Hippocampus Segmentation Through Distance Field Fusion

Shumao Pang, Zhentai Lu, Wei Yang, Yao Wu, Zixiao Lu, Liming Zhong, and Qianjin Feng(✉)

School of Biomedical Engineering, Southern Medical University, Guangdong, China
qianjinfeng08@gmail.com

Abstract. In this paper, we propose a novel automatic hippocampus segmentation framework called distance field fusion (DFF). We perform a distance transformation for each label image of training subjects and obtain a distance field (DF), which contains the shape prior information of hippocampus. Magnetic resonance (MR) and DF dictionaries are constructed for each voxel in the testing MR image. We combine the MR dictionary through local linear representation to present the testing sample and the DF dictionary by using the corresponding coefficients derived from local linear representation to predict the DF of the testing sample. We predict the label of testing images through threshold processing for DF. The proposed method was evaluated on brain images derived from the MICCAI 2013 challenge dataset of 35 subjects. The proposed method is proven effective and yields mean Dice similarity coefficients of 0.8822 and 0.8710 for the right and left hippocampi, respectively.

Keywords: Hippocampus segmentation · Distance field

1 Introduction

Hippocampus, a brain structure situated in the temporal lobe, is associated with learning and memory. Comprehensive atrophy of brain, particularly in hippocampus and medial temporal lobe, is the pathological characteristic of Alzheimer's disease [1]. For accurate disease diagnosis and prognosis, the size and shape of hippocampus should be compared between healthy and diseased subjects. Magnetic resonance imaging (MRI) is a method used to observe brain structures and delineate hippocampus to calculate its volume. Therefore, accurate and reliable segmentation of hippocampus in MR images is important for clinical applications; this technique has attracted significant attention with the widespread use of MRI.

Various methods have been proposed for hippocampus segmentation. Atlas-based methods exhibit superior performance over other state-of-the-art techniques [2]. In atlas-based methods, an anatomical image from an atlas is registered to the target image to be segmented; the target image is then segmented by warping the manual label of the atlas to the target image space via the deformation field derived from the registration procedure [3]. Biased segmentation is likely to occur when using only one atlas but can be reduced using multiple atlases. In multi-atlas segmentation methods, we register the

G. Wu et al. (Eds.): Patch-MI 2015, LNCS 9467, pp. 104–111, 2015.
DOI: 10.1007/978-3-319-28194-0_13

anatomical image intensity of each atlas to the target image to be segmented; the manual label of each atlas is then warped to the target image space via the corresponding deformation field derived from the registration procedure. A fused label is generated by combining the warped labels from all atlases and regarded as the segmentation of the target image [3].

Atlas-based methods present two limitations. On one hand, segmentation accuracy in these methods is dependent on the registration procedure. On the other hand, segmentation accuracy can be improved with development of accurate and reliable label-fusion techniques, which exhibit an inherent constraint in segmentation accuracy because they do not utilize the shape prior information of hippocampus. To overcome these limitations, we propose a method in the present study.

In this study, a patch-based method is developed for hippocampus segmentation in MR brain images. Considering that the distance field (DF) of the training dataset contains sufficient shape prior information, we combine DFs, rather than the labels of the dataset. The proposed method was evaluated on 35 subjects, including 20 training subjects and 15 testing subjects. Results show that the proposed distance field fusion (DFF) is a promising technique for hippocampus segmentation.

2 Methods

2.1 Basic Idea of DFF

A segmentation problem can be described as follows. Given a training dataset, $T = \left\{x_i^{MR}, x_i^{DF}\right\}_{i=1}^{N}$, which consists of N MR/DF image patch pairs, we calculate the patch x^{DF} of a testing MR image patch x^{MR}, which is sampled from the MR image centered at point x. The proposed method is based on two assumptions:

Assumption I*: Image patches from MR image and DF are located on different nonlinear manifolds, and a patch can be approximately represented as a linear combination of several nearest neighbors from its manifolds.*

Assumption II: *Under a local constraint, the mapping from MR manifold to DF manifold f:* $M^{MR} \rightarrow M^{DF}$ *approximates a diffeomorphism.*

Assumption I has been verified in several studies [4, 5]. In this paper, the manifolds of MR and DF patches are denoted as M^{MR} and M^{DF}, respectively, and M^{MR} and M^{DF} are assumed to be spanned by the patches in the training dataset T. To simplify the calculation, we transform the patch x^{MR} to a column vector, $\overrightarrow{x^{MR}}$. The column vector $\overrightarrow{x^{MR}}$ can be linearly represented as follows:

$$\begin{aligned} \overrightarrow{x^{MR}} &= D^{MR}\overrightarrow{w} + \varepsilon = \sum_{i=1}^{n} \overrightarrow{x_l^{MR}} w_i + \varepsilon \\ & s.t.\ \|\varepsilon\| < \tau \\ \forall\, \overrightarrow{x_l^{MR}} &\notin N_k\left(\overrightarrow{x^{MR}}\right),\ w_i = 0 \end{aligned} \tag{1}$$

where $D^{MR} = \left[\overrightarrow{x_1^{MR}}, \overrightarrow{x_2^{MR}}, \ldots \overrightarrow{x_n^{MR}}\right]$ is a matrix called MR dictionary. The column vectors of D^{MR} are derived from the training dataset T. $\overrightarrow{w} = [w_1, w_2, \ldots, w_n]^T$ is a column vector, whose elements are the coefficients of linear combination. ε is the reconstruction error of the sample x^{MR} and is lower than a small non-negative number τ. $N_k\left(\overrightarrow{x^{MR}}\right)$ is a set that consists of k-nearest neighbors (i.e., red circles filled with yellow in Fig. 2) of the sample $\overrightarrow{x^{MR}}$ in the dictionary D^{MR}.

We use f to denote the mapping between M^{MR} and M^{DF}. According to assumption II, f is locally linear; thus, the DF patch x^{DF} of the testing MR patch x^{MR} can be calculated as follows:

$$\begin{aligned}\overrightarrow{x^{DF}} &= f\left(\overrightarrow{x^{MR}}\right) \approx f\left(\sum\nolimits_{i=1}^{n} \overrightarrow{x_l^{MR}} w_i\right) = \sum\nolimits_{i=1}^{n} f\left(\overrightarrow{x_l^{MR}}\right) w_i \\ &= \sum\nolimits_{i=1}^{n} \overrightarrow{x_l^{DF}} w_i = D^{DF}\overrightarrow{w}\end{aligned} \tag{2}$$

where $\overrightarrow{x^{DF}}$ is the DF corresponding to the testing sample $\overrightarrow{x^{MR}}$, $D^{DF} = \left[\overrightarrow{x_1^{DF}}, \overrightarrow{x_2^{DF}}, \ldots, \overrightarrow{x_n^{DF}}\right]$ is a matrix called DF dictionary, and $\overrightarrow{x_l^{DF}}$ is the DF of $\overrightarrow{x_l^{MR}}$; thus, the DF of point x is defined as:

$$M(x) = \begin{cases} -dist(x, B), & x \text{ outside } C \\ dist(x, B), & otherwise \end{cases} \tag{3}$$

where C is the boundary of the hippocampus, B is the nearest point from x, and $B \in C$. $dist(x, B)$ denotes the distance between x and B.

The proposed method contains three main parts: pre-processing (including registration, distance transformation, etc.), fusion (including local dictionary construction, local linear representation, and DF prediction), and threshold segmentation. The framework of the proposed method and detailed procedures of fusion in DFF are shown in Figs. 1 and 2, respectively.

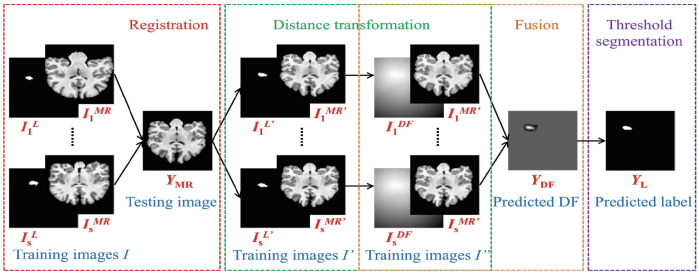

Fig. 1. The framework of the proposed DFF

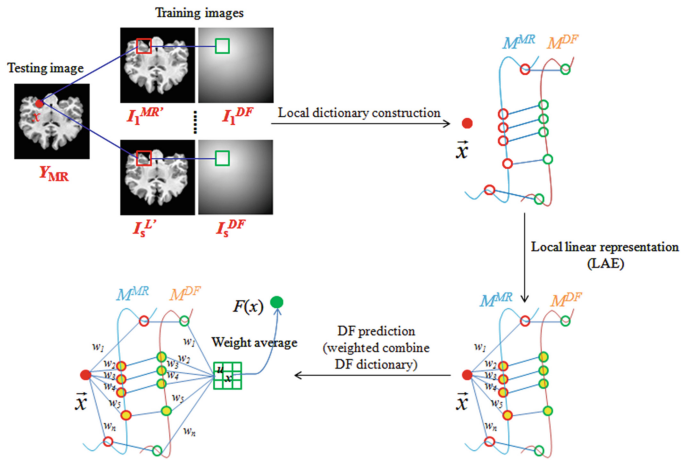

Fig. 2. The detailed procedures of fusion in DFF (Color figure online)

Local Dictionary Construction. We use $I'' = \left\{ \left(I_j^{MR'}, I_j^{DF} \right) \mid j = 1, 2, \ldots, s \right\}$ to denote the pre-processed training dataset, which is registered to the testing image and undergoes distance transformation procedure; in the formula, s denotes the number of training subjects and I_j^{DF} is the DF corresponding to the MR image $I_j^{MR'}$. For a testing image Y_{MR}, we aim to construct local dictionaries for each point x.

For a point x on the testing subject, we extract patch pairs x_i^{MR}/x_i^{DF} on the training dataset I'' in the search window centered at point x. We vectorize patches x_i^{MR} and x_i^{DF} to column vectors $\overrightarrow{x_l^{MR}} \in R^m$ and $\overrightarrow{x_l^{DF}} \in R^m$, respectively. We collect vectors $\overrightarrow{x_l^{MR}}$ and $\overrightarrow{x_l^{DF}}$ and to build dictionaries $D^{MR} = \left[\overrightarrow{x_1^{MR}}, \overrightarrow{x_2^{MR}}, \ldots, \overrightarrow{x_n^{MR}} \right] \in R^{m \times n}$ and $D^{DF} = \left[\overrightarrow{x_1^{DF}}, \overrightarrow{x_2^{DF}}, \ldots, \overrightarrow{x_n^{DF}} \right] \in R^{m \times n}$, respectively.

Local Linear Representation. Several methods have been proposed to linearly represent a testing sample by combining training samples. Sparse coding with L1 (least absolute shrinkage and selection operator) [6] emphasizes the sparsity of coefficients. This coding represents a testing sample with the least training samples and minimum construction error. By contrast, local linear classification (LCC) [7] accentuates locality, rather than sparsity. LCC represents a testing sample by using several training samples located in a local region, where the training samples are similar to the testing sample. Compared with LLC, local anchor embedding (LAE) [8] combines a non-negative constraint to coefficients. Therefore, in the current paper, we solve the linear representation problem using LAE because the testing sample in the proposed method is represented by convex combination of its closest neighbors.

DF Prediction. For a testing sample $\overrightarrow{x^{MR}}$, we predict the DF $\overrightarrow{x^{DF}}$ via Eq. (2). Each testing sample $\overrightarrow{x^{MR}}$ has a predicted DF $\overrightarrow{x^{DF}}$, and several overlapping points exist for different testing samples. Hence, the DF of the point x for a testing sample $\overrightarrow{x^{MR}}$ is determined using several predicted DFs near the point x. As such, a weighted-average strategy is introduced. First, we reshape the vector $\overrightarrow{x^{DF}} \in R^m$ to a patch $x^{DF} \in R^{a\times b\times c}$, where $m = a \times b \times c$. The weight for a point u in the region $P(x)$, which denotes the area centered at point x, with size similar to that of patch x^{DF}, is defined as follows:

$$w_u^x = \frac{1}{m} \tag{4}$$

Second, we obtain the average DF in the region $P(x)$ to predict DF for point x by using the following formula:

$$F(x) = \frac{\sum_{u\in P(x)} w_u^x I_u^x}{\sum_{u\in P(x)} w_u^x} \tag{5}$$

where I_u^x denotes the DF of x predicted by patch x^{MR} centered at the point u using Eq. (2).

2.2 Threshold Segmentation

Basing on the definition of DF in Eq. (3), we can predict the label for a point x by using a threshold:

$$L(x) = \begin{cases} 1, & F(x) > 0 \\ 0, & F(x) \le 0 \end{cases} \tag{6}$$

where $F(x)$ is the predicted DF of the point x in Eq. (5). After predicting the label of each voxel, we obtain a binary image and the point with the value of one belongs to hippocampus.

3 Experiment Results

A dataset was obtained from MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications ("SATA")[1]. We applied the proposed method to the subdataset with 35 subjects, including 20 subjects as the training dataset and 15 subjects as the testing dataset. The BET approach [9] was used to remove the skull in the MR images, and the N4 algorithm [10] was utilized to remove bias field artifacts from the images. All training MR images were non-rigidly registered to a testing MR image

[1] https://masi.vuse.vanderbilt.edu/workshop2013/.

via different deformation fields, which were applied to all corresponding label training images. The images were registered through DRAMMS [11], and the registered label training images were transformed to DFs via Eq. (3). The parameters were set as follows: a patch size of 5 × 5 × 5, a search window size of 7 × 7 × 7 for constructing dictionaries, and 30 nearest neighbors in LAE.

We used dice similarity coefficient (DSC), a common evaluation measure, to quantify the effectiveness of the proposed method for hippocampus segmentation and compare with other state-of-the-art segmentation algorithms. All experiments were conducted with MATLAB.

3.1 Influence of Distance Field Fusion Compared with Label Fusion

Two sets of experiments were performed to verify the effectiveness of DFF. First, we used the proposed DFF method. Second, after obtaining the coefficient vector $\vec{w}$ of the local linear representation via Eq. (1), we combined the label corresponding to the center point of the element of the dictionary D^{MR}. Table 1 shows the mean DSC values of 15 testing subjects by using DFFs; these values were improved by 1.21 % and 1.11 % for the right and left hippocampi, respectively. The results show that DFF can improve the accuracy of hippocampus segmentation.

Table 1. Mean and standard deviation of Dice similar coefficient obtained using LF and DFF.

Subject\Method	Label fusion(LF)	Distance field fusion(DFF)
Right hippocampus	0.8701 ± 0.0160	0.8822 ± 0.0154
Left hippocampus	0.8599 ± 0.0207	0.8710 ± 0.0185

3.2 Comparison with the Relevant Methods

To investigate the contribution of DFF, we compared the proposed method with several state-of-the-art fusion algorithms, such as major voting [12], weight voting [13], SIMPLE [14], STAPLE [15], and spatial STAPLE [16]. Comparison was performed using MASI Label Fusion toolbox[2] with default parameters. Table 2 shows the mean

Table 2. Mean and standard deviation of Dice similar coefficient for 15 testing subjects using the proposed method and five relevant methods.

Method\Subject	Right hippocampus	Left hippocampus
Major voting	0.8521 ± 0.0373	0.8387 ± 0.0502
Weight voting	0.8531 ± 0.0369	0.8384 ± 0.0523
SIMPLE	0.8525 ± 0.0373	0.8394 ± 0.0479
STAPLE	0.8481 ± 0.0434	0.8351 ± 0.0610
Spatial STAPLE	0.8526 ± 0.0374	0.8384 ± 0.0538
Our	0.8822 ± 0.0154	0.8710 ± 0.0185

[2] http://www.nitrc.org/projects/masi-fusion.

and standard deviation of DSC for 15 testing subjects by using the proposed method and the five relevant methods. The mean and standard deviations of DSC obtained by the proposed method are 0.8822 ± 0.0154 for the right hippocampus and 0.8710 ± 0.0185 for the left hippocampus. The results show that the proposed method is more accurate than the five relevant fusion algorithms.

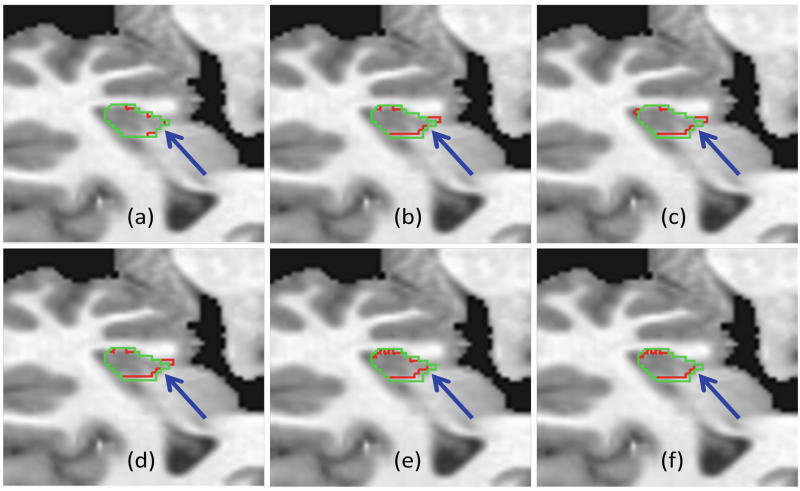

Fig. 3. Coronal view of the right hippocampus segmentation results for the testing subject by our method (Fig. 3. (a)), and five other methods (Fig. 3. (b)–(f)): Major Voting, Weight Voting, SIMPLE, STAPLE and Spatial STAPLE, along with the manual segmentation. Red contour and green contour denote the automatic segmentation results and manual segmentation results, respectively (Color figure online).

For visual comparison, we present the coronal view of the segmentation of the right hippocampus of the testing subject by using the proposed method (Fig. 3. (a)) and the five other methods (Fig. 3. (b)–(f)). The hippocampus segmented by the proposed method is similar to the ground truth than any of the five methods. In particular, in the blue arrow region, the proposed method can delineate the boundary of hippocampus, whereas the five other methods lead to large estimation errors.

4 Conclusion

In this study, we propose a novel DFF-based method to segment hippocampus in MR images. The method was established using the following procedures: (1) we used a distance transformation for label training images to obtain DF and utilize the shape information of hippocampus in training subjects; and (2) the local diffeomorphic mapping from the MR manifold was introduced to the DF manifold to predict the DF for a testing subject. The proposed method was evaluated for MR brain images on a dataset including 35 subjects. The proposed method exhibits superior performance over other fusion algorithms.

References

1. Wurtman, R.: Biomarkers in the diagnosis and management of Alzheimer's disease. Metabolism **64**, S47–S50 (2015)
2. Babalola, K.O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., Smith, S., Cootes, T., Jenkinson, M., Rueckert, D.: An evaluation of four automatic methods of segmenting the subcortical structures in the brain. Neuroimage **47**, 1435–1447 (2009)
3. Pipitone, J., Park, M.T., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos, A.N., Chakravarty, M.M., Alzheimer's Disease Neuroimaging Initiative: Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. Neuroimage **101**, 494–512 (2014)
4. Huang, M., Yang, W., Jiang, J., Wu, Y., Zhang, Y., Chen, W., Feng, Q., Alzheimer's Disease Neuroimaging Initiative: Brain extraction based on locally linear representation-based classification. Neuroimage **92**, 322–339 (2014)
5. Zhang, P., Wee, C.-Y., Niethammer, M., Shen, D., Yap, P.-T.: Large deformation image classification using generalized locality-constrained linear coding. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 292–299. Springer, Heidelberg (2013)
6. Tibshrani, R.: Regression shrinkage and selection via the LASSO. J. Roy. Stat. Soc. **58**, 267–288 (1996)
7. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained Linear Coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3360–3367 (2013)
8. Liu, W., He, J., Chang, S.F.: Large graph construction for scalable semi-supervised learning. In: International Conference on Machine Learning, pp. 679–686 (2010)
9. Smith, S.M.: Fast robust automated brain extraction. Hum. Brain Mapp. **17**, 143–155 (2002)
10. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**, 1310–1320 (2010)
11. Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C.: DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. Med. Image Anal. **15**, 622–639 (2011)
12. Cabezas, M., Oliver, A., Llado, X., Freixenet, J., Cuadra, M.B.: A review of atlas-based segmentation for magnetic resonance brain images. Comput. Meth. Programs Biomed. **104**, e158–e177 (2011)
13. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., Shen, D.: A generative probability model of joint label fusion for multi-atlas based brain segmentation. Med. Image Anal. **18**, 881–890 (2014)
14. Langerak, T.R., van der Heide, U.A., Kotte, A.N., Viergever, M.A., van Vulpen, M., Pluim, J.P.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans. Med. Imaging **29**, 2000–2008 (2010)
15. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging **23**, 903–921 (2004)
16. Asman, A.J., Landman, B.A.: Formulating spatially varying performance in the statistical fusion framework. IEEE Trans. Med. Imaging **31**, 1326–1336 (2012)