Segmentation of Brain MR Images via Sparse Patch Representation

Tong Tong¹, Robin Wolz¹, Joseph V. Hajnal², and Daniel Rueckert¹

¹ Department of Computing, Imperial College London, London, UK
² MRC Clinical Sciences Center, Imperial College London, London, UK
t.tong11@imperial.ac.uk

Abstract. Recently, patch-based segmentation has been proposed for brain MR images. However, the segmentation accuracy of this method depends on similarities over small image patches, which may not be an optimal estimator. In this paper, we propose a new segmentation strategy based on patch reconstruction rather than patch similarity. In the proposed method, the training patch library is considered as a dictionary, and the target patch is modeled as a sparse linear combination of the atoms in the dictionary. The sparse representation is naturally discriminative, which presents an entirely data-driven approach to patch-selection and label definition. This Sparse Representation Classification (SRC) strategy produces segmentation results that compare favourably to existing approaches. In addition, a smoothing term is added to the cost function of the sparse coding technique, making the proposed method more robust. To the best of our knowledge, the sparse representation technique has never been used in brain segmentation. In a leave-one-out validation, the proposed method yields a median Dice coefficient of 0.871 for hippocampus on 202 ADNI images, which is competitive compared with state-of-the-art methods.

1 Introduction

Magnetic resonance imaging (MRI) is the primary imaging modality for the analysis of brain structures. It enables us to describe how brain anatomy changes during aging or disease progression. For the extraction of biomarkers for diseases like Alzheimer’s Disease (AD) or schizophrenia, the accurate and robust segmentation of subcortical structures is an essential step. Since manual labeling by clinical experts is a highly laborious task, an automated technique is desirable to allow a routine analysis of brain MRIs in clinical use. However, it still remains a challenging task to develop fast and accurate automated segmentation methods due to the complexity of subcortical structures.

Several automated methods have been reported to extract subcortical structures. Among them, atlas-based methods have been shown to outperform other state-of-the-art algorithms [1]. To avoid bias by using a single atlas, several similar atlases can be used to improve the segmentation performance [2, 3]. However, the segmentation performance of multi-atlas techniques is directly affected by
the selection of similar atlases, the quality of registration methods and the label fusion strategy. In addition, multi-atlas segmentation requires pairwise, accurate registrations between atlas and target, which can result in a significant computational burden.

Recently, a nonlocal patch-based segmentation technique has been proposed [4–6]. Instead of fusing propagated label maps as in multi-atlas segmentation, this method obtains a label for every voxel by using its surrounding patches in coarsely aligned atlas images. First, similar patches are selected in the predefined neighborhood across the training atlases. Then weights are given to these patches according to the distances between the target patch and the selected patches. The final label of the target voxel is estimated by fusing the labels of the central voxels in the patch library using a nonlocal label fusion strategy. Such a technique does not require a dense one-to-one correspondence between the target image and candidate atlases, and a validation on hippocampus segmentation [4, 5] demonstrates a high accuracy of this approach.

Although this patch-based technique achieves good performance, this method highly depends on the similarity of small image patches extracted from each atlas, which may be affected by intensity variations [3]. In addition, the kernel width that transforms patch-distances into weights crucially influences segmentation outcome and strongly depends on the heterogeneity of the target dataset. Moreover, the selected patches can be very similar to each other and can contain redundant information. In this paper, instead of using a nonlocal means fusion strategy, we proposed to use a novel classification strategy via sparse representation by taking advantage of redundant information in the patch library. In the proposed method, the target patch will be represented by the training patches in a sparse way. The sparse representation is naturally discriminative [7], which presents an entirely data-driven approach to patch-selection and label assignment. In addition, considering the number of training patches is much bigger than the number of patches selected for representation, we added a smoothing term to the cost function of the sparse coding method to obtain more robust sparse representations. Our method was applied to hippocampus segmentation on the ADNI dataset [8] and the effect of sparsity was studied. Finally we compared the performance of the proposed method with that of the nonlocal patch-based technique.

2 Methods

2.1 Background: Nonlocal Means Patch-based Segmentation

Construction of Patch Library: First, atlas selection is carried out for every target subject based on the sum of squared intensity differences (SSD) across the initialization mask in a template space [4, 5]. Then for every voxel in the target image, the surrounding patch is extracted. Next, similar patches are selected in a defined cubic area across a set of similar atlases based on the structure similarity measure (SSIM)[4,5,9]. If the value of SSIM is bigger than a given
threshold, the patch is selected. On average, the final patch library typically contains thousands of patches after the preselection process.

Nonlocal Means Label Fusion: Suppose that the patch library contains \( n \) patches. Each patch in the library is given a weight based on the intensity distance to the target patch. The weighting function is defined as follows [4, 5]:

\[
w(x_i, x_j) = e^{-\frac{||p(x_i) - p(x_j)||^2}{h^2}}
\]  

(1)

where \( p(x_i) \) is the surrounding patch of the target voxel \( x_i \) and \( p(x_j) \) is the patch in the library. \( ||.||^2 \) is the normalized L2 norm for computing the distance between \( p(x_i) \) and \( p(x_j) \). \( h^2 \) is the smoothing parameter, which has been set to a global, heuristically defined value [6] or the minimal distance between the target patch \( p(x_i) \) and the patch library [4]. Then the estimation for the final label of voxel \( x_i \) is based on the nonlocal means fusion strategy:

\[
v(x_i) = \frac{\sum_{j=1}^{n} w(x_i, x_j)y_j}{\sum_{j=1}^{n} w(x_i, x_j)}
\]  

(2)

where \( y_j \) is the label for the central voxel of patch \( p(x_j) \) in the library, and the value \( v(x_i) \) belongs to \( \{0, 1\} \). Voxels are considered as foreground if the value of \( v(x_i) \) is bigger than 0.5, otherwise as background.

From Equations (1) and (2), we can see that the weighting function is an important factor in the nonlocal patch-based segmentation technique as it determines the label fusion scheme. The smoothing parameter \( h^2 \) plays a crucial role because it determines how this method assigns different weights for different patches. This smoothing parameter is challenging to set globally, especially on heterogeneous datasets. When \( h^2 \) is low, only a few patches contribute to the final labeling. In this case, the labels of the most similar patches are fused in a similarly sparse way. When \( h^2 \) is high, all the patches in the library tend to have similar weights and are used to estimate the final label, which is close to a classical average. However, the estimation using the final average of a large number of patches is less robust and may lead to mislabellings [4].

2.2 Sparse Representation Classification (SRC)

The basic assumption of non-local means patch-based segmentation is that the central voxels of similar patches are considered to belong to the same structure [4]. Therefore, this methods highly depends on the similarity of small image patches extracted from each atlas, which may be affected by intensity variations [3]. In addition, there exists a large amount of redundant information in the patch library. Since patches in the library can be very similar to each other, a small number of patches may be sufficient for estimating the final label of the target voxel. Inspired by work in face recognition [7], we propose to use a sparse representation classification strategy for patch selection and weighting.

The sparse representation technique has been successfully applied to different problems in face recognition [7, 10, 11]. This technique abandons the conventional idea to compare the similarity between patches in a neighborhood. Instead,
the algorithm calculates a representation of the target patch by considering
the patch library as a dictionary. Moreover, this method imposes a constraint
that the optimal representation should use the smallest number of training
patches, which means that the representation is sparse. With a patch library
\( P_L = [p_1, p_2, \cdots, p_n] \in \mathbb{R}^{m \times n} \), the target patch \( p_i \) can be considered as a linear combination of the training patches in the library \( P_L \):

\[
p_i = a_1p_1 + a_2p_2 + \cdots + a_np_n
\]

(3)

where most of the coefficients \( a_j \) are zero. Let \( A = [a_1, a_2, \cdots, a_n] \in \mathbb{R}^n \), then the coding coefficients can be obtained by solving the equation \( p_i = P_LA \). This linear system \( p_i = P_LA \) is underdetermined since \( n > m \), so this equation does not have a unique solution. In general, the sparse solution of this equation is equal to the solution of the following \( L1 \)-minimization problem:

\[
\hat{A} = \min_A \|A\|_1 \quad \text{subject to} \quad \|p_i - P_LA\|_2^2 \leq \varepsilon
\]

(4)

Equation (4) can be solved efficiently by several sparse coding methods [12].

The main idea of SRC is that the target patch \( p_i \) can be reproduced by the
patches from the same class. Therefore, the labeling of the target voxel is done
by comparing which class of training patches gives the minimal reconstruction
error. In our case, the patches associated with non-zero coefficients are divided
into two groups: patches belonging to the hippocampus and patches belonging
to the background. If the patches belonging to the hippocampus can represent
the target patch \( p_i \) with a smaller reconstruction error, then the target voxel is
labelled as hippocampus, and vice versa.

In [7], Equation (4) was solved by using the Lasso method for obtaining
sparse representations. However, when the number of predictors (\( n \)) is much
bigger than the number of observations (\( k \)), the Lasso method cannot produce a
very satisfactory variable selection [13]. In this \( n \gg k \) case, the Elastic Net (EN)
always outperforms Lasso [13]. Considering that the number of patches in the
library is much bigger than the number of patches selected for representation,
our case belongs to this ‘large \( n \) small \( k \)’ problem. To achieve robust sparse
representations, we used EN [13] for obtaining the sparse coding coefficients:

\[
\hat{A} = \min_A \frac{1}{2}\|p_i - P_LA\|_2^2 + \lambda_1\|A\|_1 + \frac{\lambda_2}{2}\|A\|_2^2
\]

(5)

Equation (5) adds a coefficient magnitude penalty to the objective function in
Equation (4), which is a convex combination of \( L1 \) lasso and \( L2 \) ridge penalties.
EN encourages a grouping effect while keeping a similar sparsity of representation
[13]. This grouping effect, which select groups of highly correlated variables,
is helpful for the final classification and could thus improve the segmentation
performance.

The major difference between our method and the nonlocal means patch-
based method is that the proposed method is based on reconstruction while the
patch-based technique is based on similarity. Also, all the patches in the library
have weights by using the patch-based technique while only a small number of patches have weights by using the proposed method.

3 Experiments and Results

Images used in our work were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). The dataset we used consists of 202 subjects obtained from different scanners (68 normal subjects, 93 subjects with Mild Cognitive Impairment (MCI) and 41 patients with AD). These 202 images were selected because their reference segmentations are available. A commercially available high dimensional brain mapping tool (Medtronic Surgical Navigation Technologies, Louisville, CO) was used to carry out semi-automated hippocampal volumetry for defining these reference segmentations. The mean age of all subjects is $75.7 \pm 6.5$ and the average Mini Mental State Examination (MMSE) score is $27.03 \pm 2.63$. All images were preprocessed by the ADNI pipeline described in [14]. All subjects were then linearly registered to the MNI152 template space by using an affine registration. Image intensities were normalized by using the method in [15]. After that, intensities were rescaled to $[0 \ 100]$. We used the union of all labels followed by a dilation as the initial mask for segmentation. All images were segmented in a leave-one-out procedure. 10 similar subjects were selected by comparing the SSD across the initial mask in the MNI152 template space for all experiments. The search volume for constructing the patch library was set to $9 \times 9 \times 9$ voxels. To solve Equation (5), $\lambda_1$ and $\lambda_2$ were set to $0.15$ for obtaining the sparse representation. These two parameters were determined via cross validation following the parameter settings described in [16].

Considering our reference segmentations of the hippocampus are defined in native space, interpolation artefacts from transforming the labels into MNI152 template space may lead to poor label overlaps. So we affinely transformed 10 atlases and labels to the native space of the query image to perform the segmentation in target coordinate system. With all atlas-template transformations $T_{AT}$ and target-to-template transformation $T_{QT}$, the selected atlases can be transformed to the native space of the target image by using the transformation $T_{AQ} = T_{AT} \circ T_{QT}^{-1}$. Finally, all the experiments were evaluated by computing the Dice coefficient between reference segmentations and automated segmentations.

3.1 Effect of Sparsity

The influence of sparsity on the segmentation performance was studied. The selection of these patches was performed in the iterative optimization process of sparse coding, where a defined number of selected patches was used as a stopping criteria. After sparse coding, only these patches would have non-zero weights and thus be used. Figure 1 demonstrates the Dice coefficient on the ADNI dataset when using different numbers of selected patches for representation and classification. The best results were obtained with 80 patches selected for classification and the median Dice coefficient was 0.868. These results indicate that a small
number of patches in the neighborhood is sufficient to obtain an accurate segmentation. One can note that our method can achieve a median Dice value of more than 0.8 by only using 5 patches from the library for classification.

![Figure 1](image.png)

**Fig. 1.** Effect of sparsity on segmentation accuracy. The results were obtained by using a patch size of $5 \times 5 \times 5$ voxels and a search volume of $9 \times 9 \times 9$ voxels, extracted from the 10 most similar atlases.

### 3.2 Comparison with the Patch-based Method

The proposed method was compared with the nonlocal patch-based technique proposed in [4]. Hippocampus segmentation was performed on the ADNI dataset in the native image space. For a fair comparison, the nonlocal patch-based method was carried out in the same settings (a patch size of $7 \times 7 \times 7$ voxels and a search volume of $9 \times 9 \times 9$ voxels) as described in [4], except that 10 atlases were selected for both methods as this produces comparable results with a significantly lower computational burden. For SRC, 80 patches were selected for representation and classification. Table 1 presents the segmentation accuracy of these two approaches. The nonlocal patch-based method obtained a median Dice coefficient of 0.844 for hippocampus segmentation. By comparison, the median Dice coefficient was 0.871 by using the proposed SRC method, which demonstrates an improved segmentation performance of the proposed method.

Figure 2 provides a visual comparison of the segmentation results in native space.

### 4 Conclusion

In this work, we proposed an alternative strategy to the nonlocal mean-s patch-based segmentation technique. By taking advantage of local variability
<table>
<thead>
<tr>
<th>Method</th>
<th>right hippocampus</th>
<th>left hippocampus</th>
<th>whole hippocampus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patch-based</td>
<td>0.848(0.032)</td>
<td>0.842(0.029)</td>
<td>0.844(0.027)</td>
</tr>
<tr>
<td>SRC</td>
<td>0.873(0.027)</td>
<td>0.869(0.026)</td>
<td>0.871(0.022)</td>
</tr>
</tbody>
</table>

Table 1. Median Dice overlaps for 202 subjects. The results were obtained by using a patch size of $7 \times 7 \times 7$ voxels and a search volume of $9 \times 9 \times 9$ voxels, extracted from the 10 most similar atlases.

![Best subject](image1)  $k = 0.9058$  $k = 0.8767$

![Median subject](image2)  $k = 0.8721$  $k = 0.8233$

![Worst subject](image3)  $k = 0.7509$  $k = 0.7088$

Reference Segmentations
| SRC | Patch-based |

**Fig. 2.** Method comparison. Segmentation results were obtained by the patch-based and the SRC methods for the subjects with the best, a median and the worst Dice coefficients.

and redundant information in the neighborhood, we used a sparse representation-based classification strategy instead of using similarity between patches. A smoothing term was also added to our cost function, resulting in more accurate segmentation results. The validation on the ADNI dataset and the comparison with the patch-based method demonstrate a high accuracy of the proposed method. Although Coupé [4] reported a median Dice coefficient of 0.884, their experiments were limited to 80 healthy subjects imaged on the same scanner. When segmenting 10 subjects with Alzheimer’s disease, the nonlocal patch-based method yielded much lower segmentation accuracy with a median Dice coefficient of 0.838 [5]. By comparison, our proposed method still yields a median Dice coefficient of 0.871 on 202 ADNI images. In future work we plan to apply the proposed method to the measurement of hippocampal atrophy and patient classification proposed in [5].
References