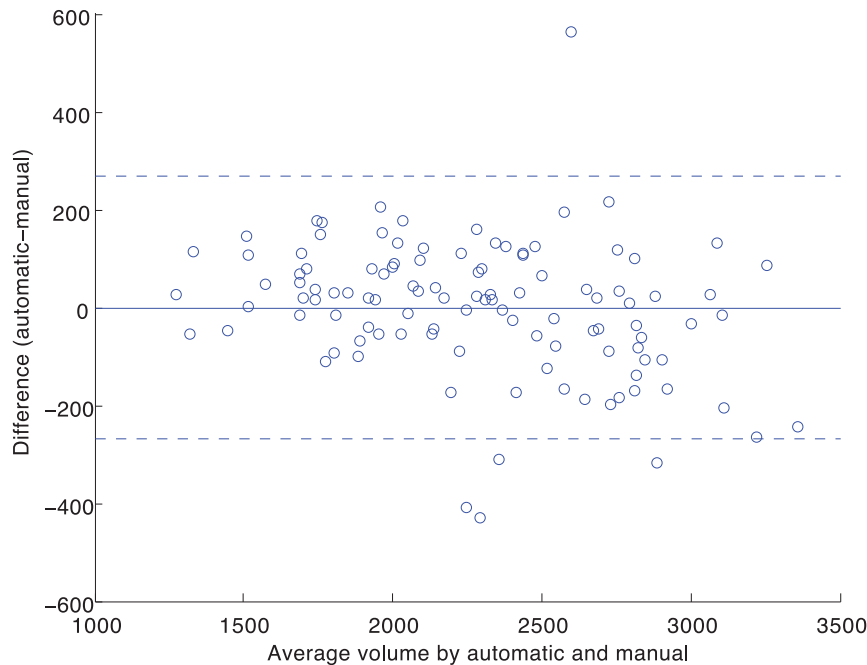


Figure 2. Bland-Altman plot. Each point corresponds to an hippocampal segmentation. The difference between automatic and manual estimates is plotted against their average. The solid horizontal line corresponds to the average difference, and the dashed lines are plotted at average ± 1.96 standard deviations of the difference.
doi:10.1371/journal.pone.0070059.g002

1.6 Segmentation by Fusion Strategy

STAPLE [31] is used to combine multiple segmentations generated from the most similar atlases. We found in our previous study [7] that STAPLE gives better results compared to a voting rule or shape-based averaging method when using the MIRIAD data set. It simultaneously computes a probabilistic estimate of the true segmentation and a measure of the performance level (sensitivity and specificity) represented by each segmentation in an expectation-maximization framework. An iterative Markov random field optimized with mean field approximation is used to provide spatial consistency in the probabilistic estimate of neighbouring voxels. The STAPLE algorithm is solved only in the non-consensus area in order to reduce bias as suggested by [28]. We denote by k_d the number of atlases used for label propagation.

1.7 Atlas Data Set of 110 Hippocampi

The MIRIAD data set is used as the atlas data set. It is a database of volumetric MRI brain scans of patients suffering from Alzheimer's disease and healthy elderly people. The data set is publicly available (www.ucl.ac.uk/drc/research/miriad) in anonymised form to aid researchers in developing new techniques for the analysis of serially acquired MRI. The atlas data set consists of 55 subjects who were recruited from the Cognitive Disorders Clinic at The National Hospital for Neurology and Neurosurgery, into a longitudinal neuroimaging study. All subjects underwent clinical assessment including the Mini-Mental State Examination (MMSE) [32]. All subjects gave written informed consent to take part in this study. Imaging data were used to create an average atlas using the groupwise registration algorithm described in § 1.4 and in the parameter optimization process in § 2.1. Subjects included 36 clinically diagnosed probable AD patients and 19 age-matched healthy controls. All patients fulfilled standard NINCDS/ADRDA criteria [33] for the diagnosis of probable

AD. Subject demographics can be seen in Table 1. T1-weighted volumetric MR brain scans were performed on the same 1.5-T Signa unit (General Electric, Milwaukee), using an inversion recovery prepared fast SPGR sequence and a 256×256 image matrix with the field of view being 18 cm (acquisition parameters: repetition time = 15 ms; echo time = 5.4 ms; flip angle = 15° ; inversion time = 650 ms). The volumetric scans were reconstructed as 124 contiguous 1.5-mm coronal images. T1-weighted volumetric scans were evaluated by one rater. All scans were N3 corrected [34] and bias correction was performed.

The left and right hippocampal regions were manually segmented by an expert segmentor S. The segmentation protocol is presented in the Appendix S1. The intra-rater variability measured by an ICC is 0.98. The left hippocampal segmentations from all 55 subjects are flipped along the mid-sagittal plane. This flipping effectively doubles the size of the data set by allowing, for example, the left hippocampus of a target image to be matched to the right hippocampus in the atlas data set. Therefore, the final atlas data set consists of 110 hippocampal images.

1.8 ADNI Data Set of 30 Subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.adni.loni.ucla.edu). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 5-year public-private partnership. The aims of ADNI included assessing the ability of imaging and other biomarkers to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California - San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and

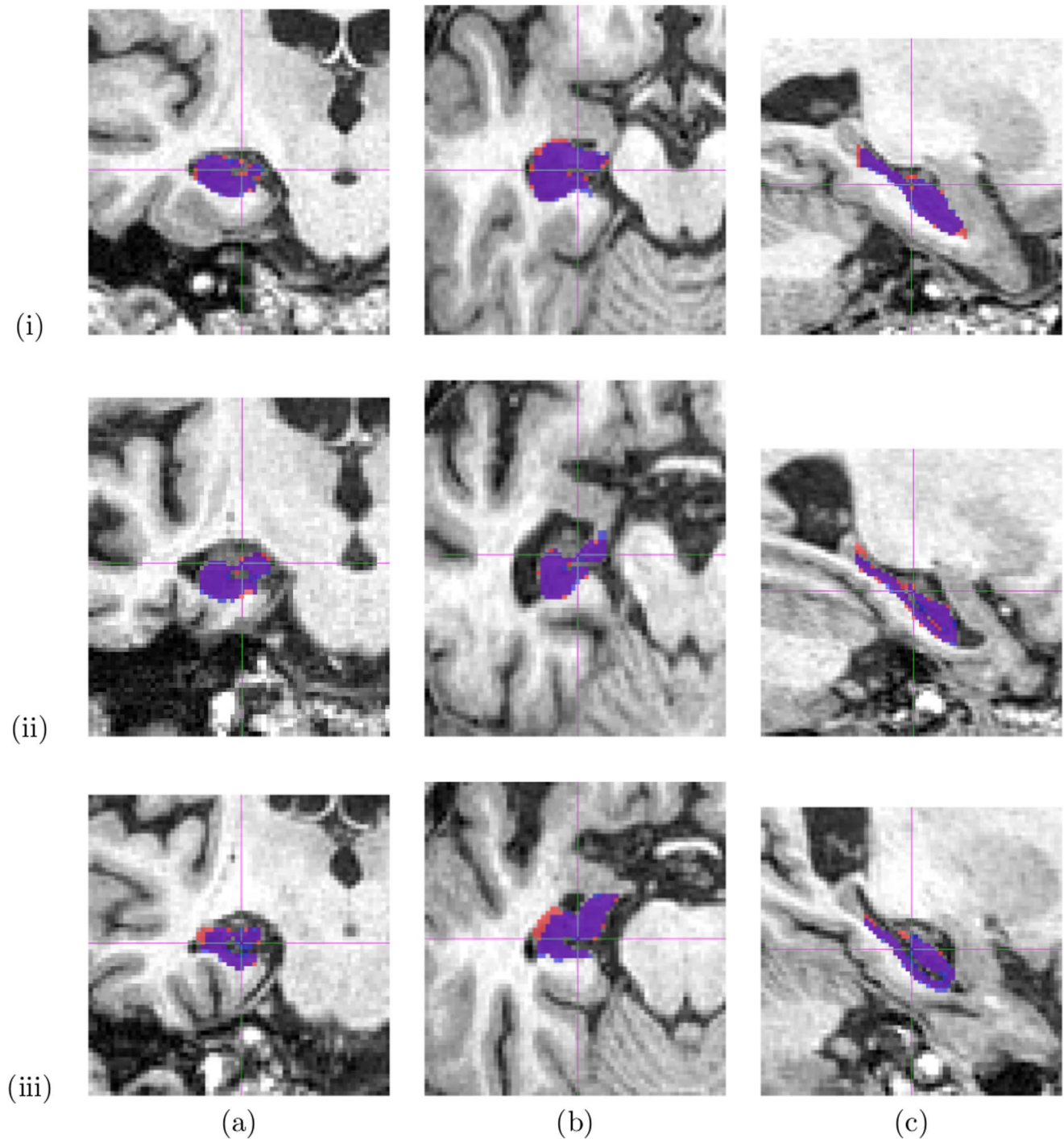


Figure 3. Hippocampal segmentation: automated (blue) vs manual (red). Overlapping area in purple. Row: (i) High case (Dice = 0.9398), (ii) Typical case (Dice = 0.9073), (iii) Low case (Dice = 0.8614). Column: (a) Coronal view, (b) Sagittal view, (c) Axial view. doi:10.1371/journal.pone.0070059.g003

private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals, 400 people with MCI and 200 people with early AD. For up-to-date information, see www.adni-info.org.

The 30 ADNI subjects (10 AD, 10 MCI and 10 controls) used for method validation consist of preprocessed baseline volumetric

T1-weighted MR images acquired using 1.5T scanners (GE Healthcare, Philips Medical Systems or Siemens Medical Solutions) at multiple sites from the ADNI website. Representative imaging parameters were TR = 2400 ms, TI = 1000 ms, TE = 3.5 ms, flip angle = 8° , field of view = 240×240 mm and 160 sagittal 1.2 mm-thick slices and a 192×192 matrix yielding a voxel resolution of $1.25 \times 1.25 \times 1.2$ mm³, or 180 sagittal 1.2 mm-thick slices with a 256×256 matrix yielding a voxel resolution of

Table 4. Mean (SD) of the volumes (in mm³) in the left hippocampus in the baseline images of the atlas library of 110 images used to assess optimal methods and parameters.

| | Control (n = 19) | AD (n = 36) |
|--|---------------------|----------------|
| Manual (SD) | 2749 (273) | 2054 (424) |
| Automated (SD) | 2722 (249) | 2066 (387) |
| Man vs Auto mean of difference (p-value) | 27 (p = 0.19) | -12 (p = 0.14) |
| SD of differences | 129 | 150 |

doi:10.1371/journal.pone.0070059.t004

0.94 × 0.94 × 1.2 mm³. The details of the ADNI MR imaging protocol are described in [35], and listed on the ADNI website (www.loni.ucla.edu/ADNI/Research/Cores/). Each scan underwent a quality control evaluation at the Mayo Clinic (Rochester, MN, USA). Quality control included inspection of each incoming image file for protocol compliance, clinically significant medical abnormalities, and image quality. The T1-weighted volumetric scans that passed the quality control were processed using the standard ADNI image processing pipeline, which included post-acquisition correction of gradient warping [36], B1 non-uniformity correction [37] depending on the scanner and coil type, intensity non-uniformity correction [34] and phantom based scaling

correction [38] with the geometric phantom scan having been acquired with each patient scan.

Table 2 shows the clinical and demographic data of the 30 ADNI subjects. The same expert segmentor S as previously mentioned manually delineated the left hippocampus of those subjects. A segmentor S2 also manually delineated the left hippocampus on the same baseline images. The segmentation protocol is presented in the Appendix S1. The inter- and intra-rater reliability correspond to a Dice’s similarity index of 0.93 and 0.96 respectively.

Experiments

2.1 Optimizing Manifold Learning Parameters Using a Manually Segmented Data Set of 110 Atlases

A leave-one-out approach that excludes both the left and right hippocampi of the target image from the library of 110 atlases is used to optimize the parameters for each manifold learning technique. The following 4-step procedure is repeated for each atlas a_{out} in the library. (i) After excluding a_{out} and its flipped image from the library, an average atlas M is built from the remaining 108 images in the data set. Distances between remaining atlases are computed based on the non-rigid transformations that align them to M as described in § 1.4. (ii) A manifold is computed from the remaining 108 atlases. (iii) The embedding is extended with a_{out} . Distances between a_{out} and the remaining atlases are derived by registering it to M and performing

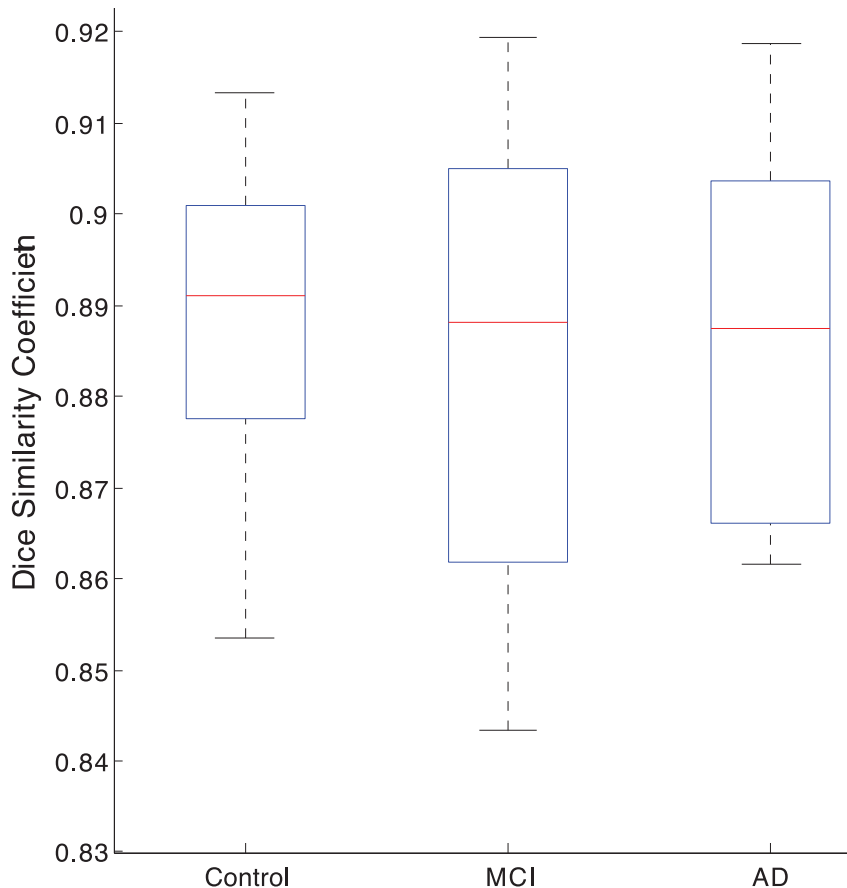


Figure 4. Average Dice’s similarity index for NC, MCI and AD group obtained by fusing top 7 atlases with STAPLE. Atlases were selected with manifold learning.

doi:10.1371/journal.pone.0070059.g004

Table 5. Mean (SD) of the volumes (in mm³) in the left hippocampus in the baseline images of the labelled ADNI data set of 30 images for method validation.

| | Control (n = 10) | MCI (n = 10) | AD (n = 10) |
|---|------------------------|----------------------|-----------------------|
| Manual (SD) | 2531 (336) | 2331 (410) | 1994 (478) |
| Automated (SD) | 2642 (360) | 2334 (431) | 2018 (387) |
| Man. vs Auto. mean of diff. (<i>p</i> -value, paired <i>t</i> -test) | -111 (<i>p</i> =0.33) | -3 (<i>p</i> =0.47) | -24 (<i>p</i> =0.29) |
| SD of differences | 168 | 155 | 130 |

doi:10.1371/journal.pone.0070059.t005

subtraction of displacement fields. (iv) Its k_d -nearest neighbours are identified on the manifold using the L2 norm and combined in STAPLE to yield an estimated segmentation of a_{out} .

Dice’s similarity index [39] is used for evaluation and is computed by measuring the overlap between the estimated segmentation and the manual segmentation. Dice’s similarity index is defined as $DS(A,B) = 2|A \cap B| / (|A| + |B|)$, where A is the set of voxels in the automated region and B is the set of voxels in the manual region. A Dice’s similarity index is calculated for each a_{out} and a mean Dice’s similarity index \overline{DS} is calculated by averaging all 110 scores.

There is no defined procedure to establish the number of dimensions d of a learned manifold, and the number of neighbours k_D to build the connected graph in the high-dimensional space is often determined empirically. Results are evaluated for 3 different techniques: Isomap, LLE and LEM with dimension $d \in [1,25]$ and a neighbourhood number of $k_D \in [3,25]$ for each manifold technique. Using STAPLE with a MRF strength of 0.2, segmentations are generated by combining the closest $k_d \in [1,25]$ neighbours to a_{out} in the lower dimensional space. For LEM, σ is set to 1. A 4D matrix of mean Dice’s similarity indexes is then computed with the following axes: manifold type $\in \{ISO, LLE, LEM\}$, $d \in [1,25]$, $k_D \in [3,25]$, and $k_d \in [1,25]$. The coordinates in this matrix that give the highest \overline{DS} indicate the best manifold learning technique with optimized parameters for this data set.

In order to compare atlas selection *with* manifold learning to atlas selection *without* manifold learning, we also compute the results given by a plain k_d -nearest neighbour selection in the high-dimensional space D . For each a_{out} , its k_d -nearest neighbours in the high-dimensional space D are computed using the metric defined in § 1.4 and combined in STAPLE to yield an estimated segmentation. As before, a Dice’s similarity index is calculated for each a_{out} and a mean Dice’s similarity index \overline{DS} is calculated by averaging all 110 scores. We refer to this selection method as BASE and results are computed for $k_d \in [1,25]$.

2.2 Method Validation Using a Manually Segmented Data Set of 30 ADNI Subjects

For method validation, the left hippocampus in the baseline images of 30 randomly selected subjects in the ADNI database (10

AD, 10 MCI and 10 controls) were segmented. Those images differ from the MIRIAD data set of atlases used for parameter optimization. The atlas data set of 110 images is used to segment each of the ADNI target images. The optimal parameters determined in § 2.1 are used to generate left hippocampal regions. Since the right hippocampus segmentations for this set of 30 subjects were not available, we only evaluate the accuracy of our method on the left hippocampus.

Results

3.1 Results from Method Optimization Using a Manually Segmented Data Set of 110 Atlases

The best combination of manifold learning technique and parameters is Locally Linear Embedding with a manifold dimension of $d = 11$, a neighbourhood size $k_D = 23$ and combining the top $k_d = 7$ matches in STAPLE, giving a mean (SD) Dice’s similarity index \overline{DS}_{max} of 0.9077 (0.0211). In contrast, Isomap and Laplacian Eigenmaps resulted in Dice’s similarity indexes of 0.8995 (0.0228) and 0.8971 (0.0245) with $d = 21$, $k_D = 23$ and $k_d = 9$ and $d = 13$, $k_D = 21$ and $k_d = 19$ respectively. Each graph in Figure 1 shows the mean Dice’s similarity index for each manifold learning technique when d , k_D and k_d are fixed to their respective optimal parameters. It is interesting to note that all 3 manifold learning techniques result in a very high mean Dice’s similarity index (>0.89). Using a 2-tailed paired *t*-test, Locally Linear Embedding gives a significantly ($p = 0.0216 < 0.05$ and $p = 0.0275 < 0.05$) higher average Dice’s similarity index compared to Isomap and Laplacian Eigenmaps, whereas the difference between Isomap and Laplacian Eigenmaps is not statistically significant ($p = 0.3250 > 0.05$). The accuracy achieved by fusing multiple segmentations quickly rises to a maximum and then gradually declines as the number of segmentations increases. This is in line with results published in [5] and [7]: the gradual decline corresponds to adding dissimilar images into the combination process, resulting in segmentation errors. The accuracy also flattens out for manifolds of 3 or more dimensions. This suggests that our data set of hippocampi can be described mostly by 3 main modes of variation, and this is consistent across all manifold learning techniques presented. The number of neighbours k_D used to build the connected graph has little effect on the accuracy when using Isomap and Laplacian Eigenmaps. In contrast, increasing k_d increases the accuracy achieved with Locally Linear Embedding.

Table 3 compares the mean Dice’s similarity index (SD) obtained by selecting atlases with manifold learning and using the BASE method. The results show that all 3 manifold learning selection methods significantly outperform ($p < 0.05$) the plain selection method.

Table 4 shows the mean (SD) of the manual and automated hippocampal volumes. The automated volumes were computed using Locally Linear Embedding with the optimized parameters.

Table 6. Effect size.

| | ES_{AD} | ES_{MCI} |
|----------------|-----------|------------|
| Manual (SD) | -1.124 | -0.490 |
| Automated (SD) | -1.614 | -0.720 |

doi:10.1371/journal.pone.0070059.t006

The mean (SD) of differences between the manual and automated hippocampal volumes by baseline diagnostic group was 27 (129) mm^3 (automated<manual) for controls and -12 (150) mm^3 (automated>manual) for AD subjects. In order to test the validity of our method, we compare the proposed method to a state-of-the-art method for hippocampus segmentation based on a similar atlas library approach [7]. Using the same library of 110 hippocampus images and optimal parameters defined in [7], a similar leave-one-out method is performed. The mean Dice's similarity index was 0.8955 (0.0172) compared to 0.9077 (0.0211) in our method. Even though these values differ by 0.01 point only, the difference is statistically significant ($p < 0.001$). Figure 2 plots the volume correlation between the manual segmentation and our automatic segmentation method. The volume differences between manual segmentation and automatic segmentation are similar to zero-mean random noise. Figure 3 shows an example of segmentation obtained with our method.

Overall, these results show that registering atlases that have been selected by manifold learning (i.e. selection in the lower-dimensional space) produces accurate and robust segmentation in the framework of multi-atlas based segmentation and gives better results compared to atlas selection without manifold learning (i.e. selection in the high-dimensional space). Also, given our data set of atlases, Locally Linear Embedding gives significantly better results than Isomap and Laplacian Eigenmaps.

3.2 Results from Method Validation Using a Manually Segmented ADNI Data Set of 30 Subjects

We use Locally Linear Embedding with the optimal parameters found in § 3.1 to generate automatic segmentation of the 30 ADNI subjects. The mean (SD) Dice's similarity indexes of the left hippocampus segmentations of the baseline ADNI images are 0.887 (0.020) for controls, 0.886 (0.025) for MCI, 0.878 (0.038) for AD and 0.883 (0.028) across the three groups. These are summarized in Figure 4. The difference in accuracy compared to the previous experiment can be explained by the fact that the atlases and the 30 ADNI subjects belong to different data sets. Also the high shape variability and the possible presence of cysts in the hippocampus can explain lower scores in AD subjects. Table 5 shows the means (SD) of the manual and automated hippocampal volumes. The mean (SD) of differences in the manual and automated hippocampal volumes by baseline diagnostic group are -111 (168) mm^3 for controls, -3 (155) mm^3 for MCI, and -24 (130) mm^3 for AD subjects with automated volumes higher than manual volumes in all the three groups. Overall, the mean (SD) of differences in the manual and automated hippocampal volumes is -45 (154) mm^3 . We also calculate the effect size $ES_{AD} = (\mu_{AD} - \mu_C) / \sigma_{AD}$ and $ES_{MCI} = (\mu_{MCI} - \mu_C) / \sigma_{MCI}$ in Table 6, where μ_C , μ_{MCI} , μ_{AD} are the average volumes in the control, MCI and AD groups respectively, and σ_{MCI} , σ_{AD} are the standard deviations in the MCI and AD groups respectively.

Conclusions

We compared Isomap, Locally Linear Embedding and Laplacian Eigenmaps for the selection of atlases to use in multi-atlas segmentation of the hippocampus of normal controls and patients with Alzheimer's disease in MR images.

We found that Locally Linear Embedding generated the best hippocampal segmentation ($DS = 0.9077$) on a leave-one-out experiment using our data set of 110 atlases. The mean volumes and SDs of the generated segmentations were similar to those produced using manual segmentation. Overall, the mean difference between our automated volumes and the manual measurements was 7.5 mm^3 or around 0.01% of the mean of all volumes.

We found good accuracy of our method on unseen data, achieving a mean Dice's similarity index of 0.883 (0.028) when comparing the automated and manual segmentations of a set of 30 subjects (10 AD, 10 MCI and 10 controls). Overall, the mean (SD) of differences in the manual and automated hippocampal volumes was 45 (154) mm^3 with manual<automated.

Our results are consistent with those in [40]. They found that large number of k_d -nearest neighbours leads to higher Dice's similarity index for large database size M and that Dice's similarity index decreases as k_d approaches the value of M . In our study, the Dice's similarity index quickly rises to a maximum when the number of k_d -nearest neighbours increases for all the manifold learning techniques. The Dice's similarity index then gradually declines as the number of k_d -nearest neighbours increases.

Not only is the choice of manifold learning important but also the parameters used to compute the embedding. For instance, most studies have represented the embedding with 2 or 3 dimensions as it enables spacial visualization of the embedding. However the optimal embedding could have been of higher dimensions. Indeed, in our study, we found that the best results arose when using 11 dimensions. Also all manifold learning techniques presented in this paper require the choice of a neighbourhood size either for the calculation of the geodesic distance in Isomap, or reconstructing a data point with its closest points in Locally Linear Embedding or Laplacian Eigenmaps. The choice of the optimal dimension and best parameters is often made empirically.

The results showed that selection of atlases with manifold learning is beneficial in the framework of multi-atlas based segmentation. The optimal accuracy can be found by fine tuning the manifold learning process. It also turned out that our atlas data set of hippocampi can be described by 3 main modes of variation regardless of the manifold learning technique used.

We found that Locally Linear Embedding gave best results for our data set of the hippocampus but it might not yield optimum results for a different anatomical structure. There is no consensus on which manifold learning technique to use for a given data set. A legitimate question that arises is which manifold learning algorithm is best suited for which data set. As demonstrated in this study, different manifold learning techniques produce different low-dimensional embeddings even for the same data set. This can be explained by the fact that the cost function to optimize associated with a manifold learning technique differs from one method to another.

The lower Dice's similarity index obtained when segmenting the 10 AD subjects from the ADNI data may also illustrate the issue of manifold sampling. Since the manifold is directly learned from points (i.e. images) in the data set, the sampling of the manifold is highly correlated with the density of points in the high-dimensional space. For example, if certain areas in the high-dimensional space are too sparse, the resulting manifold is likely to be a poor approximation of the true manifold structure. Since the atlas data set did not contain any MCI subjects, the manifold derived from this atlas data set is not representative of a population containing NC, MCI and AD subjects. It would have been preferable to derive a manifold from NC only subjects in the atlas data set to segment the 10 NC from the ADNI data set, and similarly for the 10 AD in the ADNI data set.

An important aspect in manifold learning is the metric used to relate pairs of images in the high-dimensional space. The most commonly used metrics are based on voxel intensity such as the Euclidean distance, cross correlation or mutual information. Similarly to [17] and [14], we used a metric derived from non-rigid transformation. In theory, the metric used should reflect the

information relating pairs of images [24,41]. However, there is currently no research investigating the influence of the metric on the resulting embedding. In the future, we are planning to compare the effects of several metrics such as the geometric median and the geodesic estimation proposed by [42] and [43] respectively on low-dimensional embeddings.

We have obtained one of the best accuracies reported to date for automated hippocampal segmentation when compared with gold standard manual segmentations from a set of 30 randomly chosen subjects (10 AD, 10 MCI and 10 controls) from ADNI. Our Dice's similarity index is equal to 0.88 with the previous highest Dice's similarity indexes (N = number of hippocampi in the study) being 0.86 (N = 14) [44], 0.83 (N = 60) [45], 0.81 (N = 100) [46], 0.86 (N = 54) [47], 0.87 (N = 30) [48], 0.88 (N = 5) [49] (from a cohort of 2 year old children), 0.86 (N = 40) [50], 0.85 (N = 30) [51], 0.86 (N = 40) [52], 0.83 (N = 550) [5], 0.89 (N = 160) [53], 0.89 (N = 30) [7], 0.89 (N = 120) [8] and 0.85 (N = 364) [12]. Our intra-rater variability corresponds to a Dice's similarity index of 0.96. Comparing this to the results from using our automatic method with different training and test data (0.88) suggests that the method has not been over-trained, and that there is potential to improve it further.

Overall, our technique is most similar to that reported by [12]. However it fundamentally differs in the following ways: (i) [12] used a similarity measure derived from voxel intensities, whereas we used a metric derived from registration. (ii) We embedded target images using the out-of sample extension instead of embedding all images in a single manifold. This method effectively scales with the number of atlases and not the number of images to segment. (iii) We used STAPLE as a fusion method, whereas statistical voxel classification and graph cuts was used in [12].

We developed a suitable method for segmenting large data sets by extending the manifold with an out-of-sample image. Indeed, in

our method: (i) the low-dimensional manifold learned from the space spanned by the set of atlases, (ii) the average atlas M and (iii) the registrations between the atlases and M are precomputed and stored, thus making our method very computationally efficient. We only need to perform one non-rigid registration between M and a new unseen target image x to select its most similar images from the atlases. This method is therefore scalable and extremely computationally efficient, making it suitable for segmenting large data sets and for clinical use. For instance, in the context of radiotherapy treatment, we are planning to apply our method to CT images of head and neck, where segmentations of tumor regions and organs at risk (such as the parotid glands and lymph nodes) show low agreement within and between raters due to poor boundary definition on CT images.

To conclude, manifold learning produces accurate segmentation in the framework of multi-atlas segmentation by improving atlas selection. Our method shows that Locally Linear Embedding gave better results in our experiments, however using a different atlas data set with a different density distribution will probably require the re-computation of the optimized parameters and manifold for segmentation.

Supporting Information

Appendix S1 Hippocampus segmentation protocol. (PDF)

Author Contributions

Conceived and designed the experiments: AKHD MM SO. Performed the experiments: AKHD. Analyzed the data: AKHD. Contributed reagents/materials/analysis tools: JB KKL MJC. Wrote the paper: AKHD. Made critical comments on the manuscript: TK.

References

- Klein S, van der Heide U, Lips I, van Vulpen M, Staring M, et al. (2008) Automatic Segmentation of the Prostate in 3D MR Images by Atlas Matching using Localized Mutual Information. *Medical Physics* 35: 1407–1417.
- Artaechevarria X, Munoz-Barrutia A, de Solorzano C (2009) Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans Med Imaging* 28: 1266–1277.
- Isgum I, Staring M, Ruitten A, Prokop M, Viergever M, et al. (2009) Multi-atlas-based segmentation with local decision fusion application to cardiac and aortic segmentation in CT scans. *IEEE Trans Med Imag* 28: 1000–1010.
- Langerak T, van der Heide U, Kotte A, Viergever M, van Vulpen M, et al. (2010) Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple). *Medical Imaging, IEEE Transactions on* 29: 2000–2008.
- Aljabar P, Heckemann R, Hammers A, Hajnal J, Rueckert D (2009) Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage* 46: 726–738.
- Wu M, Rosano C, Lopez-Garcia P, Carter CS, Aizenstein HJ (2007) Optimum template selection for atlas-based segmentation. *Neuroimage* 34: 1612–1618.
- Leung K, Barnes J, Ridgway G, Bartlett J, Clarkson M, et al. (2010) Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 51: 1345–1359.
- Lojtonen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, et al. (2010) Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49: 2352–2365.
- Sabuncu M, Yeo B, Leemput K, Fischl B, Golland P (2010) A generative model for image segmentation based on label fusion. *IEEE Trans Med Imaging* 29: 1714–1729.
- Skika M (2010) Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote. *Med Image Anal* 14: 219–226.
- van Rikxoort EM, Isgum I, Arzhaeva Y, Staring M, Klein S, et al. (2010) Adaptive local multiatlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis* 14: 39–49.
- Wolz R, Aljabar P, Hajnal J, Hammers A, Rueckert D (2010) LEAP: Learning embeddings for atlas propagation. *Neuroimage* 49: 1316–1325.
- Zhang Q, Souvenir R, Pless R (2006) On manifold structure of cardiac mri data: Application to segmentation 1: 1092–1098.
- Hamm J, Ye D, Verma R, Davatzikos C (2010) GRAM: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis* 14: 633–642.
- Wachinger C, Yigitsoy M, Rijkhorst EJ, Navab N (2012) Manifold learning for image-based breathing gating in ultrasound and mri. *Medical Image Analysis* 16: 806–818.
- Aljabar P, Rueckert D, Crum W (2008) Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. *NeuroImage* 43: 225–235.
- Gerber S, Tasdizen T, Thomas Fletcher P, Joshi S, Whitaker R (2010) Manifold modeling for brain population analysis. *Medical Image Analysis* 14: 643–653.
- Aljabar P, Wolz R, Srinivasan L, Counsell S, Boardman JP, et al. (2010) Combining morphological information in a manifold learning framework: application to neonatal MRI. In: Metaxas, D, Axel, L, Fichtinger, G, Szekely, G (Eds), *Medical Image Computing and Computer-Assisted InterventionMICCAI 2010* : 1–8.
- Tenenbaum J, de Silva V, Langford J (2000) A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290: 2319–2323.
- Roweis ST and Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290: 2323–2326.
- Belkin M, Niyogi P (2003) Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation* 15: 1373–1396.
- Wachinger C, Yigitsoy M, Navab N (2010) Manifold learning for multi-modal image registration. *MICCAI*.
- Wolz R, Aljabar P, Hajnal J, Rueckert D (2010) Manifold Learning for Biomarker Discovery in MR Imaging. *MICCAI* 6357: 116–123.
- Souvenir R and Pless R (2005) Isomap and Nonparametric Models of Image Deformation. In: *Proceedings of the IEEE Workshop on Motion and Video Computing*. IEEE Computer Society, volume 2, 195–200.
- Pan Y, Ge SS, Al Mamun A, Tang FR (2008) Detection of seizures in EEG signal using weighted locally linear embedding and SVM classifier. *IEEE International Conference on Cybernetics and Intelligent Systems* : 358–363.
- Bengio Y, Paiement JF, Vincent P, Delalleau O, Le Roux N, et al. (2004) Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. *Advances in Neural Information Processing Systems* 16: 177–184.
- Commowick O and Malandain G (2007) Efficient selection of the most similar image in a database for critical structures segmentation. In: *Proceedings of the*

- 10th international conference on Medical image computing and computer-assisted intervention. Berlin, Heidelberg: Springer-Verlag, MICCAI'07, 203–210.
28. Rohlfing T, Brandt R, Menzel R, Maue C (2004) Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage* 21: 1428–1442.
 29. Modat M, Ridgway G, Taylor Z, Lehmann M, Barnes J, et al. (2010) Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine* 98: 278–284.
 30. Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, et al. (1999) Nonrigid registration using freeform deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18: 712–721.
 31. Warfield S, Zou K, Wells W (2004) Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 23: 903–921.
 32. Folstein M, Folstein S, McHugh P (1975) The Mini mental state: a practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res* 12: 189–198.
 33. McKhann G, Drachman D, Folstein M, Katzman R, Price D, et al. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34: 939–944.
 34. Sled J, Zijdenbos A, Evans A (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imag* 17: 87–97.
 35. Jack C, Bernstein M, Fox N, Thompson P, Alexander G, et al. (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* 27: 685–691.
 36. Jovicich J, Czanner S, Greve D, Haley E, van der K, et al. (2006) Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 30: 436–443.
 37. Narayana P, Brey W, Kulkarni V, Sievenpiper C (1988) Compensation for surface coil sensitivity variation in magnetic resonance imaging. *Magn Reson Imaging* 6: 271–274.
 38. Gunter J, Bernstein M, Borowski B, Felmlee J, Blezek D, et al. (2006) Validation testing of the MRI calibration phantom for the Alzheimer's Disease Neuroimaging Initiative Study. *ISMRM*.
 39. Dice LR (1945) Measures of the Amount of Ecologic Association Between Species. *Ecology* 26: 297–302.
 40. Awate SP, Zhu P, Whitaker RT (2012) How many templates does it take for a good segmentation?: error analysis in multiatlas segmentation as a function of database size. *Proceedings of the Second international conference on Multimodal Brain Image Analysis* : 103–114.
 41. Pless R (2004) Differential Structure in non-Linear Image Embedding Functions. *Conference on Computer Vision and Pattern Recognition Workshop* 1.
 42. Fletcher PT, Venkatasubramanian S, Joshi S (2009) The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage* 45: S143–S152.
 43. Avants B, Gee JC (2004) Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage* 23, Supplement 1: S139–S150.
 44. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, et al. (2002) Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron* 33: 341–355.
 45. Heckemann R, Hajnal J, Aljabar P, Rueckert D, Hammers A (2006) Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* 33: 115–126.
 46. Pohl K, Bouix S, Shenton M, Grimson W, Kikinis R (2007) Automatic segmentation using non-rigid registration. *MICCAI* 26: 1201–1212.
 47. Barnes J, Schill R, Frost C, Schott J, Rossor M, et al. (2008) Increased hippocampal atrophy rates in AD over 6 months using serial MR imaging. *Neurobiol Aging* 29: 1199–1203.
 48. Chupin M, Chetelat G, Lemieux L, Dubois B, Garnero L, et al. (2008) Fully automatic hippocampus segmentation discriminates between early alzheimer's disease and normal aging. *5th IEEE International Symposium on Biomedical Imaging Computer* : 97–100.
 49. Gousias IS, Rueckert D, Heckemann RA, Dyet LE, Boardman JP, et al. (2008) Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *Neuroimage* 40: 672–684.
 50. Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, et al. (2008) Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage* 43: 59–68.
 51. Powell S, Magnotta VA, Johnson H, Jammalamadaka VK, Pierson R, et al. (2008) Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *NeuroImage* 39: 238–247.
 52. van der Lijn F, den Heijer T, Breteler MM, Niessen WJ (2008) Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *Neuroimage* 43: 708–720.
 53. Collins D, Pruessner J (2010) Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage* 52: 1355–1366.