Inference With Collaborative Model for Interactive Tumor Segmentation in Medical Image Sequences

Liang Lin, Wei Yang, Chenglong Li, Jin Tang, and Xiaochun Cao

***Abstract*—Segmenting organisms or tumors from medical data (e.g., computed tomography volumetric images, ultrasound, or magnetic resonance imaging images/image sequences) is one of the fundamental tasks in medical image analysis and diagnosis, and has received long-term attentions. This paper studies a novel computational framework of interactive segmentation for extracting liver tumors from image sequences, and it is suitable for different types of medical data. The main contributions are twofold. First, we propose a collaborative model to jointly formulate the tumor segmentation from two aspects: 1) region partition and 2) boundary presence. The two terms are complementary but simultaneously competing: the former extracts the tumor based on its appearance/texture information, while the latter searches for the palpable tumor boundary. Moreover, in order to adapt the data variations, we allow the model to be discriminatively trained based on both the seed pixels traced by the Lucas–Kanade algorithm and the scribbles placed by the user. Second, we present an effective inference algorithm that iterates to: 1) solve tumor segmentation using the augmented Lagrangian method and 2) propagate the segmentation across the image sequence by searching for distinctive matches between images. We keep the collaborative model updated during the inference in order to well capture the tumor variations over time. We have verified our system for segmenting liver tumors from a number of clinical data, and have achieved very promising results. The software developed with this paper can be found at http://vision.sysu.edu.cn/projects/med-interactive-seg/**

***Index Terms*—Collaborative model, medical image analysis, spatio-temporal inference, tumor segmentation.**

I. Introduction

FOR decades, medical imaging technologies, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US), have played a central role in tumor detection and diagnosis as well as surgery planning [1]. Since radiologists usually consider diagnosis and planning based on analyzing tumors' properties with image slices, accurate tumor extraction is generally essential and quite beneficial. Manual annotation of tumors, however, is time-consuming especially with a large amount of image sequences, and the delineation quality often depends on the operators. Hence, computational medical image segmentation receives long-term attentions for clinical analysis, and various segmentation methods have been proposed.

Liver cancer is the second most frequent cause of cancer death worldwide in men and the sixth in women, according to the report in [2]. Tumor segmentation is critical for liver clinical diagnosis and surgery. Therefore, this paper aims to develop an interactive segmentation system[1] for extracting liver tumors from medical image sequences (e.g., volumetric images or videos), and it is suitable for different types of medical data (e.g., CT, MRI, or US). Particularly, we consider the following two difficulties to build such a system.

1) There are a great diversity of liver tumor types (e.g., hemangioma, focal nodular hyperplasia, and hepatocellular carcinoma) with various modalities [1], [3], [4], and tumors are sometimes difficult to be distinguished from healthy tissues. In addition, the quality of medical imaging is highly affected by devices and individual variances. Thus, it is very challenging to construct a universal model (or detector) to separate well the tumors against background tissues, even a few user intervention could be allowed.
2) Medical image sequences usually include many slices (say more than 100), and it is impractical to operate on every slice to assist segmentation. Given one segmented slice, the system is required to automatically propagate the segmentation into consecutive image slices. It is a nontrivial task because the shape or appearance of tumors probably vary over the image sequence because of the organ's physiological deformation. In addition, the imaging condition sometimes temporally changes, which alters the brightness and the contrast of images.

We address the above issues from two aspects. First, we present a collaborative model to capture the tumor variations by tightly integrating region and structure (shape) information. More precisely, the model comprises two coupled terms: 1) region partition and 2) boundary presence. The two terms

Manuscript received May 27, 2014; revised March 27, 2015, June 22, 2015, and October 1, 2015; accepted October 8, 2015. Date of publication October 29, 2015; date of current version November 15, 2016. This work was supported in part by the Natural Science Foundation of China under Grant 61472002, in part by the Guangdong Natural Science Foundation under Grant S2013050014548 and Grant 2014A030313201, in part by the Program of Guangzhou Zhujiang Star of Science and Technology under Grant 2013J2200067, and in part by the Science and Technology Program of Guangzhou under Grant 1563000439. This paper was recommended by Associate Editor M. Shin.

L. Lin and W. Yang are with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China, and L. Lin is also with SYSU-CMU Shunde International Joint Research Institute, Shunde, China (e-mail: linliang@ieee.org; platero.yang@gmail.com).

C. Li and J. Tang are with the School of Computer Science and Technology, Anhui University, Hefei 230601, China (e-mail: lcl1314@foxmail.com; ahhftang@qq.com).

X. Cao is with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100864, China (e-mail: caoxiaochun@iie.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2015.2489719

[1]The software and testing data are available at: http://vision.sysu.edu.cn/projects/med-interactive-seg/

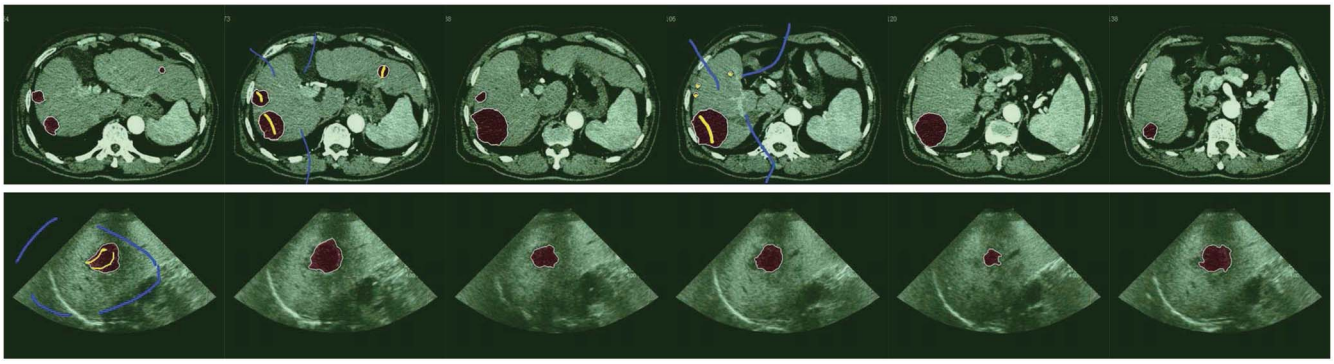


Fig. 1. Segmentation results generated by our system in a noncontrast CT volumetric images (first row) and a US image sequence (second row). There are 73 slices in the CT volumetric images, and three foreground scribbles plus three background scribbles have been placed in the tenth slice. To polish the segmentation, four foreground scribbles and three background scribbles are further added in the 42nd slice. The US image sequence contains 36 slices, where we draw two foreground scribbles and four background scribbles in the first slice. For illustration, results in this figure are sampled from the sequences. The tumors are covered by red masks and the user scribbles are highlighted by yellow (foreground) and blue (background), respectively.

not only provide the complementary information for segmentation but also compete to each other to conduct the consistent solution. The region term is defined based on the appearance/texture difference between the tumor and background, while the boundary term is defined based on the local structure/edge features. In the previous works on medical image segmentation, most of these methods consider only region appearance [4], [5] or only boundary shape [6], [7]. A number of methods had been suggested to combine region appearance with the boundary shape [8]–[10] and showed very promising results. Compared with these methods, our model is allowed to be discriminatively trained with user placed scribbles, and the pixels annotated by the scribbles are treated as either positive or negative examples. Unlike some interactive segmentation systems with fixed parameters [11], [12], we adjust the model parameters to adapt the variations of tumors with different types of medical data.

Second, we propose an efficient iterative algorithm for the inference of image sequence segmentation comprising two following main steps.

1) We optimize the model for segmenting the key slice that can be either the first slice of a sequence or an arbitrary slice specified by the radiologist. In our representation, the two terms (i.e., region partition and boundary presence) are tightly coupled, and each of them is defined to assign a discrete label (i.e., tumor or nontumor) for pixels, resulting in a nonconvex formulation. The model is thus intractable to be analytically optimized. To overcome this problem, we relax the formulation into an L_1 regularized convex optimization, and then propose an alternating direction method of multipliers (ADMMs) to solve the model [13], [14], which can guarantee to converge to global optimal solution and each subproblem has the closed form solution.
2) We propagate the segmentation from the key slice into the rest of image slices by extracting the correspondences over the images. A batch of discriminative pixels are selected as seed points from the segmented tumor area and nontumor area of the key slice. We trace these points across the consecutive image slices while removing some matching outliers. The traced points at each image slice can be then treated as training samples to incrementally update the collaborative model (i.e., to make the model adapt to the new data). Afterward, we generate the accurate segmentations for the images by applying the optimization method in the first step. Similar techniques for segmentation propagation were studied in video segmentation [15]. The main difference of our approach with theirs is that we keep our model updated during the propagation process to adapt appearance variation of the tumors against surrounding tissues.

The inference procedure can be flexibly assisted and refined by hand. Radiologists are allowed to add/remove scribbles to correct the segmentation on any single slices; the refinement is also available during the propagation by selecting a new key slice and retriggering the process. One can also fix segmentations on some slices to avoid modification during the refinement process. Two examples of liver tumor segmentation generated by our system are presented in Fig. 1, where one is processed on a nonenhanced CT volumetric images, and the other on an US image sequence. For illustration, we sample a few results from the image sequence, and the tumors are covered by red masks and the background by green masks, while the user scribbles are highlighted by yellow (foreground) and blue (background), respectively.

We have applied our system with several clinical medical data and shown very promising results. To show the effectiveness of our model, we first compare our method on single slice segmentation with two well-applied interactive segmentation method (i.e., the GrabCut [11], and the geodesic active contour (GAC) [12]), and one fully supervised algorithm (i.e., the semantic texton forest [16]). We then conduct the experiments on image sequence segmentation and show the superior performance of our method over the active contour model [17] and the hybrid level-set method (HLM) [18]. Moreover, we empirically evaluate how the main components contributing to the framework, as well as the robustness of user interactions.

The main contributions of our system are summarized as follows. First, we propose an effective collaborative model to represent tumors in terms of coupling region and boundary information, which can be incrementally updated during segmentation. Second, a novel inference framework is developed to efficiently and accurately segment the images and image sequences with very few user interactions. Last, the proposed system is very applicable for clinic applications due to its robustness and flexibility.

A. Literature Review

In medical informatics, CT and MRI provide an exceptional resolution to tumor analysis, and many computational segmentation approaches on CT and MRI have been extensively studied [1], [3], [5], [19]–[23]. Compared with CT and MRI, US imaging is cheaper, safer, and less laborious, but is more challenging for segmentation due to the speckle noise, low contrast, and low resolution of imaging [24]. By far, not many works focus on US data [24], and very few approaches are designed to handle multiple types of medical data due to the appearance diversity caused by different imaging mechanism.

Generally, there are several types of methods for medical image segmentation, e.g., deformable models, thresholding approaches, statistical models, and interactive methods. Deformable models, especially active contour models [12], [25] and level set methods [26], were widely used for liver tumor segmentation [1], [3], [6], [7], [27]. Lu *et al.* [6] solved CT volumetric images segmentation by combining the active contour segmentations in CT slices. Wong *et al.* [7] further incorporated the size and shape constraints to obtain more accurate segmentations. Smeets *et al.* [3] proposed a 3-D tumor image sequence segmentation method by adopting the 3-D level set method and fuzzy pixel classification method. Li *et al.* [1] extended the traditional level set method by incorporating image gradient, region competition and prior information, and achieved impressive results.

For improving segmentation accuracy, thresholding and morphological techniques were further introduced [20]. Hame [19] segmented the liver tumors with the thresholding and morphological operations and refined by spatial fuzzy clustering. Despite acknowledged successes, the performances of these methods are not always satisfactory in practice, as they rely on some strict assumptions on tumor appearances (e.g., close boundaries or distinct regions against surrounding tissues), and the parameters of the models are often tuned and fixed.

Recently, the methods based on statistical learning techniques are introduced to model the variations of tumor appearance and shape. Corso *et al.* [28] proposed a Bayesian tumor segmentation framework for incorporating soft model assignments into the calculation of affinities to take a step toward bridging the gap between bottom-up affinity-based segmentation and top-down generative model. Kubota [5] localized candidate tumors by training a detector on small voxels of CT volumetric images, and obtained the final segmentation by postprocessing. Stawiaski *et al.* [21] introduced the MRF model to solve the tumor segmentation together with an unsupervised watershed method. These approaches performed well on challenging scenarios, but often relied on a large amount of labeled data for offline supervised learning [29], and might be problematic in extracting tumors with large within-class variance.

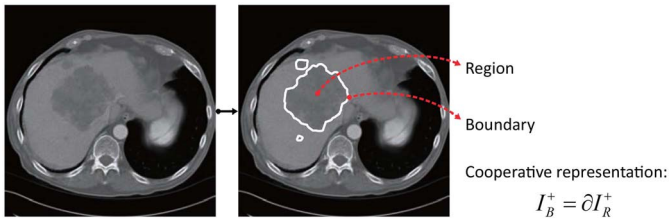


Fig. 2. Collaborative representation of a CT image with liver tumor. The image domain is decomposed into foreground region I_R^+ and background region I_R^-. We also distinguish boundaries I_B^+ from the rest of image I_B^-. We strictly constrain that $I_B^+ = \partial I_R^+$ to couple the two terms in the collaborative representation.

Besides, many user-aided (i.e., interactive) approaches [1], [3], [4], [6], [20], [21], [30] have been exploited and are generally able to produce more reliable segmentation or decomposition results than the automatic methods, at the cost of utilizing user interactions. The ways of interaction were analogous, such as specifying seed points of tumor areas [3], [4], [31], tumor-centered regions of interest (ROIs) [1], [6], [20], and drawing scribbles on the tumor and nontumor area [21], [32]. As the radiologists can flexibly manipulate the segmentation, these semi-automatic systems are very suitable for clinic applications. However, the user interactions are often very heavy and laborious for some challenging cases, particularly for a long slice sequence, e.g., placing scribbles carefully in every slice.

The rest of this paper is organized as follows. Section II describes the representations for our model. Section III proposes the inference method for image sequence segmentation. The experimental results and comparisons are presented in Section IV. This paper is concluded in Section V with the discussions on future extensions of this paper.

II. Collaborative Model

In this section, we present the formulation of our collaborative model for tumor segmentation and discuss the model parameter learning method.

We formulate the segmentation from two different aspects: 1) to partition the image I into foreground (i.e., tumor) region I_R^+ and background (i.e., nontumor) region I_R^- and 2) to localize tumor boundary I_B^+ from the rest of image I_B^-. These two terms are constrained to consistently conduct the identical segmentation solution, so that we make these two terms strictly coupled in definition, namely $I_B^+ = \partial I_R^+$. We present an example for illustration in Fig. 2.

For a pixel x in I, we adopt two indicator functions $\phi(x)$ and $\theta(x)$ to define the solutions of region partition and boundary presence, respectively, as

$$\phi(x) = \begin{cases} 1; x \in I_R^+ \\ 0; x \in I_R^- \end{cases}; \theta(x) = \begin{cases} 1; x \in I_B^+ \\ 0; x \in I_B^-. \end{cases} \quad (1)$$

The coupling of the two solutions can be then defined by $\theta = |\nabla\phi|$ over the whole image domain, where ∇ represents the gradient operator. Thus, we present the dual form energy function as

$$E(\phi, \theta) = \sum_x \phi(x) T_R(x) + \alpha \sum_x \theta(x) T_B(x)$$
$$\text{s.t.} \quad \theta = |\nabla\phi| \tag{2}$$

where α is a tuning parameter weighing the two tests, and generally we choose some $\alpha < 1$. $T_R(x)$ and $T_B(x)$ are the region term and boundary term, respectively, and defined the probability form as

$$T_R(x) = -\log \frac{P_R(\phi(x) = 1|x)}{P_R(\phi(x) = 0|x)} \tag{3}$$

$$T_B(x) = -\log \frac{P_B(\theta(x) = 1|x)}{P_B(\theta(x) = 0|x)}. \tag{4}$$

The formulation is partially motivated by the classical region competition model proposed by Zhu and Yuille [33]. For computational efficiency, we define $T_R(x)$ and $T_B(x)$ as binary classifiers that can be discriminatively trained with diverse image features. In implementation, we train the region term $T_R(x)$ online according to the foreground and background scribbles placed by the radiologists and allow it to be incrementally updated during processing. $T_B(x)$ is trained offline, similar to learning-based edge detector [34], as it is expensive to let the user to label tumor boundary during the segmentation process.

For training the two terms T_R and T_B, we specify the posterior probabilities $P_R(\phi(x)|x)$ and $P_B(\theta(x)|x)$ by the exponential family distribution, as

$$P_R(\phi(x) = 1|x) = \frac{\exp\left(\sum_m \beta_m h_m(x; \gamma_m)\right)}{1 + \exp\left(\sum_m \beta_m h_m(x; \gamma_m)\right)} \tag{5}$$

where $h_m(x; \gamma_m)$ is a weak classifier (i.e., a specific appearance feature from the feature vector) with the parameter (threshold) γ_m, and β_m is the corresponding coefficient. The form of $P_B(\theta(x)|x)$ is defined exactly the same as $P_R(\phi(x)|x)$ except for the image feature. Equation (3) can be thus formulated as

$$\log \frac{P_R(\phi(x) = 1|x)}{P_R(\phi(x) = 0|x)} = \sum_m \beta_m h_m(x; \gamma_m). \tag{6}$$

In this paper, the Adaboost algorithm [35] is employed to efficiently train this model in an incremental fashion. Specifically, we aim to update only parts of the parameters (i.e., γ_m and β_m) rather than retraining the complete model when the model needs to be adjusted duconsecutive ones.

A. Feature Descriptors

For capturing the region appearance in $P_R(\phi(x)|x)$, we concatenate a local gray-level histogram and the dense SIFT vector [36] to describe each pixel x. In implementation, the gray-level histogram is a eight-bin vector computed on the 5×5 image domain covering x. Dense SIFT descriptor is extracted by dividing the 12×12 image domain into a 4×4 cell array, and in each cell we use a vector including eight digits to characterize image gradient in different directions. By combining the two features together, each pixels is represented as a 136-bin vector. For the boundary feature in $P_B(\phi(x)|x)$, we compute the generic Haar wavelets on the image domain covering the pixel x across multiple scales. It is worth mentioning that the wavelets extracted from different scales implicitly incorporate the mid-level and context information and greatly facilitate the boundary model training.

III. Inference

In this section, we conduct the inference algorithm for image sequence segmentation by two main steps: 1) optimize the model to accurately segment the current slice and 2) propagate the segmentation to the unprocessed slices.

A. Model Optimization for Segmentation

The collaborative model in (2), which tightly couples the region and boundary information, is intractable to be optimized analytically as the discrete indicator ϕ and θ make the formulation nonconvex. On the other hand, some nonconvex optimization methods [37] usually converge very slow or require strict constraints, and might be unsuitable for our task. In this paper, we seek an equivalent form of the model for inducing a convex formulation.

Thus, we relax the binary indicator ϕ into a continuous function ranging in $[0, 1]$, and θ is also transformed into the continuous form according to the constraint in (2), i.e., $\theta = |\nabla\phi| \in [0, 1]$. This relaxation transforms the original optimization objective into an equivalent continuous convex formula

$$\min_{\phi,\theta} \phi T_R + \alpha|\theta|T_B, \ \text{s.t.}\ \theta = \nabla\phi \tag{7}$$

where

$$\phi T_R = \sum_x \phi(x) T_R(x)$$
$$|\theta|T_B = \sum_x |\theta(x)| T_B(x) \tag{8}$$

and ∇ denotes the gradient operator.[2]

Even the model in (7) is convex, the L_1 term makes this target nonsmooth and impractical to simultaneously update ϕ and θ using the classical optimization methods such as conjugate gradient or Gauss–Seidel methods [39]. To overcome this difficulty, we exploit the ADMM algorithm to update ϕ and θ alternatively. ADMM is closely related with the split-Bregman [14], Douglas–Rachford splitting [40], and inexact augmented Lagrangian (AL) method [41], and can converge to the global optimal solution [42]. Given the optimal solution of this target, the final segmentation can then be generated by thresholding ϕ on pixels.

We adopt the ADMM [13] to solve the convex optimization problem in (7), where the AL function is defined as

$$\mathcal{L}(\phi, \theta, \lambda, \mu) = \phi T_R + \alpha|\theta|T_B + \frac{\lambda}{2}||\theta - \nabla\phi||_2^2 + \mu^T(\theta - \nabla\phi) \tag{9}$$

[2]As in [38], ∇ can be treated as either a gradient filter or a circulant matrix. To save notation, we use the same ∇ to the gradient filter and its matrix representation, and this should not cause ambiguity by referring to the context.

where $\lambda > 0$ is the AL penalty parameter and μ is a vector of Lagrangian multiplier. With minor algebra, the AL function can be equivalent represented as

$$\mathcal{L}(\phi, \theta, \lambda, b) = \phi T_R + \alpha|\theta|T_B + \frac{\lambda}{2}||\theta - \nabla\phi - b||_2^2 \quad (10)$$

where $b = \mu/\lambda$.

Using the ADMM algorithm, the subproblem on ϕ can be formulated as

$$\phi^{k+1} = \min_{\phi} \phi T_R + \frac{\lambda}{2}\left\|\theta^k - \nabla\phi - b^k\right\|_2^2. \quad (11)$$

Using the strategy adopted in [38], we can derive the closed form solution on ϕ. The optimal solution ϕ^{k+1} should satisfy

$$\phi^{k+1} = \left(\lambda\nabla^T\nabla\phi\right)^{-1}\left(\nabla^T\phi\left(\theta^k - b^k\right) - T_R\right). \quad (12)$$

With the fast Fourier transform (FFT), the closed form solution on ϕ can be rewritten as

$$\phi^{k+1} = \mathrm{FFT}^{-1}\left(\frac{\mathrm{FFT}\left(\lambda\nabla^T\phi\left(\theta^k - b^k\right) - T_R\right)}{\lambda\mathrm{FFT}\left(\nabla^T\nabla\phi\right)}\right) \quad (13)$$

where "$\frac{\cdot}{\cdot}$" denotes the element-wise division. Note that $\mathrm{FFT}(\nabla^T\nabla\phi)$ is zero in $(0, 0)$. In our implementation, we exclude the $(0, 0)$th element in the element-wise division.

The subproblem on θ can be formulated as

$$\theta^{k+1} = \min_{\theta} |\theta|T_B + \frac{\lambda}{2}\left\|\theta - \nabla\phi^k - b^k\right\|_2^2 \quad (14)$$

which can be analytically solved by the soft-thresholding operator

$$\begin{aligned}\theta_x^{k+1} &= \mathcal{T}_\delta\left(\nabla\phi_x^k + b_x^k\right) \\ &= \frac{\nabla\phi_x^k + b_x^k}{\left|\nabla\phi_x^k + b_x^k\right|} * \max\left(\left|\nabla\phi_x^k + b_x^k\right| - \delta, 0\right)\end{aligned} \quad (15)$$

where δ is the parameter of shrinkage method.

In ADMM, b is updated using the following rule:

$$b^{k+1} = b^k + \left(\nabla\phi^{k+1} - \theta^{k+1}\right). \quad (16)$$

To accelerate the convergence speed, the penalty parameter λ is initialized with a small value $\lambda_{\min}$ and is updated after each iteration using

$$\lambda^{k+1} = \min\left(\rho\lambda^k, \lambda_{\max}\right) \quad (17)$$

where $\rho > 1$ is a positive scalar.

Eckstein and Bertsekas [43] proved the convergence of the ADMM algorithm.

Theorem 1 [43]: Consider the problem $\min_{\theta,\phi} f(\theta) + g(\phi)$, s.t. $\theta = G\phi$, where G is full column rank, and f and g are closed, proper convex functions. Consider arbitrary $\lambda > 0$ and θ^0, ϕ^0, and b^0. Let $\eta^k \geq 0, k = 0, 1, 2, \ldots$ and $\nu^k \geq 0, k = 0, 1, 2, \ldots$ be two sequence with

$$\sum_{k=0}^{\infty}\eta^k \leq \infty \quad \text{and} \quad \sum_{k=0}^{\infty}\nu^k \leq \infty.$$

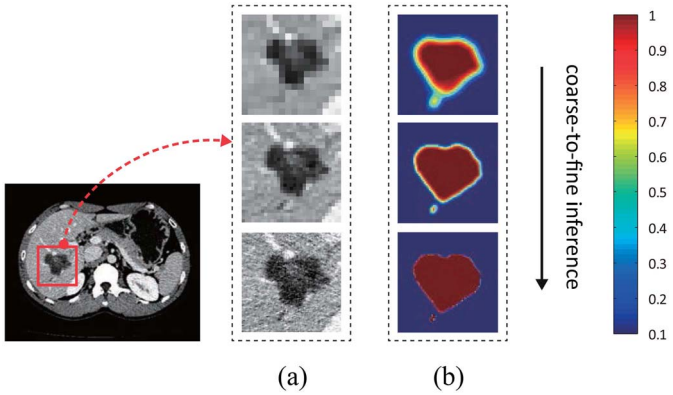


Fig. 3. (a) Inference is conducted at different image scales in a coarse-to-fine manner. An image pyramid is first built by repeated smoothing and subsampling processing. (b) Energy defined in (2) is optimized from the coarser layers to the finer layers.

Consider the three sequences $\{\theta^k\}$, $\{\phi^k\}$, and $\{b^k\}$ $(k = 0, 1, \ldots)$ that satisfy

$$\begin{aligned}&\left\|\theta^{k+1} = \arg\min_{\theta} f(\theta) + \frac{\lambda}{2}\left\|G\phi - \theta^k - b^k\right\|_2^2\right\| \leq \eta_k \\ &\left\|\phi^{k+1} = \arg\min_{\phi} g(\phi) + \frac{\lambda}{2}\left\|G\phi^{k+1} - \theta - b^k\right\|_2^2\right\| \leq \nu_k \\ &b^{k+1} = b^k - \left(G\phi^{k+1} - \theta^{k+1}\right).\end{aligned}$$

If the problem has the global optimal solution θ^*, then the sequence θ^k converges to θ^*.

It is obvious that the (7) satisfies the requirement of Theorem 1, and the subproblems on ϕ and θ have the closed form solutions. Thus, the proposed optimization algorithm converges to the global minimum.

1) Coarse-to-Fine Computation: In order to further improve the efficiency, we conduct the inference at different scales of image in a coarse-to-fine manner. This multiscale solving also makes the system more robust against noise. In implementation, an image pyramid is first built by repeated smoothing and subsampling operations. Then the inference starts from the coarser layers to the finer layers. Each pixel x in a specific layer obtains a response $\phi(x) \in [0, 1]$, indicating the confidence of predicting pixel x as foreground (tumor) or background (nontumor). We then perform thresholding on these responses. Pixels with responses greater than the upper threshold (or smaller than the lower threshold) are predicted as foreground (or background) with high confidences and are fixed for optimization at the finer image layers. The rest of pixels with ambiguous responses will be further repredicted in the finer layers. The multiscale processing is illustrated in Fig. 3.

B. Segmentation Propagation

For automatically propagating the segmentation from the segmented slice $\hat{I}$ to its consecutive slices, we propose a simple yet effective algorithm for solving this problem by selecting and tracing feature correspondences over consecutive slices

while keeping the collaborative model updated. Our algorithm relies on the following physiological observations.

1) The variations of tumor shape (boundary) are usually smooth between two consecutive slices.
2) The rough location of the tumor is generally stable with respect to the background organism (i.e., the liver).

Thus, we first select a batch of seed pixels from the nontumor area of segmented slice $\hat{I}$ and also enforce them apart from the tumor boundary. We denote I_R^- and I_B^+ as the nontumor area and the tumor boundary of $\hat{I}$, respectively. For each pixel x belonging to I_R^-, we compute its distance to the boundary I_B^+ as

$$d\big(x, I_B^+\big) = \min \mathrm{U}(x, y), \quad \text{s.t.} \quad y \in I_B^+ \tag{18}$$

where $\mathrm{U}(x, y)$ is the Euclidean distance between two pixels x and y. By thresholding, we prune pixels with small $d(x, I_B^+)$ and construct the set of seed pixels (denoted as $\mathcal{S}_b$) by randomly sampling from the remaining pixels.

Then we search for the seed pixels from the tumor area with respect to the selected $\mathcal{S}_b$. The pixels having distinct differences against the background nontumor areas are encouraged to be chosen as matching correspondences, according to the research of visual tracking [44], [45]. Hence, we define the following criteria for determining the set of seed pixels in the tumor area (denoted by $\mathcal{S}_f$):

$$\begin{aligned} \mathcal{S}_f^* &= \arg\max_{\mathcal{S}_f} \frac{h(\mathcal{S}_f)}{h(\mathcal{S}_b)} \\ &= \arg\max_{\mathcal{S}_f} \sum_{x \in \mathcal{S}_f} \sum_{y \in \mathcal{S}_b} \mathrm{KL}(h(x) \parallel h(y)) \\ &\quad \text{s.t.} \quad S_f \in I_R^+ \end{aligned} \tag{19}$$

where $h(\cdot)$ denotes the feature representation of pixels. Note that we use the same feature as we discussed in Section II. And the KL distance is defined as

$$\mathrm{KL}(h(x) \parallel h(y)) = \sum_m h_m(x) \ln \frac{h_m(x)}{h_m(y)}. \tag{20}$$

Afterward, we trace the pixels in both S_b and S_f across the consecutive image slices by Lucas–Kanade algorithm [46]. Note that some outliers during tracing can be easily removed by voting. In these consecutive slices, the traced seed pixels (i.e., the matches of S_b and S_f) can be treated as either positive or negative samples for updating the region term of our model in (3). At last, we segment these consecutive slices by optimizing the energy with the updated model as we discussed in the last section. The overall inference procedure is sketched in Algorithm 1.

Algorithm 1 Inference Sketch for Image Sequence Segmentation

Input:

A sequence of medical image slices $\{I_k\}_{k=1}^K$, $K \geq 1$.

Output:

Segmentation results of all images in the sequence.

1 Select one key slice $\hat{I}$ at first.
2 Draw scribbles on tumor and non-tumor areas in $\hat{I}$.
3 Extract training examples from the scribbles, and train the model defined in Eq. (3).
4 Perform segmentation on $\hat{I}$ by transferring the model into a new convex formulation in Eq. (7).

WHILE not all slices segmented
1 Set $\tilde{I}$ as the unsegmented slice adjacent to $\hat{I}$.
2 Propagate segmentation from $\hat{I}$ to $\tilde{I}$:
a) Select discriminative seed pixels from $\hat{I}$ (i.e., S_f from the tumor area and S_b from the non-tumor area) by Eq. (19).
b) Trace S_f and S_b into $\tilde{I}$ by the Lucas-Kanade algorithm;
c) Update the model parameters by treating the traced points as training samples.
3 Perform segmentation on $\tilde{I}$ by optimizing the model.
4 $\hat{I} \leftarrow \tilde{I}$.
END

IV. Experiments

We test our system on several data sets: SYSU-CT, SYSU-US, and ITKs liver tumor (ILT) data set. The SYSU-CT and SYSU-US are both provided by the first affiliated hospital, Sun Yat-sen University, and ILT is a public data set available online.[3] The SYSU-CT data set is constructed by seven CT volumetric images of liver tumor from different patients; all the patients were scanned using a 64 detector row CT machine (Aquilion64, Toshiba Medical System). The ILT data set consists of six different patient CT volumetric images of liver; all CT slices are 512×512 pixels with an in-plane resolution of 0.6–0.8 mm, and are 5-mm slice spacing. The SYSU-US data set consists of 20 US image sequences of abdomen with liver tumor. The used equipment was Aplio SSA-770A (Toshiba Medical System). The ground truths are carefully annotated by experts. All the experiments are carried out on an Intel Dual-Core E6500 (2.93 GHz) CPU and 8 GB RAM PC. The tuning parameter α in (2) is set to be five for all experiments. The penalty function weight λ is initialized as $\lambda_{\min} = 0.1$, and we choose $\rho = 1.2$ and $\lambda_{\max} = 10$. The δ in shrinkage function (15) is 0.15.

In this paper, we use the segmentation accuracy as evaluation criteria

$$\frac{I_R^+(S) \bigcap I_R^+(G)}{I_R^+(S) \bigcup I_R^+(G)}.$$

This criteria measures overlap between the segmented tumor area ($I_R^+(S)$) and the ground-truth area ($I_R^+(G)$).

We evaluate the proposed system from three aspects: 1) the capability of the collaborative model on single slice segmentation; 2) the image sequence segmentation; and 3) user study for the robustness of interaction.

At the beginning, we present a brief introduction for the operation flow of our system, and the system interface is exhibited in Fig. 4. The radiologists first select one key slice from the image sequence. Scribble type (foreground for tumor

[3] http://public.kitware.com/pub/itk/Data/LiverTumor/

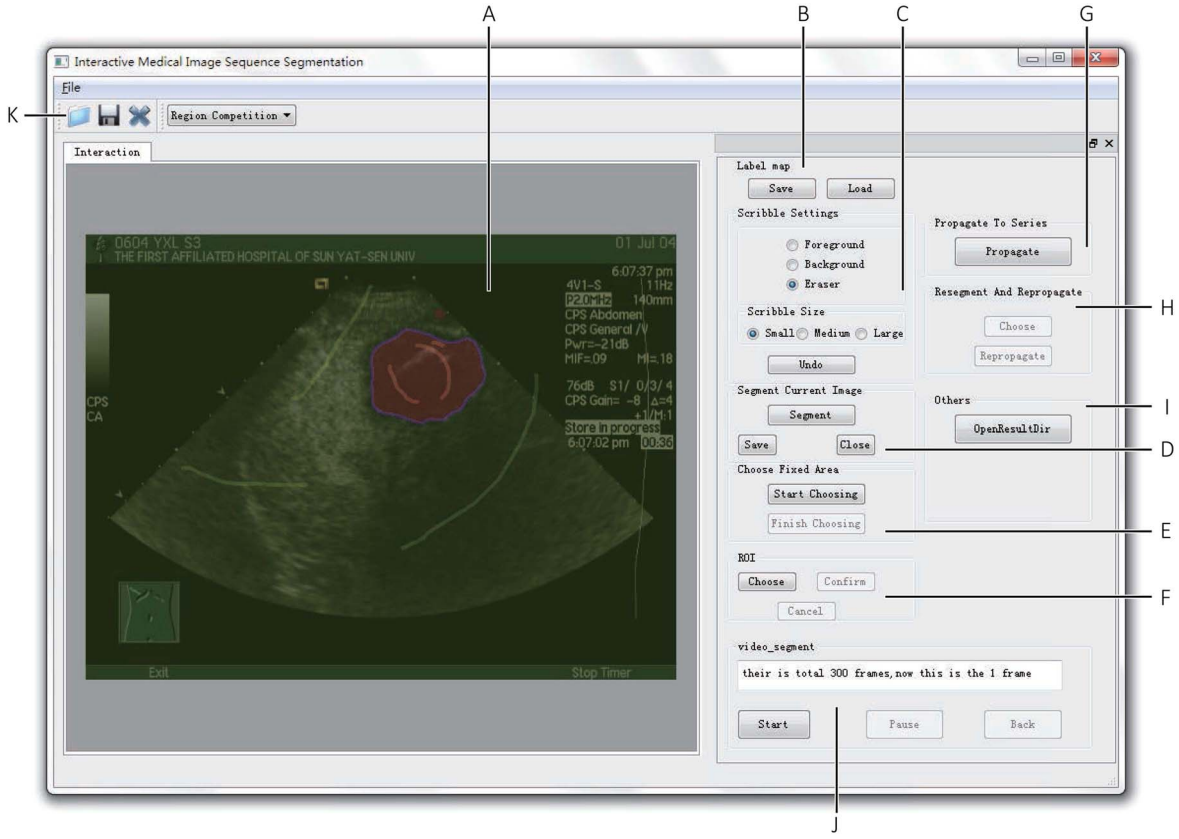


Fig. 4. User interface of our system. (A) Document window. (B) Save/load scribbles. (C) Scribble settings. (D) Segment the current image. (E) Fix a segmented area. (F) Select an ROI. (G) Start segmentation propagation. (H) Restart propagation from the choosing slice. (I) Open the results folder. (J) Browse slices in videos. (K) Options bar.

area and background for nontumor area) and size can be set in the scribble settings panel. After scribbles indicating tumor/nontumor areas have been drawn, the tumor can be fast extracted by calling segment function. Further refinement can be achieved by adding or removing scribbles until we find the result is satisfactory. Note that some segmented area can be chosen to be fixed during the refinement through choose fixed area panel. By calling propagate, the rest slices of the sequence can be segmented. Radiologists are allowed to polish the tumor segmentation at any slice and restart the segmentation propagation procedure.

A. Experiment I: Capability of Collaborative Model on Single Slice Segmentation

We first evaluate the effectiveness of the proposed model by applying it for single slice segmentation. We compare our system with two state-of-the-art interactive segmentation algorithm (i.e., GrabCut [11] and GAC [12]), one automatic method (i.e., distance regularized level set evolution (DRLSE) [47]), and one fully supervised algorithm (i.e., semantic texton forest (STF) [16]). We adopt the public available implementations of these algorithms. In addition to show the benefit of the dual term representation, we simplify our model by removing either the region term T_R or the boundary term T_B as the baselines.

A CT images subset (subCT) contains 118 images (66 from ILT and 52 from SYSU-CT), and a US images subset (subUS) contains 144 images are collected for testing.

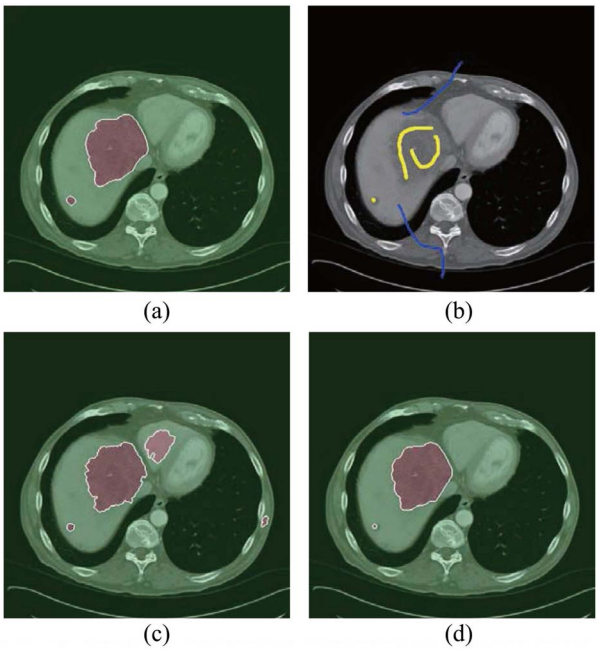


Fig. 5. Example of segmentations obtained by the simplified model. (a) Ground-truth segmentation. (b) User scribbles. (c) We remove the boundary term T_B in the model and achieve segmentation. (d) Result obtained by the complete model.

For fair comparison, we adopt the same user scribbles for all interactive methods. And each image is tested on three different interactions drew by different users, which ensures that the

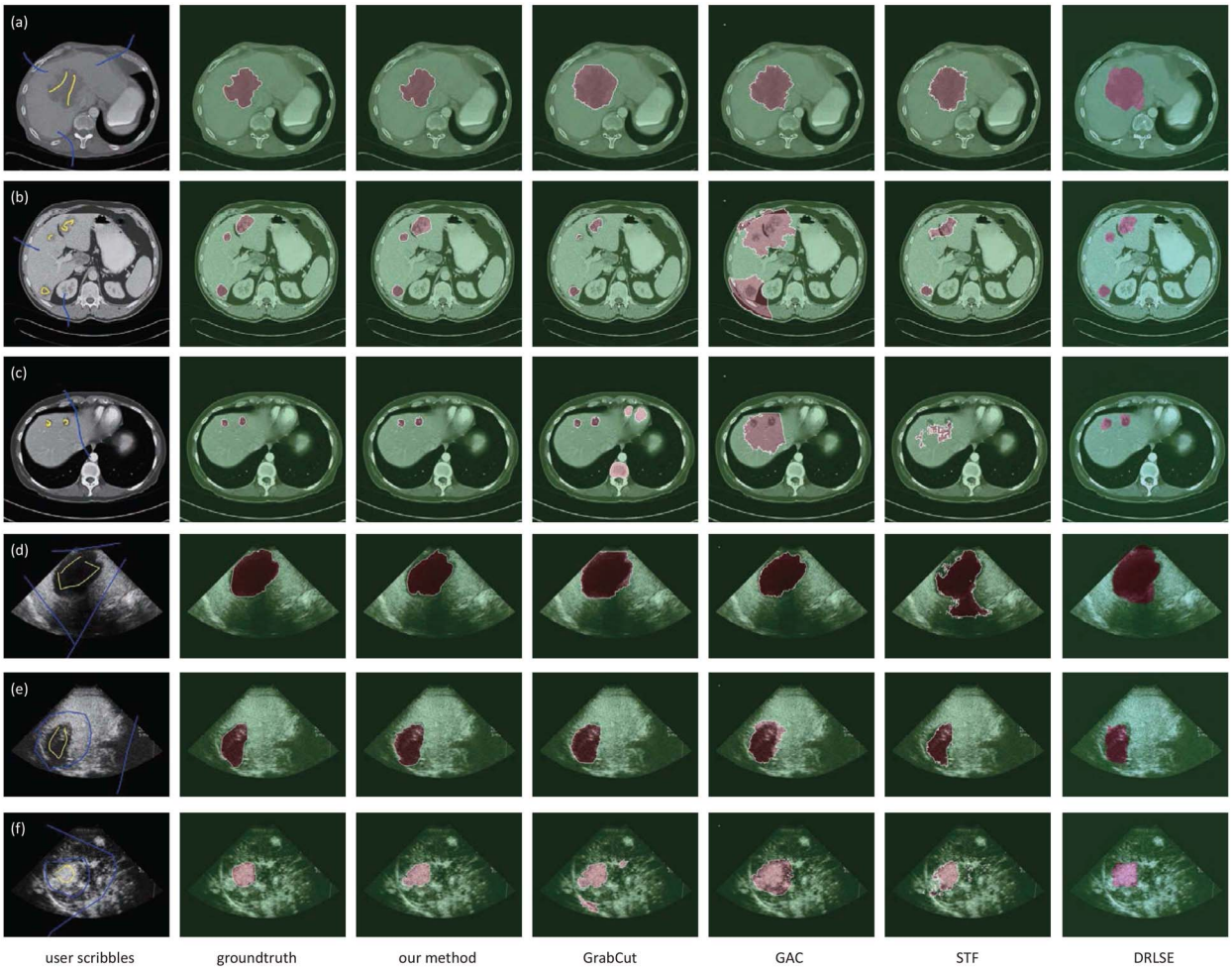


Fig. 6. Single slice segmentation results generated by our method, GrabCut [11], GAC [12], STF [16], and DRLSE [47]. One can see that our method achieves promising segmentation results with few user scribbles and qualitatively outperforms other methods. (a)–(c) show three segmentation instances of CT sequence and (d)–(f) show three segmentation instances of US video.

TABLE I
SEGMENTATION ACCURACY AND AVERAGE RUNNING TIME (SECONDS) ON SINGLE IMAGE SLICES. THE BOLD FONTS INDICATE THE BEST PERFORMANCE

	GrabCut [11]	GAC [12]	STF [16]	DRLSE [47]	Ours-R	Ours-B	Ours-full
subCT accuracy	0.6322	0.4870	0.4910	0.6808	0.6837	0.0268	**0.8469**
subUS accuracy	0.6089	0.5680	0.5818	0.6995	0.7057	0.0378	**0.7631**
average accuracy	0.6194	0.5315	0.5546	0.6901	0.6910	0.0304	**0.8008**
Code Type	C++	C++	C#	MATLAB	C++	C++	C++
Dependence	OpenCV	–	VISION.NET SVM.NET	–	OpenCV	OpenCV	OpenCV
Average Runtime	2.494	6.635	19.716	65.700	0.565	**0.334**	0.684

results are less dependent to the user interactions. Moreover, as STF is a fully supervised algorithm, we randomly separate half of the images for training and the other half for testing, and repeat this procedure ten times to ensure randomness.

As Table I reports, the collaborative model (ours-full) achieves the segmentation accuracy of 84.7% on subCT, and 76.3% on subUS. The quantitative results show that our proposed model is capable of handling different types of medical image data and achieves generally superior performances over other competing approaches. Specifically, we observe that GrabCut cannot accurately distinguish those tumors which have similar appearances/gray scales with the surrounding tissues, because this algorithm utilizes GMM-based color model in the segmentation process. GAC and DRLSE (two gradient-based level set methods) often fail when the tumors' boundaries are blurry, which are mainly caused by the boundary leaking issue. Moreover, due to large intraclass variations and noisy, STF (i.e., the fully supervised learning method) is limited to lacking of generalization performance, especially on dealing with diverse medical image data. And the fully

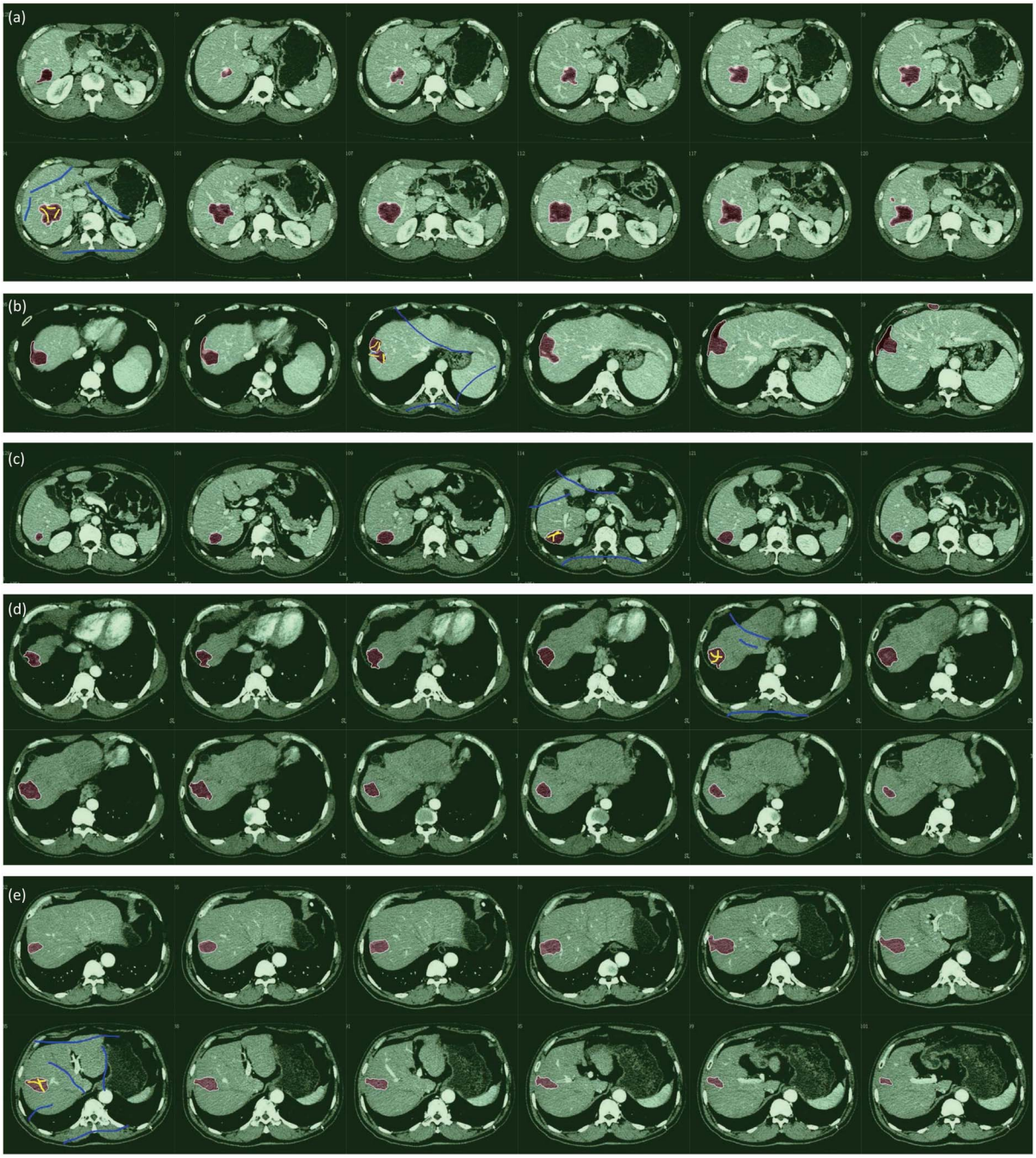


Fig. 7. Sample segmentation results generated by our method on five CT volumetric images. The tumors are covered by red masks while the background is covered by green masks. One can see that our method achieves accurate spatial segmentation and temporal propagation with few user scribbles on CT volumetric images in the context of drastic tumor changes. (a)–(e) show different segmentation instances.

supervised learning relies on a large number of manually annotated samples for off-line model training. Several typical results are provided in Fig. 6. Based on the results, our method is demonstrated to be more robust to some challenging scenarios, such as blurry boundaries [Fig. 6(a)] and tumors very similar to surrounding tissues [Fig. 6(f)].

Besides, the collaborative model also achieves a significant improvement compared with the baseline methods ours-*R* and ours-*B*. Ours-*R* denotes the model with only the region term and ours-*B* with only the boundary term. An example of comparison with the simplified model is shown in Fig. 5. It is worth mentioning that our full system (ours-full) requires less running time (0.52 s) compared with GrabCut (2.49 s), GAC (6.63 s), STF (19.71 s), and DRLSE (65.70 s). The efficiency is mainly due to the fast convex optimization.

B. Experiment II: Image Sequence Segmentation

This experiment demonstrates the performance of our approach on image sequence segmentation. We compare our framework with two well-applied approaches in medical image analysis: the PDE-based active contour model, namely

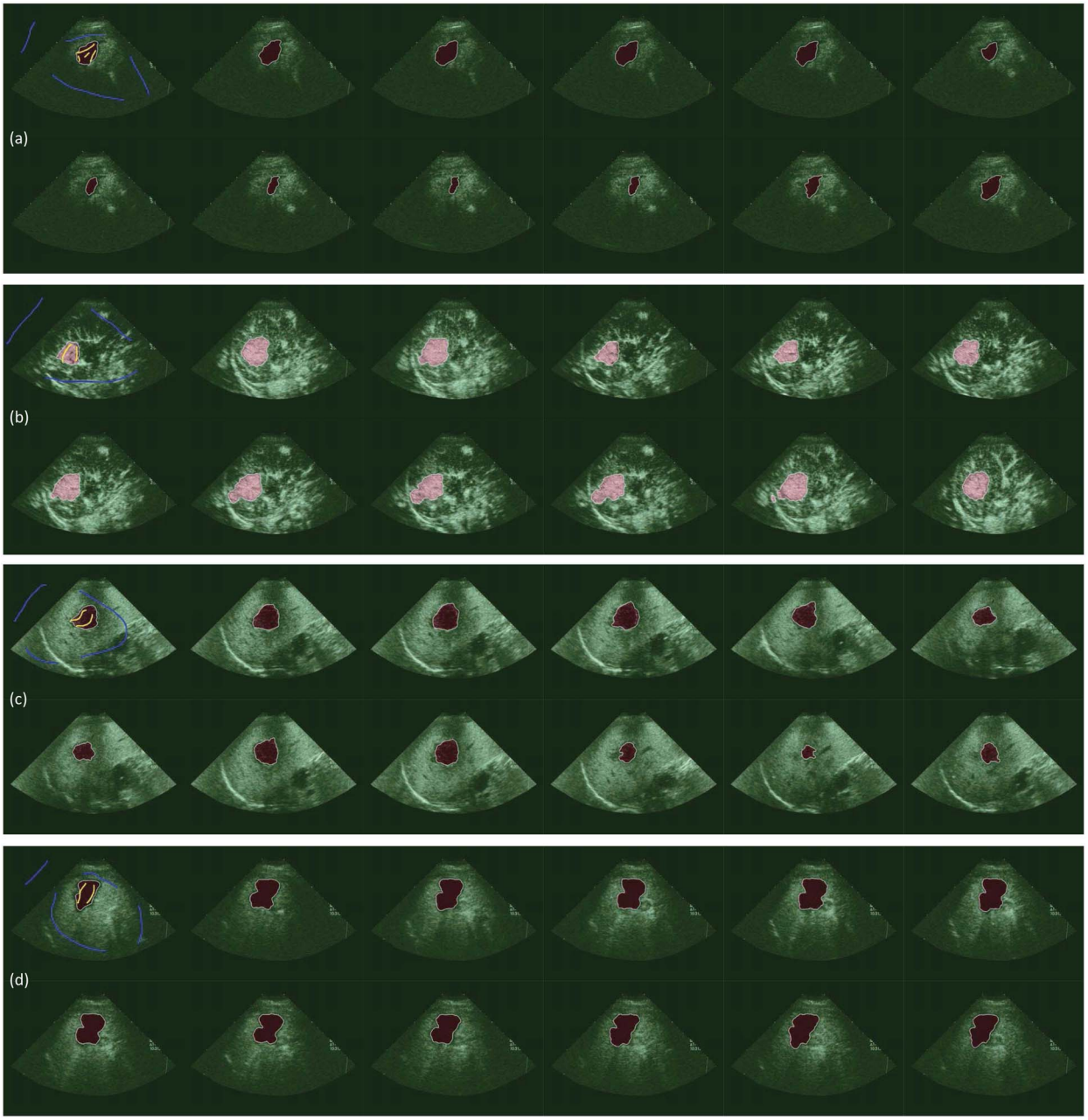


Fig. 8. Sample segmentation results generated by our method on four US image sequences. The tumors are covered by red masks while the background is covered by green masks. One can see that our method achieves accurate spatial segmentation and temporal propagation with few user scribbles on US image sequences. (a)–(d) show different segmentation instances.

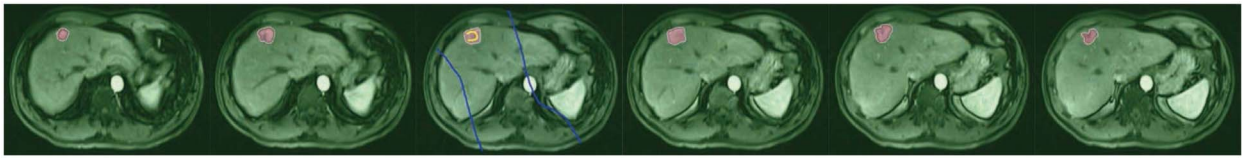


Fig. 9. Sample segmentation results generated by our method on one MRI image sequence. The tumors are covered by red masks while the background is covered by green masks. One can see that our method achieves accurate spatial segmentation and temporal propagation with few user scribbles on MRI image sequence.

Chan–Vese model (CVM) [17], and the HLM [18]. We also compare with the state-of-the-art interactive video segmentation method [48] (called RotoBrush in this paper), which is a local classifier-based segmentation method and has been included in Adobe After Effects CS5 as the roto-brush tool, and introduce the supervised method (i.e., STF [16]) as baseline. For assessing the significance for updating model during propagation, we also present the results by disabling the model update over image sequences. Similar to the last experiment, for the STF method, we randomly select half images from

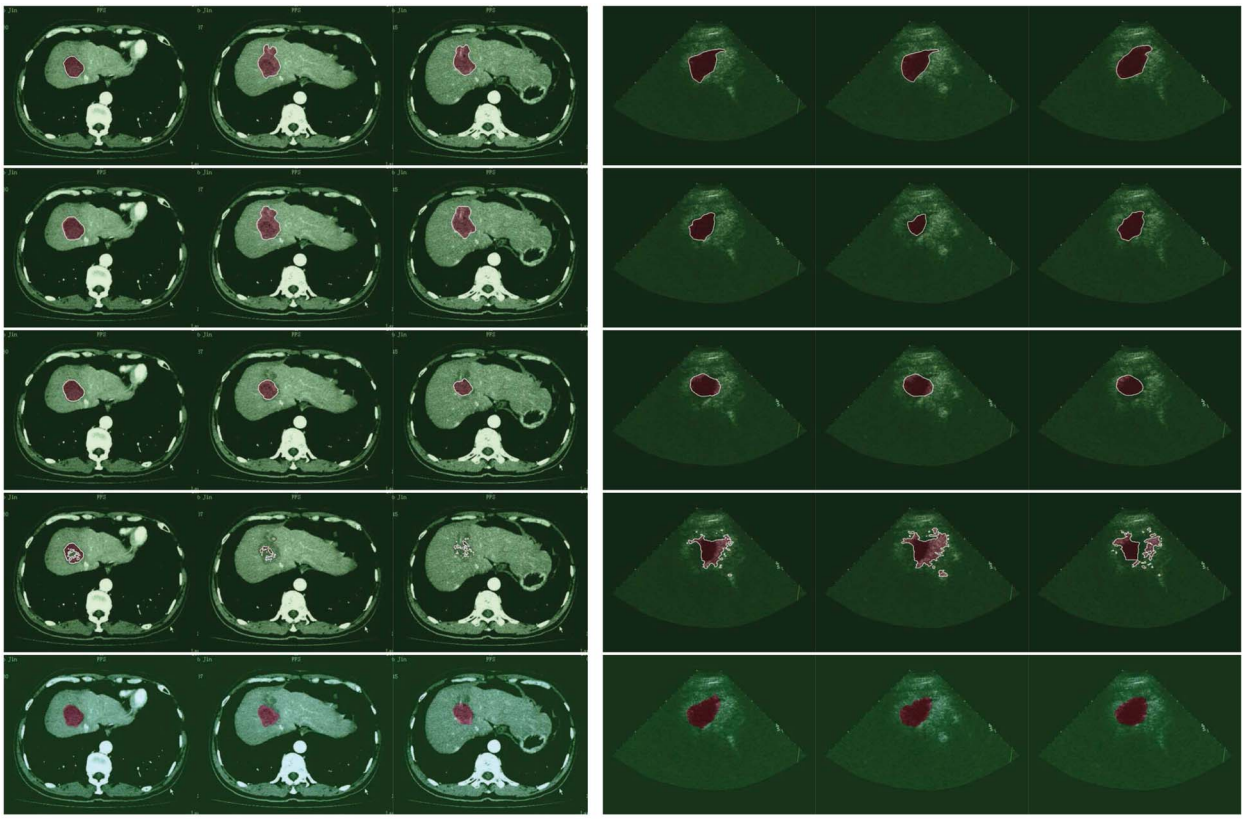


Fig. 10. Comparison of the proposed method (second row) with CVM [17] (third row), HLM [18] (fourth row), and RotoBrush [48] (fifth row). The ground-truth annotations are shown in the top row. (a) 8th, 15th, and 23rd slices of the CT volumetric images. (b) 238th, 242nd, and 249th frames of the US image sequence.

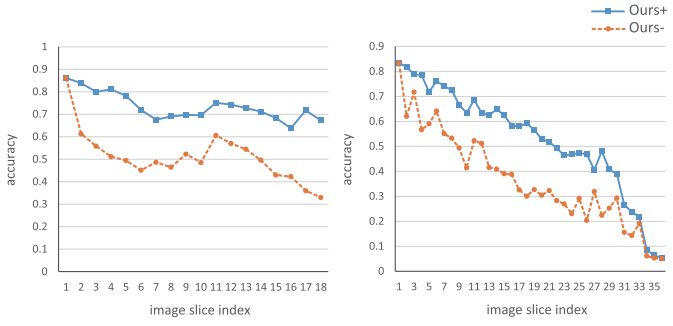


Fig. 11. Segmentation accuracy with the propagation procedure over the image sequences. We place scribbles on the first slice and automatically propagate segmentation over the rest of the slices. The left figure shows the result on SYSU-CT data set, while the right on SYSU-US data set. The horizontal axis represents the index of image slice and the vertical axis represents the average segmentation accuracy. Ours+ denotes the result generated by making the model adapted, while ours− denotes the result with fixed model parameters.

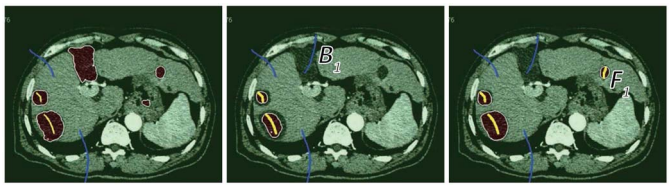


Fig. 12. Example of segmentation refinement by stepwise user interactions. The initialization segmentation is shown in the left image. The center image shows that one background scribble B_1 is added to correct the wrongly segmented tumor area. The right image shows that one foreground scribble F_1 is added to extract a tumor area that is missed in previous steps.

each image sequence for training and the other half for testing, and repeat this procedure ten times to calculate the average segmentation accuracy.

We execute this experiment on SYSU-CT and SYSU-US. As Table II reports, our framework (ours+) achieves the segmentation accuracy of 85.6% on SYSU-CT and 65.7% on SYSU-US, outperforming other methods. In the case of turning off the model updating during propagation, the segmentation accuracies (ours−) seriously decrease on the both data sets, (e.g., around 28% and 30%, less than the results by ours+, respectively). In Fig. 11, we further present the quantitative comparisons for segmenting image sequences with or without model updating. For each sequence, we propagate the segmentation from the first slice and show how the segmentation accuracy decreases over the sequence. We observe that the model updating is quite necessary to maintain the accurate segmentations in the sequences, as the tumors and tissues often distinctly vary in appearance and shape over image sequences. In clinical application, we usually achieve satisfactory results by allowing refinement every 10–20 slices.

Figs. 7–9 present some results on the CT, US, and MRI data, respectively. With very coarse scribbles, our system is able to generate accurate results in the context of drastic tumor

TABLE II
AVERAGE SEGMENTATION ACCURACY ON SYSU-CT AND SYSU-US. THE BOLD FONTS INDICATE THE BEST PERFORMANCE

	Ours+	Ours-	CVM [17]	HLM [18]	STF [16]	RotoBrush [48]
SYSU-CT	**0.8561**	0.5762	0.4448	0.3869	0.6620	0.6076
SYSU-US	**0.6571**	0.3530	0.3368	0.3388	0.5704	0.6226

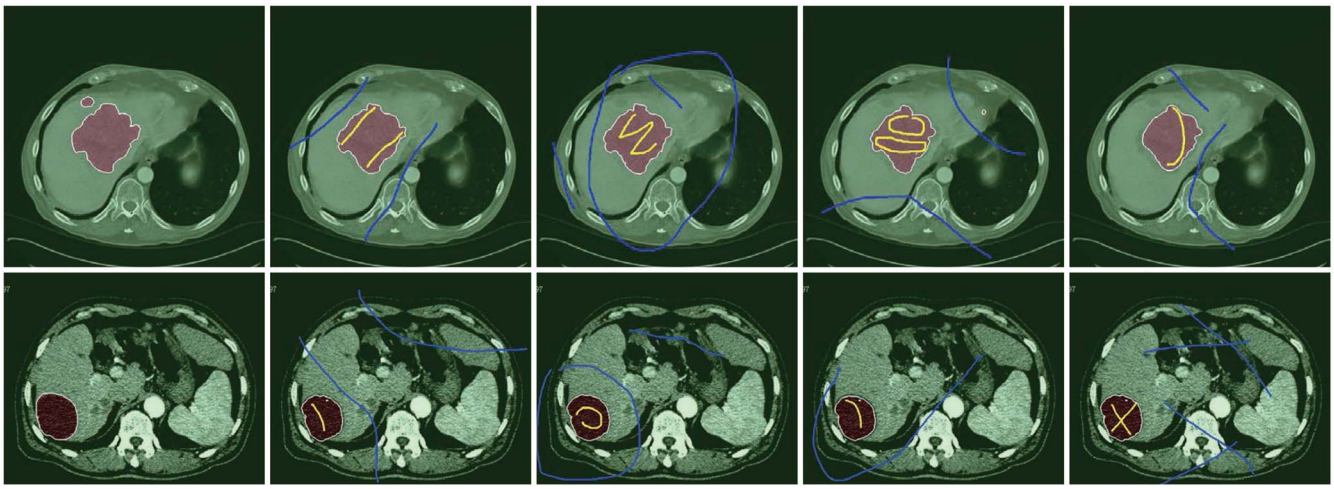


Fig. 13. Illustration of the robustness of the proposed system to different styles of user-assisted scribbles. The first column is the ground-truth segmentation. The remaining columns are segmentations obtained by placing scribbles in four different styles.

changes, e.g., Figs. 7(a) and 8(a). For tumors similar to healthy tissues, our framework can also conduct satisfactory results, e.g., Fig. 8(b). Several segmentation results in comparison with CVM [17], HLM [18], and RotoBrush [48] are exhibited in Fig. 10, and those generated by our approach clearly achieve better accuracies. These results further demonstrate the effectiveness of our model, while lack of the boundary models in CVM and RotoBrush can lead to inaccurate tumor boundary localization in the case of cluttered surrounding background. HLM relies on the assumption that the regions to be segmented exhibit homogeneous intrinsically, which may not be always satisfactory such as in the presented examples.

C. Experiment III: User Study

In the system, the radiologists are allowed to correct the segmented areas at any slice by adding or removing new scribbles in a stepwise manner. One example is illustrated in Fig. 12, where we first remove a wrongly segmented tumor area and extract one missing tumor area with two steps of adding scribbles.

We also show that our system is very robust to diverse user scribbles with the same intention. As shown in Fig. 13, scribbles are placed in four different patterns, but all achieve visually very similar segmentation results. These results show that our collaborative model less relies on the user scribbles, and thus makes the system more applicable for different radiologists.

V. CONCLUSION

In this paper, we study a general inference framework for extracting liver tumors from medical image sequences. A collaborative formulation of tumor segmentation is discussed by jointly integrating region and boundary information. The inference algorithm iterates to solve single slice segmentation and propagate the segmentation to consecutive slices. The implementation and system details are presented as well. The experiments are carried out on several very challenging liver tumor data sets with different imaging technologies (e.g., CT, MRI, and US), and our system outperforms the existing methods. Even though our approach was developed for liver tumor segmentation, it is suitable for other types of tissue where we can extract distinct region or boundary features.

There are several directions in which we intend to extend this paper. The first is to incorporate knowledge priors (e.g., tumor shapes, locations, and other attributes) into our framework, thereby improving the segmentation performance while further reducing user interactions. Second, we could utilize deep learning techniques (e.g., convolutional neural nets [49]) to replace the handcraft features. Another potential extension is to generalize our system in the context of intelligent diagnosis. Specifically, we could develop pattern classification techniques to recognize tumor categories or types, e.g., benign or malign, together with the segmentation process.

REFERENCES

[1] B. N. Li, C. K. Chui, S. Chang, and S. H. Ong, "A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9661–9668, 2012.

[2] A. Jemal *et al.*, "Global cancer statistics," *CA Cancer J. Clin.*, vol. 61, no. 2, pp. 69–90, 2011.

[3] D. Smeets *et al.*, "Semi-automatic level set segmentation of liver tumors combining a spiral-scanning technique with supervised fuzzy pixel classification," *Med. Image Anal.*, vol. 14, no. 1, pp. 13–20, 2010.

[4] Y. Häme and M. Pollari, "Semi-automatic liver tumor segmentation with hidden Markov measure field model and non-parametric distribution estimation," *Med. Image Anal.*, vol. 16, no. 1, pp. 140–149, 2012.

[5] T. Kubota, "Efficient automated detection and segmentation of medium and large liver tumors: CAD approach," in *Proc. MICCAI Workshop 3-D Segmentat. Clin. Grand Challenge II*, 2008.

[6] R. Lu, P. Marziliano, and C. H. Thng, "Liver tumor volume estimation by semi-automatic segmentation method," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, Shanghai, China, 2006, pp. 3296–3299.

[7] D. Wong *et al.*, "A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints," *Midas J.*, 2008.

[8] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1684–1699, Dec. 1998.

[9] K. Wang, L. Lin, J. Lu, C. Li, and K. Shi, "PISA: Pixelwise image saliency by aggregating complementary appearance contrast measures with edge-preserving coherence," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3019–3033, Oct. 2015.

[10] L. Lin, R. Zhang, and X. Duan, "Adaptive scene category discovery with generative learning and compositional sampling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 2, pp. 251–260, Feb. 2015.

[11] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[12] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.

[13] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[14] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[15] X. Bai and G. Sapiro, "Geodesic matting: A framework for fast interactive image and video segmentation and matting," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 113–132, 2009.

[16] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, 2008, pp. 1–8.

[17] T. F. Chan and L. A. Vese, "Active contour and segmentation models using geometric PDE's for medical imaging," in *Geometric Methods in Bio-Medical Image Processing*. Berlin, Germany: Springer, 2002, pp. 63–75.

[18] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *Proc. IEEE Int. Conf. BioMed. Vis.*, London, U.K., 2008, pp. 71–76.

[19] Y. Hame, "Liver tumor segmentation using implicit surface evolution," *MIDAS J.*, pp. 1–10, Jul. 2008.

[20] J. H. Moltz, L. Bornemann, V. Dicken, and H.-O. Peitgen, "Segmentation of liver metastases in CT scans by adaptive thresholding and morphological processing," in *Proc. MICCAI Workshop 3-D Segmentat. Clin. Grand Challenge II*, vol. 41, 2008, p. 195.

[21] J. Stawiaski, E. Decenciere, and F. Bidault, "Interactive liver tumor segmentation using graph-cuts and watershed," in *Proc. MICCAI Workshop 3-D Segmentat. Clin. Grand Challenge II*, 2008.

[22] L. Zhao, W. Wu, and J. J. Corso, "Semi-automatic brain tumor segmentation by constrained MRFs using structural trajectories," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, Berlin, Germany, 2013, pp. 567–575.

[23] W. Wu, A. Y. C. Chen, L. Zhao, and J. J. Corso, "Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 241–253, 2013.

[24] M. Cvancarova, F. Albregtsen, K. Brabrand, and E. Samset, "Segmentation of ultrasound images of liver tumors applying snake algorithms and GVF," *Int. Congr. Ser.*, vol. 1281, pp. 218–223, May 2005.

[25] F. Meng, H. Li, G. Liu, and K. N. Ngan, "Image cosegmentation by incorporating color reward strategy and active contour model," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 725–737, Apr. 2013.

[26] S. Balla-Arabe, X. Gao, and B. Wang, "A fast and robust level set method for image segmentation using fuzzy clustering and lattice Boltzmann method," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 910–920, Jun. 2013.

[27] K. Zhang, Q. Liu, H. Song, and X. Li, "A variational approach to simultaneous image segmentation and bias correction," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1426–1437, Aug. 2015.

[28] J. J. Corso *et al.*, "Efficient multilevel brain tumor segmentation with integrated Bayesian model classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 629–640, May 2008.

[29] L. Lin, X. Wang, W. Yang, and J.-H. Lai, "Discriminatively trained and-or graph models for object shape detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 959–972, May 2015.

[30] J. Shen, X. Yang, X. Li, and Y. Jia, "Intrinsic image decomposition using optimization and user scribbles," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 425–436, Apr. 2013.

[31] M. Freiman *et al.*, "An iterative Bayesian approach for nearly automatic liver segmentation: Algorithm and validation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 3, no. 5, pp. 439–446, 2008.

[32] M. Freiman, O. Cooper, D. Lischinski, and L. Joskowicz, "Liver tumors segmentation from CTA images using voxels classification and affinity constraint propagation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 6, no. 2, pp. 247–255, Mar. 2011.

[33] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.

[34] P. Dollar, Z. Tu, and S. Belongie, "Supervised learning of edges and object boundaries," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, vol. 2. New York, NY, USA, 2006, pp. 1964–1971.

[35] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*. Berlin, Germany: Springer, 1995, pp. 23–37.

[36] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[37] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comput. Phys.*, vol. 79, no. 1, pp. 12–49, 1988.

[38] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[39] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[40] P. L. Combettes and J. Pesquet, "A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Select. Topics Signal Process.*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[41] A. Ganesh *et al.*, "Fast algorithms for recovering a corrupted low-rank matrix," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adap. Process. (CAMSAP)*, 2009, pp. 213–216.

[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[43] J. Eckstein and D. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, 1992.

[44] X. Liu, L. Lin, S. Yan, H. Jin, and W. Jiang, "Adaptive object tracking by learning hybrid template on-line," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1588–1599, Nov. 2011.

[45] L. Lin, Y. Lu, C. Li, H. Cheng, and W. Zuo, "Detection-free multiobject tracking by reconfigurable inference with bundle representations," *IEEE Trans. Cybern.*, DOI: 10.1109/TCYB.2015.2478515, 2015.

[46] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 2. San Francisco, CA, USA, 1981, pp. 674–679.

[47] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, Dec. 2010.

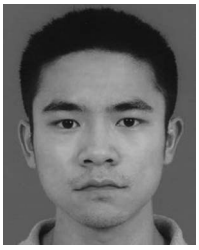
[48] X. Bai, J. Wang, D. Simons, and G. Sapiro, "Video SnapCut: Robust video object cutout using localized classifiers," *ACM Trans. Graph.*, vol. 28, no. 3, Aug. 2009, Art. ID 70.

[49] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4766–4779, Dec. 2015.

Liang Lin received the B.S. and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, in 1999 and 2008, respectively. From 2006 to 2007, he was a joint Ph.D. student with the Department of Statistics, University of California, Los Angeles (UCLA), Los Angeles, CA, USA.

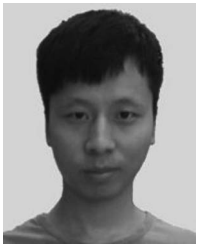
He is a Professor with the School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China. He was a Post-Doctoral Research Fellow with the Center for Vision, Cognition, Learning, and Art, UCLA. He has published over 80 papers in top-tier academic journals and conferences. His current research interests include new models, algorithms, and systems for intelligent processing and understanding of visual data such as images and videos.

Prof. Lin was a recipient of the Best Paper Runners-Up Award in NPAR 2010, the Google Faculty Award in 2012, the Hong Kong Scholars Award 2014, and the Best Student Paper Award in IEEE ICME 2014. He currently serves as an Associate Editor of *Neurocomputing* and *The Visual Computer*. He was supported by several promotive programs or funds for his works such as Guangdong NSFs for Distinguished Young Scholars in 2013.



Wei Yang received the B.S. degree in software engineering and the M.S. degree in computer science from Sun Yat-sen University, Guangzhou, China, in 2011 and 2014, respectively.

His current research interests include computer vision and machine learning.

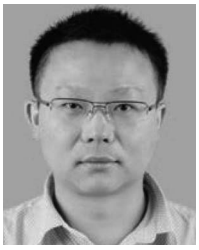


Chenglong Li received the B.S. degree in applied mathematics and the M.S. degree in computer science from Anhui University, Hefei, China, in 2010 and 2013, respectively, where he is currently pursuing the Ph.D. degree in computer science.

His current research interests include computer vision, machine learning, and intelligent media technology.

Jin Tang received the B.Eng. degree in automation and the Ph.D. degree in computer science from Anhui University, Hefei, China, in 1999 and 2007, respectively.

He is currently a Professor with the School of Computer Science and Technology, Anhui University. His current research interests include computer vision, pattern recognition, and machine learning.



Xiaochun Cao (SM'XX) received the B.E. and M.E. degrees from Beihang University, Beijing, China, and the Ph.D. degree from the University of Central Florida, Orlando, FL, USA, all in computer science.

He is a Professor of the Institute of Information Engineering, Chinese Academy of Sciences, Beijing. He was a Research Scientist with ObjectVideo Inc., Reston, VA, USA, for three years. From 2008 to 2012, he was a Professor with Tianjin University, Tianjin, China. He has authored and co-authored over 100 journal and conference papers.

Prof. Cao was a recipient of the Piero Zamperoni Best Student Paper Award at the International Conference on Pattern Recognition, from 2004 to 2010. He was nominated for the university level Outstanding Dissertation Award for his dissertation. He is a fellow of IET.