Medical Image Analysis 24 (2015) 297–312

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Structured patch model for a unified automatic and interactive segmentation framework

Sang Hyun Park[a], Soochahn Lee[b,*], Il Dong Yun[c,*], Sang Uk Lee[a]

[a] Department of Electrical Engineering, ASRI, INMC, Seoul National University, Seoul, Republic of Korea
[b] Department of Electronic Engineering, Soonchunhyang University, Asan-si, Republic of Korea
[c] Department of Digital Information Engineering, Hankuk University of Foreign Studies, Yongin, Republic of Korea

ARTICLE INFO

Article history:
Received 24 February 2014
Revised 5 January 2015
Accepted 19 January 2015
Available online 29 January 2015

Keywords:
Structured patch model
Interactive segmentation
Adaptive prior
Markov random field
Incremental learning

ABSTRACT

We present a novel interactive segmentation framework incorporating a priori knowledge learned from training data. The knowledge is learned as a structured patch model (StPM) comprising sets of corresponding local patch priors and their pairwise spatial distribution statistics which represent the local shape and appearance along its boundary and the global shape structure, respectively. When successive user annotations are given, the StPM is appropriately adjusted in the target image and used together with the annotations to guide the segmentation. The StPM reduces the dependency on the placement and quantity of user annotations with little increase in complexity since the time-consuming StPM construction is performed offline. Furthermore, a seamless learning system can be established by directly adding the patch priors and the pairwise statistics of segmentation results to the StPM. The proposed method was evaluated on three datasets, respectively, of 2D chest CT, 3D knee MR, and 3D brain MR. The experimental results demonstrate that within an equal amount of time, the proposed interactive segmentation framework outperforms recent state-of-the-art methods in terms of accuracy, while it requires significantly less computing and editing time to obtain results with comparable accuracy.

1. Introduction

With the advancement of medical imaging technology, high quality medical images have significantly increased. Accordingly, the demand for effective techniques to analyze medical images has increased as well. Segmentation of target objects is an especially important task required to study pathological changes of organs, compare inter-subject variability, monitor disease progression and analyze clinical trials. However, manual segmentation is laborious and time-consuming. Thus, various approaches to enable efficient segmentation have been proposed. The wide variety of segmentation methods can be loosely classified as either interactive or automatic.

Interactive methods require the user to provide annotations to incrementally refine the segmentation. To update the segmentation efficiently, most interactive segmentation methods are based on low-level statistics of appearance, including live-wire (Barrett and Mortensen, 1997), region growing (Pohle and Toennies, 2001), interactive graph cut (Boykov and Funka-Lea, 2006; Shim et al., 2009a,b), random walk (Grady, 2006; Kim et al., 2008) and geodesic segmentation (Bai and Sapiro, 2007). Although these methods are fast, flexible, and facilitate intuitive editing, extremely detailed user annotations may be required for noisy images or target objects with obscure boundaries. Many different approaches are taken to overcome this problem. One is to leverage simpler user interactions such as the bounding box of the target object (Rother et al., 2004; Lempitsky et al., 2009). Another is to reduce the amount of required user annotations by using sophisticated graphs which reflect the relationships between annotated and unknown regions (Kim et al., 2010). Yet another approach is to introduce multiple categories of annotations that better represents user intention (Yang et al., 2010). Although these methods may reduce the amount of required annotations, the changes in annotation often affect the segmentation results significantly. To receive more informative user annotations on ambiguous regions, methods based on the active learning strategy have been proposed. In the method by Wang et al. (2012), the expected confidence change of superpixels is measured to inform the user about the regions where annotation is more desired. In the method of Top et al. (2011), the two-dimensional plane having the highest uncertainty among a three dimensional image space is provided to the user at each editing step for the next annotation. Although these methods

* Corresponding authors.
E-mail addresses: shpark13135@gmail.com (S.H. Park), soochahn.lee@gmail.com (S. Lee), yun@hufs.ac.kr (I.D. Yun), sanguk@ipl.snu.ac.kr (S.U. Lee).

http://dx.doi.org/10.1016/j.media.2015.01.003
1361-8415/

guide the user to provide effective annotations, a significant amount of annotations is nonetheless necessary for accurate segmentation of ambiguous regions because the methods still rely on low-level statistics, such as intensity distributions and gradients. Basically, all the aforementioned methods do not utilize a priori knowledge of the object of interest. While this enables the methods to be generalized to various target objects, it hinders use of informative cues of target objects. Moreover, this restricts the reproducibility of the methods. Therefore, the user will need to laboriously repeat similar annotations when segmenting a common target object within many similar images. Clinicians burdened in these situations, which occur very often in clinical practices, will indeed benefit from more automated methods.

On the other hand, most automatic segmentation methods take advantage of a priori knowledge of target objects learned from training data, based on the assumption that target images share the similar appearance and structure. These methods can be divided into example-based and model-based. Example-based methods (Heckemann et al., 2006; Aljabar et al., 2009; van der Lijn et al., 2008; Lotjonen et al., 2010; Coupe et al., 2011; Rousseau et al., 2011; Park et al., 2013a; Asman and Landman, 2013; Bai et al., 2013; Tong et al., 2013) search for relevant example images and their labels from the training set which are directly used to guide the segmentation of the target image. In works of Heckemann et al. (2006) and Aljabar et al. (2009), the segmentation is determined by majority voting of aligned manual segmentation labels. In works of van der Lijn et al. (2008) and Lotjonen et al. (2010), the label information is incorporated into the graph cut framework of Boykov and Funka-Lea (2006) to further deal with local variations. Unlike the methods which directly use the labels of aligned training images, in works of Coupe et al. (2011) and Rousseau et al. (2011), the labels are determined by non-local weighted voting of labels of local atlas patches according to appearance similarity. Tong et al. (2013) proposed a similar patch based label fusion method, but used sparse representation to determine the weight for fusion. Though these methods are effective with a small number of training images, highly complex registration or per-patch similarity computation is required. Thus they are not scalable to the size of training set. Model-based methods overcome the limitation of example-based methods by modeling the target object from training data (Cootes et al., 1995; Duta and Sonka, 1998; van Ginneken et al., 2002; Sukno et al., 2007; Gleason et al., 2002; Seghers et al., 2007; Ibragimov et al., 2012; Yang and Ramanan, 2011; Zhang et al., 2012). For example, in the active shape model (ASM) by Cootes et al. (1995), the average and variations of the shape are modeled by statistics of object boundary landmarks. However, these methods require a large enough training set for the model to be sufficiently generalizable, which is hard to obtain in many clinical tasks where it is common to only have a small number of images. They also require laborious tasks during training such as manual extraction of landmarks or annotation of local parts. While the example-based and model-based methods have reduced the user efforts for many clinical applications, the segmentation results can often be inaccurate due to aforementioned weaknesses, especially at ambiguous regions. Although user editing is necessary in these cases, most automatic methods cannot be easily extended to include an effective interactive editing process.

Recently, several example-based and model-based interactive methods that incorporate a priori knowledge of the target objects or images have been proposed. In works of Barnes et al. (2009) and Barnes et al. (2010), the example-based method *PatchMatch* is proposed for labeling problems by efficiently searching for image patch correspondences and propagating their manual annotations. These methods have been extended to super-resolution of cardiac MRI (Shi et al., 2013) and hippocampus segmentation (Ta et al., 2014) for medical image analyses. Nonetheless, they may not be applicable for target objects with specific shape, since spatial relationships between adjacent patches are neglected. In works of Branson et al. (2011) and Wah et al. (2011), the deformable part model (DPM) (Felzenszwalb et al., 2010) is utilized in an interactive recognition framework supporting seamless learning. However, since the DPM is based on part labels, it is not easily extended to interactive segmentation framework which relies on user annotations, often given as detailed pixelwise labels. In the work by Schwarz et al. (2008). ASM is incorporated into the interactive segmentation framework by enabling the user to edit the positions of landmark points in the determined boundary. Whenever incorrect landmark points are edited by the user, adjacent landmarks are accordingly modified by Gaussian interpolation and the whole boundary is regularized based on the ASM. In the work by Sun et al. (2013), the segmentation boundary is determined by the optimal surface finding (OSF) method, based on an initial segmentation using ASM. The user can correct errors in OSF results by marking points on the correct boundary, which are used as constraints to recompute the OSF. While these methods also incorporate interactive editing with prior information, the required user interaction of 3D point positions can be difficult to achieve with only a common 2D interface.

In this paper, we present an efficient model-based interactive framework using the structured patch model (StPM)[1] for segmentation of target objects within a large number of medical images acquired in a common environment. The proposed StPM is an *example-based-model*, comprising sets of corresponding local patch priors and their pairwise spatial distribution statistics compiled from the example images and their segmentations in the training set. When a test image is given, the optimal local patch priors are adaptively selected and localized through a global probabilistic optimization based on the user annotations, local patch similarity, and the likelihood of global structure based on the pairwise spatial distribution. Then, voxel-wise segmentation labels are computed through a global probabilistic optimization based on the selected StPM and the user annotations.

The key advantages of the proposed framework based on StPM are as follows: First, we enforce the example-based multiple patch priors, which encapsulate a wide variety of specific local instances, into a model structure. It enables the method to use the optimal examples, in terms of both local adaptiveness and global consistency, as priors for segmentation. Second, user annotations are easily incorporated into the segmentation framework. Since the StPM is compatible with all types of annotations, the user can freely insert annotations on ambiguous regions without any restrictions based the model, making efficient segmentation possible for any image. Third, since interactive segmentation is constrained by the StPM as well as user annotations, the segmentation result is robust to the quantity and placement of the annotations. Compared to the previous interactive methods, the proposed method requires fewer annotations and is more robust to their changes due to the StPM. Finally, the StPM can easily be expanded by directly incrementing the local image and segmentation patch set and the pairwise distribution with the results obtained from the proposed framework for a new test image. This incremental learning system is particularly effective when constructing the training image set since the required laborious manual annotation is significantly reduced.

We note that this paper is based on our preliminary work presented by Park et al. (2013b). The preliminary method was sensitive to initial user annotations and could not handle the drifting

[1] We use the acronym *StPM* to avoid confusion with *statistical parametric mapping (SPM)*.

problem where localization errors of adjacent patches accumulated as their distance from the annotations increased. In this paper, we present a more comprehensive framework with the advanced model including the spatial distribution between neighboring local regions and a new method based on Markov random field (MRF) structure to localize the StPM within a test image. The MRF structure allows us to alleviate the sensitivity of the user interactions as well as avoid the drifting problem. Furthermore, we extend the method to handle segmentation of multiple objects by adopting a multi-label optimization solver.

The proposed structured patch model and the segmentation framework are described in Section 2. The framework is evaluated on various target objects in chest CT images, knee MR images, and brain MR images. The experimental settings and the results are described in Section 3. Finally, the paper is concluded in Section 4.

2. Interactive segmentation framework based on structured patch model

Our framework is comprised of the following steps: (1) Construct the StPM from a small number of training images offline; (2) localize the StPM within a given test image to perform initial segmentation; (3) interactively correct the specific state of the localized StPM and the segmentation of incorrect regions with manual annotations until necessary; (4) add the test image and its segmentation result to the training set to incrementally update the StPM. As the number of training sets increases, the StPM is able to handle more general images. Thus, better initial segmentations will be generated in step 2 and the amount of annotations required in step 3 will be reduced. Fig. 1 provides a visual description of the proposed framework.

The StPM is comprised of the set of patch sets $\mathbb{P}$ and their structure, represented by pairwise distance and orientation statistics between the patch sets. $\mathbb{P} = \{\mathbf{P}_j | j = 1, \ldots, n\}$ comprises n patch sets, where each patch set $\mathbf{P}_j$ represents a corresponding local region centered at the target object boundary. Each $\mathbf{P}_j = \{P_j^i | i = 1, \ldots, m\}$ comprises m patch pairs from the training data $\mathbb{T} = \{T^i | i = 1, \ldots, m\}$, where each pair includes the image and its segmentation labels. We note that subscripts i and j denote the index of the particular training image and local region, respectively. The patch structure comprises histograms of distances and angles between $\mathbf{P}_j$ and $\mathbf{P}_{j'}$. For both the initial segmentation and the interactive editing, the StPM is first localized and configured by optimization of a patch-level MRF. Then the segmentation is computed by optimizing a second voxel-level MRF based on the configured StPM.

Given a target volume V, appropriate patches $\mathbf{P}^{\mathbf{x}} = \{P^{x_j} | j = 1, \ldots, n\}$ are selected among each $\mathbf{P}_j$ and are transferred to optimal locations $\mathbf{v} = \{v_j | j = 1, \ldots, n\}$ within V. That is, x_j denotes the index of the most suitable training patch among $\mathbf{P}_j$, and $\mathbf{x}$ is the set of all x_j. This problem is formulated on a patch-level MRF as:

$$E(\mathbf{v}, \mathbf{x}) = \sum \phi(v_j, x_j | \mathbf{P}_j, U_j) + \lambda_{\mathbf{x}} \sum \psi(v_j, v_{j'} | \mathbf{P}_j, \mathbf{P}_{j'}), \quad (1)$$

where U_j denotes the user annotation labels, within the jth local region. In our framework, the user can provide object and background scribbles for the annotation. The first voxelwise potential term $\phi(v_j, x_j | \mathbf{P}_j, U_j)$ is based on the similarity between the test volume patch at v_j and $P^{x_j} \in \mathbf{P}_j$ along with the consistency between labels of P^{x_j} at v_j and U_j. The second pairwise potential term is based on the StPM spatial patch distribution statistics. Optimization of (1) is efficiently done in the interactive editing stage, since only the portion of $\phi(v_j, x_j | \mathbf{P}_j, U_j)$ depending on the updated annotations U_j needs to be recomputed.

Next, segmentation is done using the prior patches $\mathbf{P}^{\mathbf{x}}$ localized at $\mathbf{v}$ and user annotations U. The problem is formulated on a voxel-level MRF as:

$$E(L) = \sum_{v \in V} \phi(l(v) | \mathbf{P}^{\mathbf{x}}, U) + \lambda_L \sum_{u,v \in \Gamma} \delta_{l(u) \neq l(v)} \cdot \exp \frac{|I(u) - I(v)|}{2\beta}, \quad (2)$$

where

$$\delta_{l(u) \neq l(v)} = \begin{cases} 1 & \text{if } l(u) \neq l(v) \\ 0 & \text{if } l(u) = l(v) \end{cases} \quad (3)$$

is the Kronecker delta. $l(v)$ is the random variable representing the label of voxel v and L is the label variable set. Here, the first term

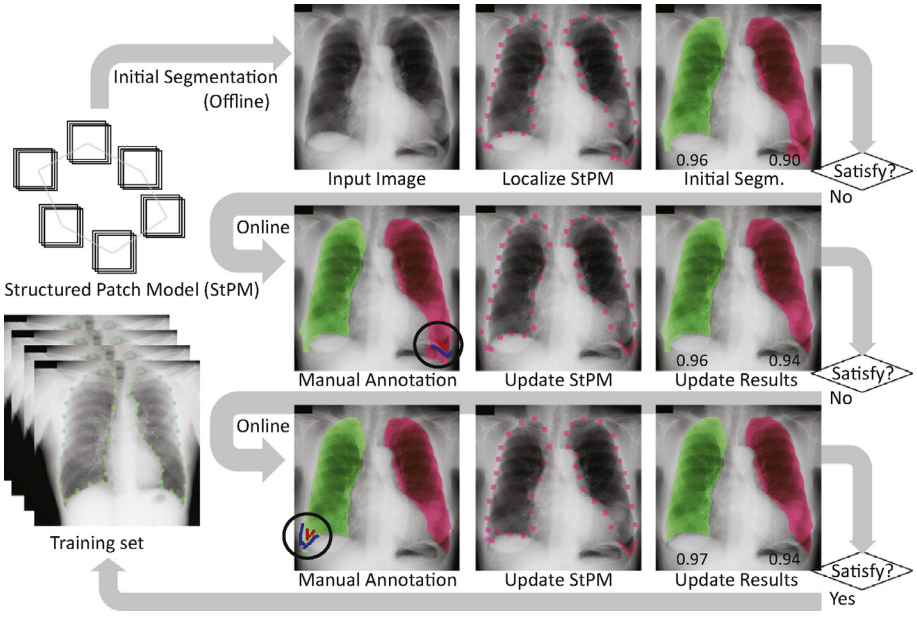

Fig. 1. Proposed framework based on the structured patch model (StPM). Dots around the boundaries of right and left lungs represent the centers of patches. The DSC values for each lung are also given below the segmentation results. Erroneous regions (black circles) are repeatedly corrected with respect to the user annotations (red and blue denote the object and background scribbles, respectively) until the segmentation results are satisfactory. The final result can be directly used to increment the StPM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

represents the unary potential representing the likelihood of v for each label and the second term represents the pairwise potential which enforces the smoothness between labels of neighboring voxel pair (u, v) in the neighbor set Γ. $I(u)$ is the intensity of voxel u and β is the average square-distance of intensities between adjacent voxels in V (Boykov and Funka-Lea, 2006).

In the initial step, segmentation is performed automatically by optimizing (1) and (2). Here, specific prior locations $\mathbf{v}$ are searched within a predetermined search range which is assumed to contain the optimal locations. The user is then requested to provide annotations at erroneous regions. When additional user annotation is given, the segmentation is immediately updated by recomputing (1) and (2) with updated U. This interactive editing is repeated until the user is satisfied. Further details are described in the following subsections.

2.1. Preprocessing

Images of different cases must be aligned in order to construct a relevant model of the target object. We randomly select an instance $T^{i'} = (V^{i'}, M^{i'})$ from the training data $\mathbb{T}$, where $V^{i'}$ and $M^{i'}$ denote the image volume and ground truth labels, respectively, to assign as the reference. Then, all other training images and labels are aligned to this reference in the training step.

In the test step, only the reference $T^{i'}$ is aligned to test volume V to constrain the possible localized patch prior coordinates. We use the non-rigid registration method proposed by Glocker et al. (2008), namely, the Drop method based on a pairwise MRF energy model. We note that this alignment takes less than one minute because only the reference training data is aligned to V with coarse registration parameters. More accurate alignment is performed in subsequent steps. Also, intensity distributions of the aligned volumes are normalized to that of $V^{i'}$ by histogram matching.

2.2. Construction of structured patch model

The StPM includes the sets of corresponding patch across training sets and their pairwise connections. First, n patches and their connections are extracted from the reference data $T^{i'}$, and then corresponding patches are found from the other training data. After all patches are extracted from $\mathbb{T}$, the priors of corresponding patch sets and the spatial relationship between the adjacent patch sets are learned. Fig. 2 shows a visual description of the StPM.

The n patches $\mathbf{P}^{i'} = \{P_j^{i'} | j = 1, \ldots, n\}$ are sampled from the surface of the target object with even distribution in $T^{i'}$. $P_j^{i'}$ includes the sub-volume $V_j^{i'}$ and the sub-label $M_j^{i'}$ centered at voxel $v_j^{i'}$. Patches within a certain distance τ are set as the adjacent patches. Here, the patches are sampled so that the surface of target object is included in the combined sampled patches.

To search for the corresponding patches from the other $m-1$ training sets, $\mathbf{P}^{i'}$ is first aligned using the Drop registration method (Glocker et al., 2008). Since the aligned location is often inaccurate, the correspondence is searched again within a search range. Within the search range centered at the initial aligned location, the patch having the highest similarity with $V_j^{i'}$ is determined as the corresponding patch. In our experiments, we set the search range as half of the patch size and measure the similarity by normalized cross correlation (NCC).

After $n \times m$ patches are determined, the priors are learned. For each patch P_j^i, mean $\mu(V_j^i)$ and covariance $\sigma(V_j^i)$ of voxel intensities in V_j^i are computed to accelerate comparison of the patch similarity. The spatial distribution between each adjacent patch pair P_j^i and $P_{j'}^i$ is learned as histograms of distance and direction angles between v_j^i and $v_{j'}^i$. In our experiments, distance histograms have 20 bins representing a range from 0 to 120 in voxel coordinates, while the angle histograms contains 10 bins representing 360° for each axis.

To determine the segmentation, M_j^i and intensity histograms are used as the shape and appearance priors, respectively. For each patch P_j^i, intensity histograms $\mathbf{H}_j^i = \{H_j^i(l(v)) | l(v) = 1, \ldots, K\}$ of voxels with K segmentation labels are constructed, respectively; *e.g.*, for binary segmentation problem with $K = 2$, separate histograms are constructed from voxels with label 0 and 1, respectively. In addition, we follow the method of Park et al. (2013a) to adaptively emphasize the shape and appearance priors according to the properties of local regions. Specifically, the weight $w_j^i(l(v))$, which controls the emphasis between shape and appearance priors, is computed by the distance $d_s(v)$ from the target object surface and the appearance confidence $f_j^i(l(v))$ as:

$$w_j^i(l(v)) = 1 - \exp\left(-d_s^2(v)/f_j^i(l(v))^2\right). \tag{4}$$

A higher weight is assigned to the appearance prior as the voxel coordinate is closer to the surface or the appearance of target object is clearly distinguished from other tissues. For details, we refer the reader to Park et al. (2013a).

2.3. MRF-based patch localization

$\mathbf{P}^{\mathbf{x}}$ among $\mathbb{P}$ and $\mathbf{v}$ in V are determined by optimizing (1). The graph model consists of n nodes representing random variables representing specific patch index and location along with their pairwise edges. The patch index and location are determined by assigning a label among a set of size $m \times$ the volume of localization search range.

Since the inference of (1) respect to the all labels regarding $\mathbf{v}$ and $\mathbf{x}$ needs considerable computation due to the large label set size, we approximate (1) as:

$$E(\mathbf{v}) = \sum \phi(v_j | \mathbf{P}_j, U_j) + \sum \psi(v_j, v_{j'} | \mathbf{P}_j, \mathbf{P}_{j'}), \tag{5}$$

where

$$\phi(v_j | \mathbf{P}_j, U_j) = \arg\min_{x_j}(\phi(v_j, x_j | \mathbf{P}_j, U_j)). \tag{6}$$

That is, potentials of m training patches are computed at each coordinate in the search range and the lowest is set as the potential for that location. Therefore, the number of labels for each node is reduced to the volume of the localization search range. Also, we set the search range around the initial aligned position for each region and sample coordinates at a regular interval within the search range. For each sampled location, the NCC similarity $S_{ncc}(V_j, V_j^{x_j})$ with the reference patch $V_j^{x_j}$ is computed as:

$$S_{ncc}(V_j, V_j^{x_j}) = \frac{1}{\eta(V_j)} \sum_{v \in V_j} \frac{(V_j(v) - \mu(V_j))(V_j^{x_j}(v) - \mu(V_j^{x_j}))}{\sigma(V_j)\sigma(V_j^{x_j})}, \tag{7}$$

where $\eta(V_j)$ is the number of voxels in V_j. Among all sampled locations, q locations having the highest NCC scores are set as candidates. The potential of each location candidate is computed as:

$$\phi(v_j, x_j | \mathbf{P}_j, U_j) = \begin{cases} -\log(S_{ncc}(V_j, V_j^{x_j})), & \text{if } S_{ncc}(V_j, V_j^{x_j}) > 0 \\ \infty. & \text{otherwise} \end{cases} \tag{8}$$

$\psi(v_j, v_{j'} | \mathbf{P}_j, \mathbf{P}_{j'})$ reflects the spatial constraints between the candidate locations v_j and $v_{j'}$ of adjacent patches. First, the differences of distance and direction angles between pairs of the candidates are computed. Then, the probabilities regarding the differences are defined by using the learned histograms. Since only decoupled

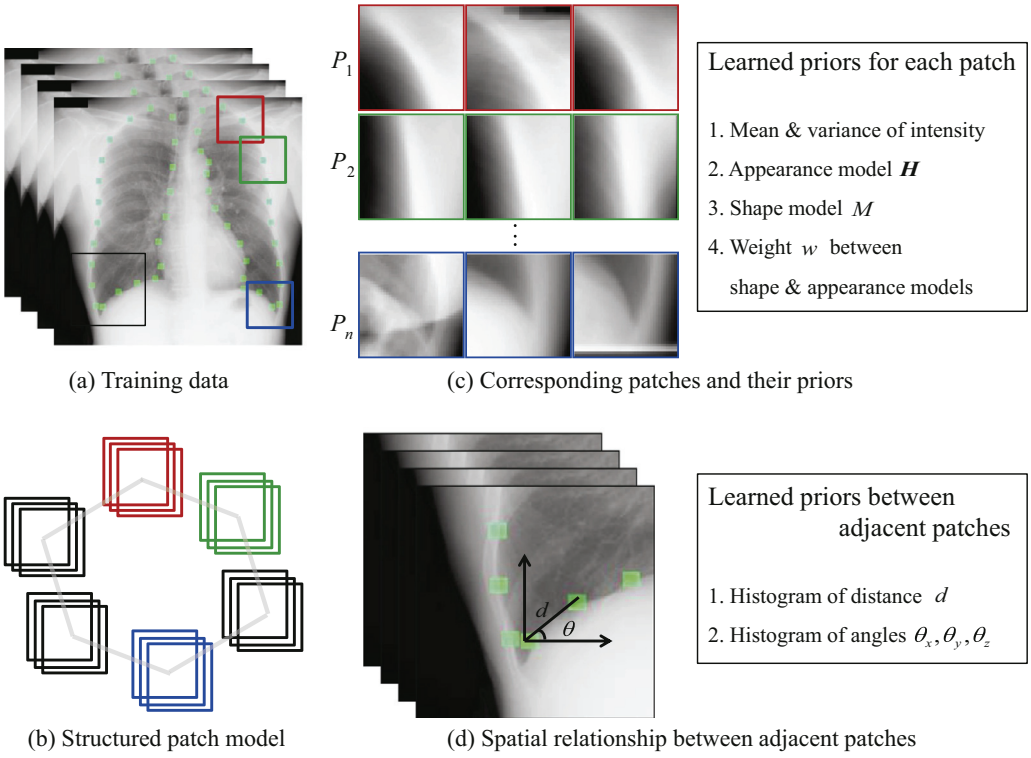

Fig. 2. The structured patch model (StPM). (a) Training data; (b) description of StPM comprising sets of corresponding local patches and their spatial distribution; (c) examples of corresponding local patch sets corresponding to the red, green, blue boxes of (a) and their priors; (d) the spatial relations between the adjacent patches on region shown by the black box in (a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

statistics for distance and angles of patch pairs have been constructed to reduce storage, we approximate the joint pairwise probability by averaging the marginal probabilities based on distance and angles. $\psi(v_j, v_{j'}|\mathbf{P}_j, \mathbf{P}_{j'})$ is computed as the negative log of the approximated probability.

Since (5) is not sub-modular, the global optimal solution cannot be determined efficiently. We use the α-expansion method (Boykov et al., 2001; Kolmogorov and Zabih, 2004) to compute approximate values of $\mathbf{v}$ and $\mathbf{x}$ from (5).

In the editing steps, the user will provide annotations to erroneous regions. At each step during editing, only the portion near the annotation is modified while the rest is fixed to the previous segmentation result L^{t-1} in order to avoid unwanted alterations. To achieve this, we set the patches near the user annotations as the activated patches. Then, the unary potentials within the activated patches are recomputed based on the consistency between the user annotation U_j and training set labels $M_j^{x_j}$, while the unary and pairwise potentials are fixed to 0 for all other patches. Specifically, the unary potentials of q position candidates in the activated patches are recomputed as:

$$\phi(v_j, x_j|\mathbf{P}_j, U_j) = \exp(-S_{ovl}(U_j, M_j^{x_j})), \tag{9}$$

where the overlapping similarity $S_{ovl}(U_j, M_j^{x_j}))$ is defined as:

$$S_{ovl}(U_j, M_j^{x_j}) = \sum_{v \in U_j} (\delta_{U_j(v)=M_j^{x_j}(v)} + w_U \delta_{L_j^{t-1}(v)=M_j^{x_j}(v)}). \tag{10}$$

The more the training patch labels are consistent with the user annotation, $\delta_{U_j(v)=M_j^{x_j}(v)}$ is increased, while the more the training patch is consistent with L^{t-1}, $\delta_{L_j^{t-1}(v)=M_j^{x_j}(v)}$ is increased. w_U controls the weight between the user annotation and the previous segmentation. As w_U is increased, the effect of the user annotation is damped by L^{t-1}.

2.4. MRF-based segmentation

The segmentation based on the priors of $\mathbf{P^x}$ is formulated as an MRF energy minimization framework as (2). In this case, the nodes and the labels indicate the voxels of V and the segmentation labels, respectively. The unary potential $\phi(l(v)|\mathbf{P^x}, U)$, which is based on the likelihood probability $\Pr(l(v)|\mathbf{P^x}, U)$ for v to be labeled as $l(v)$, is defined as follows:

$$\phi(l(v)|\mathbf{P^x}, U) = -\log(\Pr(l(v)|\mathbf{P^x}, U)). \tag{11}$$

The likelihood $\Pr(l(v)|\mathbf{P^x}, U)$ is determined by aggregating the likelihood of local regions $\Pr(l(v)|P_j^{x_j}, U_j)$, which are defined as:

$$\Pr(l(v)|P_j^{x_j}, U_j) = \begin{cases} 1, & \text{if } U_j(v) = l(v) \\ 0, & \text{if } U_j(v) \neq l(v),\ U_j(v) \neq \varnothing \\ \Pr(l(v)|P_j^{x_j}), & \text{if } U_j(v) = \varnothing \end{cases} \tag{12}$$

where

$$\Pr(l(v)|P_j^{x_j}) = w_j^{x_j}(l(v)) \cdot \Pr(l(v)|M_j^{x_j}) + (1 - w_j^{x_j}(l(v))) \cdot \Pr(l(v)|\mathbf{H}_j^{x_j}). \tag{13}$$

The likelihood $\Pr(l(v)|M_j^{x_j})$ based on the shape model is computed as:

$$\Pr(l(v)|M_j^{x_j}) = \begin{cases} 1, & \text{if } M_j^{x_j}(v) = l(v) \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

while the likelihood $\Pr(l(v)|\mathbf{H}_j^{x_j})$ based on the appearance model is computed as:

$$\Pr(l(v)|\mathbf{H}_j^{x_j}) = \frac{P(I(v)|H_j^{x_j}(l(v)))}{P(I(v)|H_j^{x_j}(l(v))) + \sum_{\tilde{l}(v)} P(I(v)|H_j^{x_j}(\tilde{l}(v)))}. \quad (15)$$

Here, $\tilde{l}(v)$ is the label indices except $l(v)$. Since there is no user annotations ($U(v) = \varnothing$) in the initial step, the likelihoods of all voxels are computed as (12).

In regions where adjacent patches overlap, multiple likelihoods computed from different reference patches are averaged to determine the global likelihood probability $\Pr(l(v)|\mathbf{P}^{\mathbf{x}}, U)$. The likelihoods of voxels, not covered by any localized StPM patch, are defined based on the labels of aligned reference set $T^{i'}$. Specifically, if the aligned $T^{i'}$ label is $l(v)$, $\Pr(l(v)|\mathbf{P}^{\mathbf{x}}, U) = 1$, otherwise $\Pr(\tilde{l}(v)|\mathbf{P}^{\mathbf{x}}, U) = 0$. To obtain L, (2) is optimized by the α-expansion method (Boykov et al., 2001; Kolmogorov and Zabih, 2004). While optimization is conducted on the whole image in the initial step, it is only conducted on the local regions patches activated by user annotations in the editing steps.

Algorithm 1. Algorithm of the proposed framework.

Input: the structured patch model $\mathbb{P}$ (Section 2.2 and a target volume V.

1: Preprocess V. (Section 2.1)
2: Transfer appropriate patches of $\mathbb{P}$ to V by optimizing Eq. (1). (Section 2.3)
3: Segment the target objects in V by optimizing Eq. (2). (Section 2.4)
4: **Iterate** 5–7 steps,
5: Insert user scribbles on erroneous regions.
6: Update the training patches and their positions on the local regions where the user scribbles were given by optimizing Eq. (1).
7: Update the segmentation on the local regions by optimizing Eq. (2).
8: **Until** the segmentation is satisfied.
9: (Optional) Add V and the final result to $\mathbb{P}$ (Section 2.2)

The overall framework is presented in Algorithm 1.

3. Experimental evaluation

The proposed framework is evaluated for three public datasets comprising 2D chest CT data (Shiraishi et al., 2000), 3D knee MR data (Heimann et al., 2010), and brain MR data.[2] We perform segmentation of the left and right lungs in chest CT images, the femur, tibia, femoral and tibial cartilages in knee MR images, and fourteen parts of the diencephalon in brain MR images. For the brain MR images, we specifically focus on segmentation of the left and right hippocampus and the thalamus proper among the fourteen parts. For a comprehensive validation of the proposed framework, we aim to evaluate both the accuracy of the StPM-based initial automatic segmentation (Auto-StPM) and the efficiency and robustness of the StPM-based interactive framework (IA-StPM).

3.1. Experimental setting

For the experimental results presented here, the parameters were set as follows: local patch size of the StPM was manually determined as a single value for each dataset so that all possible local variations are covered, 141×141 for the chest CT data, $41 \times 41 \times 21$ for the knee MR data, and $15 \times 15 \times 15$ for the brain MR data, respectively; the number of StPM patches were subsequently determined by setting the patch sampling interval so that a certain portion, 0.6–0.7, of adjacent patches overlapped. Based on the target object boundary size, the number of StPM patches were 94, 843 and 523 for the chest CT, knee MR and brain MR data, respectively. The search range for StPM localization was set so that the true optimal localization position was included, while avoiding excess computation. Specifically, 500–600 positions sampled in regular intervals within a cubic volume which was half the patch size with each side. After NCC was computed for all sample positions in the search range, only the 200 coordinates with the highest NCC were retained as the candidates ($q = 200$) in the optimization of (1). $\lambda_{\mathbf{x}}$ and λ_L were empirically determined as 0.1 and 5.0 for the chest CT data, 1.0 and 0.3 for the knee MR data, and 0.1 and 0.15 for the brain MR data, respectively. w_U, was set as 0.01 for large objects like lungs and bones and 0.1 for small or thin objects like cartilages and hippocampus.

All experiments were conducted on a PC with a 2.93 GHz Intel quad-core i7 CPU, and 16 GB of RAM. Computation of patch similarity was accelerated using OpenMP parallelization. Segmentation performance was measured by Dice similarity coefficient (DSC), which is defined as the ratio of the overlapping volume to the combined volume of manual label M and segmentation label L as $\mathrm{DSC}(M, L) = \frac{2\cdot|M\cap L|}{|M|+|L|}$.

3.2. Evaluation of automatic segmentation method using StPM

For each dataset, the evaluation of initial automatic segmentation is performed five times on ten training and ten test subjects randomly selected from the dataset. To compare the performances of initial automatic segmentation, we provide comparison of DSC values obtained by the proposed method (Auto-StPM) to that obtained by the label fusion method (LF) of Heckemann et al. (2006) based on majority voting of all aligned training labels, the label fusion method (PLF) of Coupe et al. (2011) based on non-local weighted voting according to the appearance similarity of local patches, and the label fusion method (SLF) of Tong et al. (2013) based on non-local weighted voting according to the sparse representation of local patches. Here, we used our own implementations for the LF, PLF, and SLF methods. Specifically, the Drop registration (Glocker et al., 2008) with parameter tuning, which took 2 min on average for one-to-one matching, was used to align training data to a test volume. The number of examples for the LF, PLF, and SLF methods as well as the patch sizes for the PLF and SLF methods have been determined empirically by cross validation on the training data. Among the ten training data, seven examples with the highest appearance similarity measured by sum of square distance (SSD), were used for the fusion in the experiments. The patch size was set as 13×13 for the chest CT data, $9 \times 9 \times 9$ for the knee MR data, and $9 \times 9 \times 9$ for the brain MR data, respectively. The SLEP software[3] was used for the sparse coding of SLF method. The patch similarity computation in the PLF method was accelerated by using the OpenMP like the StPM method, while the SLF method was not accelerated because the SLEP software did not provide the parallelization.

[2] https://masi.vuse.vanderbilt.edu/workshop2013/index.php/Main_Page/.

[3] http://www.public.asu.edu/jye02/Software/SLEP/.

Table 1
Average (standard deviation) DSC values and computational time for four automatic methods tested on fifty different test subjects. LF, PLF, SLF, and Auto-StPM denote the label fusion with majority voting (Heckemann et al., 2006), the label fusion with non-local weighted voting (Coupe et al., 2011), the label fusion method based on sparse representation (Tong et al., 2013), and the initial segmentation of proposed method, respectively. The highest DSC and the lowest standard deviation values are highlighted as boldface.

	LF	PLF	SLF	Auto-StPM
Chest CT data set				
R. lung	0.957 (0.0381)	0.958 (0.0387)	0.958 (0.0316)	**0.96** (**0.0206**)
L. lung	0.954 (0.0306)	0.954 (0.0346)	**0.954** (0.0269)	0.952 (**0.0157**)
Avg	0.956 (0.0344)	0.956 (0.0367)	**0.956** (0.0293)	0.956 (**0.0182**)
Time (min.)	1	11	135	**0.3**
Knee MR data set				
Femur	0.928 (0.0187)	0.957 (0.0172)	0.954 (0.013)	**0.959** (**0.0126**)
Tibia	0.916 (0.0219)	0.953 (0.0175)	0.95 (0.0124)	**0.967** (**0.0097**)
F. cartilage	0.415 (0.1054)	0.609 (0.0808)	0.642 (0.0686)	**0.671** (**0.0599**)
T. cartilage	0.272 (0.1146)	0.517 (0.1057)	0.528 (**0.092**)	**0.531** (0.0979)
Avg.	0.633 (0.0652)	0.759 (0.0553)	0.768 (0.0465)	**0.782** (**0.0451**)
Time (min.)	20	110	6200	**4**
Brain MR data set				
L. hippo.	0.778 (0.0407)	0.836 (0.0285)	0.836 (**0.0236**)	**0.842** (0.0259)
R. hippo.	0.776 (0.0461)	0.834 (0.0348)	0.828 (**0.0248**)	**0.835** (0.0321)
L. thalamus	0.885 (0.0236)	**0.912** (**0.0087**)	0.904 (0.0169)	0.904 (0.0115)
R. thalamus	0.891 (0.0215)	0.913 (0.0129)	**0.919** (0.0143)	0.909 (**0.0125**)
Avg.	0.832 (0.033)	**0.874** (0.0212)	0.87 (**0.0199**)	0.873 (0.0205)
Time (min.)	20	35	210	**2**

Table 1 presents the average and standard deviation of DSC values of fifty subjects and the corresponding computational time for each method. Fig. 3 shows the box plots which represent the variance of the DSC values of these results. Generally, the Auto-StPM method outperformed the LF method, while being comparable with the PLF and SLF methods for most cases in terms of accuracy. For the left lung in the chest CT dataset, the median DSC value of the Auto-StPM method was less than the label fusion based methods. Nonetheless, the Auto-StPM method showed to be more stable, in that, it had smaller standard deviation. This is due to the large shape variations of the left lung compared to other objects, which cause large variance for the registration in the label fusion based methods. For the thalamus proper in the brain MR dataset, the PLF and SLF methods slightly outperformed the Auto-StPM method, by less than 0.01, on average. Since the boundary of thalamus proper was often very unclear and had small shape variations, the finely aligned labels might better guide segmentation than the adaptive priors of the StPM on some cases. However, for the other target objects, the proposed method outperformed the label fusion based methods without complex non-rigid registration. In terms of computational time, the Auto-StPM method was 3–5 times faster than the LF method, more than 17 times faster than the PLF method, and more than 105 times faster than the SLF methods. The Auto-StPM method would be more than 20 times faster than the SLF method even if the SLF method was parallelized in the same setting.

We also present the statistical significance between the automatic segmentation results based on p-values obtained by paired t-tests of the DSC scores in Table 2. Except for the lung dataset, the Auto-StPM method obtained results with higher statistical significance than them of the LF method, rejecting the null hypothesis beyond the 95% of confidence level. The Auto-StPM method also statistically outperformed the PLF method for the tibia and femoral cartilage cases and the SLF method for the femur, tibia and femoral cartilage cases. For other target objects, the Auto-StPM method was neither better nor worse with statistical significance, except for the left thalamus proper where the PLF method outperformed the StPM. Overall, the Auto-StPM method was statistically comparable with that by PFL and SFL methods.

Finally, we present the performance change of the Auto-StPM results depending on the StPM with different training dataset size to validate the model-based framework. The experiment was conducted on twenty test volumes. The average DSC values presented in Fig. 4 show that the DSC performance is generally improved as the number of training data increases. This demonstrates the validity of the example-based-model framework, in which the relevant information encapsulated by the StPM becomes better by incrementing the number of examples.

3.3. Evaluation of interactive editing using StPM

To measure the effectiveness of the proposed interactive framework, denoted as IA-StPM, we evaluate the quantitative segmentation accuracy of the results obtained by performing IA-StPM on the results of Auto-StPM segmentation with an StPM constructed from thirty training subjects. We denote this full framework combining Auto-StPM and IA-StPM as Auto+IA-StPM. We compare these results to results obtained from (1) different fully automatic methods, namely, the LF, PLF, SLF and Auto-StPM methods, (2) different editing frameworks, namely, manual correction (Manual), interactive graph cuts specifically modified for local editing (GC-Edit), and IA-StPM, applied to the results of the PLF method, denoted as PLF+Manual, PLF+GC, and PLF+IA-StPM, respectively, and (3) different interactive segmentation frameworks, namely, the graph cuts (GC) method of Shim et al. (2009a) and the *TurtleSeg* (TS) method of Top et al. (2011) based on active learning. In this context, an *editing* framework is one used to correct errors in a precomputed segmentation while a *segmentation* framework is one to compute a clinically satisfactory segmentation from the image.

For interactive editing frameworks, editing comprised both user annotation and re-computing the segmentation was conducted for a fixed amount of time. The correction time of each subject were set as 40 s, 8 min and 8 min for the chest CT images, knee MR images and the brain MR images, respectively. This respectively translated to two, five and six user corrections, on average, for each lung in the chest CT, each bone and cartilage in the knee MR, and each hippocampus and thalamus proper in the brain MR. Computational time to update the segmentation given an additional user

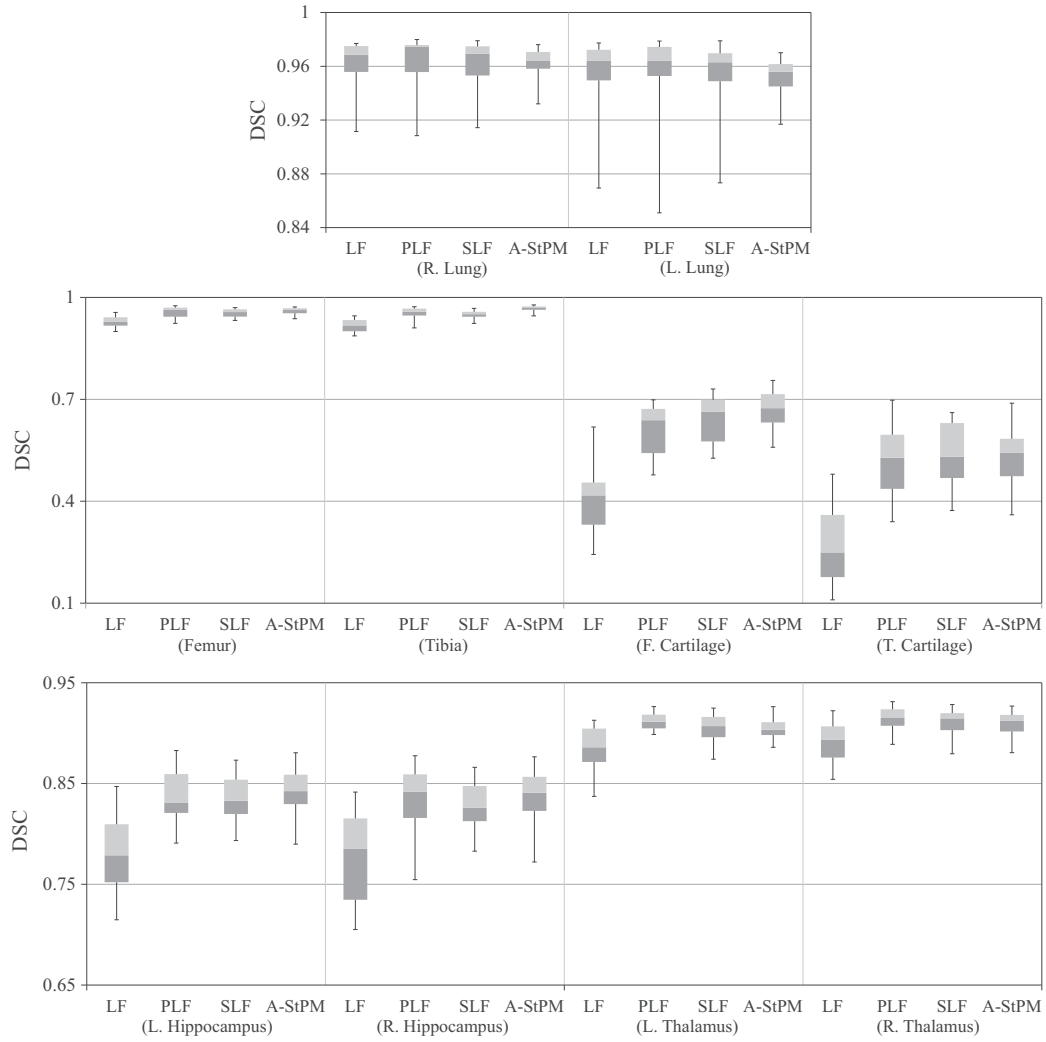

Fig. 3. DSC performance of automatic methods for fifty test volumes. LF, PFL, SPL, and A-StPM denote the label fusion with majority voting (Heckemann et al., 2006), the label fusion with non-local weighted voting (Coupe et al., 2011), the label fusion with sparse representation (Tong et al., 2013), and the initial automatic segmentation results of the proposed method (Auto-StPM), respectively. The DSC values are represented as boxes with top and bottom positions representing the upper and lower quartile and a subdivision representing the median value. The whiskers connected to each box indicate the DSC values of top and bottom 5% subjects.

Table 2
Statistical significance (paired *t*-test) between the DSC values obtained from three label fusion based methods (LF, PLF, SLF) and the initial automatic segmentation of proposed method (Auto-StPM) for fifty different test subjects.

	LF	PLF	SLF
Chest CT data set			
R. lung	0.6047	0.813	0.6739
L. lung	0.5962	0.6263	0.5833
Knee MR data set			
Femur	6.99e−16	0.4199	0.012
Tibia	9.70e−27	3.12e−06	2.06e−13
F. cartilage	1.41e−26	3.30e−05	0.0235
T. cartilage	7.37e−21	0.5086	0.4072
Brain MR data set			
L. hippo.	2.62e−15	0.24	0.2176
R. hippo.	4.17e−11	0.888	0.2462
L. thalamus	1.20e−06	1.64e−04	0.9101
R. thalamus	1.30e−06	0.1113	0.5892

annotation took less than one second for the chest CT and brain MR images, and three seconds for the knee MR due to larger local patch size. The majority of the editing time was spent on finding ambiguous regions and providing the user annotations.

3.3.1. Comparison with fully automatic segmentation methods

The DSC values of automatic segmentation results by the LF, PLF, SLF and Auto-StPM methods with 30 training images tested on 5 test images are measured. For the three label fusion methods, the optimal number of training data was empirically determined as twenty. The average and standard deviation of DSC values and overall computational time of these methods evaluated on five test subjects are presented in Table 3. Since segmentation errors mostly occurred in relatively small ambiguous portions of the image, the segmentation accuracy was largely improved by the IA-StPM even with the small numbers of user annotations, from DSC value 0.961 to 0.975 for lung, 0.813 to 0.833 for knee, 0.89 to 0.905 for brain. In terms of computational time, even when considering the time taken for user annotations, the IA-StPM method was 1.6–5.5 times faster than the LF method, more than 7 times faster than the PLF method, and more than 30 times faster than the SLF methods. The comparison in terms of statistical significance is presented in Table 4. Due to the small number of test subjects, most of the *p*-values are relatively higher than the *p*-values of Table 2. Nonetheless, the IA-StPM framework statistically outperformed the LF, PLF, and SLF methods for all cases, except for the femur in knee MR images and thalamus proper in brain MR images. Although there was no statistical significance for the femur and thalamus proper

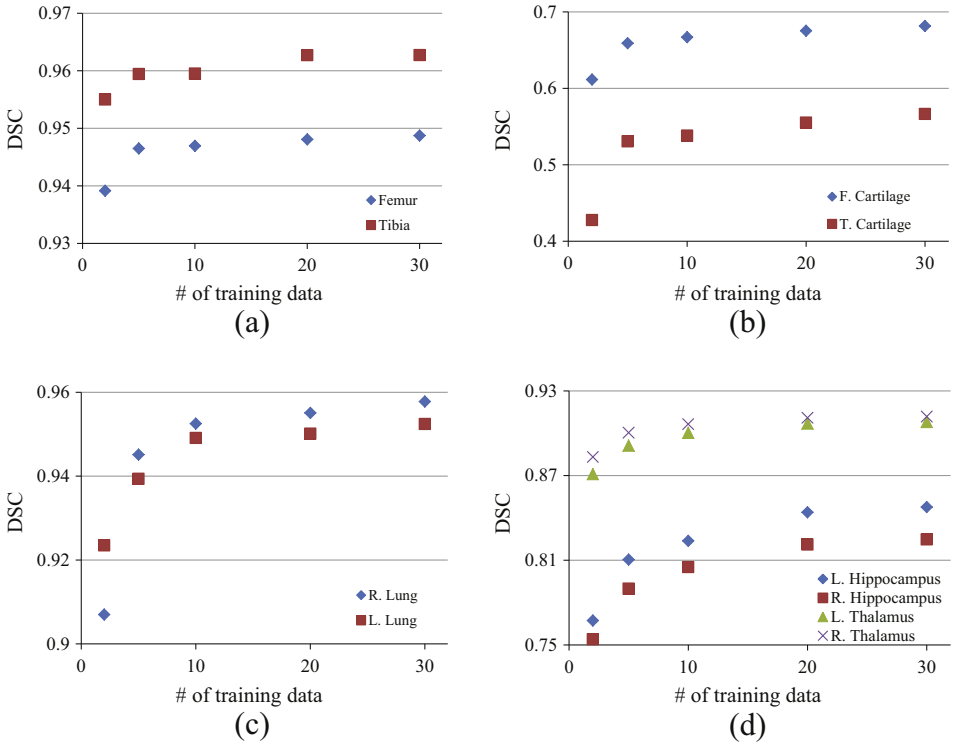

Fig. 4. Change of the initial automatic segmentation accuracy of the Auto-StPM method with different training dataset size 2, 5, 10, 20 and 30 measured by DSC. Values for (a) femur and tibia, (b) femoral and tibial cartilages, (c) right and left lungs, and (d) right and left hippocampus and thalamus proper.

Table 3
Average (standard deviation) DSC values and computational time for three label fusion methods (LF (Heckemann et al., 2006), PLF (Coupe et al., 2011), SLF (Tong et al., 2013)), the initial automatic segmentation of StPM (Auto-StPM) and the interactive segmentation based on StPM (IA-StPM) evaluated on five test subjects. IA-StPM represents results obtained after interactively editing the results of Auto-StPM for a fixed time. The highest DSC and the lowest standard deviation values are highlighted as boldface.

	LF	PLF	SLF	Auto-StPM	Auto+IA-StPM
Chest data					
R. lung	0.967 (0.0088)	0.967 (0.0087)	0.969 (0.0064)	0.972 (0.0047)	**0.977** (**0.0022**)
L. lung	0.957 (0.0082)	0.959 (0.007)	0.959 (0.0071)	0.949 (0.0213)	**0.973** (**0.0045**)
Avg.	0.962 (0.0085)	0.963 (0.0078)	0.964 (0.0067)	0.961 (0.013)	**0.975** (**0.0033**)
Time (min.)	2	22	270	**0.5**	1.2
Knee data					
Femur	0.94 (0.0212)	0.966 (0.0091)	0.966 (0.0114)	0.965 (0.0074)	**0.971** (**0.0031**)
Tibia	0.923 (0.0278)	0.964 (0.0069)	0.964 (0.0078)	0.972 (0.0081)	**0.974** (**0.0065**)
F. cart.	0.523 (0.0802)	0.68 (0.0505)	0.701 (0.0375)	0.719 (0.0183)	**0.746** (**0.0133**)
T. cart.	0.277 (0.1203)	0.584 (0.0283)	0.588 (0.0302)	0.595 (**0.0244**)	**0.641** (0.0259)
Avg.	0.666 (0.0624)	0.798 (0.0237)	0.805 (0.0215)	0.813 (0.0145)	**0.833** (**0.0122**)
Time (min.)	60	240	12,500	**8**	16
Brain data					
L. Hippo.	0.8 (0.0497)	0.859 (0.0207)	0.862 (0.0159)	0.866 (0.0168)	**0.89** (**0.0086**)
R. Hippo.	0.79 (0.046)	0.848 (0.0237)	0.852 (0.0155)	0.858 (0.0134)	**0.882** (**0.009**)
L. Tha.	0.909 (**0.0071**)	0.92 (0.0079)	0.919 (0.0093)	0.917 (0.0099)	**0.922** (0.0076)
R. Tha.	0.907 (0.0087)	0.922 (0.0081)	0.921 (**0.0076**)	0.918 (0.0106)	**0.925** (0.0088)
Avg.	0.852 (0.0279)	0.887 (0.0151)	0.889 (0.0121)	0.89 (0.0127)	**0.905** (**0.0085**)
Time (min.)	60	90	520	**5**	13

cases, the DSC values of edited results were nonetheless larger than the comparison methods for all test subjects (five subjects for each dataset). We expect that the statistical significance would be better as the number of test subjects increase.

3.3.2. *Comparison with different interactive editing methods*

The segmentation accuracy of results obtained by the PLF+Manual, PLF+GC-Edit, and PLF+IA-StPM are measured. We note that as long as an appropriate StPM has been constructed, the IA-StPM can be applied to results obtained by any other method. The IA-StPM is initialized by optimizing Eq. (1) and localizing the StPM among candidate positions sampled from the surface boundary of the given precomputed segmentation. After initialization, the IA-StPM is performed identically as when following Auto-StPM. For the GC method, here, we applied a modified version of the method of Shim et al. (2009a) so that the segmentation result is changed only in

Table 4
Statistical significance (paired t-test) between the DSC values obtained from automatic methods (LF, PLF, SLF, Auto-StPM) and interactive StPM method (IA-StPM) for five test subjects.

	LF	PLF	SLF	Auto-StPM
Chest CT data set				
R. lung	0.0375	0.0373	0.0252	0.0451
L. lung	0.0049	0.0041	0.005	0.0389
Knee MR data set				
Femur	0.0171	0.3529	0.4281	0.3049
Tibia	0.0037	0.0344	0.0449	0.5899
F. cartilage	0.0003	0.0226	0.0357	0.0286
T. cartilage	0.0002	0.0109	0.0186	0.021
Brain MR data set				
L. Hippo.	0.0041	0.0137	0.0083	0.022
R. Hippo.	0.0023	0.0167	0.0054	0.0094
L. Thalamus	0.0222	0.6475	0.5884	0.3295
R. Thalamus	0.0118	0.5121	0.447	0.2552

Table 5
Average (standard deviation) DSC values for different interactive editing methods tested on five test subjects. The editing was conducted for a fixed time on results obtained by fully automatic methods – the patch based label fusion (PLF) method for the left three columns and Auto-StPM method for the rightmost column. Manual, GC-Edit, and IA-StPM denote the manual editing method, the modified local graph cut method, and the proposed interactive method, respectively. The highest DSC and the lowest standard deviation values are highlighted as boldface.

	PLF+Manual	PLF+GC	PLF+IA-StPM	Auto+IA-StPM
Chest data				
R. lung	0.971 (0.0119)	0.974 (0.0054)	0.977 (0.0026)	**0.977** (**0.0022**)
L. lung	0.962 (0.0057)	0.965 (0.0061)	**0.974** (**0.0044**)	0.973 (0.0045)
Avg.	0.967 (0.0088)	0.97 (0.0058)	**0.976** (0.0035)	0.975 (**0.0033**)
Knee data				
Femur	0.968 (0.0078)	0.97 (0.0043)	**0.973** (**0.0028**)	0.971 (0.0031)
Tibia	0.965 (0.0055)	0.969 (0.0063)	0.973 (**0.0046**)	**0.974** (0.0065)
F. cart.	0.706 (0.042)	0.711 (0.0318)	0.739 (0.0266)	**0.746** (**0.0133**)
T. cart.	0.601 (0.0258)	0.622 (0.0265)	0.632 (0.0221)	**0.641** (**0.0259**)
Avg.	0.81 (0.0203)	0.818 (0.0172)	0.829 (0.014)	**0.833** (**0.0122**)
Brain data				
L. Hippo.	0.871 (0.0078)	0.869 (0.0115)	0.885 (**0.0074**)	**0.89** (0.0086)
R. Hippo.	0.87 (0.0072)	0.865 (0.0104)	0.88 (**0.0055**)	**0.882** (0.009)
L. Tha.	0.922 (0.0079)	0.919 (0.0096)	**0.924** (0.0082)	0.922 (**0.0076**)
R. Tha.	0.924 (0.0088)	0.923 (0.0091)	**0.926** (**0.0087**)	0.925 (0.0088)
Avg.	0.897 (0.0079)	0.894 (0.0102)	0.904 (**0.0075**)	**0.905** (0.0085)

regions near user annotations, hence the notation GC-Edit. Without this modification, unwanted change may occur, frequently, in correct regions far from the annotations. Also, unary likelihoods are computed based on intensity histograms of annotated voxels together with voxels of the initial segmentation for foreground and background, respectively. For Manual method, only voxels that receive user annotations are updated directly. For all methods, annotations are given as scribbles using a mouse as in common drawing applications, where brush size and image zoom can be controlled.

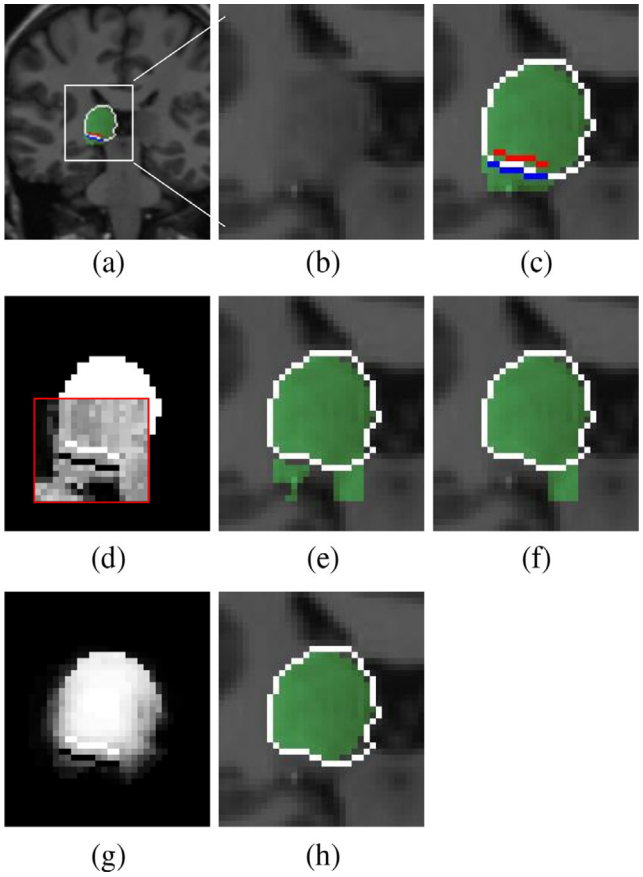

Fig. 5. (a) Initial segmentation by the Auto-StPM method (green), the ground truth boundary (white line), and foreground (red) and background (blue) user annotations. (b) Enlarged view of image region in white square of (a) and (c) enlarged view including label information. (d) The local likelihood computed by the appearance model for the modified graph cut (GC-Edit) method (Shim et al., 2009a). The GC-Edit is conducted on the local region (red box) near the user annotations. (e) GC-Edit result with low smoothness cost and (f) GC-Edit result with high smoothness cost. (g) The local likelihood computed by adaptive priors of StPM. (h) The result of IA-StPM editing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5 presents the performances of editing results. Since the Manual editing method required accurate and fine user annotations in erroneous regions, the performance was limited compared to the other methods. Though the GC-Edit method gave better results for regions with clearly distinguishable boundaries, it often gave much worse results for boundaries with low contrast as shown in Fig. 5. Especially, due to the similar foreground and background appearance of the hippocampus and thalamus proper, the performance of GC-Edit was even worse than that of Manual method. On the other hand, the PLF+IA-StPM and Auto+IA-StPM methods gave robust results even on ambiguous regions. Also, editing can be more effectively performed for cases where large errors occur on a small number of local regions. Specifically, a few local user scribbles sufficed to correct the errors for the lung, bones and hippocampus (0.008–0.03 DSC gain) because their errors occurred on small number of ambiguous weak boundary regions. On the other hand, the gains of GC-Edit and IA-StPM editing methods were relatively small (0.005–0.007 DSC gain) for the thalamus proper because the automatic segmentation boundary was mostly close to the true boundary. In this case, erroneous segmentation regions in the form of thin strips are distributed evenly

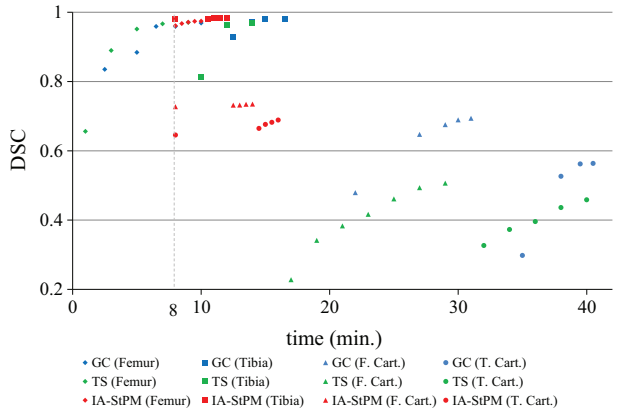

Fig. 6. DSC performance versus the cumulative processing time for segmentation of femur, tibia, femoral cartilage, and tibial cartilage from a knee MR image. Blue, green, red represent the results of the graph cut (GC) method (Shim et al., 2009a), *TurtleSeg* method (TS) (Top et al., 2011), and the proposed method (StPM), respectively. The processing times for GC and TS linearly increase for the number of objects. On the other hand, for the StPM method, editing is performed on the initial segmentation results obtained automatically for the four compartments within 8 min (gray line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

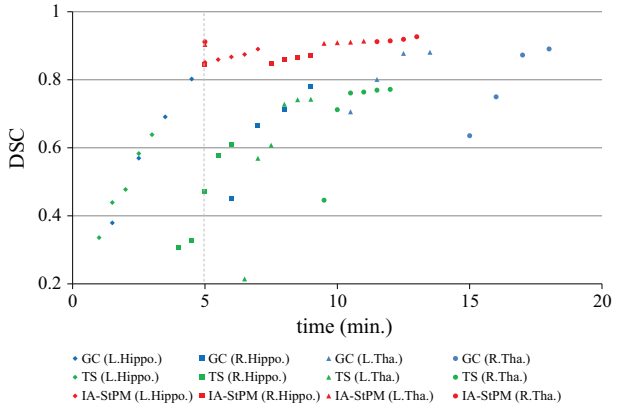

Fig. 7. DSC performance versus the cumulative processing time for segmentation of the right and left hippocampus and thalamus proper from a brain MR image. Blue, green, red represent the results of the graph cut (GC) method (Shim et al., 2009a), *TurtleSeg* method (TS) (Top et al., 2011), and the proposed method (StPM), respectively. The processing times for GC and TS linearly increase for the number of objects. On the other hand, for the StPM method, editing is performed on the initial segmentation results obtained automatically for the four compartments within 5 min (gray line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

along the boundary, making it difficult for the user to insert scribbles.

3.3.3. *Qualitative comparison of segmentation efficiency*

We compare the Auto+IA-StPM method with two interactive methods, namely, the graph cuts (GC) method of Shim et al. (2009a) and the *TurtleSeg* (TS) method of Top et al. (2011) based on active learning, as well as the different editing frameworks applied to the results of the PLF method (PLF+Manual, PLF+GC, PLF+IA-StPM) in the previous subsection. The comparison is performed only on three-dimensional image data sets since the TS method is not applicable to two-dimensional images. For the GC and TS methods, implementations by the original authors were used. Since the GC method requires a considerable amount of user scribbles, especially for background regions, the software provides

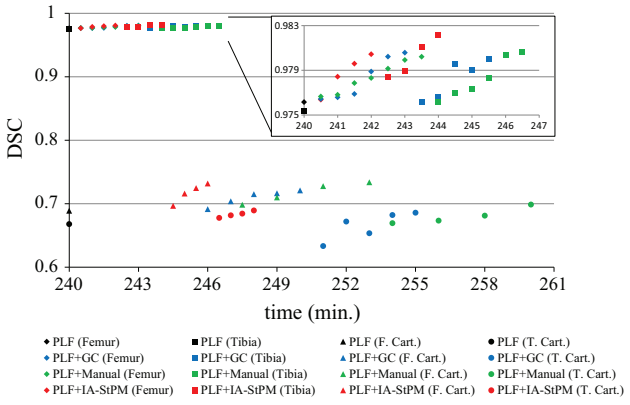

Fig. 8. DSC performance versus the cumulative processing time for segmentation of femur, tibia, femoral cartilage, and tibial cartilage from a knee MR image. Editing was performed on the segmentation results obtained by patch based label fusion (PLF) method. Black, green, blue, red represent the results of the PLF, PLF+Manual method, PLF+GC-Edit method (Shim et al., 2009a), and PLF+IA-StPM method, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

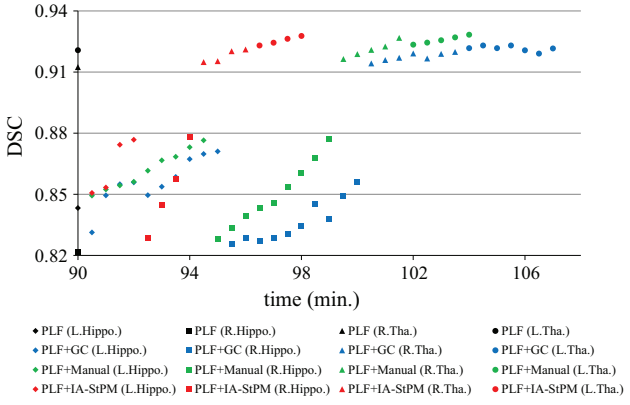

Fig. 9. DSC performance versus the cumulative processing time for segmentation of the right and left hippocampus and thalamus proper from a brain MR image. Editing was performed on the segmentation results obtained by patch based label fusion (PLF) method. Black, green, blue, red represent the results of the PLF, PLF+Manual method, PLF+GC-Edit method (Shim et al., 2009a), and PLF+IA-StPM method, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a crop function so that the segmentation is conducted on a specific region when the target object is small within a large image. To limit computation and provide a fair comparison, the GC optimization was conducted on a cropped region of interest covering the target objects. For the TS method, we used the default setting provided in the authors website.[4] Figs. 6 and 7 show the DSC values of the GC, TS, and StPM methods versus the cumulative annotation and processing time for segmenting multiple objects in a test subject. Here, the multiple objects were sequentially segmented. The computational time of GC and TS methods increased proportionally to the number of target objects of the test subject. On the other hand, the proposed StPM was much more efficient due to the initial automatic segmentation which simultaneously computed results for all objects. Specifically, the average computational time required for the Auto-StPM segmentation was 8 and 5 min, for MR images of the knee joint comprising four compartments and MR images of the brain

[4] http://www.turtleseg.org.

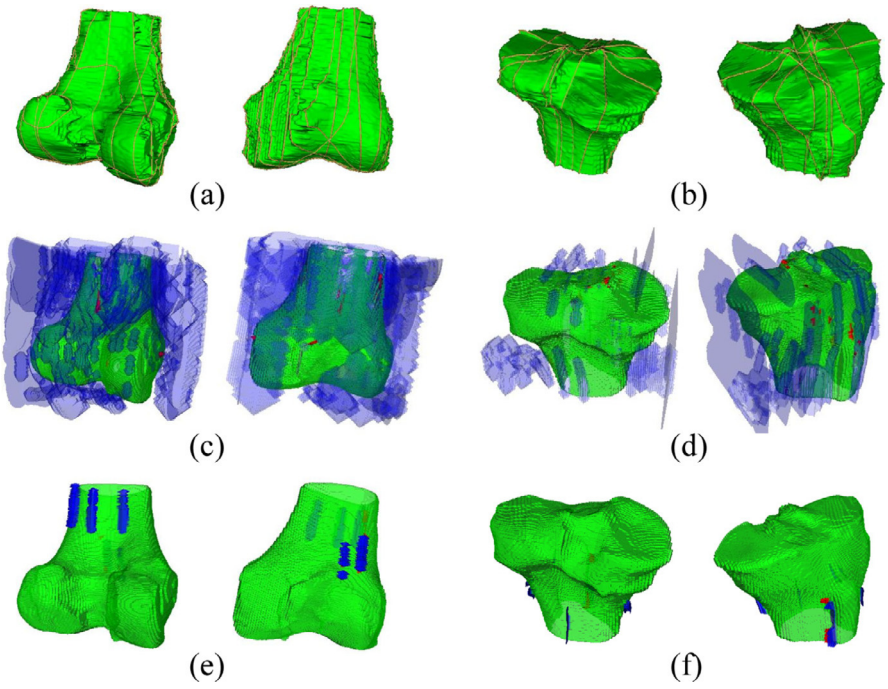

Fig. 10. Comparison of required user annotations to obtain similar segmentation results (green) within front and back views of femur and tibia. (a) and (b): User annotations for the *TurtleSeg* (TS) method (Top et al., 2011) represented as orange lines; (c) and (d): user annotations for the graph cut (GC) method (Shim et al., 2009a) represented as red (object) and blue (background) scribbles, respectively; (e) and (f): user annotations for the proposed IA-StPM method represented as red (object) and blue (background) scribbles, respectively. More than 10 delineations and 20 pairs of object and background scribbles are required for the TS and GC methods, respectively. On the other hand, 5 pairs of scribbles are required for the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

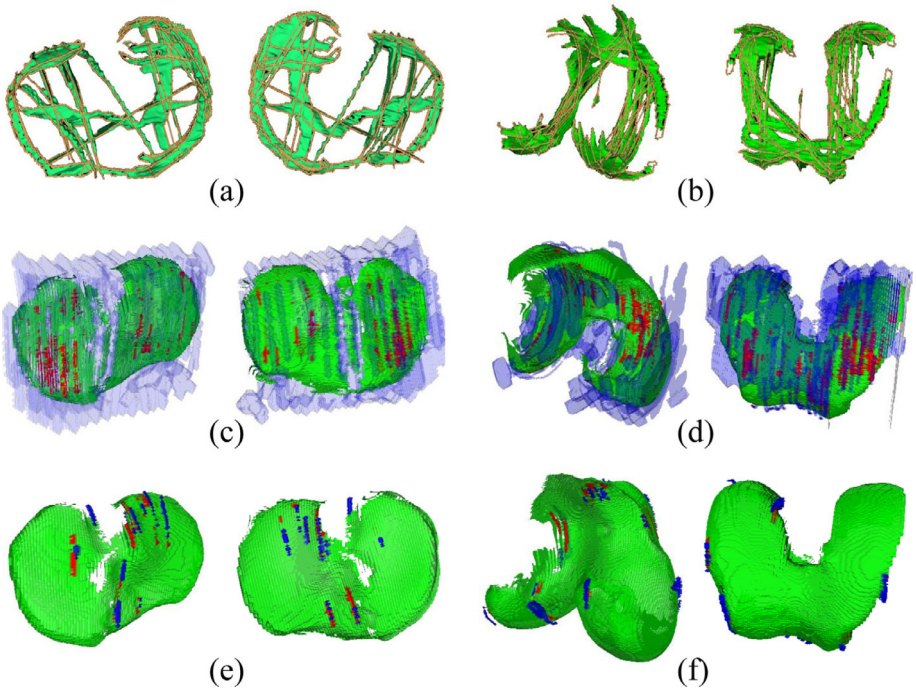

Fig. 11. Comparison of required user annotations to obtain similar segmentation results (green) within front and back views of femoral and tibial cartilage. (a) and (b): User annotations for the *TurtleSeg* (TS) method (Top et al., 2011) represented as orange lines; (c) and (d): user annotations for the graph cut (GC) method (Shim et al., 2009a) represented as red (object) and blue (background) scribbles, respectively; (e) and (f): user annotations for the proposed IA-StPM method represented as red (object) and blue (background) scribbles, respectively. More than 10 delineations and 20 pairs of object and background scribbles are required for the TS and GC methods, respectively. On the other hand, 8 pairs of scribbles are required for the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

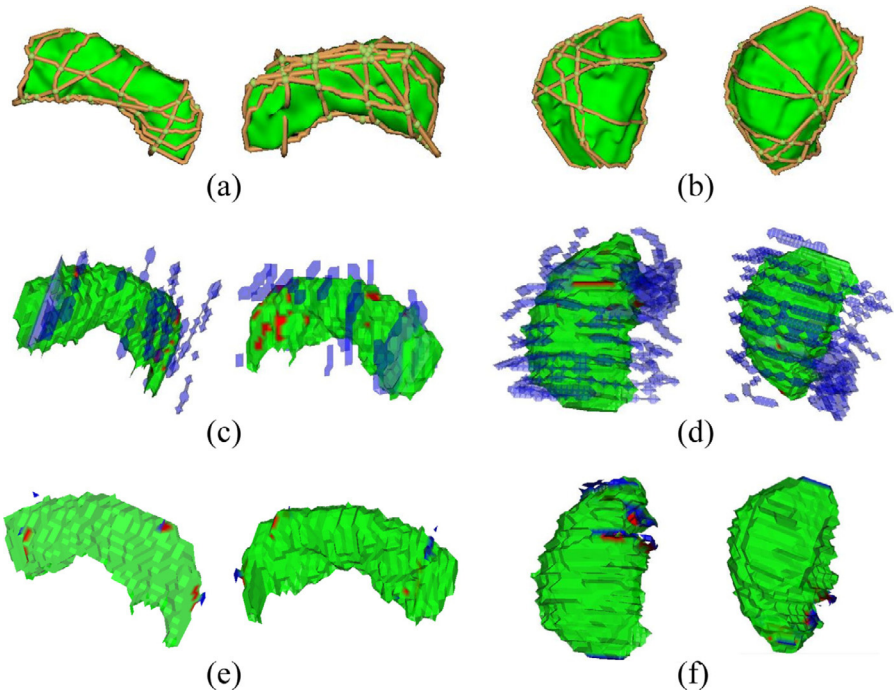

Fig. 12. Comparison of required user annotations to obtain similar segmentation results (green) within front and back views of hippocampus and thalamus proper. (a) and (b): User annotations for the *TurtleSeg* (TS) method (Top et al., 2011) represented as orange lines; (c) and (d): user annotations for the graph cut (GC) method (Shim et al., 2009a) represented as red (object) and blue (background) scribbles, respectively; (e) and (f): user annotations for the proposed IA-StPM method represented as red (object) and blue (background) scribbles, respectively. More than 7 delineations and 10 pairs of object and background scribbles are required for the TS and GC methods, respectively. On the other hand, 4 pairs of scribbles are required for the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

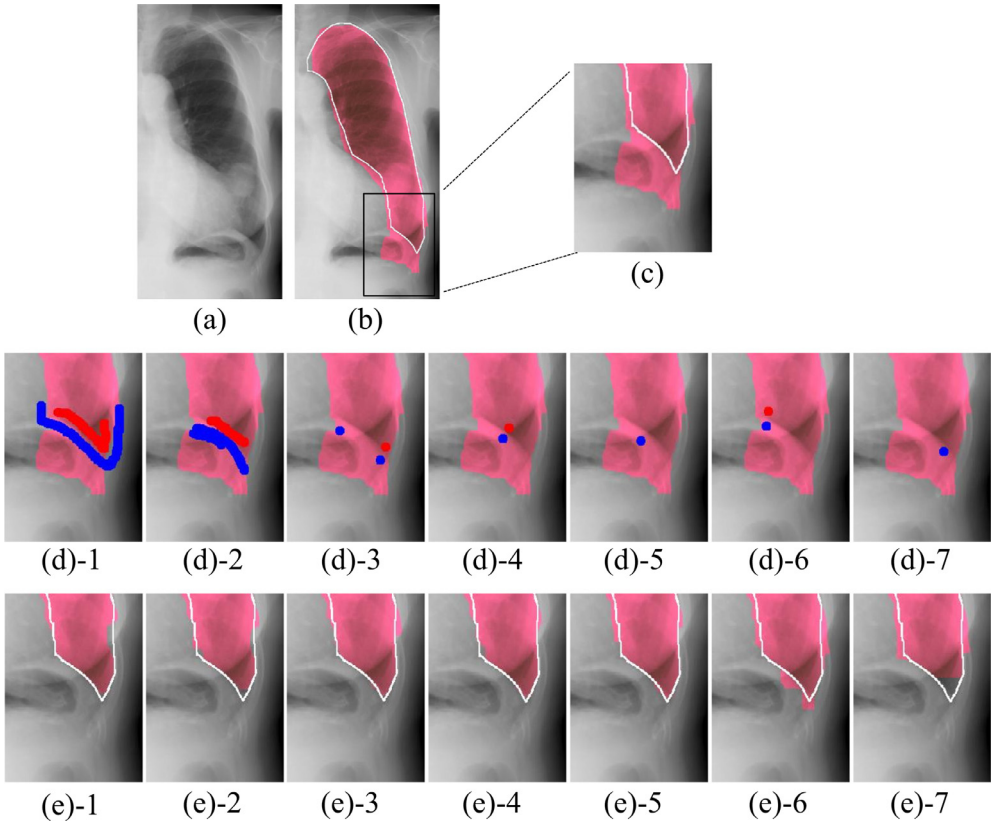

Fig. 13. Robustness of the proposed IA-StPM method regarding different user annotations. (a) Test image, (b) initial segmentation (magenta) and ground truth boundary (white line), (c) enlarged view of region corresponding to the black square in (b), (d)-1 to (d)-7: different user annotations (red, blue markings represent the object and background, respectively) in the region of (c), and (e)-1 to (e)-7: corrected results based on the StPM corresponding to the annotations in (d)-1 to (d)-7. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

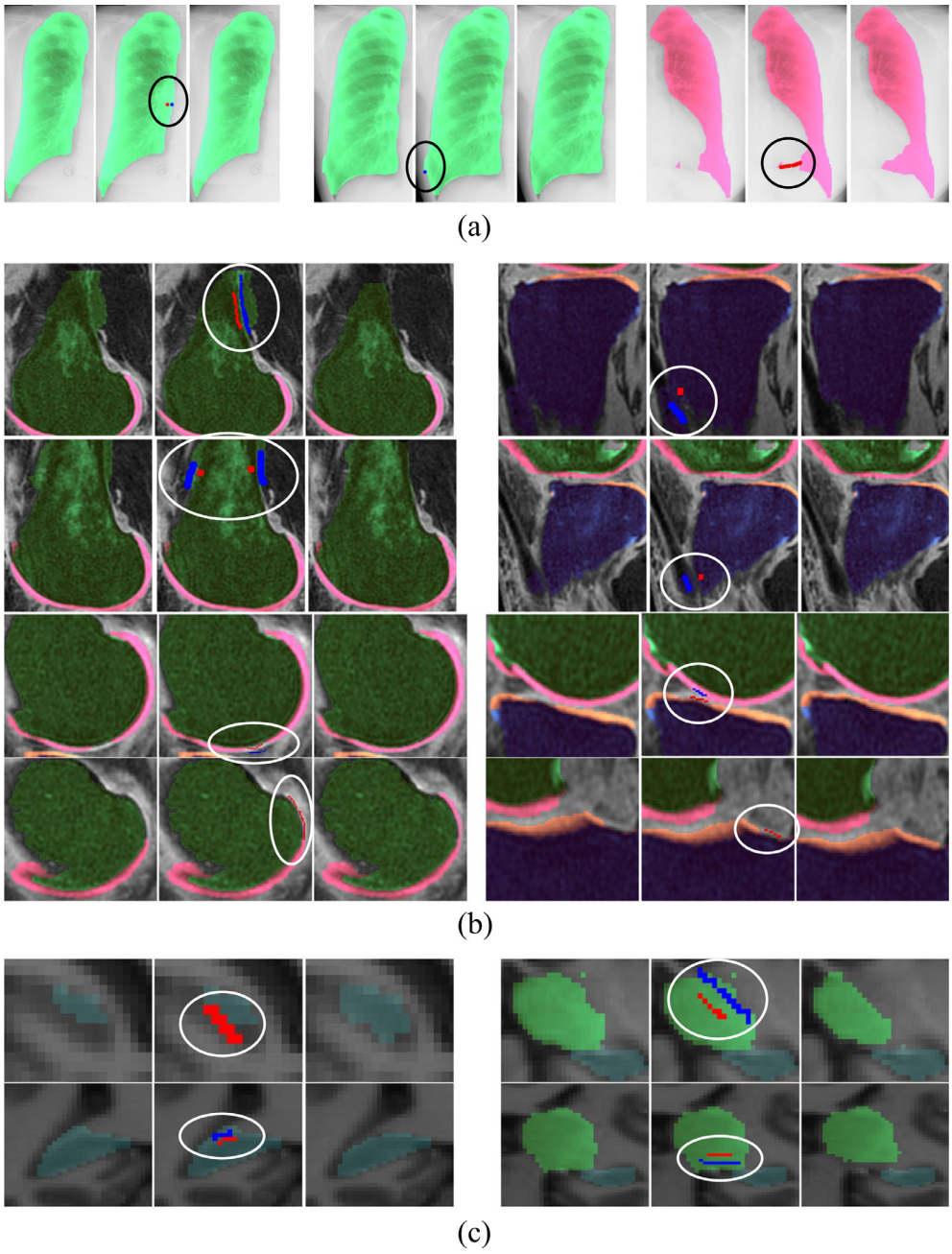

Fig. 14. Qualitative segmentation results of (a) lungs from 2D lung CT images, (b) bone (left) and cartilage (right) from 3D knee and (c) hippocampus (left) and thalamus proper (right) from 3D brain MR images. Initial results from Auto-StPM, user scribbles (highlighted in circles), and corrected results are shown in first, second, third column of each set. Segmentation is successfully corrected from small amount of user annotations.

comprising fourteen compartments, respectively. To obtain the same level of accuracy with the Auto-StPM results, the GC method required over 40 and 18 min and the TS method over 40 and 15 min, respectively. Moreover, the accuracy of the TS method often converged on a lower DSC value than the other methods, depending on the target object shape. On the other hand, since the editing of StPM started from the initial segmentation results, laborious initial user annotation was largely reduced.

Figs. 8 and 9 show the DSC values of PLF+Manual, PLF+GC, and PLF+IA-StPM methods versus the cumulative annotation and processing time for the same test subject. Since these methods also assume an initial segmentation like the Auto+IA-StPM method, user interaction can be reduced significantly. However, the starting time of editing was relatively slow due to the high complexity of the PLF method. Excluding the starting time, the performance change of PLF+IA-StPM method was similar with that of Auto-IA-StPM method shown in Figs. 6 and 7. On the other hand, much more time was required to improve the segmentation accuracy using the Manual or GC method. Furthermore, when using the GC method, segmentation accuracy even became worse for some editing steps because new errors were introduced on ambiguous regions, as in the example shown in Fig. 5. On the other hand, since the user annotations can be given as a small number of dots or rough scribbles in the IA-StPM method, the correction time to obtain the same level of accuracy, 8 min, was less than the PLF+Manual and PLF+GC methods, which were over 20 min and 16 min for knee image and 14 min and 20 min for brain image, respectively (see Figs. 8 and 9).

Fig. 10, Fig. 11 and Fig. 12 show the difference of user annotations for the three interactive methods. Specifically, for the GC method, the user was required to provide scribbles surrounding most of the true boundary to prevent the segmentation 'leaking' over, especially on boundaries with weak image gradient such as upper part of femur, bottom part of tibia, boundaries between cartilages, and the thalamus proper. Relatively, the TS method required a smaller amount of user delineations since they were required only on uncertain 2-D planes. Although the TS method was effective to simple shaped objects like tibia and thalamus proper, we observed that the required number of delineations significantly increased for thin or deformable objects like the knee cartilages, causing the user to repeatedly delineate the boundary in planes with similar orientations. On the other hand, the required annotation of the StPM method was much less than the other methods and was also much less dependent on the shape of the target object.

Fig. 13 shows an example of the robustness of the StPM method to the placement and quantity of user annotations. When user scribbles were roughly given in regions that require modification, the segmentation was correctly updated for most cases, even with small amounts, as small as a dot, and for substantially different amounts and positioning. Moreover, we can see in Fig. 13(d)-6/(e)-6 and (d)-7/(e)-7 that the StPM enables correction of the segmentation error even though the user annotations are placed disproportionately relative to the where the boundaries require modification. We argue that this property helps to reduce the inter variability between different users. Fig. 14 shows further qualitative results.

4. Conclusion

In this paper, we have proposed a new interactive framework for robust and efficient segmentation of target objects from a large number of medical images. To address this problem, we incorporate the high-level prior knowledge of training data represented as the structured patch model (StPM) into the interactive framework. The proposed framework is flexible and effective and is perhaps more useful for clinical applications compared to previous fully automatic methods in which the performance can be critically effected by target object characteristics or parameter settings. This is made possible since the global shape structure as well as the local shape and appearance are well represented by the proposed StPM. Within the interactive framework, configuration of the priors and structure of the StPM and the segmentation results are repeatedly updated whenever more user annotation is given. The performance was compared with the three label fusion based methods (LF (Heckemann et al., 2006), PLF (Coupe et al., 2011), and SLF (Tong et al., 2013)) and the interactive methods based on graph cuts (Shim et al., 2009a) and active learning (Top et al., 2011) for various target objects from chest CT, knee MR, and brain MR datasets. In terms of accuracy and statistical significance, the Auto-StPM method outperformed the LF method for all target objects except lung, while was comparable with the PLF and SLF methods. On the other hand, the IA-StPM method with few user annotations outperformed the LF method for all target objects, and the PLF and SLF methods for the most target objects except femur and thalamus proper. Furthermore, the proposed framework was considerably more efficient, requiring approximately one fifth for knee joint MR images containing femur, tibia, and cartilages and one third for brain MR images containing hippocampus and thalamus proper, respectively, of the time required by the compared interactive methods. In our evaluation, each editing step, consisting of updating both the StPM configuration and local segmentation, was conducted in less than three seconds. Since appropriate priors were initially determined overall, segmentations can be efficiently corrected and are robust to variations in user annotations. As the size of patch sets in the StPM increases, the accuracy of segmentation results will most likely improve as well. Thus, we believe that the proposed framework can be applied to facilitate construction of larger databases and to conduct longitudinal studies more efficiently.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2A10004550) and the Soonchunhyang University Research Fund (No. 20140227).

References

Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images atlas selection and its effect on accuracy. NeuroImage 46, 726–738.

Asman, A.J., Landman, B.A., 2013. Non-local statistical label fusion for multi-atlas segmentation. Med. Image Anal. 17, 194–208.

Bai, X., Sapiro, G., 2007. A geodesic framework for fast interactive image and video segmentation and matting. In: Proceedings of IEEE International Conference on Computer Vision, pp. 1–8.

Bai, W., Shi, W., O', D.P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N.S., Rueckert, D., 2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans. Med. Imag. 32, 1302–1315.

Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B., 2009. PatchMatch: a randomized correspondence algorithm for structural image editing. In: Proceedings of SIGGRAPH, pp. 24:1–24:11.

Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A., 2010. The generalized PatchMatch correspondence algorithm. In: Proceedings of European Conference on Computer Vision, pp. 29–43.

Barrett, W.A., Mortensen, E.N., 1997. Interactive live-wire boundary extraction. Med. Image Anal. 1, 331–341.

Boykov, Y., Funka-Lea, G., 2006. Graph cuts and efficient N-D image segmentation. Int. J. Comput. Vision 70, 109–131.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23, 1222–1239.

Branson, S., Perona, P., Belongie, S., 2011. Strong supervision from weak annotation: Interactive training of deformable part models. In: Proceedings of IEEE International Conference on Computer Vision, pp. 1832–1839.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models: their training and application. Comput. Vis. Image Underst. 61, 38–59.

Coupe, P., Manjon, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L., 2011. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage 54, 940–954.

Duta, N., Sonka, M., 1998. Segmentation and interpretation of MR brain images: an improved active shape model. IEEE Trans. Med. Imag. 17, 1049–1062.

Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., 2010. Object detection with discriminatively trained part based models. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1627–1645.

Gleason, S., Sari-Sarraf, H., Abidi, M., Karakashian, O., Morandi, F., 2002. A new deformable model for analysis of X-ray CT images in preclinical studies of mice for polycystic kidney disease. IEEE Trans. Med. Imag. 21, 1302–1309.

Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N., 2008. Dense image registration through MRFs and efficient linear programming. Med. Image Anal. 12, 731–741.

Grady, L., 2006. Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1768–1783.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33, 115–126.

Heimann, T., Morrison, B.J., Styner, M.A., Niethammer, M., Warfield, S.K., 2010. Segmentation of knee images: a grand challenge. In: Proceedings of Medical Image Computing and Computer Assisted Intervention, pp. 207–214.

Ibragimov, B., Likar, B., Pernu, F., Vrtovec, T., 2012. A game-theoretic framework for landmark-based image segmentation. IEEE Trans. Med. Imag. 31, 1761–1776.

Kim, T.H., Lee, K.M., Lee, S.U., 2008. Generative image segmentation using random walks with restart. In: Proceedings of European Conference on Computer Vision, pp. 264–275.

Kim, T.H., Lee, K.M., Lee, S.U., 2010. Nonparametric higher-order learning for interactive segmentation. In: Proceedings of IEEE Computer Vision and Pattern Recognition, pp. 3201–3208.

Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26, 147–159.

Lempitsky, V., Kohli, P., Rother, C., Sharp, T., 2009. Image segmentation with a bounding box prior. In: Proceedings of IEEE International Conference on Computer Vision, pp. 277–284.

Lotjonen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage 49, 2352–2365.
Park, S.H., Lee, S., Yun, I.D., Lee, S.U., 2013. Hierarchical MRF of globally consistent localized classifiers for 3D medical image segmentation. Pattern Recogn. 46, 2408–2419.
Park, S.H., Yun, I.D., Lee, S.U., 2013b. Data-driven interactive 3D medical image segmentation based on structured patch model. In: Information Processing in Medical Imaging, pp. 196–207.
Pohle, R., Toennies, K.D., 2001. Segmentation of medical images using adaptive region growing. In: Proceedings of SPIE Medical Imaging, pp. 1337–1346.
Rother, C., Kolmogorov, V., Blake, A., 2004. "GrabCut" – Interactive foreground extraction using iterated graph cuts. In: Proceedings of SIGGRAPH, pp. 309–314.
Rousseau, F., Habas, P.A., Studholme, C., 2011. Human brain labeling using image similarities. In: Proceedings of IEEE Computer Vision and Pattern Recognition, pp. 1081–1088.
Schwarz, T., Heimann, T., Tetzlaff, R., Rau, A.M., Wolf, I., Meinzer, H.P., 2008. Interactive surface correction for 3d shape based segmentation. In: Proceedings of SPIE Medical Imaging, pp. 69143O-1–69143O-8.
Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P., 2007. A new deformable model for analysis of X-ray CT images in preclinical studies of mice for polycystic kidney disease. IEEE Trans. Med. Imag. 26, 1115–1129.
Shi, W., Caballero, J., Ledig, C., Zhuang, X., Bai, W., Bhatia, K., de Marvao, A.M.S.M., Dawes, T., ORegan, D., Rueckert, D., 2013. Cardiac image super-resolution with global correspondence using multi-atlas PatchMatch. In: Proceedings of Medical Image Computing and Computer Assisted Intervention, pp. 9–16.
Shim, H., Chang, S., Tao, C., Wang, J., Kwoh, C., Bae, K., 2009. Knee cartilage: efficient and reproducible segmentation on high-spatial-resolution MR images with the semi-automated graph-cut algorithm method. Radiology 251, 548–556.
Shim, H., Kwoh, C., Yun, I., Lee, S., Bae, K., 2009b. Simultaneous 3-d segmentation of three bone compartments on high resolution knee MR images from osteoarthritis initiative (OAI) using graph-cuts. In: Proceedings of SPIE Medical Imaging, pp. 72593P-1–72593P-8.
Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., Doi, K., 2000. Development of a digital image database for chest radiographs with and without a lung nodule. Am. J. Roentgenol. 174, 71–74.
Sukno, F., Ordas, S., Butakoff, C., Cruz, S., Frangi, A., 2007. Active shape models with invariant optimal features: application to facial analysis. IEEE Trans. Med. Imag. 29, 1105–1117.
Sun, S., Sonka, M., Beichel, R.R., 2013. Lung segmentation refinement based on optimal surface finding utilizing a hybrid desktop/virtual reality user interface. Comput. Med. Imag. Graph. 37, 15–27.
Ta, V.T., Giraud, R., Collins, D.L., Coupe, P., 2014. Optimized PatchMatch for near real time and accurate label fusion. In: Proceedings of Medical Image Computing and Computer Assisted Intervention, pp. 105–112.
Tong, T., Wolz, R., Coup, P., Hajnal, J.V., Rueckert, D., 2013. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage 71, 11–23.
Top, A., Hamarneh, G., Abugharbieh, R., 2011. Active learning for interactive 3D image segmentation. In: Proceedings of Medical Image Computing and Computer Assisted Intervention, pp. 603–610.
van der Lijn, F., den Heijer, T., Breteler, M.M., Niessen, W.J., 2008. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. NeuroImage 43, 708–720.
van Ginneken, B., Frangi, A.F., Staal, J.J., ter Haar Romeny, B.M., Viergever, M.A., 2002. Active shape model segmentation with optimal features. IEEE Trans. Med. Imag. 21, 924–933.
Wah, C., Branson, S., Perona, P., Belongie, S., 2011. Multiclass recognition and part localization with humans in the loop. In: Proceedings of IEEE International Conference on Computer Vision, pp. 2524–2531.
Wang, D., Yan, C., Shan, S., Chen, X., 2012. Active learning for interactive segmentation with expected confidence change. In: Proceedings of Asian Conference on Computer Vision, pp. 790–802.
Yang, Y., Ramanan, D., 2011. Articulated pose estimation with flexible mixtures-of-parts. In: Proceedings of IEEE Computer Vision and Pattern Recognition, pp. 1385–1392.
Yang, W., Cai, J., Zheng, J., Luo, J., 2010. User-friendly interactive image segmentation through unified combinatorial user inputs. IEEE Trans. Image Process. 19, 2470–2479.
Zhang, S., Zhan, Y., Metaxas, D.N., 2012. Deformable segmentation via sparse representation and dictionary learning. Med. Image Anal. 16, 1385–1396.