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

Saif Dawood Salman Al-shaikhli [a,b,*], *Michael Ying Yang* [c], *Bodo Rosenhahn* [b]

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

ARTICLE INFO

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 caudate nucleus in the training data. The classification is based on the measure of the similarity between the sparse representation of region-based 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.

© 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

---

* *Corresponding author.* Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany. Fax: +49 511 762-5333.
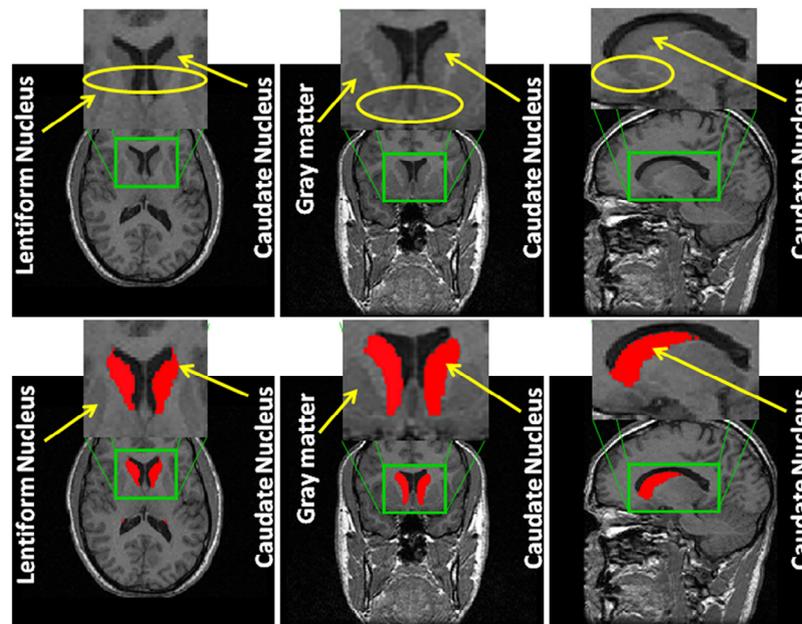 E-mail address: shaikhli@tnt.uni-hannover.de (S.D.S. Al-shaikhli).

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.)

interesting method for Alzheimer's disease detection using support vector machine. Zhang et al. [6–9] proposed four classification methods. The methods [6,9] 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
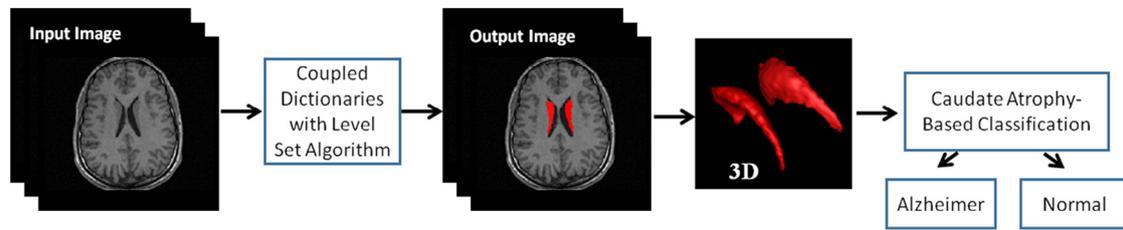
Fig. 2 – The key idea of the proposed framework. The caudate nucleus in the test image volume is segmented. Then, based on the volume reduction of the segmented caudate nucleus, the subject is classified as normal or Alzheimer.

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
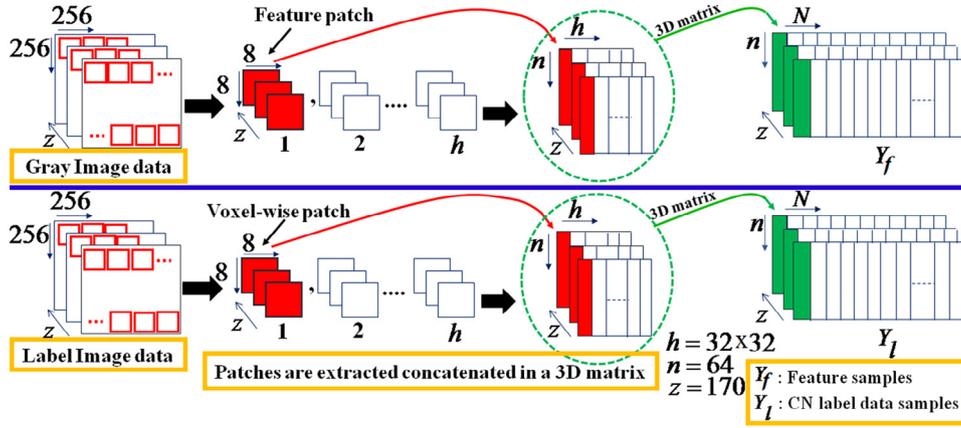
Fig. 3 – Schematic illustration of the matrix dimension explains how the patches are extracted and concatenated in Y_f and Y_l matrices of the training samples [31,35].

label patches of label image data. From each slice of image data, $h = 32^2$ patches with size of $8 \times 8 \times 1$ are extracted, as illustrated in Fig. 3.

Due to the difficulties to compute texture features of a single voxel v, we represent a texture feature of the voxel v as a texture feature of a 26-neighborhood window with the center of the voxel v. The contrast texture feature is computed using a gray level co-occurrence matrix (GLCM) [37]. The GLCM is an important method for textural feature extraction. It represents the histogram of co-occurring gray level values at a given offset over the image data. The number of rows and columns in the GLCM represents the number of gray levels in the image.

The contrast feature measures the change in the intensity value of voxels in the image, which provides better information than directly using the voxel values.

The feature patches of each image data are concatenated as a matrix with size $64 \times h \times 170$. The voxel-wise label patches of each label image data are concatenated in a matrix with size $64 \times h \times 170$. Then, both of feature and label patches are concatenated in Y_f and Y_l matrices respectively. The size of Y_f and Y_l is $n \times N \times z$, where $n = 64$, $N = 200$, and $z = 170$. Fig. 3 illustrates how the feature and label patches are extracted and concatenated in a 3D matrix.

2.1.2. Dictionary learning

Let D_f be a feature dictionary $n \times K \times z$ matrix $D_f = (d_1, d_2, \ldots, d_K)$, which consists of K atoms (columns), $\{d_i \in R^{n \times h \times z} : i = 1, \ldots, K\}$. D_f is learned from the feature sample matrix Y_f, where $Y_f = (y_1, y_2, \ldots, y_N)$ is a matrix with size $n \times N \times z$, where $\{y_i \in R^{n \times h \times z} : i = 1, \ldots, N\}$ of N feature samples and $(K \ll N)$, as illustrated in Fig. 3. The sparse representation $A_f = (a_1, a_2, \ldots, a_N) \in R^{K \times N \times z}$ is computed s.t. $\|y_i - D_f a_i\|_2^2 \le \epsilon$ and $\|a_i\|_0 \ll K, i = 1, \ldots, N$, ϵ is a small value. Thereby, each feature sample of the training data is represented by a linear combination of few atoms in the dictionary D_f according to the non-zero element of A_f.

To build the label dictionary, let D_l be a label (voxel-wise) dictionary matrix $D_l \in R^{n \times K \times z}$. D_l is built and learned from the CN ground truth segmentation in the same manner as explained above (see Fig. 3). The above notations can be formulated as follows:

$$\arg\min_{D_f, A_f} \|Y_f - D_f A_f\|_F^2 \quad s.t. \quad \forall 1 \le i \le N, \quad \|a_i\|_0 \ll K \tag{1}$$

and

$$\arg\min_{D_l, A_l} \|Y_l - D_l A_l\|_F^2 \quad s.t. \quad \forall 1 \le i \le N, \quad \|a_i\|_0 \ll K \tag{2}$$

In Eqs. (1) and (2), the sparse coefficients A_f and A_l are assumed to be equal in order to achieve the coupling between the dictionaries D_f and D_l. Thus, under this assumption, the sparse coefficients A_l could be used to encode the feature patches in D_f (see Alg. 1):

$$\arg\min_{D_f} \|Y_f - D_f A_l\|_F^2 \quad s.t. \quad \forall 1 \le i \le N, \quad A_f = A_l \tag{3}$$

In the label image, $l = 1$ for the CN label and $l = 0$ for the background. Let v be a set of voxels in the label atom. In a perfect representation of the CN label in D_l, each atom in D_l is represented by voxels which have a maximum probability for the foreground label (CN label) and a minimum probability for the background label:

$$\hat{D}_l(v) = \begin{cases} 1 & \text{if} \quad l_1 = max_l\, d_i(v) \\ 0 & \text{if} \quad background \end{cases} \tag{4}$$

To optimize the requirement of containing one label information in each voxel, the optimized label dictionary D_l^* is computed by minimizing the following equation:

$$D_l^* = \arg\min_{D_l, A_l} \left\| \hat{D}_l - D_l A_l \right\|_F^2 \tag{5}$$

Alg. 1 illustrates the procedure to build the coupled dictionaries. The dictionary learning procedure is done simultaneously for both D_f and D_l to approximate a solution of the error matrix. Thus, each atom d_f in D_f has a coupled atom d_l in D_l, i.e. the label information in D_l can be inferred on the feature information in D_f. For the update dictionary and sparse coding steps, we propose the use of the online dictionary learning method [38]. This method is robust and efficient when the training data are large and it is suitable for dynamically changing data.

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 $iter = 0$
2: Compute D_l^* (Eq. (5)) using online dictionary method.
3: At each iteration: Update d_{l_i} and a_{l_i}.
4: Use a_{l_i} to update d_{f_i} in D_f according to Eq. 3.
5: Update $D_f^{t+1} = D_f^t$, and $D_l^{t+1} = D_l^t$.
6: **end if**
7: **if** ($\|Y_l - D_l A_l\|_F^2 > \epsilon$) **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]:

$$E(u, C) = \underbrace{\int_\Omega (I(v) - u(v))^2 \, dx}_{\text{data term}} + \underbrace{\lambda_1 |C|}_{\text{regularization term}} \tag{6}$$

where $u : \Omega \to \mathcal{R}^3$ is a constant approximation of the observed image I. C is the surface of u, and λ_1 is a positive real constant. Ignoring the regularization term, the data term is a k-mean clustering problem:

$$|I(v) - p_1| < |I(v) - p_2| \tag{7}$$

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 ℓ_2 distance measure. This notation can be represented as follows:

$$\min_{P, X} \|Y - PX\|_2^2 \quad subjected\ to\ x_i = x_k \tag{8}$$

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 $\{d_i\}_{i=1}^K$ 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 ℓ_2 norm by considering the feature dictionary D_f as follows:

$$Data = \left(\min_{A_f} \|Y_f^{test} - D_f A_f\|_2^2\right) \tag{9}$$

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

$$Data = \left(\min_{A_l} \|Y_f^{test} - D_f A_l\|_2^2\right) \tag{10}$$

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), $\hat{D}_l$ can be considered as a label state of the CN in the test image. This means that $\hat{D}_l$ is equivalent to the Heaviside function $H(\varphi)$ in the level set equation:

$$Data = \underbrace{\hat{D}_l(v)\left(\min_{A_l} \|Y_f^{test} - D_f A_l\|_2^2\right)}_{\text{Date term}} \tag{11}$$

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

$$E_{total}(A_l) = \underbrace{\hat{D}_l(v)\left(\min_{A_l} \|Y_f^{test} - D_f A_l\|_2^2\right)}_{\text{Date term}} + \underbrace{\lambda_1 |C| + \lambda \|A_l\|_1}_{\text{Regularization term}} \tag{12}$$

where λ is the sparsity parameter. In Eq. (12), the data term is constrained by D_f and $\hat{D}_l$. As we explained in Section 2.1, the atoms in D_f and D_l are coupled. When $\hat{D}_l = 1$ (Eq. (4)), $\|A_l\|_1$ has one non-zero element. Thus, the term $\left(\min_{A_f} \|Y_f^{test} - D_f A_f\|_2^2\right)$ is optimized. This means that Y_f^{test} has a best representation in D_f. As mentioned above, according to Eqs. (4) and (12), $\hat{D}_l$ 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 ℓ_1 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 tth iteration:

$$W_1^t = \hat{D}_l(v) \|Y_f^{test} - D_f A_l^{t-1}\|_2^2 \tag{13}$$

At each iteration t, $|C|$ represents the surface of the segmented volume. We consider a weighted difference W_2^t of the
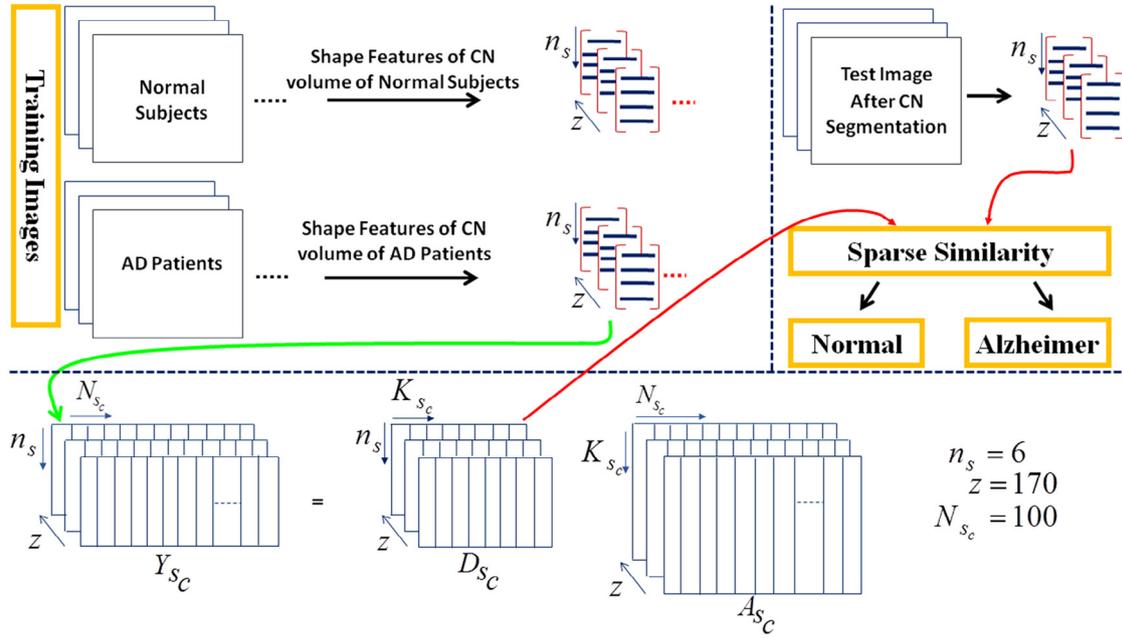
Fig. 4 – The classification step of our method. From the sets of the training data (normal and AD sets) shape features are extracted. These features are concatenated in Y_{s_c} and represented in D_{s_c} according to the sparse coefficients A_{s_c}. The classification of the test image is based on the sparse similarity between the shape features of the segmented CN in test image and D_{s_c}.

average surface distance of the surface C at iterations t and $t-1$:

$$W_2^t = \lambda_1 |C^t - C^{t-1}| \qquad (14)$$

where C^t and C^{t-1} are the surface voxels of the segmented volume at iterations t and $t-1$ respectively. W_2^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 ϵ, i.e.

$$\left\| Y_f^{test} - D_f A_i^{t-1} \right\|_2^2 < \epsilon$$

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

which means that there is no further change with time (reach the convergence). The same stopping criteria (ϵ) 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:

$$W_1 + W_2 < \epsilon \qquad (15)$$

Using the above notations, the sparse coefficients A_l are iteratively optimized:

$$A_i^t \leftarrow A_i^{t-1} \qquad (16)$$

In all experiments, the regularization parameters λ and λ_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_{s_c} be a region-based feature dictionary $n_s \times K_{s_c} \times z$ matrix $D_{s_c} = (d_1, d_2, \ldots, d_{K_{s_c}})$ which consists of K_{s_c} atoms (columns), $\{d_i \in R^{n_s \times z} : i = 1, 2, \ldots, K_{s_c}\}$ and each atom represents the key features, which are extracted from Y_c, where $(K_{s_c} \ll N_{s_c})$. $Y_c = (y_1, y_2, \ldots, y_{N_{s_c}})$ is a $n_s \times N_{s_c} \times z$ matrix which consists of feature matrices $\{y_i \in R^{n_s \times z} : i = 1, 2, \ldots, N_{s_c}\}$. $n = 6$ is the number of features used to learn D_{s_c}, and $z = 170$.

The sparse representation $A_{s_c} = (a_1, a_2, \ldots, a_{N_{s_c}}) \in R^{K_{s_c} \times N_{s_c} \times z}$ is computed s.t. $y_i = D_{s_c} a_i$ and $\|a_i\|_0 \ll K_{s_c}, i = 1, \ldots, N_{s_c}$. In such a way that each feature matrix y_i in Y_c is represented by linear combination of a few atoms in the dictionary D_{s_c} according to the non-zero elements in A_{s_c}, as illustrated in Fig. 4. The problem can be formulated as the following minimization:
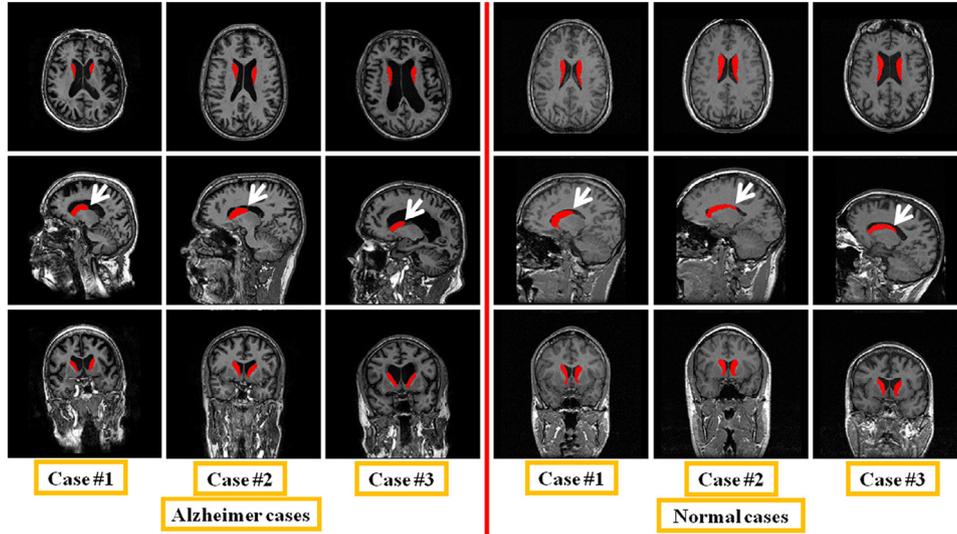
Fig. 5 – CN segmentation examples of both AD and normal cases. Each column represents three different planes of one case. In the second row, the white arrows indicate the atrophy of the tail part of the CN in AD cases compared to the normal.

$$\min_{D_{sc}, A_{sc}} \|Y_{sc} - D_{sc} A_{sc}\|_F^2 \quad s.t. \quad \forall 1 \le i \le N_{sc}, \quad \|a_i\|_0 \ll K_{sc} \tag{17}$$

$$D_s = \{D_{sc}\}, \quad c = 1, 2 \tag{18}$$

where D_s is the global feature dictionary. We propose the use of the online dictionary learning method [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_{sc}. 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_{sc}. 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 $D_{s_{c^{th}}}$, then it is classified as a cth class:

$$\|X_s - D_{sc} A_{sc}\|_F^2 \le \epsilon, \quad \|A_{sc}\|_0 = min\{\|A_{s_b}\|_0 : b = 1, 2\} \tag{19}$$

In Eq. (19), we obtain the sparse representation A_{sc} for $c = 1$, 2. Then, the sample X_s is classified to class c^{th}, when X_s appears more sparse with respect to $D_{s^{th}}$ (min A_{sc} 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) [41], internet brain segmentation repository (IBSR) database (20 normal subjects) [4], and brain web database (20 normal subjects) [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 (ϵ) and the sparsity (λ) parameter. λ is important to control the trade-off between the reconstruction error ϵ and the sparsity. λ_1 has less effect than λ. To optimize the selection of these parameters, many experiments are conducted. We test the value of λ and λ_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 on 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 [31,35]. The improvement of
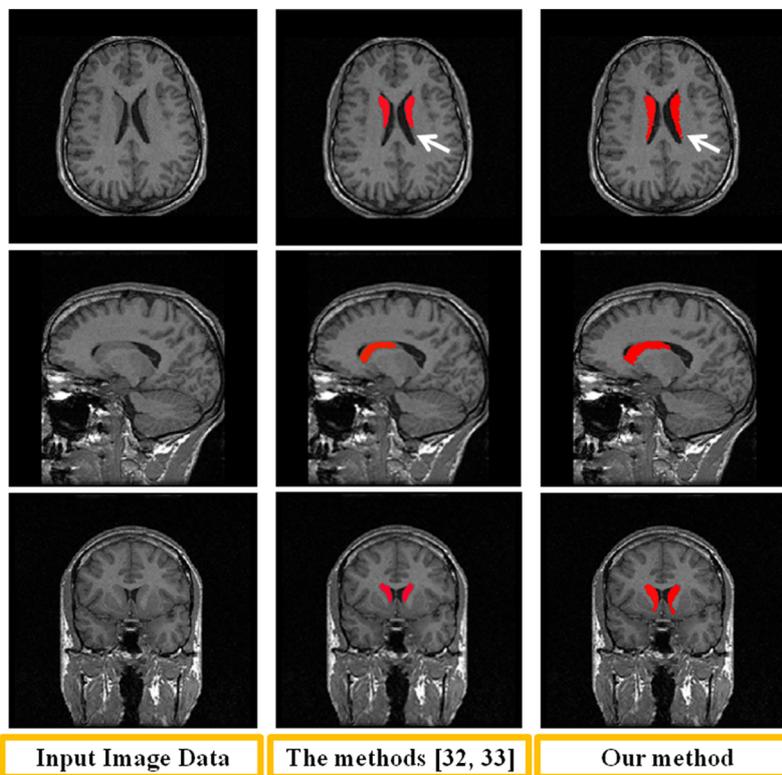
Fig. 6 – An example of CN segmentation results of the proposed method and the results of the methods [31,35]. The first column is the original image. The second column is the segmentation results of the methods [31,35]. The third column is the segmentation results of the proposed method.

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
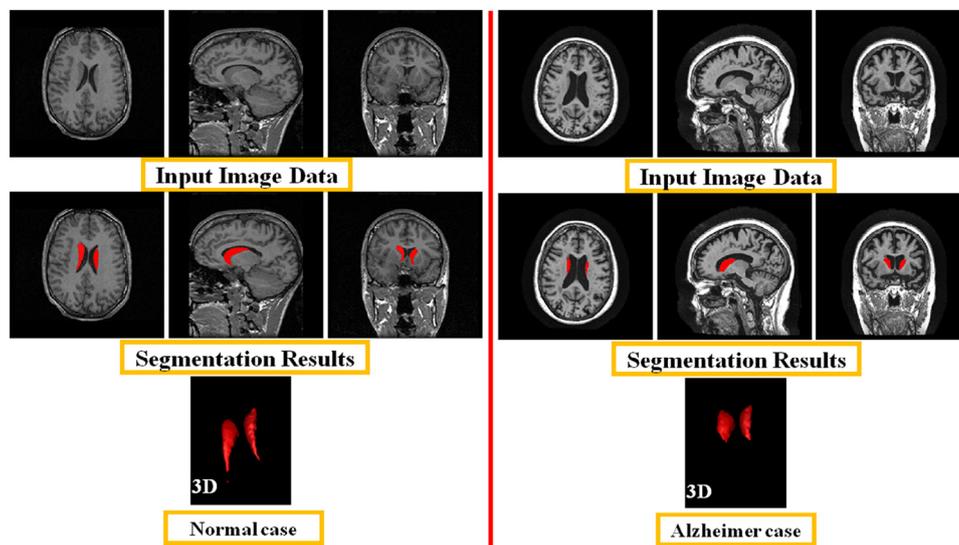

Fig. 7 – Two CN segmentation examples of both normal and AD cases. In each example, the first row shows the input image data in three different planes. The second row is the segmentation results of our method. The third row shows the 3D segmentation of the proposed method. The significant CN volume reduction 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 ($SEN = TP/(TP + FN)$), specificity ($SPE = TN/(TN + FP)$), positive predictive value ($PPV = TP/(TP + FP)$), negative predictive value ($NPV = TN/(TN + FN)$), and accuracy ($AA = TP + TN/(TP + TN + FP + FN)$) 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 ($DSC = 2(A \cap B)/(absolute(A) + absolute(B))$) 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.

Method	SEN	SPE	PPV	NPV	AA	MAD	HD	DSC
Proposed	**92.4%**	90.4%	91.1%	**91.9%**	**91.5%**	**0.8 mm**	3.5 mm	**93.1%**
[§]Proposed	92.0%	**90.8%**	**91.3%**	91.3%	91.0%	0.84 mm	**3.0 mm**	92.8%
Method [33]	91.3%	89.2%	90.6%	91.0%	90.6%	0.9 mm	**3.5 mm**	92.8%
Method [10]	90.0%	88.2%	87.7%	89.1%	88.6%	0.96 mm	3.8 mm	92.0%
Method [31]	90.0%	87.8%	87.6%	89.3%	88.4%	1.1 mm	4.5 mm	90.7%
Method [26]	88.9%	87.2%	87.7%	88.3%	87.8%	1.3 mm	4.1 mm	90.1%

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.

Method	SEN	SPE	PPV	NPV	AA
Proposed	**94.5%**	90.5%	90.86%	**94.3%**	**92.5%**
[§]Proposed	**93.7%**	90.7%	90.0%	**93.8%**	**92.1%**
Method [10]	90.0%	**92.5%**	**92.3%**	90.24%	91.25%
Method [32]	91.5%	88.0%	88.4%	91.2%	89.75%
Method [34]	89.0%	90.0%	89.9%	89.1%	89.5%

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.

classification step is based on using individual dictionaries (a dictionary for each class). The results are improved compared to method that is based on measuring the ratio between different brain structures [10], which implies that not only this classifier performs better than the rest, but also that the shape features of the segmented CN provide more discriminative information than the ratio itself, improving the AD diagnostic accuracy. The SPE and PPV of the method [10] are better than ours, since this method is a semi-automatic method.

It should be noted that our method is able to recognize the mild cognitive impairment. Because the proposed method is based on the measure of the CN atrophy which is sometimes difficult to recognize in mild cognitive impairment cases, the proposed method recognizes the mild cognitive impairment with lower accuracy than AD cases. From the results, we found that the atrophy of CN in all images is ranged from 12% to 20%. The atrophy mostly occurred in the tail part rather than in the body of the CN.

Compared to the methods [31,35], from the results in Table 2, we observe that the coupling of image features and label information gives an improvement, in terms of segmentation accuracy, compared to the coupling of the voxel values of real and label image information. Moreover, embedding the coupled dictionaries in the level set formulation gives an advantage that its results are smoother than the results of dictionary-learning-based segmentation methods [31,33].

Compared to the method presented by Zhang et al. [7], which is based on the estimation of the displacement field between normal and AD brains, this method requires a registration step, which is considered as time-consuming step. Moreover, the use of the dictionary learning method gives better sparse representation of image features than the use of PCA.

In Figs. 5 and 7, behind the reduction of CN volume, there is a significant enlargement in the ventricles in AD patients. Although this enlargement may be considered as useful information to detect AD, it adds a complexity to the segmentation step due to the gray level similarity between the ventricles and the background in MRI-T1 modality. In other words, the ventricle tends to be dark region in MRI-T1 modality. Because the level set formulation has the property of topological changes, embedding the sparse representation of image information, via dictionary learning approach, into the level set formulation gives an advantage of making the segmentation results smoother than the results obtained using voxel-based approaches.

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

[1] R. Barber, I. McKeith, C. Ballard, J. O'Brien, Volumetric MRI study of the caudate nucleus in patients with dementia with Lewy bodies, Alzheimer's disease, and vascular dementia, J. Neurol. Neurosurg. Psychiatry 72 (3) (2002) 406–407.

[2] O.P. Almeida, E.J. Burton, I. McKeith, A. Gholkar, D. Burn, J.T. O'Brien, MRI study of caudate nucleus volume in Parkinson's disease with and without dementia with Lewy bodies and Alzheimer's disease, Dement. Geriatr. Cogn. Disord. 16 (2) (2003) 57–63.

[3] J.C. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette, et al., In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease, Neuroimage 14 (2) (2001) 298–309.

[4] Internet brain segmentation repository. http://www.nitrc.org/projects/ibsr/.

[5] S. Wang, Y. Zhang, G. Liu, P. Phillips, T.-F. Yuan, Detection of Alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging, J. Alzheimers Dis. 50 (1) (2016) 233–248.

[6] Y. Zhang, S. Wang, P. Phillips, J. Yang, T.-F. Yuan, Three-dimensional eigenbrain for the detection of subjects and brain regions related with Alzheimer's disease, J. Alzheimers Dis. 50 (4) (2016) 1163–1179.

[7] Y. Zhang, S. Wang, Detection of Alzheimer's disease by displacement field and machine learning, PeerJ 3 (2015) e1251, http://dx.doi.org/10.7717/peerj.1251.

[8] Y. Zhang, S. Wang, Z. Dong, Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree, Prog. Electromagn. Res. 144 (2014) 171–184.

[9] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, et al., Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning, Front. Comput. Neurosci. 9 (66) (2015) http://dx.doi.org/10.3389/fncom.2015.00066.

[10] J. Jiji, K.A. Smitha, A.K. Gupta, V.P.M. Pillai, R.S. Jayasree, Segmentation and volumetric analysis of the caudate nucleus in Alzheimer's disease, Eur. J. Radiol. 82 (2013) 1525–1530.

[11] S.K. Madsen, A.J. Ho, X. Hua, P.S. Saharan, A.W. Toga, C.R. Jack Jr., et al., 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects, Neurobiol. Aging 31 (8) (2010) 1312–1325.

[12] S.A. Rombouts, F. Barkhof, M.P. Witter, P. Scheltens, Unbiased whole-brain analysis of gray matter loss in Alzheimer's disease, Neurosci. Lett. 285 (3) (2000) 231–233.

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

Huntington's disease, Cell. Mol. Biol. (Noisy-Le-Grand) 46 (4) (2000) 821–833.

[14] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.O. Habert, et al., Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database, Neuroimage 56 (2) (2011) 766–781.

[15] M.W. Weiner, P.S. Aisen, C.R. Jack, W.J. Jagust, J.Q. Trojanowski, L. Shaw, The Alzheimer's disease neuroimaging initiative: progress report and future plans, Alzheimers Dement. 6 (3) (2010) 202–211, e7.

[16] Caudate segmentation challenge (CAUSE07). http://mbi.dkfz -heidelberg.de/grand-challenge2007/sites/proceed.htm, 2007.

[17] S. Gouttard, M. Styner, S. Joshi, R.G. Smith, H.C. Hazlett, G. Gerig, Subcortical structure segmentation using probabilistic atlas priors. MIICAI Workshops, pp. 37–46, 2007.

[18] J.H. Levy, K. Gorczowski, X. Liu, S.M. Pizer, M. Styner, Caudate segmentation using deformable M-reps. MIICAI Workshops, pp. 47–55, 2007.

[19] K.O. Babalola, V. Petrovic, T.F. Cootes, C.J. Taylor, C.J. Twining, T.G. Williams, et al., Automatic segmentation of the caudate nuclei using active appearance models. MIICAI Workshops, pp. 57–64, 2007.

[20] J. Liu, C. Smith, H. Chebrolu, Automatic subcortical structure segmentation using local likelihood-based active contour. MIICAI Workshops, pp.91–98, 2007.

[21] M. Wels, M. Huber, J. Hornegger, Fully automated knowledge-based segmentation of caudate nuclei in 3D MRI. MIICAI Workshops, pp. 19–27, 2007.

[22] Y. Arzhaeva, E.V. Rikxoort, B.V. Ginneken, Automated segmentation of caudate nucleus in MR brain images with voxel classification. MIICAI Workshops, pp. 65–72, 2007.

[23] D. Nain, S. Haker, A. Bobick, A. Tannenbaum, Shape-driven 3D segmentation using spherical wavelets, Med. Image Comput. Comput. Assist. Interv. 9 (2006) 66–74.

[24] Y. Xia, K. Bettinger, L. Shen, A.L. Reiss, Automatic segmentation of the caudate nucleus from human brain MR images, IEEE Trans. Med. Imaging 26 (4) (2007).

[25] E.M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M.A. Viergever, et al., Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus, Med. Image Anal. 14 (2010) 39–49.

[26] L. Igual, J.C. Soliva, R. Gimeno, S. Escalera, O. Vilarroya, P. Radeva, Automatic internal segmentation of caudate nucleus for diagnosis of attention-deficit/hyperactivity disorder, in: A. Campilho, M. Kamel (Eds.), 9th International Conference, ICIAR 2012, Aveiro, Portugal, June 25-27, 2012. Proceedings, Part II, vol. 7325, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, pp. 222–229.

[27] E.E. Bron, M. Smits, W.M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, et al., Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge, Neuroimage 111 (2015) 562–579, http://dx.doi.org/10.1016/ j.neuroimage.2015.01.048.

[28] L. Sø rensen, A. Pai, C. Anker, I. Balas, M. Lillholm, C. Igel, et al., Dementia diagnosis using MRI cortical thickness, shape, texture, and volumetry. MICCAI-CAD Dementia Workshop, pp.111–118, 2014.

[29] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Commun. Pure Appl. Math. 42 (1989) 577–685.

[30] S.D.S. Al-Shaikhli, M.Y. Yang, B. Rosenhahn, Brain tumor classification using sparse coding and dictionary learning, in: Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 2752–2756.

[31] S.D.S. Al-Shaikhli, M.Y. Yang, B. Rosenhahn, Coupled dictionary learning for automatic multi-label brain tumor segmentation in Flair MRI images, in: Advances in Visual Computing Part I, vol. 8887, Lecture Notes in Computer Science, 2014, pp. 489–500.

[32] H. Deshpande, P. Maurel, C. Barillot, Detection of multiple sclerosis lesions using sparse representations and dictionary learning. STMI-MICCAI, 2014.

[33] T. Tong, R. Wolz, P. Coupé, J.V. Hajnal, D. Rueckert, Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling, Neuroimage 76 (2013) 11–23.

[34] L. Manhua, Z. Daoqiang, S. Dinggang, Ensemble sparse classification of Alzheimer's disease, Neuroimage 60 (2) (2012) 1106–1116.

[35] S.D.S. Al-Shaikhli, M.Y. Yang, B. Rosenhahn, Brain tumor classification and segmentation using sparse coding and dictionary learning, Biomed. Tech. (Berl) 61 (4) (2015) 413– 429, http://dx.doi.org/10.1515/bmt-2015-0071 in press.

[36] A. Myronenko, MIRT—Medical Image Registration Toolbox for Matlab. https://sites.google.com/site/myronenko/ research/mirt, 2010.

[37] R.M. Haralick, K. Shanmugam, I. Dinstein, Texture feature for image classification, IEEE Trans. Syst. Man Cybern. 3 (6) (1973) 610–621.

[38] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: ICML, 2009, pp. 689–696.

[39] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.

[40] D. Legland, K. Kiêu, M.-F. Devaux, Computation of Minkowski measures on 2D and 3D binary images, Image Anal. Stereol. 26 (2007) 83–92.

[41] Alzheimer's disease neuroimaging initiative. http://adni .loni.usc.edu/.

[42] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, A.C. Evans, BrainWeb: online interface to a 3D MRI simulated brain database, Neuroimage 5 (4) (1997) S425.

[43] J. Wang, F. Li, Q. Li, Automated segmentation of lungs with severe interstitial lung disease in CT, Med. Phys. 36 (10) (2009) 4592–4599.

[44] K.H. Zou, S.K. Warfield, A. Bharatha, C.M.C. Tempany, M.R. Kaus, S.J. Haker, et al., Statistical validation of image segmentation quality based on a spatial overlap index, Acad. Radiol. 11 (2004) 178–189.