Local manifold learning for multiatlas segmentation: application to hippocampal segmentation in healthy population and Alzheimer's disease

Xin-Wei Li,¹,² Qiong-Ling Li,¹,² Shu-Yu Li,¹,² & De-Yu Li² for the Alzheimer's Disease Neuroimaging Initiative*

¹ State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
² Key Laboratory for Biomechanics and Mechanobiology of Ministry of Education, School of Biological Science & Medical Engineering, Beihang University, Beijing, China

Keywords
Hippocampal segmentation; Local label fusion; Manifold learning; Multiatlas segmentation.

SUMMARY
Aims: Automated hippocampal segmentation is an important issue in many neuroscience studies. Methods: We presented and evaluated a novel segmentation method that utilized a manifold learning technique under the multiatlas-based segmentation scenario. A manifold representation of local patches for each voxel was achieved by applying an Isomap algorithm, which can then be used to obtain spatially local weights of atlases for label fusion. The obtained atlas weights potentially depended on all pairwise similarities of the population, which is in contrast to most existing label fusion methods that only rely on similarities between the target image and the atlases. The performance of the proposed method was evaluated for hippocampal segmentation and compared with two representative local weighted label fusion methods, that is, local majority voting and local weighted inverse distance voting, on an in-house dataset of 28 healthy adolescents (age range: 10–17 years) and two ADNI datasets of 100 participants (age range: 60–89 years). We also implemented hippocampal volumetric analysis and evaluated segmentation performance using atlases from a different dataset. Results: The median Dice similarities obtained by our proposed method were approximately 0.90 for healthy subjects and above 0.88 for two mixed diagnostic groups of ADNI subjects. Conclusion: The experimental results demonstrated that the proposed method could obtain consistent and significant improvements over label fusion strategies that are implemented in the original space.

Introduction
Accurate segmentation of subcortical structures from brain magnetic resonance (MR) images plays an increasingly important role in many medical applications. For example, the hippocampus, a deep-brain structure known for its involvement in learning and memory, is damaged in many psychiatric disorders, such as temporal lobe epilepsy [1], schizophrenia [2], and Alzheimer's disease (AD) [3,4]. The size and morphology of the hippocampus constitutes a valuable tool in clinical diagnosis, treatment, and assessment of these diseases. The current gold standard for hippocampus segmentation is manual delineation by experts. However, manual segmentation is extremely time-consuming and is subject to inter- and intrarater differences [5], thus limiting its clinical application. Therefore, automated segmentation techniques are desirable, particularly when the dataset is large.
Multiatlas-based segmentation is currently one of the most accurate automatic approaches [6–8]. In this method, an atlas is defined as an image with manual labels. The segmentation starts by registering each intensity atlas image to the target image and then warping the corresponding label image to the target space using the same transformation. Each warped label image is regarded as candidate segmentation for the target image. Subsequently, label fusion is implemented by integrating all candidate segmentations into the definitive segmentation for the target. Several label fusion methods were developed to further improve the quality of multiatlas-based segmentation. One class of label fusion methods is based on Simultaneous Truth and Performance Level Estimation (STAPLE) framework, which does not use the intensities of the atlases after registration [9–12]. Another class of combination strategy is based on ad hoc voting. The simplest form is majority voting, where the final label is determined to be the one in which most segmentations agree [13]. Recently, it was demonstrated that weighted label fusion approaches using global [14], local [15,16], and nonlocal [17] intensity similarity metrics are successful in practice. The criterion of weighted voting is to assign larger weights to the atlases that are more similar with the target image. Atlas weights derived from local or nonlocal similarity outperformed the global-based methods [14,17,18]. Additionally, more sophisticated fusion methods were proposed, such as joint label fusion, which models the pairwise dependency between the atlases as the joint probability of the two atlases producing the same segmentation error, that can be used to reduce the influence of redundancy information on the atlas dataset [19,20].

It is worth noting that the label fusion methods mentioned above were conducted in original high-dimensional space (on the order of the number of image/patch voxels). However, not all dimensions are useful for image analysis. Learning the low-dimensional representation of high-dimensional data can reduce the complexity and improve the interpretability of image data [21]. Manifold learning assumes that images are points on a low-dimensional manifold embedded in a high-dimensional space and provide a possibility to obtain the underlying manifold. First, a graph representation of the image data was obtained, where each vertex of the graph represents an image, and each edge represents similarity between the images. Then, an embedded manifold representation was derived from the graph matrix. Manifold learning has been useful in many medical imaging applications [22], such as segmentation [23], registration [24,25], and classification [21,26,27]. In the multiatlas segmentation scenario, manifold learning has been successfully applied in atlas selection, that is, to select atlases that are near the target image on the learning manifold to implement segmentation [28–33]. In their work, the geodesic distances used to construct the graph were based on global image similarities.

Herein, we proposed a multiatlas segmentation scheme that employs manifold learning to determine spatially local weights for label fusion in a low-dimensional coordinate space. A local search strategy was adopted to select a candidate patch for each atlas. Then, the manifold coordinates of local patches were achieved by applying a manifold learning algorithm. Finally, label fusion was conducted in the low-dimensional space. The performance of the proposed method has been evaluated on one in-house dataset and two ADNI datasets using leave-one-out cross-validations. We also investigated the influences of different parameters and then compared the results with two widely used label fusion methods, that is, local majority voting (LdMV, the majority voting fusion strategy based on local patches) and local weighted inverse distance voting (LWINV), which we implemented in the original high-dimensional space. In addition, we implemented hippocampal volumetric analysis and evaluated segmentation performance using atlases from different datasets.

**Materials and Methods**

**Datasets**

We utilized three datasets in this study. First, we applied our method to segment the hippocampus in 28 healthy adolescents. Second, we demonstrated the ability of our method to address pathological variability using two ADNI datasets.

**In-House Dataset**

This dataset contains sagittal T1-weighted MR images of 28 healthy adolescents (15 boys, 13 girls) with an average age of 13.8 years (age range: 10–17 years). All the MR images were acquired using a 3.0T Siemens Trios MRI scanner in the Imaging Center for Brain Research, Beijing Normal University. The imaging parameters were as follows: repetition time [TR] = 2530 ms, echo time [TE] = 3.39 ms, flip angle = 7°, slice thickness = 1 mm, field of view = 256 × 256 mm², matrix size = 256 × 256, and number of slices = 176.

All scans were corrected for intensity inhomogeneity by the N3 algorithm [34] and then aligned to the MNI152 template space using a linear transformation. Subsequently, the hippocampus was manually delineated in stereotaxic space by two blinded raters according to previously described protocols [35]. The manual labeling was performed using ITK-SNAP software (http://www.itksnap.org). A subset of five images was randomly chosen to perform reliability tests. The intrarater reliability in terms of Dice overlap was 0.89. Approximately 3 months later, one rater repeated the hippocampal boundary tracing to access intrarater reliability, which was 0.91.

**ADNI 1.5T Dataset and ADNI 3.0T Dataset**

These images were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu/). The detailed information of ADNI is shown in the Appendix. Many previous studies have used the Medtronic Surgical Navigation Technologies (SNT) semi-automated segmentations provided by ADNI as benchmarks for validation [8,36]. However, over- and underestimation and even misalignments of the entire hippocampal segmentation were found in the SNT segmentations when inspected by an expert manual rater [37]. Thus, we selected all 100 subjects for whom the manual hippocampal masks were provided by EADC project (www.hippocampal-protocol.net). The sample consisted of 58 1.5T and 42 3.0T images, and they were used as two datasets. There were four diagnosis groups in this sample, including normal controls (NC), mild cognitive impairment (MCI), late MCI (LMCI), and patients...
with AD. The demographic information of these subjects is listed in Table 1.

The raw images were oriented along the anterior and posterior commissures line (AC-PC), and then manual hippocampal labeling was performed by expert tracers according to the Harmonized Hippocampal Protocol (http://hippocampal-protocol.net/SOPs/LINK_PAGE/HarmonizedProtocol_ACPC_UserManual_biblio.pdf). The image resolution is 1 × 1 × 1 mm³. For each participant, an N3 bias field correction was used. All MR images were transformed into MNI152 template space by linear registration.

**Local Manifold Learning Method**

The flow chart of local manifold learning (LML)-based segmentation method is schematically illustrated in Figure 1. We first selected a subset of atlases in the MNI152 space and linearly warped the atlases to the target image (based on cropped images around the hippocampus). Subsequently, a local patch search strategy was performed to obtain the most similar patch with the target patch for each atlas. Next, low-dimensional embedding of patches was implemented using manifold learning. Finally, the label of the target voxel was identified by fusing atlas labels in the low-dimensional space.

**Atlas Selection and Registration**

Because using the most similar atlases with the target image $T$ can improve the segmentation accuracy [6], a subset of $n$ atlases ($A_1$, $A_2$, . . . , $A_n$) was selected based on the sum of the squared intensity differences (SSD) between the atlases and the target image as suggested by Aljabar et al. [6]. The selection was performed in the MNI152 template space. Each selected atlas MR image was linearly registered to the target image and then transformed the corresponding hippocampal label image using the same transformation matrix with nearest neighbor interpolation. The image affine registration was performed using the FLIRT tool [38].

**Local Patch Search**

A local patch search strategy was conducted to reduce the influence of registration error as suggested by the literature [19]. We selected the local patch search strategy rather than nonlocal mean patch strategy [17] because the nonlocal method employs all patches within the searching neighborhood and causes a computational burden in the next manifold learning step. A patch was defined as a cube centered on the location in consideration (radius was denoted as $R$). A search volume (SV) was defined as a cube around the target voxel (radius was denoted as $S$). A total number of $(2S + 1)^3$ patches in the SV for each atlas were computed for its similarity with the target patch $P_t$. The intensity vector obtained from each patch was normalized first. Then, the most similar atlas patch $P_a$ was selected for each warped atlas as follows:

$$P_a = \arg\min_{i \in SV} \|I(P_t) - I(P_a(i))\|^2$$

(1)

where $I(x)$ represents the normalized intensity vector of patch $x$.

**Patch-Based Manifold Embedding**

For each voxel of the target image, manifold learning was implemented with an Isomap algorithm [39]. The Isomap is a commonly used nonlinear dimension reduction method, which attempts to estimate the intrinsic geometry of the underlying manifold based on pairwise distances derived from high-dimensional data. In this study, we assumed that a total number of $n + 1$ patches (i.e., the target patch and $n$ atlas patches) were lying on a manifold with dimensionality $d$ embedded in the $D = (2R + 1)^3$ dimension space, where $d \ll D$. The similarity for each pair of patches was coded by the L2 or Euclidean distance derived from the voxel intensities as follows:

$$S(i, j) = \|I(P_i) - I(P_j)\|_2 \leq 1, i, j \leq (n + 1)$$

(2)

where $I(P)$ represents the intensity vector of patch $P$. Then, a neighborhood graph, $G$, was constructed by linking each patch to its $K$ nearest neighbors (kNN). This graph allows an approximate calculation of the geodesic distance $D_G(i, j)$ between all pairs of patches computed as the shortest path distance connecting patches, $i$ and $j$, in the neighborhood graph, $G$. Subsequently, Isomap uses classical multidimensional scaling to obtain the $d$-dimensional coordinates, which is implemented by minimizing the reconstruction error as follows:

$$R(Y) = \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} (D_G(i, j))^2 - \|y_i - y_j\|^2$$

(3)

where $y_i$ is the corresponding newly defined low-dimensional coordinate vector of patch $i$, and $Y$ represents a set of $[y_1, y_2, ..., y_{n+1}]$.  

© 2015 John Wiley & Sons Ltd  
CNS Neuroscience & Therapeutics (2015) 1–11
Segmentation by Label Fusion

Finally, the label fusion step was performed in the low-dimensional space. The weight of each atlas patch was estimated based on its local appearance similarity with the target patch. A label fusion strategy similar to the local weighted inverse distance (LWINV) method [14] was employed as follows:

\[
 w_i = \frac{[|w_i - y_i|^2]^{-\beta_{\text{rev}}}}{Z}, \quad i = 1, 2, \ldots, n
\]

(4)

where \( y_i \) and \( y_i \) are the low-dimensional coordinates of the target patch and atlas patch, respectively, \( \beta_{\text{rev}} \) is a model parameter controlling the weight distribution, and \( Z \) is a normalization constant. While in the LWINV, \( y_i \) and \( y_i \) in this formula represent the intensity vector of the target patch and atlas patch, respectively. Once the weight, \( w_i \), of each atlas patch was determined, the final label of the target voxel was assigned as the one with the maximum voting probability:

\[
 \hat{l} = \arg\max_{l \in \{1, \ldots, L\}} \sum_{i=1}^{n} w_i \cdot P_i(l)
\]

(5)

where \( L \) represents the number of all possible labels (\( L = 2 \) in this work) and \( P_i(l) = 1 \) when the label of the atlas \( i \) is \( l \) (otherwise \( P_i(l) = 0 \)).

Experiments and Results

We chose a widely used evaluation metric, Dice similarity coefficient (DSC), between the manual label, \( S_1 \), and the automatic segmentation, \( S_2 \), to quantitatively assess the performance of segmentation as follows:

\[
 \text{Dice}(S_1, S_2) = 2 \cdot \frac{V(S_1 \cap S_2)}{V(S_1) + V(S_2)}
\]

(6)

where \( S_1 \cap S_2 \) represents the overlapping volume between two segmented regions and \( V(x) \) is the volume of a segmentation \( x \). The Dice index ranges from 0 to 1, and a higher value indicates better segmentation. In the experiments, the leave-one-out cross-validation technique was adopted, and for each target image, we chose the top 20 similar atlases based on SSD [40]. To reduce the computational burden, the images were cropped around the hippocampus. More-
over, the intensities were linearly rescaled to the interval $[0, 100]$. The image selection and registration were based on the cropped images.

**Choosing the Parameters**

The influence of model parameters on segmentation accuracy was studied via leave-one-out cross-validation on the left hippocampus of the in-house dataset according to the parameter tuning method described in [41]. Our method has five free parameters: $K$, the number of neighbors in KNN to construct a neighborhood graph; $d$, the embedding low dimension; $\beta$, the parameter controlling the weight distribution; $S$, the radius of the local searching window; and $R$, the radius of the local patch. The DSC values over varying model parameters are presented in Figure 2. When studying the impact of a certain parameter on segmentation accuracy, the other parameters were held fixed. For example, we first selected the search patch radius $S$ among $\{0, 1, 2, 3, 4, 5\}$, with the other parameters conservatively set as $K = 5$, $d = 3$, $\beta = 5$, and $R = 2$. The performance begins to converge when $S = 3$ or more. Due to the greater time needed for local patch search with larger searching radii, we chose $S = 3$ in this study. Additionally, the search range for other parameters was empirically chosen as $R \in \{1, 2, 3\}$, $K \in \{2, 5, 7, 10\}$, $d \in \{1, 2, 3, 4, 5\}$, and $\beta \in \{1, 2, 3, 4, 5, 6\}$. Finally, we chose a combination of parameters for LML in the following experiments ($S = 3$, $R = 2$, $K = 2$, $d = 3$, and $\beta = 4$).

**Comparison With Other Methods**

Once the optimal parameters were selected, the proposed LML method was used to segment the hippocampus in all three datasets, and its accuracy was evaluated using a leave-one-out procedure. In each cross-validation step, one image was treated as the target image, and the top 20 most similar images selected from the remaining images were used as atlases. This process was repeated until each image was served as the target image once. The LML method was also compared with two state-of-the-art label fusion methods, including LWINV and LMV. The LMV was considered as a baseline automatic label fusion algorithm in this study, which implemented the majority voting fusion strategy on local patches. The LWINV was chosen because it is similar to our LML method described in [41]. Our method has five free parameters: $K$, $d$, $S$, $R$, and $\beta$. The parameter tuning methods would produce spatial bias that undersegments convex structures, such as the hippocampus [42]. However, the segmentations achieved by our method were the closest to manual segmentations. Consistently, a previous study demonstrated that weighted voting strategies would produce spatial bias that undersegments convex structures, such as the hippocampus [42]. However, the segmentation performance of the in-house dataset was better than the two ADNI datasets. Moreover, the performance of the ADNI 1.5T dataset outperformed the ADNI 3.0T dataset. Figure 3 is a 3D view of a representative hippocampal segmentation result (median DSC value) for each dataset to qualitatively assess the voxelwise distribution of label errors between automatic and manual segmentations. Other qualitative results are illustrated in Figure 4, where the sagittal views for the best, median, and worst results based on the ADNI 3.0T dataset for all considered algorithms are presented. The segmentations from LML are more accurate at the boundaries and have fewer hollows.

**Hippocampal Volumetric Analysis**

A hippocampal volumetric analysis was performed based on all ADNI subjects (both 1.5T and 3.0T). Because brains with a large total intracranial volume (TIV) tend to have larger hippocampi, the hippocampal volumes were corrected according to the TIVs estimated by VBM8 (http://dbm.neuro.uni-jena.de/vbm/). This correction can be simply performed by dividing the hippocampal volume by the TIV and then multiplying by the mean of TIVs of all images. The normalized volumes of the left and right hippocampi by groups of all ADNI subjects are presented in Figure 5. The volume measurements produced by automatic methods were all smaller than the reference segmentations. Consistently, a previous study demonstrated that weighted voting strategies would produce spatial bias that undersegments convex structures, such as the hippocampus [42]. However, the segmentations achieved by our method were the closest to manual segmentations. In addition, the Bland–Altman diagrams [43] plotting the hippocampal volume difference between automatic and manual segmentations for ADNI 1.5T and 3.0T datasets are presented in Figure 6. These diagrams indicate that the automatic methods tend to underestimate the hippocampal

![Figure 2](image-url)

**Figure 2** Optimal selection of five model parameters using the leave-one-out cross-validation based on left hippocampal segmentation of an in-house dataset. The optimal combination parameters $S = 3$, $R = 2$, $K = 2$, $d = 3$, and $\beta = 4$ were selected.
volumes, particularly for larger hippocampi. The Bland–Altman plots also demonstrate that our method has the smallest bias (mean difference) in volume estimation.

We also computed Cohen’s $d$ effect sizes to test the ability of each method to detect hippocampal atrophy in the LMCI, MCI, and AD groups relative to controls (Table 3). Larger values of Cohen’s $d$ indicate a greater effect size, where 0.2 indicates a small effect, 0.5 indicates a medium effect, and $>$ 0.8 indicates a large effect [44]. It was computed as follows:

$$d = \frac{\bar{V}_1 + \bar{V}_2}{s}$$  \hspace{1cm} (7)$$

where $\bar{V}_1$ and $\bar{V}_2$ are the mean volumes of two groups, and $s$ represents a pooled standard deviation for the data defined as follows:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}}$$  \hspace{1cm} (8)$$

where $n_i$ and $s_i$ represent the number of subjects and standard deviation of volumes in group $i$. As reported in Table 3, our method yields the largest effect size among the three automatic label fusion methods. The effect sizes characterizing the between-group difference of NC and LMCI are small, and the effect sizes of NC, MCI, NC, and AD are large. Moreover, it can be observed that, in general, the left hippocampus yields a better effect size than the right.

**Performance Evaluation Across Different Datasets**

The proposed method was evaluated across two ADNI datasets acquired with different field strengths. In particular, the 20 most similar images were selected as atlases from the ADNI 3.0T dataset to segment each image in the ADNI 1.5T dataset and vice versa. The results are listed in Table 4. The segmentation accuracy of the ADNI 3.0T images with atlases obtained from ADNI 1.5T images was similar with atlases obtained from its own dataset. However, the

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Left hippocampus</th>
<th>Right hippocampus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dice</td>
<td>$t$-stat</td>
<td>$P$-value</td>
</tr>
<tr>
<td>In-house</td>
<td>LMV</td>
<td>0.879 ± 0.057</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>LWINV</td>
<td>0.896 ± 0.031</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>LML</td>
<td>0.902 ± 0.026</td>
<td>–</td>
</tr>
<tr>
<td>ADNI 1.5T</td>
<td>LMV</td>
<td>0.841 ± 0.083</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>LWINV</td>
<td>0.886 ± 0.040</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>LML</td>
<td>0.891 ± 0.033</td>
<td>–</td>
</tr>
<tr>
<td>ADNI 3.0T</td>
<td>LMV</td>
<td>0.817 ± 0.087</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>LWINV</td>
<td>0.876 ± 0.029</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>LML</td>
<td>0.881 ± 0.022</td>
<td>–</td>
</tr>
</tbody>
</table>

LWINV, local weighted inverse distance voting; LMV, local majority voting; LML, local manifold learning.

**Figure 3** 3D views of hippocampal segmentation results obtained by local manifold learning, local weighted inverse distance voting, and local majority voting for a representative subject with a median Dice index from three datasets. Overlapping areas between manual segmentations (red) and automatic segmentations (green) are shown in blue.
segmentation performance of the ADNI 1.5T images with 3.0T atlases was worse than with the 1.5T atlases. This may be because the ADNI 3.0T dataset had fewer subjects and more variation than the ADNI 1.5T dataset resulting in a larger error when selecting atlases.

**Computational Complexity**

The proposed method was implemented in C++ using the Insight Toolkit (ITK). The experiments were conducted using a single core of an Intel core i7-4500 processor at 1.8 GHz with an 8GB of RAM. After the atlas’s linear registration to the target image, it took approximately 40 seconds to segment one subject with the optimized parameter settings.

**Discussion and Conclusion**

In this work, we present a novel local label fusion strategy for multiatlas segmentation. Instead of assigning voting weights to each atlas in the original space, we utilized a manifold learning technology to build low-dimensional coordinate systems based on local patches and conducted label fusion in this space. The proposed method was applied to segment the whole hippocampus on three datasets, including an in-house dataset and two ADNI data-
sets. We also compared the proposed LML method with previous label fusion methods. The results demonstrated that our method could obtain good and robust segmentation results and achieved better performance than competing methods.

Our work focused on the label fusion step in the multiatlas segmentation framework. Previous label fusion methods based on STAPLE framework [9–12] or ad hoc voting, such as majority voting [13], global [14], local [15,16], and nonlocal [17] weighted label fusion approaches, have demonstrated good segmentation results. The main difference between our method and other label fusion methods is that our label fusion is implemented in the low-dimensional space acquired by manifold learning. In addition, some studies focus on another other step, that is, image registration, in the multiatlas segmentation to reduce the need of very accurate registration, such as atlas selection [6,28], optimization image registration [8,31,45], and local and nonlocal patch searches [17,19]. We utilized the local patch search strategy in this study. We have shown that LML can achieve a significant improvement of label overlaps with manual segmentation in comparison with two representative local label fusion methods, that is, LMV and LWINV. When constructing the graph model in the manifold learning step, not only the relations of the target patch and atlas patches but also the relations between each pair of atlas patches were considered. Therefore, the weights assigned to each atlas patch may potentially consider the dependencies among the atlases, which makes them less susceptible to noise and provides a more robust measure of similarity. Two recent studies have demonstrated that good segmentation performance can be achieved by modeling the joint probability of pairwise atlases to reduce error redundancy produced by atlases in a multiatlas frame [19,20]. We speculated that manifold learning adopted in this study could discover some underlying critical structures of image patches and reduce potential redundant information among them. This might be used to explain why our method outperforms other methods that were conducted in the original high-dimensional space. In addition, it is worth noting the Dice index improvement had an upper bound approximately determined by the inter-/intrarater variability.

Note that it is difficult to directly compare segmentation results across publications due to differences in the imaging protocols, segmentation protocols, and image datasets [46]. Some recent hippocampal segmentation methods have reported DSC values >0.8 [6,47–51]. More recently, several studies have achieved high DSC values ranging from 0.88 to 0.9 [17,19,20,36]. Only two methods achieved the highest DSC values above 0.9. Wang et al. used an error correction learning to improve the performance of multiatlas segmentation and reported a DSC value of 0.908 for 57 NCs [52]. Hao et al. used local support vector machine (SVM) to learn classifiers.

Table 3 Cohen’s d effect sizes between the controls and the other three diagnosis groups from ADNI subjects

<table>
<thead>
<tr>
<th>Method</th>
<th>NC-LMCI</th>
<th>NC-MCI</th>
<th>NC-AD</th>
<th>NC-LMCI</th>
<th>NC-MCI</th>
<th>NC-AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMV</td>
<td>0.0944</td>
<td>0.7017</td>
<td>0.7820</td>
<td>0.1608</td>
<td>0.1563</td>
<td>0.6970</td>
</tr>
<tr>
<td>LWINV</td>
<td>0.2562</td>
<td>1.6584</td>
<td>1.8842</td>
<td>0.3328</td>
<td>0.9808</td>
<td>1.3297</td>
</tr>
<tr>
<td>LML</td>
<td>0.3868</td>
<td>1.7880</td>
<td>1.9097</td>
<td>0.4289</td>
<td>1.1124</td>
<td>1.4134</td>
</tr>
<tr>
<td>Manual</td>
<td>0.4345</td>
<td>1.9081</td>
<td>1.4921</td>
<td>0.5475</td>
<td>1.4340</td>
<td>1.2629</td>
</tr>
</tbody>
</table>

NC, normal control; MCI, mild cognitive impairment; LWINV, local weighted inverse distance voting; LMV, local majority voting; LML, local manifold learning; LMCI, late MCI; AD, Alzheimer’s disease.
for a target image voxel in a featured space and achieved a DSC value of 0.910 for mixed diagnostic groups of 30 ADNI subjects [53]. Our DSC index was approximately 0.9 for normal subjects and above 0.88 for the mixed diagnostic groups of ADNI subjects and demonstrated that our LML method can yield comparable results to recently published methods.

The results on the in-house dataset achieved a higher DSC value than those on both ADNI datasets, which was consistent with a previous finding [36]. This may be caused by several factors. First, the in-house dataset included only normal subjects, while the ADNI datasets included four pathological groups that produced a larger anatomical variability. Second, the protocols for hippocampal segmentation were not similar resulting in a different gold standard for the hippocampus. In addition, the segmentation performances were different between the two ADNI datasets, where the 1.5T dataset performed better than the 3.0T dataset, which could be due to fewer subjects and more varied pathological distribution in the ADNI 3.0T dataset.

Several strategies can be used to improve the segmentation accuracy of the proposed LML method. First, using nonlinear registration instead of linear registration may improve the segmentation performance as mentioned in the literature [18]. However, this will increase the computation time, and we have adopted the local patch search strategy to reduce the influence of registration errors introduced by the linear registration method. Second, the segmentation accuracy may be improved using correction learning to reduce systematic errors produced by our method [52]. Third, the proposed approach may be improved using morphological operations to fill in holes in the segmentation results. Fourth, the use of the nonlocal patch search strategy employing all patches in the SV may obtain more accurate estimations than the local patch search strategy [54]. However, the nonlocal search method may increase the computational cost and also burden the next manifold embedding step.

Future research will focus on the following aspects. It is straightforward to extend our algorithm for segmentation problems with multiple structures simultaneously rather than segment a single structure processed in this study. In the future, we plan to apply LML for hippocampal subfields and whole-brain segmentation. Moreover, we only used the Euclidean distance to reflect the relation of paired patches in the high-dimensional space in this study. We will evaluate the effects of other metrics, such as normalized cross-correlation [55] and mutual information [56], on low-dimensional embedding. Duc et al. [28] have compared the effects of different manifold learning methods, that is, Isomap, Laplacian eigenmaps [57], and locally linear embedding [58], for atlas selection in multi-atlas segmentation and found the locally linear embedding gives the best results. In future work, it would be interesting to compare these manifold learning techniques on the segmentation results under our proposed framework.

Acknowledgments

This work was supported by the National Science Foundation of China (Nos. 81471731 and 81171403). 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, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc., F. Hoffman-La Roche, Schering-Plough, and Synarc, Inc., as well as nonprofit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI 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 was coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data were 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.

Conflict of Interest

The authors declare no conflict of interest.

References


Table 4 Segmentation performances across two ADNI datasets. Dice similarity coefficients (median ± standard deviation) are listed

<table>
<thead>
<tr>
<th></th>
<th>Left hippocampus</th>
<th>Right hippocampus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5T → 3.0T (3.0T → 3.0T)</td>
<td>0.880 ± 0.029 (0.881 ± 0.022)</td>
<td>0.880 ± 0.026 (0.882 ± 0.019)</td>
</tr>
<tr>
<td>3.0T → 1.5T (1.5T → 1.5T)</td>
<td>0.877 ± 0.045 (0.891 ± 0.033)</td>
<td>0.884 ± 0.046 (0.891 ± 0.036)</td>
</tr>
</tbody>
</table>

A → B: Selected atlases from dataset A to segment images of dataset B.
Appendix

The Alzheimer’s Disease Neuroimaging Initiative

The 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 nonprofit organizations as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of...
mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness and decrease the time and cost of clinical trials. 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 coinvestigators from a broad range of academic institutions and private corporations, and the subjects have been recruited from over 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 subjects, but ADNI has been followed by ADNI-GO and ADNI-2. To date, these three protocols have recruited over 1500 adults (ages 55–90) to participate in the research consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.