Alzheimer’s disease detection via automatic 3D caudate nucleus segmentation using coupled dictionary learning with level set formulation

Saif Dawood Salman Al-shaikhli \textsuperscript{a,b,*}, Michael Ying Yang \textsuperscript{c}, Bodo Rosenhahn \textsuperscript{b}

\textsuperscript{a} School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA
\textsuperscript{b} Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany
\textsuperscript{c} ITC - Faculty of Geo-Information Science and Earth Observation, Department of Earth Observation Science, University of Twente, Netherlands

\textbf{ARTICLE INFO}

Article history:
Received 2 March 2016
Received in revised form 7 August 2016
Accepted 9 September 2016

Keywords:
3D segmentation
Caudate nucleus
Alzheimer
Dictionary learning
MRI-T1 medical image

\textbf{ABSTRACT}

Background and objective: This paper presents a novel method for Alzheimer’s disease classification via an automatic 3D caudate nucleus segmentation.

Methods: The proposed method consists of segmentation and classification steps. In the segmentation step, we propose a novel level set cost function. The proposed cost function is constrained by a sparse representation of local image features using a dictionary learning method. We present coupled dictionaries: a feature dictionary of a grayscale brain image and a label dictionary of a caudate nucleus label image. Using online dictionary learning, the coupled dictionaries are learned from the training data. The learned coupled dictionaries are embedded into a level set function. In the classification step, a region-based feature dictionary is built. The region-based feature dictionary is learned from shape features of the segmented caudate in the test image and the region-based feature dictionary.

Results: The experimental results demonstrate the superiority of our method over the state-of-the-art methods by achieving a high segmentation (91.5\%) and classification (92.5\%) accuracy.

Conclusions: In this paper, we find that the study of the caudate nucleus atrophy gives an advantage over the study of whole brain structure atrophy to detect Alzheimer’s disease.

\textcopyright{} 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The caudate nucleus (CN) is a periventricular gray matter structure in the center of the brain. It is a part of the basal ganglia which is responsible for the voluntary movement and memory. It shows up in MRI-T1 modality as a bright gray area. It is lighter compared to the cortical gray matter. Although it has a homogeneous intensity, CN segmentation is considered as one of the most challenging tasks in medical imaging due to its boundary ambiguity and the topological attachment with surrounding gray matter structures at multiple locations [1–3]. Fig. 1 illustrates the difficulties of CN segmentation in MRI-T1 images.

The detection of brain diseases such as Alzheimer’s disease (AD), schizophrenia, and epilepsy via volumetric measurements of subcortical structures in MRI has gained importance [1,2]. Among brain diseases, AD is the most common degenerative disorder of late life. Wang et al. [5] proposed an
interesting method for Alzheimer’s disease detection using support vector machine. Zhang et al. [6–9] proposed four classification methods. The methods are based on using the 3D eigenbrain to find distinguishing regions between normal subjects and AD patients. The method [7] is based on the use of displacement field estimation as a feature to classify normal and AD brains. The method [8] consists of three steps. The first step is a registration step, the second step is a principal component analysis (PCA) step for feature extraction. The third step is the classification step using kernel support vector machine decision tree (kSVM-DT).

AD is usually associated with a progressive cerebral atrophy, which can be revealed using MRI [10,11]. CN atrophy, 9–14% reduction in the CN volume, or CN abnormality in AD patients has been reported in many studies [1,3,10–13]. Many studies have reported that the atrophy of CN leads to motor learning and memory disorders such as Alzheimer’s disease (AD) [10–13]. An increasing number of research focused on automatic techniques for studying the brain region atrophy to detect AD [12–15]. Cuingnet et al. [14] evaluated ten (voxel-based, cortical thickness-based, and hippocampus shape-based) methods for the AD patient classification. The methods in this evaluation are either based on the volumetric analysis of the whole brain (such as gray and white matter, and cerebrospinal fluid), or local brain structures (such as the hippocampus).

In 2007, caudate segmentation competitions (CAUSE07) from MRI data were held in conjunction with MICCAI [16]. Between 2007 and 2010, twenty methods were proposed using different approaches such as an atlas registration [17], a shape prior [18], an active appearance model [19], a level set [20], and a voxel-based classification [21,22]. Moreover, some methods have been proposed for caudate segmentation, which are not published on CAUSE07 competitions. Nain et al. [23] proposed a shape based segmentation method for CN segmentation. This method used multiscale shape representation using spherical wavelets. Xia et al. [24] proposed a knowledge-driven method for CN segmentation. This method considered the lateral ventricle to automatically detect the CN. van Rikxoort et al. [25] proposed a method for CN segmentation. This method is a multi-atlas-based segmentation method.

In recent years a variety of methods have been proposed for CN segmentation. Igual et al. [26] proposed a method for the CN segmentation. This method is based on a multi-atlas registration with a graph-cut method. Jiji et al. [10] proposed a segmentation method of different brain structures. This method is used to diagnose the AD by evaluating the atrophy of the CN and the atrophy of the gray matter, the white matter, and the cerebrospinal fluid.

In 2014, dementia diagnosis competitions were held in conjunction with MICCAI conference [27]. Twenty nine methods were proposed to classify normal, AD, and mild cognitive impairment. These methods have used different classifiers such as support vector machine, neural network, and linear discriminant analysis (LDA). These classifiers are learned from different combinations of features (volume, shape, thickness, intensity) [27]. The best performing method, yielding an accuracy of 63.0%, is based on the texture, volume, thickness, and shape features of 7 brain structures using LDA [27,28].

Many approaches have shown that variational formulation is the most effective method to solve many image segmentation problems, but it needs a well-defined region boundary [29]. Moreover, a voxel-wise dictionary learning and sparse coding methods have been popular and useful tools in medical image segmentation and classification [30,31]. Deshpande et al. [32] used a dictionary learning and sparse coding method to detect multiple sclerosis lesions. Al-shaikhli et al. [30] proposed a global feature-based approach for the brain tumor classification using the dictionary learning. Tong et al. [33] proposed a method to

Fig. 1 – Overlapped boundaries (indicated by yellow ellipses) between the CN and surrounding brain structures in MRI-T1 of the IBSR database [4]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
segment the hippocampus using a discriminative dictionary learning. This method is based on an atlas registration and a label fusion. Manhua et al. [34] proposed a patch-based AD classification method. This method used a sparse coding classifier. Al-shaikhli et al. [31] proposed a voxel-wise coupled dictionary method for a brain tumor segmentation. This method considers the grayscale image data and its associated label image as coupled images because they have the same spatial extension, i.e., this method couples the voxel value of the grayscale image and its associated voxel value in the label image. In contrast to this method, our method couples the feature information in the grayscale image with the label information in the label image, then embeds them in a level set formulation.

Volumetric analysis-based methods of the whole brain have a limitation to classify the intermediate and advanced stages of the AD because, normally, the brain volume is reduced with the age [14]. In order to solve this limitation, and in contrast to the aforementioned methods, we present a novel AD classification method via measuring the volume atrophy of the CN.

The main contribution in this paper represents a novel level set cost function to segment the CN. The proposed level set equation is constrained by the sparse representation of the local image features. The sparse representation of local image features is embedded in the level set equation using a coupled dictionary learning approach. The coupled dictionaries are: a voxel-wise feature dictionary of the gray image data, and a voxel-wise label dictionary of the label image data (the ground truth segmentation of the training data). The segmented CN is used for the AD classification. In the classification step, a region-based feature dictionary is built. This dictionary is learned from CN shape features of both normal and AD subjects. The classification is based on the measure of the sparse representation similarity between the segmented CN in the test image and the region-based learned dictionary.

1.1. Major contribution

The method in this paper is an extension of our previous method presented in References [31,35]. The method in References [31,35] is a voxel-wise coupled dictionary learning method. This method is sensitive to the type of image modality used, i.e., although it achieves a good segmentation accuracy, it has a limitation to select correct foreground and background labels, when the gray level values, between image regions, are quite similar as in the MRI-T1 image modality. To solve this limitation, in this paper, we propose the use of local image texture features instead of voxel values to build the coupled dictionaries. Then, the coupled dictionaries are embedded in a level set formulation.

In the proposed method, the coupling procedure is modified over the method in References [31,35]. We propose a segmentation algorithm by coupling the local image features with the local label information. We embed this prior knowledge into a level set formulation to integrate a novel cost function. This method is applied to detect Alzheimer’s disease (AD) via the CN segmentation. The key idea of the proposed framework is demonstrated in Fig. 2.

The remaining sections are organized as follows. The proposed method is described in detail in Section 2. In Section 3, experimental results are presented. In Section 4, the results are discussed. Finally, this work is concluded in Section 5.

2. Method

The proposed method consists of two steps: the segmentation step and the classification step. In the segmentation step, coupled dictionaries are embedded into a level set equation to integrate a novel cost function for the CN segmentation. The classification step uses the segmentation results to classify the AD patients from normal subjects. The classification step is based on the measure of the similarity between the sparse representation of shape features of the segmented CN and a region-based feature dictionary.

2.1. Coupled dictionaries

The grayscale image data and the corresponding ground truth segmentation have the same spatial extension, and they can be considered as coupled images [31]. Thus, each voxel in the grayscale image refers to the same voxel location in the label image [31,35]. Each voxel in the grayscale image is represented by a single texture feature. This gives an advantage over using the voxel values directly as presented in References [31,35]. From each coupled image, two sets of patches are extracted. The first set is local feature patches. The second set is voxel-wise label patches. These sets of patches are coupled to build two dictionaries: a feature dictionary and a label dictionary. Alg. 14 illustrates the procedure which is considered to couple the feature and label dictionaries.

2.1.1. Patch extraction

The size of all images, which are used in this method, is set to $256 \times 256 \times 170$ using Medical Image Registration Toolbox (MIRT) [36]. Two sets of patches are extracted. The first set is local feature patches of grayscale image data, and the second set is...
Fig. 3 – Schematic illustration of the matrix dimension explains how the patches are extracted and concatenated in $Y_f$ and $Y_l$ matrices of the training samples [31,35].
Algorithm 1 Algorithm for coupled dictionary learning

Input: Given a set of grayscale and label image data.
Input: Extract two sets of associated patches \(\{y_{f_i}\}_{i=1}^{N}\) and \(\{y_{l_i}\}_{i=1}^{N}\), as in Fig. 3.

1: **Initiate** \(D_f^{(0)}\) and \(D_l^{(0)}\), iteration \(\text{iter} = 0\)
2: Compute \(D_f^*\) (Eq. (5)) using online dictionary method.
3: At each iteration: Update \(d_t\), and \(a_i\).
4: Use \(a_i\) to update \(d_{f_i}\) in \(D_f\) according to Eq. 3.
5: Update \(D_f^{*+1} = D_f^*\), and \(D_l^{*+1} = D_l^*\).
6: end if
7: if \(\|Y_l - D_l A_i\|^2 > \epsilon\) then
8: Go to step 2.
9: end if

2.2. Level-set formulation

In this section, we embed the coupled dictionaries, which are explained in Subsection 2.1, in the level set formulation using the piecewise constant approximation of the Mumford–Shah variational model [29]:

\[
E(u, C) = \int_{\Omega} \left( (I(u) - u(v))^2 + \lambda_1 I_1 \right) \quad \text{data term} + \lambda_2 I_2 \quad \text{regularization term}
\]

where \(u : \Omega \rightarrow \mathbb{R}^3\) is a constant approximation of the observed image. \(C\) is the surface of \(u\), and \(\lambda_i\) is a positive real constant. Ignoring the regularization term, the data term is a k-mean clustering problem:

\[
\|f(u) - p_i\| < \|f(u) - p_j\|
\]

where \(p\) is the cluster. Many reports have mentioned that this is an interesting relation between the clustering and sparse representation [39]. In clustering method, there is a set of vectors \(P = \{p_i\}_{i=1}^{k}\) learned from the training data \(Y\), and each training sample is represented by one of the vectors \(p_i\) that is very close to it in \(\ell_2\) distance measure. This notation can be represented as follows:

\[
\min_{Y, X} \|Y - PX\|^2 \quad \text{subjected to} \quad x_i = x_i
\]

where \(X\) is a matrix almost zero except one non-zero element, when \(p_i\) is very close to \(Y\).

In the sparse representation, each training sample is represented as a linear combination of few vectors \(\{d_i\}_{i=1}^{m}\), according to the sparse coefficients. Therefore, sparse representation could also be inferred as a clustering problem in Eq. (8). From the aforementioned notations, the data term in Eq. (6) can be reformulated as a \(\ell_2\) norm by considering the feature dictionary \(D_f\) as follows:

\[
\text{Data} = \left( \min_{A_i} \|Y_f^{*} - D_f A_i\|_2^2 \right)
\]

where \(Y_f^{*}\) represents the feature samples of the test image. Since \(A_i = A_0\), Eq. (9) can be rewritten in term of the sparse coefficients of the label data:

\[
\text{Data} = \left( \min_{A_i} \|Y_l^{*} - D_l A_i\|_1 \right)
\]

The initialization of the level set is sensitive due to the boundary ambiguity of the CN as shown in Fig. 1. This sensitivity may also affect the convergence of the level set, i.e. it converges to a non-optimal local minimum. To solve the initialization problem of the level set, the label dictionary (Eq. (4)) is used. In Eq. (4), \(\tilde{D}\) can be considered as a label state of the CN in the test image. This means that \(\tilde{D}\) is equivalent to the Heaviside function \(H(\phi)\) in the level set equation:

\[
\text{Data} = \tilde{D}(u) \left( \min_{A_i} \|Y_f^{*} - D_f A_i\|_2^2 \right)
\]

Moreover, since the label dictionary \(D\) represents the binary label of the VOI, the sparse representation \(A_i\) can be considered as a regularity term. Therefore, the proposed level set equation can be presented as follows:

\[
E_{\text{data}}(A_i) = \tilde{D}(u) \left( \min_{A_i} \|Y_f^{*} - D_f A_i\|_2^2 \right) + \lambda_1 |C| + \lambda_2 \|A_i\|_1
\]

where \(\lambda\) is the sparsity parameter. In Eq. (12), the data term is constrained by \(D_f\) and \(\tilde{D}\). As we explained in Section 2.1, the atoms in \(D_f\) and \(D_l\) are coupled. When \(\tilde{D} = 1\) (Eq. (4)), \(|A_i|\) has one non-zero element. Thus, the term \(\|Y_f^{*} - D_f A_i\|_2^2\) is optimized. This means that \(Y_f^{*}\) has a best representation in \(D_f\). As mentioned above, according to Eqs. (4) and (12), \(\tilde{D}\) is considered as a function of the foreground (CN) label state. In the regularization term, \(|C|\) represents the surface voxels of the segmented object. It is used as an indicator of the convergence during the evolution process.

2.2.1. Level-set optimization

To optimize (Eq. (12)), the best matching between the sparse representation of the target image data and coupled dictionaries \(D_f\) and \(D_l\) is computed. Once the \(\ell_2\) norm is applied to the regularization term, Eq. (12) becomes difficult to solve. Here, we adopt the concept of iteratively re-weighted (IR) algorithm to handle this problem. We adopt the general idea of an IR algorithm by reformulating the minimization problem in Eq. (12) to the weighted mean square error at the \(t^{th}\) iteration:

\[
W^t = \tilde{D}(u) \left( \|Y_f^{*} - D_f A_i^t\|_2^2 \right)
\]

At each iteration \(t\), \(|C|\) represents the surface of the segmented volume. We consider a weighted difference \(W^t\) of the
average surface distance of the surface $C$ at iterations $t$ and $t-1$:

$$W_t = \lambda_1 |C^t - C^{t-1}|$$  \hspace{1cm} (14)

where $C^t$ and $C^{t-1}$ are the surface voxels of the segmented volume at iterations $t$ and $t-1$ respectively. $W_t$ computes the average distance of all surface voxels between $C^t$ and $C^{t-1}$. Eqs. (13) and (14) are constrained by the parameter $\epsilon$, i.e.

$$|Y^{pred} - D_t A^t| < \epsilon$$

$$\lambda_1 |C^t - C^{t-1}| < \epsilon$$

which means that there is no further change with time (reach the convergence). The same stopping criteria ($\epsilon$) are used for both $W_1$ and $W_2$. $W_1$ and $W_2$ are solved alternately, i.e. fixing $W1$ when solving $W2$ and vice versa. Because $W_1$ determines the reconstruction error, $W_2$ measures the change of the surface. Thus, the weighted mean square error is minimized as follows:

$$W_1 + W_2 < \epsilon$$  \hspace{1cm} (15)

Using the above notations, the sparse coefficients $A_t$ are iteratively optimized:

$$A_t^t \leftarrow A_t^{t-1}$$  \hspace{1cm} (16)

In all experiments, the regularization parameters $\lambda$ and $\lambda_1$ are set to 0.4 and 0.3 respectively.

### 2.3. Classification

The goal of the segmentation stage is to extract the CN volume from the target image. The segmentation results are used to classify the normal and abnormal (AD patients) subjects using the fact that the CN volume in AD patient is reduced compared to the normal subjects. The classification stage is based on measure of the similarity of the region-based sparse representation of image features, as illustrated in Fig. 4. The segmented images (testing data) and the ground truth segmentation of the training data consist of two labels: $l = 1$ for foreground (CN) and $l = 0$ for background. Let $c = 1, 2$ be the class number of normal and AD classes. From the training data of both classes, a respective region-based feature dictionary is built. As illustrated in Fig. 4, we extract six shape features of the CN label from each class (volume, surface area, Euler number, mean breadth, major axis length, and minor axis length) [40].

Let $D_n$ be a region-based feature dictionary $n \times K_n \times z$ matrix $D_n = (d_1, d_2, \ldots, d_K_n)$ which consists of $K_n$ atoms (columns), $\{d_i \in \mathbb{R}^{n \times z} : i = 1, 2, \ldots, K_n\}$ and each atom represents the key features, which are extracted from $Y_n$, where $(K_n \ll N_n)$. $Y_n = (y_1, y_2, \ldots, y_{n_n})$ is a $n_n \times N_n \times z$ matrix which consists of feature matrices $\{y_i \in \mathbb{R}^{n \times z} : i = 1, 2, \ldots, N_n\}$, $n = 6$ is the number of features used to learn $D_n$, and $z = 170$.

The sparse representation $A_n = (a_1, a_2, \ldots, a_{n_n}) \in \mathbb{R}^{K_n \times N_n \times z}$ is computed s.t. $y_i = D_n a_i$ and $|a_i| \ll K_n$, $i = 1, 2, \ldots, N_n$. In such a way that each feature matrix $y_i$ in $Y_n$ is represented by linear combination of a few atoms in the dictionary $D_n$ according to the non-zero elements in $A_n$, as illustrated in Fig. 4. The problem can be formulated as the following minimization:
where $D_s$ is the global feature dictionary. We propose the use of the online dictionary learning method \cite{38} to solve Eq. (17).

To classify the testing data, the algorithm tries to find a match between the shape features $X_s$ of the segmented CN volume in the test image data and the $D_s$. This can be achieved by computing the similarity of the sparse representation of the shape features of the segmented CN in the test image data with the contents (key shape features) of $D_s$. The sparse representation of the test image data is computed using the individual dictionaries of the two classes. If $X_s$ is more sparse with $D_s$, then it is classified as a $c$th class:

$$\min_{A_s} \|Y_s - D_s A_s\|^2 \quad \text{s.t.} \quad \forall 1 \leq i \leq N_s, \quad \|A_s\| < K_s$$

$$D_s = \{D_i\}, \quad c = 1, 2$$

(17)

(18)

where $D_s$ is the global feature dictionary. We propose the use of the online dictionary learning method \cite{38} to solve Eq. (17).

To classify the testing data, the algorithm tries to find a match between the shape features $X_s$ of the segmented CN volume in the test image data and the $D_s$. This can be achieved by computing the similarity of the sparse representation of the shape features of the segmented CN in the test image data with the contents (key shape features) of $D_s$. The sparse representation of the test image data is computed using the individual dictionaries of the two classes. If $X_s$ is more sparse with $D_s$, then it is classified as a $c$th class:

$$\|X_s - D_s A_s\| \leq \epsilon, \quad \|A_s\| = \min \{\|A_s\| : b = 1, 2\}$$

(19)

In Eq. (19), we obtain the sparse representation $A_s$ for $c = 1, 2$. Then, the sample $X_s$ is classified to class $c$, when $X_s$ appears more sparse with respect to $D_s$ (min $A_s$ is selected).

3. Experimental results

3.1. Image database

Three medical databases (MRI-T1 modality) are used, namely, Alzheimer’s disease neuroimaging initiative (ADNI) database (200 AD patients and 160 normal subjects) \cite{41}, internet brain segmentation repository (IBSR) database (20 normal subjects) \cite{4}, and brain web database (20 normal subjects) \cite{42}. The total number of image data is 400 subjects, which were randomly and equally split into 200 subjects of training and 200 subjects of testing sets. ADNI provides a large normal and AD database. The age and gender are considered when we select subjects for training and testing (175 males and 185 females between 40 and 90 years old). The training data and their associated ground truth segmentation (label image) are used to build and learn the feature and label dictionaries.

3.2. Computational time and parameter selection

All experiments are conducted in MATLAB using a 2.0 GHz Intel core i3 CPU. The average computation time per subject is 3.5 minutes. The proposed method is based on measure of the reconstruction error. Therefore, the main parameters that affect the segmentation accuracy are the reconstruction error ($\epsilon$) and the sparsity ($\lambda$) parameter. $\lambda$ is important to control the trade-off between the reconstruction error $\epsilon$ and the sparsity. $\lambda_1$ has less effect than $\lambda$. To optimize the selection of these parameters, many experiments are conducted. We test the value of $\lambda$ and $\lambda_1$ for different values (0.05, 0.1, 0.15, …, 0.9). We find that the best values of these parameters are $\lambda = 0.4$ and $\lambda_1 = 0.3$. The reconstruction error parameter is tested for different values (0.0005, 0.001, …, 0.3). We find $\epsilon = 0.001$ is the best value.

3.3. Qualitative evaluation

Fig. 5 shows the segmentation results of the CN of both normal and AD cases. All segmentation examples are at the same slice level in three different planes. From the results, a significant difference in the shape of the CN was observed between AD and normal cases, i.e. reduction in the CN volume in AD cases. The CN atrophy in AD cases is mainly noticed in the tail part of the CN, which is indicated by white arrows in Fig. 5.

Fig. 6 shows a comparison between the results of the proposed method and the methods \cite{31,35}. The improvement of
the segmentation of the tail part of CN is indicated by the white arrow in the transverse section in Fig. 6.

Fig. 7 shows two CN segmentation examples (the example, on the left, is normal case and, on the right, is AD case). The significant reduction in CN volume can be noted in AD segmentation examples. Furthermore, the ambiguity of the CN boundaries and the gray level similarity between the CN and surrounding brain structures are noted in all examples. The

Fig. 6 – An example of CN segmentation results of the proposed method and the results of the methods [31,35]. The first column is the original image. The second column is the segmentation results of the methods [31,35]. The third column is the segmentation results of the proposed method.

Fig. 7 – Two CN segmentation examples of both normal and AD cases. In each example, the first row shows the input image data in three different planes. The second row is the segmentation results of our method. The third row shows the 3D segmentation of the proposed method. The significant CN volume reduction in the AD case, compared to the normal case especially in the tail part of the CN, can also be seen in the 3D segmentation.
results show the ability of our method to solve this segmentation challenge using the coupled patches via the level set formulation, which makes the CN segmentation results more accurate.

3.4. Quantitative evaluation

For quantitative evaluation, the sensitivity (\(SEN = \frac{TP}{TP+FN}\)), specificity (\(SPE = \frac{TN}{TN+FP}\)), positive predictive value (\(PPV = \frac{TP}{TP+FP}\)), negative predictive value (\(NPV = \frac{TN}{TN+FN}\)), and accuracy (\(AA = \frac{TP+TN}{TP+TN+FP+FN}\)) are calculated by computing the true positive (TP), the true negative (TN), the false positive (FP), and the false negative (FN). Dice coefficient is useful to measure the spatial overlap between the segmentation results and the ground truth segmentation.

Moreover, to emphasize the segmentation step of the proposed method, we use three evaluation metrics: dice coefficient (DSC), the symmetric mean absolute distance (MAD) and Hausdorff distance (HD) [43]. Dice coefficient (\(DSC = \frac{2|A \cap B|}{|A|+|B|}\)) is useful to measure the spatial overlap between the segmentation results and the ground truth segmentation [44]. MAD is calculated by measuring the average distance from all points on the border of the automatically segmented CN to the border of the reference segmentation. HD is calculated between the border of the automatically segmented CN and that of the reference segmentation. HD is useful to assess the maximal local discrepancy between an automatic segmentation and reference segmentation. The smaller the MAD or Hausdorff distance, the better the points aligned on the two borders and thus the better the agreement with the reference segmentation.

**Table 1** shows that the proposed method outperforms the state-of-the-art segmentation methods that are based only on level set [10], dictionary learning [31], and atlas based segmentation with dictionary learning [33]. However, the Hausdorff distance of \(>3\) mm sounds quite large. This number is related to the HD in the tail part of the CN.

In **Table 2**, we compare our method to the state-of-the-art classification methods that used different classification techniques such as computing the relative CN volume with respect to the total brain volume [10], and sparse coding classifier [32,34].

The results of the proposed method in **Tables 1 and 2** represent the average accuracies of 10 runs. In our experiments, firstly, the 400 samples are randomly divided into 200 for training and 200 for testing. Then, these groups are randomly re-selected by changing 20 samples from one group to another. This procedure is repeated 10 times.

The use of many datasets may induce a bias in the segmentation and classification results. To emphasize the robustness of the proposed algorithm, we also use a leave-5-cross-validation for the segmentation and the classification evaluation using only ADNI dataset (100 normal subjects and 100 AD patients). **Tables 1 and 2** show the evaluation results which are computed as an average of 40 runs of randomly selecting the testing data using leave-5-cross-validation. The results of this evaluation are indicated by (§) in **Tables 1 and 2**.

### 4. Discussion

In regard to SEN and SPE of both segmentation and classification results, note that the sensitivity is higher than the specificity, which means that the AD cases are better detected than the normal cases. Moreover, in our method, the

---

**Table 1** — Evaluation of segmentation results of our method compared to the state-of-the-art segmentation methods. These results are the average of 10 runs. In each run, the training and testing sample groups are randomly re-selected by changing 20 samples from one group to another.

<table>
<thead>
<tr>
<th>Method</th>
<th>SEN</th>
<th>SPE</th>
<th>PPV</th>
<th>NPV</th>
<th>AA</th>
<th>MAD</th>
<th>HD</th>
<th>DSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>92.4%</td>
<td>90.4%</td>
<td>91.1%</td>
<td>91.9%</td>
<td>91.5%</td>
<td>0.8 mm</td>
<td>3.5 mm</td>
<td>93.1%</td>
</tr>
<tr>
<td>Proposed</td>
<td>93.7%</td>
<td>90.7%</td>
<td>90.8%</td>
<td>91.0%</td>
<td>90.0%</td>
<td>0.84 mm</td>
<td>3.0 mm</td>
<td>92.8%</td>
</tr>
<tr>
<td>Method [33]</td>
<td>91.3%</td>
<td>89.2%</td>
<td>90.6%</td>
<td>91.0%</td>
<td>90.6%</td>
<td>0.9 mm</td>
<td>3.5 mm</td>
<td>92.8%</td>
</tr>
<tr>
<td>Method [10]</td>
<td>89.0%</td>
<td>87.8%</td>
<td>89.3%</td>
<td>88.4%</td>
<td>88.4%</td>
<td>1.1 mm</td>
<td>4.5 mm</td>
<td>90.7%</td>
</tr>
<tr>
<td>Method [31]</td>
<td>89.0%</td>
<td>87.2%</td>
<td>87.7%</td>
<td>88.3%</td>
<td>87.8%</td>
<td>1.3 mm</td>
<td>4.1 mm</td>
<td>90.1%</td>
</tr>
</tbody>
</table>

Note: The results are obtained using leave-5-cross-validation of ADNI dataset (100 normal subjects and 100 AD patients). Highest values are in bold.

**Table 2** — Evaluation of classification results of our method compared to the state-of-the-art classification methods. These results are the average of 10 runs. In each run, the training and testing sample groups are randomly re-selected by changing 20 samples from one group to another.

<table>
<thead>
<tr>
<th>Method</th>
<th>SEN</th>
<th>SPE</th>
<th>PPV</th>
<th>NPV</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>94.5%</td>
<td>90.5%</td>
<td>90.86%</td>
<td>94.3%</td>
<td>92.5%</td>
</tr>
<tr>
<td>Proposed</td>
<td>94.1%</td>
<td>90.5%</td>
<td>90.86%</td>
<td>94.3%</td>
<td>92.5%</td>
</tr>
<tr>
<td>Method [10]</td>
<td>92.3%</td>
<td>90.7%</td>
<td>90.86%</td>
<td>94.3%</td>
<td>92.5%</td>
</tr>
<tr>
<td>Method [32]</td>
<td>91.5%</td>
<td>88.6%</td>
<td>88.4%</td>
<td>91.2%</td>
<td>89.7%</td>
</tr>
<tr>
<td>Method [34]</td>
<td>89.0%</td>
<td>90.0%</td>
<td>89.9%</td>
<td>89.1%</td>
<td>89.5%</td>
</tr>
</tbody>
</table>

Note: The results are obtained using leave-5-cross-validation of ADNI dataset (100 normal subjects and 100 AD patients). Highest values are in bold.
5. Conclusion

This paper presents a method for AD detection via an automatic 3D CN segmentation using a coupled dictionary learning with a level set formulation. The data and regularization terms of the level set equation are integrated in a novel manner. In the segmentation step, two types of dictionaries are coupled: the feature dictionary, which represents the texture features of image patches, and the label dictionary, which represents the voxel-wise CN label image patches. In the classification step, the region-based shape features of both normal and AD cases are used to learn the dictionary. The test image, after the CN segmentation, is classified as normal or abnormal (AD) by computing the sparse similarity of the shape features of the segmented CN with the contents (shape features) of the dictionary. The experimental results show that our method outperforms the state-of-the-art methods by achieving a segmentation and classification accuracy of 91.5% and 92.5% respectively.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2016.09.007.

References

[13] G. Bartzokis, T.A. Tishler, MRI evaluation of basal ganglia ferritin iron and neuro-toxicity in Alzheimer’s and


