Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Label fusion for segmentation via patch based on local weighted voting*

Kai ZHU[1], Gang LIU[†‡1,2], Long ZHAO[1], Wan ZHANG[1]

([1]*School of Automation Engineering, Shanghai University of Electrical Power, Shanghai 200090, China*)

([2]*Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA*)

[†]E-mail: lukelg@gmail.com

Received Dec. 16, 2015; Revision accepted May 25, 2016; Crosschecked Apr. 27, 2017

Abstract: Label fusion is a powerful image segmentation strategy that is becoming increasingly popular in medical imaging. However, satisfying the requirements of higher accuracy and less running time is always a great challenge. In this paper we propose a novel patch-based segmentation method combining a local weighted voting strategy with Bayesian inference. Multiple atlases are registered to a target image by an advanced normalization tools (ANTs) algorithm. To obtain a segmentation of the target, labels of the atlas images are propagated to the target image. We first adopt intensity prior and label prior as two key metrics when implementing the local weighted voting scheme, and then compute the two priors at the patch level. Further, we analyze the label fusion procedure concerning the image background and take the image background as an isolated label when estimating the label prior. Finally, by taking the Dice score as a criterion to quantitatively assess the accuracy of segmentations, we compare the results with those of other methods, including joint fusion, majority voting, local weighted voting, majority voting based on patch, and the widely used FreeSurfer whole-brain segmentation tool. It can be clearly seen that the proposed algorithm provides better results than the other methods. During the experiments, we make explorations about the influence of different parameters (including patch size, patch area, and the number of training subjects) on segmentation accuracy.

Key words: Label fusion; Local weighted voting; Patch-based; Background analysis
http://dx.doi.org/10.1631/FITEE.1500457 **CLC number:** TP391.4

1 Introduction

Segmentation is one of the fundamental problems in biomedical image analysis. The traditional approach for segmenting a given biomedical image involves manual delineation of the regions of interest (ROIs) by a trained expert. However, due to this highly labor-intensive process and its poor reproducibility, it is often desirable to have accurate automatic segmentation techniques (Shi *et al.*, 2009). Early segmentation algorithms deal mainly with tissue classification, in which local image intensity contains important information (Cocosco *et al.*, 2003). However, these algorithms cannot guarantee accuracy.

The development of accurate, robust, and reliable segmentation techniques for automatic extraction of anatomical structures is becoming an important challenge in quantitative magnetic resonance imaging (MRI) analysis. Several automatic methods have been proposed. One family of supervised segmentation techniques that has become popular in brain MRI is multi-atlas segmentation (Rohlfing *et al.*, 2004).

Generally, the multi-atlas segmentation method sometimes requires to select a subset of best atlases for a given target image based on a certain predefined measurement of anatomical similarity. It also consists of the following two steps: first, in the

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (No. 61203224) and the Science and Technology Innovation Foundation of Shanghai Municipal Education Commission, China (No. 13YZ101)
ORCID: Gang LIU, http://orcid.org/0000-0003-3358-9299

registration step, all selected atlases and their corresponding label maps are aligned to the target image. The maturity of registration algorithms, for example, advanced normalization tools (ANTs) (Avants *et al.*, 2011), enables multi-atlas techniques to achieve very high performance. The ANTs open-source software library consists of a suite of state-of-the-art image registrations (e.g., diffeomorphisms: SyN), which provide a hierarchy of transformations with adjustable levels of complexity, regularization, degrees of freedom, and behavior as optimizers. Second, in the label fusion step, the registered label maps from the selected atlases are fused into a consensus label map for the target image by proper algorithms (Isgum *et al.*, 2009; Sabuncu *et al.*, 2010; Ghasemi *et al.*, 2012). The label fusion methods based on multiple atlases can effectively solve the problem of segmentation using prior knowledge without artificial interference, and finish the image segmentation of specific organizations automatically with high accuracy. Much attention has been put into the label fusion step, since it has a great influence on the final segmentation performance.

Although the local weighted voting strategy of label fusion based on a generative model yields statistically better segmentations than the benchmarks in Sabuncu *et al.* (2010), the image background is not regarded as a label for reducing processing time when the label prior is estimated by the logarithm of odds (LogOdds) model based on the signed distance transform (SDM). It will cause lower accuracy without regard to the background (Section 2.2).

Most of the above-mentioned methods perform segmentation on a voxel-by-voxel basis. Recently, patch-based methods (Coupé *et al.*, 2011; 2012; Eskildsen *et al.*, 2012) have been proposed for label fusion and segmentation. Their main idea is to allow multiple candidates (usually in the neighborhood) from each training image and to aggregate them based on non-local means (Manjón *et al.*, 2008). Different from multi-atlas based label fusion algorithms (Sabuncu *et al.*, 2010; Asman and Landman, 2013; Wang *et al.*, 2013), which require accurate non-rigid image registration, these patch-based methods are less dependent on the accuracy of registration, and thus even low-accuracy rigid registration can be applied. This technique has been successfully validated on brain labeling (Rousseau *et al.*, 2011) and hippocampus segmentation (Coupé *et al.*, 2011) with promising results. Wang *et al*. (2014) proposed a patch-driven level set method for the segmentation of neonatal brain MRI scans by taking advantage of sparse representation techniques.

Accordingly, in this paper, we propose a novel patch-driven level set method for label fusion by taking advantage of a probabilistic model and a local weighted voting scheme. First, the patches of the target image and training atlases are extracted. As shown in some previous work (Rousseau *et al.*, 2011; Wang *et al.*, 2014), patch-based methods can overcome the limitation of high-accuracy registration. Second, probabilistic models of label fusion are built based on patches. Bayesian inference is used to extend the popular method (local weighted voting). When calculating label prior, we analyze the label fusion procedure concerning the image background and regard it as an isolated label. The Kronecker delta function is employed as the model of label prior. Thus, not only can the processing time be reduced compared with the LogOdds model based on SDM (Sabuncu *et al.*, 2010), but accuracy segmentation can be higher. During experiments, the influences of different parameters are studied, and a comparison with other methods is performed. Finally, we discuss further improvements by a new approach.

2 Method

Generally, we assume that similar patches share the same label. That is to say, if we extract patches from the target image and training scans and find that they share similar image intensity, we will deem that they have the same label at the same location of voxels. Based on this assumption, the probabilistic model between the target patch and the training atlas patches is established. Then the labels from the training data are propagated to the target image by the label fusion segmentation algorithm of local weighted voting.

For each voxel in the test image, its intensity patch can be taken from $w\times w\times w$ neighborhood. Its patch dictionary can be adaptively built from all N aligned atlases as follows. First, let $\mathcal{N}_n(x)$ be the neighborhood of voxel x in the nth atlas, with neighborhood size $w_p\times w_p\times w_p$ (called 'patch area'

here). Then, for each voxel belonging to $\mathcal{N}_n(x)$, we can obtain its corresponding patch from the nth atlas, i.e., a $w\times w\times w$ column vector. By gathering all these patches from $w_p\times w_p\times w_p$ neighborhoods of all N training atlases, we can build a patch dictionary that contains $w_p^3 N$ training patches from all N aligned atlases.

In the label fusion step, the local weighted voting strategy is used to segment the target image. We build the probabilistic model for intensity prior and label prior, and use Bayesian inference to derive the segmentation algorithm. The specific process of the proposed algorithm is shown in Fig. 1.

2.1 Intensity prior

A voxel from a target image with a certain intensity value can be treated as a weight to perform the local weighted voting scheme. Intensity prior describes a probability that a voxel belongs to a certain training image patch or a likelihood. We adopt a Gaussian distribution for intensity information between the patch of the target image and the patch dictionary from training subjects as the intensity prior, which can be written as follows:

$$p_m\left(I(x);I_m\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}\left(I(x)-I_m(x)\right)^2\right], \tag{1}$$

where σ is the standard deviation of the Gaussian distribution, and $I(x)$ and $I_m(x)$ (m=1, 2, …, M) represent the intensity patch of the target image and the mth training patch at location x, respectively.

2.2 Label prior

Label prior is a kind of metric to describe the probability that a voxel at a certain location belongs to a certain label. If considering only these labeled anatomical structures (ROIs) as competitors to be involved in the label fusion procedure, we will more likely obtain the specific label, even if it should be background.

According to our analysis, there is no reason to say that the label prior of a voxel near the edge of one region is less than that of a corresponding voxel in the center of a region. If there is the pattern indeed, it should not be too much, though in the label fusion community, the LogOdds model based on SDM (Sabuncu *et al.*, 2010) seems a default setting for label prior. This kind of model does not fit the real problem very well, and leads to a poor topology near the edge of a region, like the signed distance transform, which plausibly models label prior but with unnecessary computational complexity.

Based on the above analysis and a large number of our experiments, the Kronecker delta function proves to be more suitable to be label prior, where the

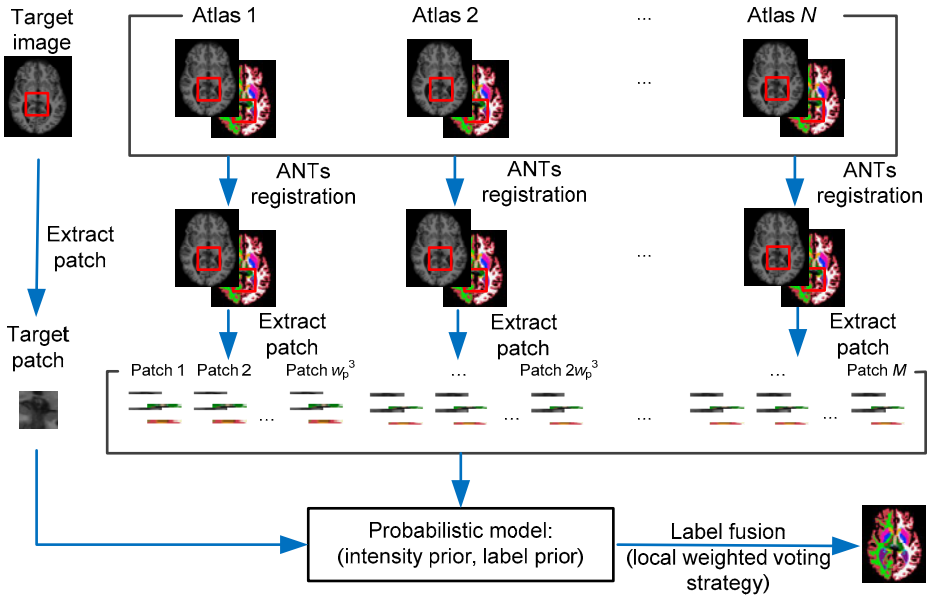


Fig. 1 The process of the label fusion algorithm via patch based on a probabilistic model and local weighted voting

We extract patches from the target image and training scans, and then establish the probability model between the target patch and the training atlas patches. Finally the segmentation result can be obtained through the proposed label fusion algorithm

function is equal to 1 if the label from the automatic segmentation method at a certain location and the manual label in training label maps at the same position are equal, and 0 otherwise. However, the Kronecker delta function is rarely used as label prior in most label fusion segmentation methods based on patch and probabilistic model. The joint label fusion technique (Wang and Yushkevich, 2013; Wang *et al.*, 2013), which won the first place of the 2012 MICCAI Multi-atlas Labeling Challenge (https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Main_Page), has been developed with the delta function for hippocampal segmentation, but the accuracy of the multi-label result is still lower than that of our proposed algorithm based on patch (Section 3.4).

Therefore, in this study we define other anatomical regions (not ROIs) to be the background label. Namely, while estimating the label prior, the background should be taken into account as an isolated label. Its value is set at 0 so that the background has the same privilege as the other labels. The Kronecker delta function is used for label prior:

$$P_m(S(x)=l;L_m(x))=\frac{1}{Z_m(x)}(\delta(S(x)=l)), \quad (2)$$

where $S(x)$ denotes the estimate of the label at location x and L_m the training patches label maps. $l\in\{0,$ labels of target$\}$. $Z_m(x)=\sum_{l=0}^{\mathcal{L}}\delta(S(x)=l)$ is the partition function that makes the probability $p_m(S(x)=l; L_m)$ between 0 and 1 at location x. L is the total number of labels including the background label. $\delta(\cdot)$ refers to the Kronecker delta function. The function is 1 if the variables are equal, and 0 otherwise:

$$\delta(S(x),l)=\begin{cases}0, & S(x)\neq l,\\ 1, & S(x)=l.\end{cases} \quad (3)$$

Sabuncu *et al*. (2010) adopted the LogOdds model based on SDM for label prior. Although it has the advantage of yielding nonzero probabilities everywhere, the image background is not treated as an isolated label when label prior is established. We find that the computer running time will be too long when calculating SDM if label l starts from 0 but not 1. As discussed in our experiments presented in Fig. 5, the proposed model produces more accurate results.

2.3 Label fusion

Through a patch-based process, we may assume that the voxels are independent and identically distributed so that the label on every location can be estimated separately. The label fusion scheme becomes a classical local weighted voting scheme, except for the intensity prior and the minor modification of the label prior. The segmentation is formulated as

$$\begin{aligned}\hat{S}(x)&=\arg\max_{S(x)} p_m(S(x)\mid I(x);\{L_m,I_m\})\\ &=\arg\max_{l\in\{0,\text{ labels of target}\}}\sum_{m=1}^{M}p_m(I(x)\mid I_m)p_m(S(x)=l;L_m).\end{aligned} \quad (4)$$

Sabuncu *et al.* (2010) gave the exhaustive derivation. The Bayesian condition and marginal distribution are used in Eq. (4). Then, we make the following corrections:

$$\hat{S}(x)=\arg\max_{l\in\{0,\text{ labels of target}\}}\sum_{m=1}^{M}p_m(I(x);I_m)p_m(S(x)=l;L_m), \quad (5)$$

where $p_m(I(x); I_m)$ serves as weights and label prior values serve as votes.

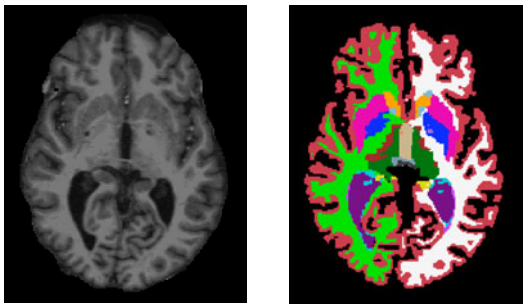
It is well known that the local weighted voting strategy is based on voxel, and the neighborhood of voxel x has no impact on the central point in one patch. So, we finally take the center location label value from the labeled patch as the segmentation result, and put it in location x on the whole label image. In the program we unfold the 3D patch matrix into a 1D array, and the position $(w\times w\times w-1)/2+1$ is corresponding to the center position. It is not difficult to see that when a patch area becomes a $1\times1\times1$ voxel, the proposed label fusion algorithm can become the local weighted voting strategy without patch.

3 Experiments and results

3.1 Dataset

During the experiments we employed 20 brain MRI scans and corresponding label maps (Fig. 2), selected from a large data set. We note that the 20 brain scans were resampled to 1 mm isotropic resolution, skull stripped, bias field corrected, and intensity normalized with FreeSurfer. Intensity

normalization is necessary because it enables us to directly compare image intensities (even across datasets). The scans were then pairwise registered with the software package ANTs. We used the default parameters for the deformation model, the number of iterations (20×30×50), and the cost function (neighborhood cross correlation).



(a) (b)

Fig. 2 Two-dimensional slices for visualization: (a) intensity image of the human brain; (b) corresponding label map

The MRI images are of dimension 256×256×256. In training label maps, each one has been labeled for 45 anatomical regions by experts. However, nine anatomical ROIs in two of the hemispheres we used are white matter (WM), cerebral cortex (CT), lateral ventricle (LV), hippocampus (HP), thalamus (TH), caudate (CA), putamen (PU), pallidum (PA), and amygdala (AM). During the experiments, considering both reasonable computational processing time and segmentation accuracy, we use union sets along the ROIs of all training label maps to fetch voxels of ROIs to segment the target image. The remaining 36 anatomical regions will be treated as background in label fusion.

3.2 Comparison of methods

To verify our theoretical analysis, massive experiments in the context of multi-atlas label fusion were performed to compare our proposed algorithm with a variety of related methods. We used a volume overlap measure known as the Dice score to quantitatively assess the accuracy of segmentations. The evaluation criterion is of the form

$$\mathrm{DS}=2\frac{V(S_{\mathrm{man}}\cap S_{\mathrm{auto}})}{V(S_{\mathrm{man}})+V(S_{\mathrm{auto}})}, \tag{6}$$

where S_{auto} is an automatic segmentation of the target image, S_{man} refers to manual segmentation, and V denotes the region of segmentation. The range of DS is between 0 and 1, with 1 meaning a perfect consistence about volume overlap between the two segmentations.

In the experiments, the proposed algorithm and other popular methods were implemented to segment brain MRI scans automatically. With the exception of FreeSurfer and joint fusion, we all concern the image background and take it as an isolated label while estimating label prior by the Kronecker delta function in the remaining methods. Details are as follows:

1. FreeSurfer

We directly use the FreeSurfer software package (http://www.freesurfer.net/) to segment all anatomical regions of the test images one by one. It does not rely on the preprocessing step of pairwise registration compared with the following label fusion methods.

2. Joint fusion (JF)

In this technique (Wang *et al.*, 2013), weighted voting is formulated in terms of minimizing the total expectation of the labeling error and pairwise dependency between atlases is explicitly modeled as the joint probability of two atlases making a segmentation error at a voxel.

3. Majority voting (MV)

MV can be seen as the most likely labeling in a probabilistic model in which the segmentation $S(x)$ is sampled randomly from one of the N atlases.

4. Local weighted voting (LWV)

This label fusion strategy takes advantage of the image intensities of the deformed atlases, which can improve segmentation quality. Compared with the proposed algorithm, the difference is that the processing is not at the patch level.

5. Majority voting based on patch (MVP)

This method does not consider image intensity, but it computes label prior at the patch level, which is an improvement of MV.

6. Local weighted voting based on patch (LWVP)

As described in Section 2, the probabilistic models of intensity prior and label prior are established based on patch. We assume that similar patches share the same label and build a patch dictionary between the target image and training data. Then label fusion will be processed on M training patches instead of N training subjects.

We investigate another representation to define the label prior term using the LogOdds model based on SDM (Sabuncu *et al.*, 2010). However, it is a compromise between segmentation accuracy and processing time, which can greatly reduce the running time but will cause lower accuracy without regard to the background. The Kronecker delta function is used to build label prior and make background analysis in this study. As will be discussed in Section 3.4, it can be clearly seen that the proposed method based on patch produces more accurate results.

3.3 Parameter optimization

The four mentioned label fusion methods (except FreeSurfer and JF) were used to segment the scans in a leave-one-out manner, where the test subject was left out of the atlas set. To determine the optimal parameter σ in the intensity prior probability, we tried many values, from 1 to 10. In LWV and LWVP, by contrast, we found that when σ=5 the segmentation result is best. The other parameter settings and optimizations are discussed in the following.

3.3.1 Impact of patch size and patch area

In MVP and LWVP label fusion schemes, we studied the impact of the patch size and patch area on segmentation accuracy. In one patch, the neighborhood of voxel x has no impact on the central point, because the strategies of local weighted voting and majority voting are both based on voxel. During experiments the patch size was set as 7×7×7 voxels (namely, w=7). In the LWVP scheme we set N=2, and when a patch area of 3×3×3 voxels was chosen (namely, w_p=3), segmentation precision was at its best (Fig. 3).

The optimal patch area seems to reflect the complexity of the anatomical structure. As can be seen from Fig. 4, when we set N=19 and w_p=3, the accuracy of segmentation results will likewise be at their best.

3.3.2 Impact of the number of training scans

We explored how the accuracy results vary with the change of N. Segmentation accuracy was studied from N=2 to N=19. The results of the LWV method are presented in Fig. 4, which reports the average Dice score as a function of N. As expected, the segmentation accuracy can be improved by increasing the number of selected training subjects.

In other experiments, we found that the other methods had the same regularity. So, the remaining results in this study are based on N=19.

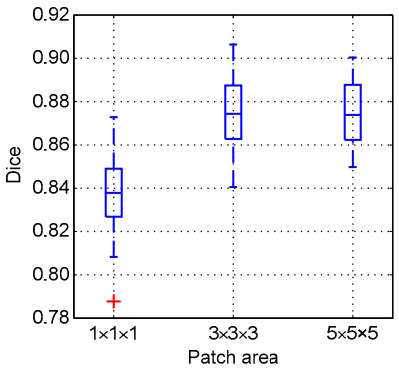


Fig. 3 Dice scores of the LWVP scheme according to patch size w=7 and three different patch areas
The segmentation results were obtained with N=2 in leave-one-out way. On each box, the central mark was the median, the edges of the box were the 25th and 75th percentiles, and the whiskers spanned the extreme data points not considered as outliers (marked with a cross)

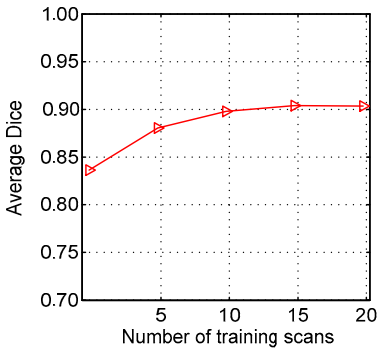


Fig. 4 The average Dice scores for the LWV method as a function of the number of training scans
It was obtained with σ=5, w=7, and w_p=3 in leave-one-out way and reached 0.908 when all 19 subjects were used

3.4 Results

We segmented the brain MRI images in a leave-one-out cross-validation fashion and by label fusion methods as described in Section 3.2. Fig. 5 shows the automated segmentation of a scan on the nine ROIs in the left hemisphere using LWVP. We first show the sagittal slice of the result in Fig. 5a. The 3D rendering of the segmentation is vividly shown in Fig. 5b. Some anatomical regions in relation to Alzheimer's disease, especially hippocampus, can be seen more clearly.

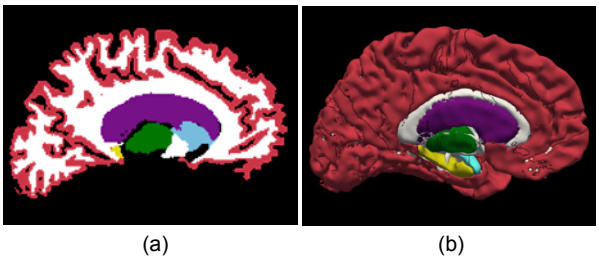


Fig. 5 Sagittal slice of a segmentation which includes only nine ROIs in the left hemisphere using the proposed algorithm (a) and 3D rendering of the segmentation (b)

To evaluate the performance of the proposed method, we measured the Dice scores for the manual and automated segmentations for each ROI.

As described in Section 2.2, establishing label prior, we adopted SDM in LWV (represented for LWV-S), and Kronecker delta in LWV (represented for LWV-K) and LWVP. In this process, the image background was analyzed and taken as an isolated label in the latter two methods. Fig. 6 shows the box plots for the Dice scores for the nine ROIs in the left hemisphere. The proposed method based on patch produced more accurate results. The mean Dice scores of LWV-K were very close to those in LWV-S, and even better in WM structure. However, LWV-K took less running time than LWV-S in program processing. Furthermore, Fig. 6 shows that making background analysis is necessary and that the Kronecker delta function is more suitable for label prior.

We also studied other methods, such as FreeSurfer, JF, MV, and MVP, to segment brain MRI scans. FreeSurfer segmented whole-brain anatomical regions, and the remaining label fusion algorithms yielded only segmentation of ROIs. Table 1 and Fig. 7 report the average Dice scores achieved by all algorithms for segmentation accuracy of the nine ROIs in the left hemisphere.

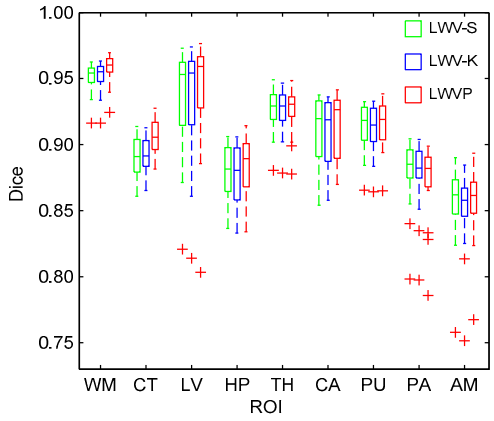


Fig. 6 Dice scores for the three methods: LWV-S (green), LWV-K (blue), LWVP (red)
On each box, the central mark was the median, the edges of the box were the 25th and 75th percentiles, and the whiskers spanned the extreme data points that were not considered outliers (marked with red crosses). References to color refer to the online version of this figure

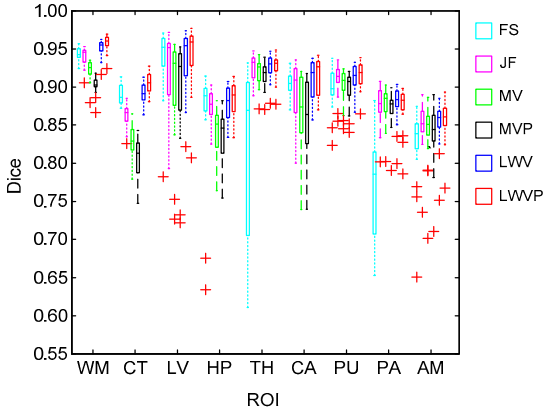


Fig. 7 Dice scores corresponding to the mentioned methods: FreeSurfer (dark green), JF (purple), MV (light green), MVP (black), LWV (blue), LWVP (red)
On each box, the central mark was the median, the edges of the box were the 25th and 75th percentiles, and the whiskers spanned the extreme data points that were not considered outliers (marked with red crosses). References to color refer to the online version of this figure

Table 1 Average Dice scores for each ROI corresponding to the five methods

Method	Average Dice score								
	WM	CT	LV	HP	TH	CA	PU	PA	AM
FS	0.943	0.890	0.935	0.864	0.815	0.904	0.898	0.760	0.825
JF	0.940	0.863	0.921	0.874	0.924	0.896	0.910	0.874	0.851
MV	0.922	0.827	0.903	0.840	0.920	0.871	0.901	0.877	0.840
LWV	0.952	0.892	0.935	0.878	**0.925**	0.909	0.913	**0.878**	0.851
MVP	0.902	0.806	0.899	0.835	0.916	0.863	0.898	0.873	0.838
LWVP	**0.958**	**0.906**	**0.941**	**0.883**	**0.925**	**0.915**	**0.916**	0.873	**0.856**

FS: FreeSurfer; JF: joint fusion; MV: majority voting; LWV: local weighted voting; MVP: majority voting based on patch; LWVP: local weighted voting based on patch. Boldface indicates the best score

LWVP, our proposed patch-driven level set method, yielded the most accurate segmentations in all ROIs but PA. LWV obtained better results than other methods including FreeSurfer, which is a state-of-the-art whole-brain segmentation tool. The segmentation of LWV and LWVP clearly benefits from the additional use of local intensity information. The difference between FreeSurfer and the above other methods is that pairwise registration of the latter is based on ANTs, which is a maturity algorithm and can improve segmentation accuracy. For this reason, in TH, PU, PA, and AM anatomical ROIs, the mean Dice scores for FreeSurfer were lower than those of other methods.

Joint fusion, a label fusion strategy with the Kronecker delta function as label prior, yielded better segmentation accuracy than MV and MVP, but worse than LWV and LWVP. Thus, the experimental results demonstrated that it is necessary to take the background as an isolated label when the label prior is built. The superiority of the proposed algorithm based on patch is apparent. Compared with FreeSurfer, joint fusion obtained higher average Dice scores in HP, TH, PU, PA, and AM.

In addition, we noted that the results of MV and MVP were worse than those of weighted label fusion methods. The reason is that the majority voting strategy ignores the image intensity, which contains a significant amount of relevant information. Also, we found that MVP performed slightly worse than MV. This might be due to the conflict between patch level and the discard of image intensity information in MVP, which has an influence on the accuracy of the segmentation.

4 Conclusions and discussion

We have presented a patch-driven level set method for label fusion by taking advantage of the probabilistic model and local weighted voting scheme. Compared with other algorithms, it can produce more accurate and robust results. The label fusion procedure concerning the image background has been analyzed and the background taken as an isolated label when the label prior was established by the Kronecker delta function in the patch level. The running time could be reduced and the segmentation performance could be much better by the use of the Kronecker delta function in the label prior term rather than the LogOdds model based on SDM. Future work will be focused on newer, more sophisticated label fusion algorithms that can produce results of higher accuracy. We will apply an expectation maximization (EM) or variational EM algorithm to segment the target image and compare it with LWVP.

References

Asman, A.J., Landman, B.A., 2013. Non-local statistical label fusion for multi-atlas segmentation. *Med. Image Anal.*, **17**(2):194-208. http://dx.doi.org/10.1016/j.media.2012.10.002

Avants, B.B., Tustison, N.J., Song, G., *et al.*, 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, **54**(3):2033-2044. http://dx.doi.org/10.1016/j.neuroimage.2010.09.025

Cocosco, C., Zijdenbos, A., Evans, A., 2003. A fully automatic and robust brain MRI tissue classification method. *Med. Image Anal.*, **7**(4):513-527. http://dx.doi.org/10.1016/S1361-8415(03)00037-9

Coupé, P., Manjón, J., Fonov, V., *et al.*, 2011. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *NeuroImage*, **54**(2): 940-954. http://dx.doi.org/10.1016/j.neuroimage.2010.09.018

Coupé, P., Eskildsen, S.F., Manjón, J.V., *et al.*, 2012. Simultaneous segmentation and grading of anatomical structures for patient's classification: application to Alzheimer's disease. *NeuroImage*, **59**(4):3736-3747. http://dx.doi.org/10.1016/j.neuroimage.2011.10.080

Eskildsen, S.F., Coupé, P., Fonov, V., *et al.*, 2012. BEaST: brain extraction based on nonlocal segmentation technique. *NeuroImage*, **59**(3):2362-2373. http://dx.doi.org/10.1016/j.neuroimage.2011.09.012

Ghasemi, J., Mollaei, M.R.K., Ghaderi, R., *et al.*, 2012. Brain tissue segmentation based on spatial information fusion by Dempster-Shafer theory. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **13**(7):520-533. http://dx.doi.org/10.1631/jzus.C1100288

Isgum, I., Staring, M., Rutten, A., *et al.*, 2009. Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. *IEEE Trans. Med. Imag.*, **28**(7):1000-1010. http://dx.doi.org/10.1109/TMI.2008.2011480

Manjón, J.V., Carbonell-Caballero, J., Lull, J.J., *et al.*, 2008. MRI denoising using non-local means. *Med. Image Anal.*, **12**(4):514-523. http://dx.doi.org/10.1016/j.media.2008.02.004

Rohlfing, T., Brandt, R., Menzel, R., *et al.*, 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, **21**(4):1428-1442. http://dx.doi.org/10.1016/j.neuroimage.2003.11.010

Rousseau, F., Habas, P.A., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. *IEEE Trans. Med. Imag.*, **30**(10):1852-1862. http://dx.doi.org/10.1109/TMI.2011.2156806

Sabuncu, M.R., Yeo, B.T.T., van Leemput, K., *et al.*, 2010. A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imag.*, **29**(10):1714-1729. http://dx.doi.org/10.1109/TMI.2010.2050897

Shi, F., Yang, J., Zhu, Y.M., 2009. Automatic segmentation of bladder in CT images. *J. Zhejiang Univ.-Sci. A*, **10**(2): 239-246. http://dx.doi.org/10.1631/jzus.A0820157

Wang, H.Z., Yushkevich, P.A., 2013. Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation. *Front. Neuroinform.*, **7**:27. http://dx.doi.org/10.3389/fninf.2013.00027

Wang, H.Z., Suh, J.W., Das, S.R., *et al.*, 2013. Multi-atlas segmentation with joint label fusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, **35**(3):611-623. http://dx.doi.org/10.1109/TPAMI.2012.143

Wang, L., Shi, F., Li, G., *et al.*, 2014. Segmentation of neonatal brain MR images using patch-driven level sets. *NeuroImage*, **84**(1):141-158. http://dx.doi.org/10.1016/j.neuroimage.2013.08.008