

INTERACTIVE CT IMAGE SEGMENTATION WITH ONLINE DISCRIMINATIVE LEARNING

Wei Yang, Xiaolong Wang, Liang Lin^{*}, Chengying Gao

School of Software, Sun Yat-Sen University, Guangzhou 510275, P.R. China

ABSTRACT

Although interactive image segmentation has been widely exploited, current approaches present unsatisfactory results in medical image processing. This paper proposes a fast method for interactive CT image segmentation in which the tumor regions should be partitioned as foreground against the healthy tissues. In contrast to natural images, we have the following observation on CT images: (1) CT images often include discontinuous silhouette or cluttered spots caused by input devices or patient corporeity; (2) Disease areas often have varying appearance and shape. We thus train a discriminative foreground/background model based on user-placed scribbles. In our method, we extract positive and negative samples according to the foreground and background scribbles respectively, and use dense SIFT descriptors plus gray-level histogram as candidate features. With online learning, segmentation can be fast solved by the Bregman iteration. We test our method on CT liver images and demonstrate the advantage by comparing to state-of-the-art approaches.

Index Terms— interactive image segmentation, CT image, online discriminative learning

1. INTRODUCTION

Interactive image segmentation has been widely exploited in computer vision. Current state-of-the-art approaches, however, present unsatisfactory results in clinical applications of medical image processing, especially for tumor segmentation. For example, Graph Cuts [1] is unable to locate the boundary of liver tumor accurately with very few user scribbles (e.g., small training set, as shown in Fig.1.(a)) under the circumstances of high noisy and complex patterns, as shown in Fig.1.(b). Although some automatic segmentation approaches have been proposed, their performance is unreliable. Moreover, under the circumstances of low contrasts and especially tumors with contact to a structure of the resemble density(intensity in images), such methods will pose problems. The objective of this paper is to propose a real-time and accurate segmentation algorithm for clinical use. A result of our approach is shown in Fig.1.(c).

In the literature, graph cut formalism is well suited for segmentation of natural images. Esneault et al. [2] adopt Graph Cuts to perform semi-automatic segmentation of the

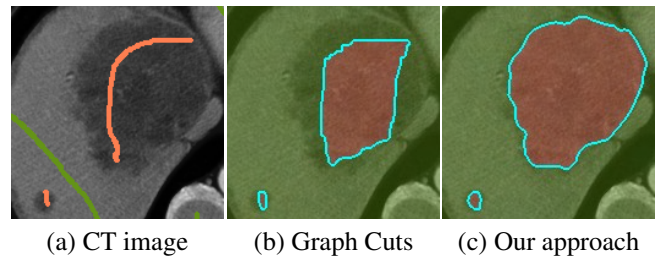


Fig. 1. Results of different segmentation methods on (a) liver tumor CT scan using (b) Graph Cuts and (c) our method.

liver, the tumor and the hepatic vascular networks. This approach uses fixed parameters, and fails to deal with the diversity and complexity of various tumor patterns. Geodesic [3] is another approach successful in natural scene segmentation based on belief propagation. However, this method is closely correlated to the initial seed, and is not reliable for clinical use. Online learning algorithms, on the other hand, are able to adapt and learn difficult patterns, and guarantee for noisy situations. Online learning has been applied by Gong et al. [4] for foreground segmentation, and has been proven to be successful for real-time applications.

In the image level, computed tomography (CT) provides an exceptional resolution to achieve good segmentation for most tumor types. Still, challenge exists. Take liver tumors for instance, first, most of liver tumors are similar to healthy tissues. Moreover, there is a great diversity of tumor types with various appearances. As liver is an organ of high vascularization, the CT scans are unavoidably noisy. What is more, the image contrast between tumors and healthy tissues may be low due to individual differences in perfusion and scan timing. All above make the segmentation of liver tumor quite a tough task.

Therefore, we propose a novel interactive segmentation approach which is real-time and accurate for clinical applications. As Fig.2 shows, we decompose an image into regions and boundaries. Then, we adopt online learning procedure and offline learning procedure to train a discriminative region model and a discriminative boundary model respectively. In the procedure of online training, we extract positive and negative samples according to the foreground and background scribbles respectively, and use dense SIFT descriptors plus quantized gray-level histogram as candidate features. Finally, Bregman iteration [5] is adopted to solve the segmentation. Our method is evaluated on a public dataset provided by the ITK's Liver Tumor dataset and the dataset from the first affiliated hospital, Sun Yat-Sen University. Compared

This work was partially supported by the Fundamental Research Funds for the Central Universities (Grant No. 2010620003162041), and SYSU-Sugon high performance computing typical application projects (Grant No. 62000-1132001).

^{*}Corresponding author: Liang Lin (linliang@ieee.org)

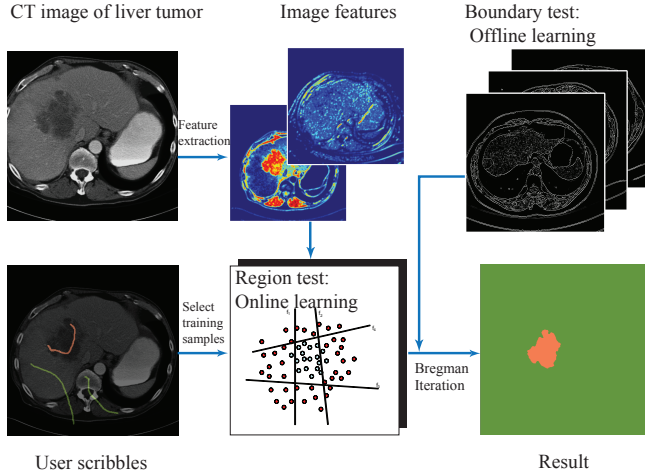


Fig. 2. A general framework of our method.

to three state-of-the-art approaches including Graph Cuts [1], geodesic [3] and CO3 [6], our method outperforms in accuracy and achieves promising results.

2. PROBLEM FORMULATION

The goal of our segmentation is to decompose the image domain R into foreground R_+ showing the liver tumors and background R_- showing the other parts of the image. Meanwhile, we also distinguish the boundaries B_+ from the rest B_- in the image to classify the pixels more precisely. In this case, our problem can be illustrated by two partitions. To combine the results of the two partitions, we introduce the discriminative probability test whose equation is:

$$E(\prod) = - \sum_{i=1}^{n \times m} [\mathbf{1}(\text{pixel}_i \in R_+) \log \frac{P_{R_+}(+|I, S)}{P_{R_+}(-|I, S)} + \alpha \mathbf{1}(\text{pixel}_i \in B_+) \log \frac{P_{B_+}(+|I)}{P_{B_+}(-|I)}] \quad (1)$$

where $n \times m$ is the number of pixels in the image; $\mathbf{1}(x) \in \{+1, -1\}$ is an indicator function for a Boolean variable x ; the parameter α is a weight between region and boundary tests. In this case, we turn the segmentation problem to the minimization of the energy $E(\prod)$.

In the region test $T_R = \log \frac{P_{R_+}(+|I, S)}{P_{R_+}(-|I, S)}$, the posterior probability P_R is based on the scribbles including the foreground scribbles S_+ and background scribbles S_- . We use the pixels on different scribbles as positive and negative training samples, so the probability P_R is learned by online discriminative classifier. In this test we use new features different from [6] and fit the logistic regression model to the additive models, which can be solved by boosting algorithm we will describe later in chapter 3.

In the boundary test $T_B = \log \frac{P_{B_+}(+|I)}{P_{B_+}(-|I)}$, there are many offline learning methods to detect the edges in the image. Here we use the algorithms mentioned in [3] as edge detectors.

3. ONLINE DISCRIMINATIVE LEARNING

To minimize the EQ.(1), we need to calculate the conditional probability $P(+|I, S)$ and $P(-|I)$ in region and boundary test respectively. Since the method of using offline learning algorithm in boundary test has been mentioned in many researches before, here we focus on introducing the online discriminative learning method in the region test. By transforming the log-posterior ratio form to additive model, we can use GentleBoost [7] algorithm to learn a strong classifier with the help of user's scribbles. During the procedure of learning, we choose features which are more appropriate for CT images.

3.1. Online discriminative model

In the problem of classifying the pixels between background and foreground region, we express the probability in a log-posterior ratio form, and it can fit to the additive logistic model:

$$\log \frac{P(+|I)}{P(-|I)} = \sum_{m=1}^M f_m(x) \quad (2)$$

where $f_m(x)$ represents the weak classifier in boosting. To use the boosting algorithm to classify the pixels in the image, we first choose the pixels which have been classified by the scribbles drawn manually. Then we apply the features (supposing each of them has τ bins) of these pixels as a training set.

By training the input image features, we can combine the weak classifier to the strong one, which can be used to classify each pixel between background and foreground. Here we show that the boosting algorithm fits the additive model:

$$F(x) = \sum_{m=1}^M f_m(x) \quad (3)$$

where $F(x)$ represents the strong classifier. During boosting we train out the classifier $F(x)$ by optimizing an exponential cost which has a sensible population:

$$J = \sum_{i=1}^N e^{-y_i F(x_i)} \quad (4)$$

where x_i indicates the feature in the training data and y_i represents a quantitative number selecting from $\{-1, +1\}$ indicating background and foreground respectively.

To be specific, during the m -th iteration in boosting, we add a new weak classifier $f_m(x)$ to $F_{m-1}(x)$ trained in the last iteration to form a new strong classifier $F_m(x)$:

$$F_m(x) = F_{m-1}(x) + f_m(x) \quad (5)$$

We choose the new $f_m(x)$ by minimizing the exponential cost. By combining (3) and (4) we get:

$$J(F + f_m) = \sum_{i=1}^N e^{-y_i (F(x_i) + f_m(x_i))} \quad (6)$$

To optimize exponential criterion, we apply the Gentle-

Boost [7] algorithm so that we do not have to calculate the exact optimization. Instead, we transform EQ.(6) to EQ.(7):

$$J(F) \propto \sum_{i=1}^N e^{-y_i F(x_i)} (y_i - f_m(x_i))^2 \quad (7)$$

so that we simply need to solve a weighted least squares problem during each iteration.

Furthermore, we define the format of the update as:

$$f_m(x) = a \times (x > \text{threshold}) + b \quad (8)$$

where the parameters $(a, b, \text{threshold})$ are calculated in the way shown as followed:

(I) Knowing that we have τ bins in the feature of each pixel, we select one bin each time and get the values of the training pixels in this bin to form a set of input numbers.

(II) We use regression stumps to calculate the parameters $(a, b, \text{threshold})$ by minimizing the exponential error.

(III) We get τ sets of data $(a, b, \text{threshold}, \text{error})$ after enumerating each bin. Among these τ sets of numbers, we choose the one including the minimum error and apply $(a, b, \text{threshold})$ in it as parameters in the function $f_m(x)$.

3.2. Feature design

In our problem, the features of each pixel are applied as training samples in boosting. Here we select two types of features including gray-level histogram and dense SIFT descriptor [8] of each pixel and combine them in the form of a histogram.

The gray-level histogram is a basic and useful feature especially when it is applied in the gray level image. Since each pixel in our CT image contains only one channel, the gray level histogram is very significant in distinguishing “black” pixels from “white” pixels. In this phase, we divide all gray level values into a 12-bin histogram.

Shape, size, and position of livers may vary a lot due to individual differences in perfusion and timing. Thus we need transform-invariant feature to represent the pixels. Motivated by siftflow algorithm [9], we adopt a SIFT descriptor as a feature for each pixel. We extract the feature by dividing each pixel’s neighborhood (including 12×12 pixels) into a 4×4 cell array. In each cell, we use a vector which includes 8 numbers characterizing gradient information in different directions to represent the feature of the cell. As a result, we obtain a $4 \times 4 \times 8 = 128$ -dimensional vector as a feature for each pixel, the procedure is shown in Fig.3. After that, we store these features as a 128-bin histogram.

Combining the two histograms we mentioned above together, we get a 136-bin histogram as input training samples in boosting.

4. INFERENCE WITH BREGMAN ITERATION

After classifying the pixels in the region test and boundary test separately, we have to combine the results in these two tests aiming at minimizing the EQ.(1).

As it is said in [10], by transforming the indicator functions to continuous interval $\alpha, \beta \in [0, 1]$, we can obtain a convex formulation which is equivalent to EQ.(1). So the EQ.(1) can be minimized through optimizing the convex formulation.

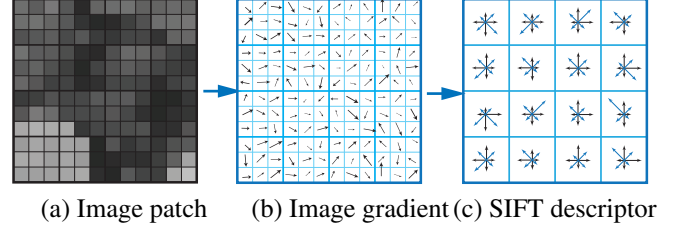


Fig. 3. A SIFT descriptor is obtained by computing the orientation and gradient magnitude of the sample points in a region over 4×4 cells around the key point, as shown in (b). In each cell, we represent gradient magnitudes of 8 orientations by 8 values, and finally obtain a $4 \times 4 \times 8 = 128$ -dimensional feature vector for each pixel, as shown in (c).

In this case, we can use Bregman iteration[5] which is an efficient and reliable minimization algorithm. During each iteration, the convex formulation is optimized by minimizing the cumulative error. As a result, we will obtain the functions α and β which bring the minimization of EQ.(1).

5. EXPERIMENTS

The proposed method is tested on two different sources of liver tumor CT images. 10 CT images are selected from ITK’s Liver Tumor dataset, each of which has the size of 512×512 pixels. 20 CT images are provided by the first affiliated hospital, Sun Yat-Sen University, each of which has the size of 500×500 pixels. All these CT images are taken from different patients under different conditions. In our experiment, we evaluate both the performance and the accuracy of our algorithm based on these data.

5.1. Qualitative evaluation

Fig.4 shows some results of our segmentation algorithm on the images from the dataset in the first affiliated hospital, Sun Yat-Sen University and ITK’s Liver Tumor dataset.

In Fig.4(a, b, c), tumor regions are easily segmented from backgrounds with few scribbles. In more complicated cases like Fig.4 (d, e), ambiguous patterns exist, which might be challenges for other algorithms. However, our method with more scribbles is able to obtain the ideal results.

5.2. Quantitative evaluation

In this paper, we compare our method to other 3 approaches: Graph Cuts [1], geodesic [3], CO3 [6] by doing quantitative evaluation on them. We apply our methods on the CT images from the first affiliated hospital, Sun Yat-Sen University dataset and the ITK’s Liver Tumor dataset separately by using two different sets of scribbles. As a result, the first four rows in Table 1 show the precisions of the algorithms which defined as:

$$\text{Precision} = (TP) / (TP + FP + FN) \quad (9)$$

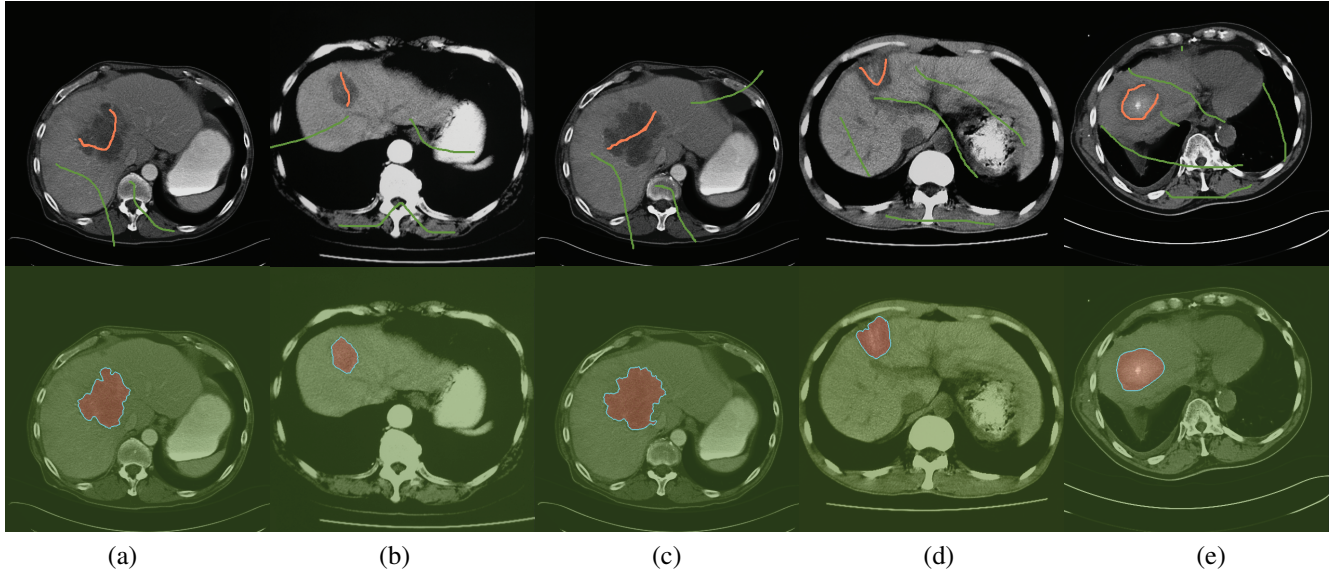


Fig. 4. Some segmentation results on ITK's Liver Tumor dataset and dataset from the first affiliated hospital, Sun Yat-Sen University.

	Graph Cuts	Geodesic	CO3	Ours
$P_{1_{scr1}}$	0.7366	0.5676	0.7144	0.8037
$P_{1_{scr2}}$	0.7844	0.6900	0.7675	0.8115
$P_{2_{scr1}}$	0.6517	0.4752	0.8084	0.8273
$P_{2_{scr2}}$	0.6576	0.3714	0.8054	0.8185
Average	0.7076	0.5261	0.7739	0.8152

Table 1. Quantitative evaluation on two different datasets. $P_{1_{scr1}}$ and $P_{1_{scr2}}$ are precisions evaluated on dataset from the first affiliated hospital, Sun Yat-Sen University using two sets of scribbles. $P_{2_{scr1}}$ and $P_{2_{scr2}}$ are precisions evaluated on ITK's Liver Tumor dataset. The last row shows the average precision of all the tests.

where TP (true positive), FP (false positive) and FN (false negative) are obtained by comparing the results of segmentation by different algorithms to the ground-truth.

By combining the results of 2 sets of tests, we get the average precision showing our method is more accurate.

6. CONCLUSIONS

We propose a fast interactive segmentation approach for CT image analysis. We introduce the discriminative probability test consists of region term and boundary term, which can be trained online and offline respectively. Experimental results demonstrate our method outperforming in accuracy compared to state-of-the-art approaches.

In the future work, we plan to apply our method for segmenting objects in a video or sequential images.

7. REFERENCES

- [1] Y.Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *ICCV*, 2001, vol. 1, pp. 105–112.
- [2] S. Esneault, N. Hraiech, E. Delabrousse, and J.L. Dillenseger, "Graph cut liver segmentation for interstitial ultrasound therapy," in *EMBS*, 2007, pp. 5247–5250.
- [3] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *IJCV*, vol. 22, pp. 61–79, 1997.
- [4] Minglun G. and Li C., "Real-time foreground segmentation on gpus using local online learning and global graph cut optimization," in *ICPR*, 2008, pp. 1–4.
- [5] T. Goldstein, X. Bresson, and S. Osher, "Geometric applications of the split bregman method: Segmentation and surface reconstruction," *Journal of Scientific Computing*, vol. 45, pp. 272–293, 2010.
- [6] Y. Zhao, S.C. Zhu, and S. Luo, "Co3 for ultra-fast and accurate interactive segmentation," in *Proceedings of the international conference on Multimedia*, 2010, pp. 93–102.
- [7] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," in *Annals of Stat.*, 2000, vol. 28, pp. 337–407.
- [8] D.G. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, 1999, vol. 2, pp. 1150–1157.
- [9] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *TPAMI*, vol. PP, no. 99, pp. 1, 2010.
- [10] X. Bresson, S. Esedoğlu, P. Vandergheynst, J. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, pp. 151–167, 2007.