15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Multiple-organ Segmentation by Graph Cuts with Supervoxel Nodes

Toshiya Takaoka*, Yoshihiko Mochizuki*, Hiroshi Ishikawa*
*Waseda University, Department of Computer Science and Engineering

Abstract

Improvement in medical imaging technologies has made it possible for doctors to directly look into patients' bodies in ever finer details. However, since only the cross-sectional image can be directly seen, it is essential to segment the volume into organs so that their shape can be seen as 3D graphics of the organ boundary surfaces. Segmentation is also important for quantitative measurement for diagnosis. Here, we introduce a novel higher-precision method to segment multiple organs using graph cuts within medical images such as CT-scanned images. We utilize supervoxels instead of voxels as the units of segmentation, i.e., the nodes in the graphical model, and design the energy function to minimize accordingly. We utilize SLIC supervoxel algorithm and verify the performance of our segmentation algorithm by energy minimization comparing to the ground truth.

1 Introduction

Recent development of precision medical diagnostic tools have improved the precision and quality of diagnosis by doctors in clinical settings. Specifically, improvement in medical imaging technologies such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) has made it possible for doctors to directly look into patients' bodies in ever finer details. However, although medical images are volumetric, only the cross-section can be directly seen. Thus, it is essential to segment the volume into organs so that their shape can be displayed and be seen as 3D graphics of the organ boundary surfaces. Segmentation is also important for quantitative measurement for diagnosis. For instance, the size of organs offers crucial diagnostic information to doctors that can be compared over time.

One popular approach for medical image segmentation is energy minimization methods such as graph cuts[1], where the problem is treated as labeling of each volumetric pixels (voxels) as one of possible organs. That is, each voxel in the volumetric image is classified into one of the organs such as heart and lung, thereby characterizing each segment as those voxels labeled by a particular label. This is done by finding a labeling that minimizes a function called the *energy*, which assigns a real number for each labeling that assesses the quality of the labeling so that the lower the value is, the better the quality of the segmentation is. In this paper, we use the X-ray CT images such as shown on the left of Fig. 1, which consists of slices of 2D images.

In the multiple-organ segmentation methods [2] utilizing energy minimization, typically every voxel is labeled. Although there is smoothing terms that prevent the resulting segmentation from becoming too rough, it still often allows isolated voxels labeled differently from its neighbors, which cannot be true segmentation.

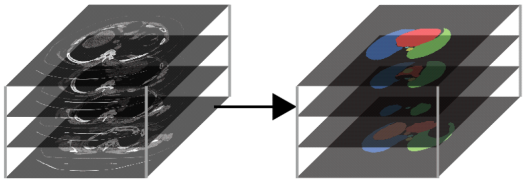

Figure 1. Left: Stacked CT slices can be seen as a 3D volumetric image. Right: the result of segmentation, which is also a 3D image of labels.

One way to alleviate this problem is to use supervoxels. A supervoxel is a set of voxels with common properties such as close intensities or voxel values. By using them as the unit to be labeled, multiple voxels are treated as a group to be labeled in the same way. As examples of segmentation methods using supervoxels, a multiple-object segmentation in 3D CT images [3] and a technique for detetion of mitochondria in electron microsscopy images [4] have been proposed.

In this paper, we utilize supervoxels and design the energy function for such a case, thereby solving the problem of rough segmentation and improve the precision of multiple-organ segmentation.

2 Energy Minimization

First, we explain the previous energy minimization approach where individual voxels are labeled.

We denote the set of all voxels in the CT image by V, whereas v denotes the individual voxels in V. We denote the CT value at voxel v by X_v, and the label assigned to v by $L_v \in \mathcal{L}$. The CT value at a voxel represents the relative absorption of X-ray by the tissue at that voxel, normalized so that the value for the air is the lower limit of -1000, and the value for water is 0. Let us denote by M the number of the labels, including the one that represents the background. Then $X = (X_1, \cdots, X_N)$ is the CT value at each voxel, while $L = (L_1, \cdots, L_N)$ is the labels assigned to each voxel, where N is the number of voxels in the CT image.

The volume is segmented by MAP-estimation of the labeling random variable L by finding the labeling that maximizes the posterior probability $P(L|X)$ given the CT-value random variable X. The L that maximizes $P(L|X)$ is obtained by solving the following energy minimization problem [5]:

$$\operatorname*{argmax}_L P(L|X) = \operatorname*{argmin}_L E(L;X) \tag{1}$$

The energy $E(L;X)$ consists of three terms as in (2): the data term $E_{\text{data}}(L;X)$, the probabilistic atlas term $E_{\text{atlas}}(L)$, and the smoothing term $E_{\text{smooth}}(L;X)$.

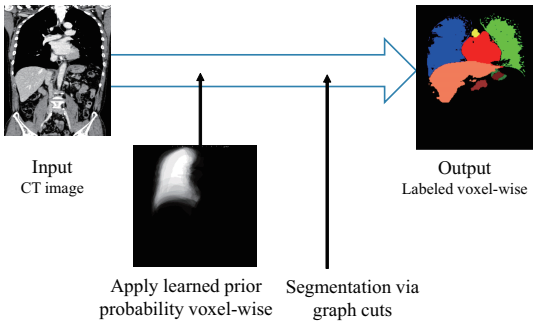

Figure 2. Previous method: voxel-based labeling

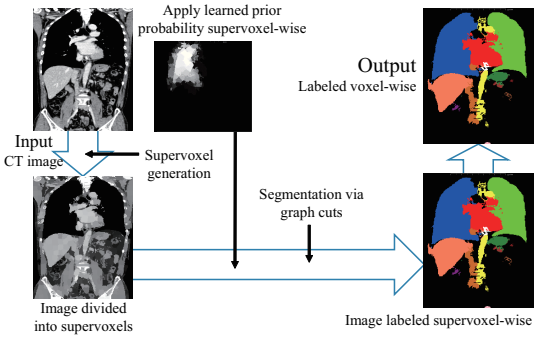

Figure 3. Our method: supervoxel-based labeling

$$E(L;X) = w_{\text{data}} E_{\text{data}}(L;X) + w_{\text{atlas}} E_{\text{atlas}}(L) + w_{\text{smooth}} E_{\text{smooth}}(L;X). \quad (2)$$

Here, w_{data}, w_{atlas}, and w_{smooth} are the weights for the terms, which are the parameters we calibrate so that $E(L;X)$ is minimized. We use graph cuts [1] to minimize the energy.

We define the terms in (2) as follows:

$$E_{\text{data}}(L;X) = \sum_{v \in V} f(L_v; X_v), \quad (3)$$

$$f(L_v; X_v) = -\log P(X_v | L_v), \quad (4)$$

where $P(x|l)$ is the conditional probability that CT value at a voxel takes x when it is assigned label l which, as a physical property, should be independent of position;

$$E_{\text{atlas}}(L) = \sum_{v \in V} g(L_v), \quad (5)$$

$$g(L_v) = -\log P(v, L_v), \quad (6)$$

where $P(v,l)$ is the prior probability that voxel v is assigned label l; and

$$E_{\text{smooth}}(L;X) = \sum_{(u,v) \in C_2} h(L_u, L_v, X_u, X_v), \quad (7)$$

$$h(L_u, L_v, X_u, X_v) = \begin{cases} 0 & (L_u = L_v) \\ \frac{1}{|X_u - X_v| + 1} & (L_u \neq L_v), \end{cases} \quad (8)$$

where C_2 is the set of all neighboring pairs.

Note that, in the experiments, the probabilities $P(x|l)$ and $P(v,l)$ are estimated from training data by counting.

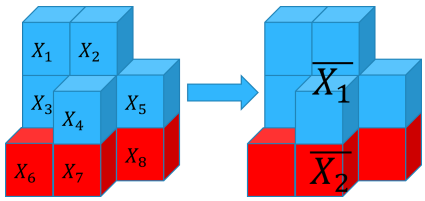

Figure 4. CT value is averaged over a supervoxel.

3 Energy Minimization on Supervoxels

Now, we modify the above method so that supervoxels, instead of voxels, are labeled. Figs 2 and 3 illustrate the processes in the two approaches, respectively. As explained in the previous section, in the original, voxel-based labeling, each voxel in the input 3D CT image is individually labeled, so that the energy (2) is minimized.

In our method, we first oversegment the input 3D CT image into supervoxels. Then we label the supervoxels using graph cuts. Accordingly, the energy is defined in terms of supervoxels, instead of voxels. Naturally, the definition of the energy must be modified, as explained below.

The relationship between the voxels and the supervoxels are as in (9) and (10):

$$V = S_1 \cup S_2 \cup S_3 \cup \cdots \cup S_K, \quad (9)$$

$$W = \{S_1, \cdots, S_K\}, \quad (10)$$

where $S_i (i = 1, \cdots, K)$ are supervoxels, which are sets of voxels and W is the set of supervoxels that jointly covers the whole set V of voxels, without overlaps.

We assign a label $\Lambda_i \in \mathcal{L}$ to each supervoxel $S_i (i = 1, \cdots, K)$ so that a modified energy $D(\Lambda; X)$ is minimized. The energy $D(\Lambda; X)$ is defined similarly to (2), as a sum of three terms:

$$D(\Lambda;X) = u_{\text{data}} D_{\text{data}}(\Lambda;X) + u_{\text{atlas}} D_{\text{atlas}}(\Lambda) + u_{\text{smooth}} D_{\text{smooth}}(\Lambda;X). \quad (11)$$

The three terms, the data term $D_{\text{data}}(\Lambda;X)$, the atlas term $D_{\text{atlas}}(\Lambda)$, and the smoothness term $D_{\text{smooth}}(\Lambda;X)$ are defined so that they make sense in terms of supervoxels, as explained in the following.

3.1 Data Term

The data term $D_{\text{data}}(\Lambda;X)$ is defined, as illustrated in Fig. 4, by:

$$D_{\text{data}}(\Lambda;X) = \sum_{S \in W} \phi(\Lambda_S; \overline{X_S}), \quad (12)$$

$$\phi(\Lambda_S; \overline{X_S}) = -\log P(\overline{X_S} | \Lambda_S). \quad (13)$$

Here, Λ_S is the label assigned to supervoxel $S \in W$, $\overline{X_S}$ is the average of the CT values at the voxels contained in supervoxel S:

$$\overline{X_S} = \frac{1}{|S|} \sum_{v \in S} X_v, \quad (14)$$

and $P(\overline{X_S}|\Lambda_S)$ is the conditional probability (likelihood) that supervoxel S's average CT value is $\overline{X_S}$ given it is assigned label Λ_S. We take $P(\overline{X_S}|\Lambda_S)$ to be the same as $P(x|l)$, since it should be independent of position.

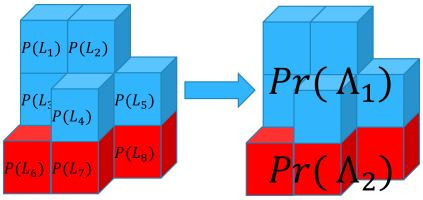

Figure 5. The prior probability that a label is assigned to a supervoxel is the average of the voxels in the supervoxel be assigned it.

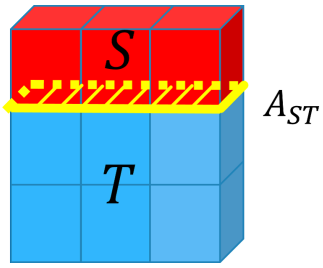

Figure 6. The area (yellow) where two supervoxels, consisting of three (red) and six (blue) voxels, respectively, touch.

3.2 Atlas Term

The atlas term D_{atlas}, as depicted in Fig. 5, corresponds to the prior probability $P(\Lambda_S)$ that supervoxel S is assigned label Λ_S, which is given by the average of the prior probabilities that the voxels in S are assigned it. Thus, the term is defined as:

$$D_{\text{atlas}}(\Lambda) = \sum_{S \in W} \psi(\Lambda_S), \tag{15}$$

$$\psi(\Lambda_S) = -\log P(\Lambda_S), \tag{16}$$

$$P(\Lambda_S) = \frac{1}{|S|} \sum_{v \in S} P(v, \Lambda_S). \tag{17}$$

3.3 Smoothness Term

In the voxel-based formulation, a voxel's neighbor is simple. Typically, six-, eighteen- or twenty-six- neighborhood system is used. In our supervoxel-based case, the arrangement of supervoxels is irregular: each supervoxel has different shape and different number of neighbors and the smoothness term must be designed accordingly. Here, we take into account both the average CT values of neighboring supervoxels and the area of the surface they touch. As illustrated in Fig. 6, we define the smoothness term $D_{\text{smooth}}(\Lambda, X)$ as follows:

$$D_{\text{smooth}}(\Lambda, X) = \sum_{S,T \in W} \eta(\Lambda_S, \Lambda_T, \overline{X_S}, \overline{X_T}), \tag{18}$$

$$\eta(\Lambda_S, \Lambda_T, \overline{X_S}, \overline{X_T}) = \begin{cases} 0 & (\Lambda_S = \Lambda_T) \\ A_{ST} \cdot \frac{1}{|\overline{X_S} - \overline{X_T}| + 1} & (\Lambda_S \neq \Lambda_T), \end{cases}$$

where S, T are neighboring supervoxels (i.e., they split at least one pair of neighboring voxels) and A_{ST} is the area they touch, which is defined as the number of neighboring pairs (v, u) such that $v \in S, u \in T$.

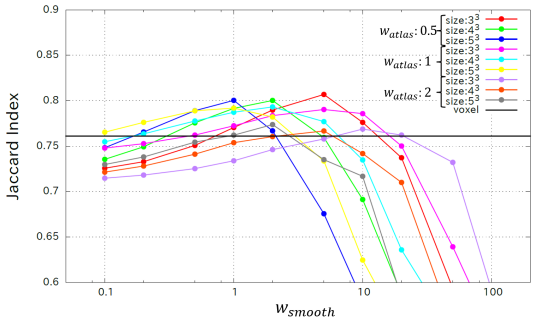

Figure 7. Plot of the average Jaccard Index obtained for 24 images using different relative weights and SLIC *size* parameters. The horizontal line indicates the performance by the baseline, voxel-based method.

4 Experimental Validation

4.1 Method

We used 24 CT images for an experimental validation of our method. We first cut out a box that roughly matches using landmarks from each image, and then resized it to $209 \times 158 \times 258$ voxel size. The 21 labels used are shown in Table 1; there are labels for 20 organs and the background.

For each CT image, we learned the prior $P(v, l)$ and the likelihood $P(x|l)$ from the other 23 images to determine the energy function, and then used it to segment it. Similar leave-one-out scheme was used for the baseline voxel-based algorithm.

We used the SLIC algorithm [6] for dividing the images into supervoxels. The algorithm has two parameters: *size* that controls the size of supervoxels and *compactness* that controls the uniformity of their shapes. Here, we fixed the *compactness* to 80 and tried three values $(3^3, 4^3, 5^3)$ of *size*.

As for the weights for the three terms in the energy, we fixed u_{data} to 1, since only the relative weights matter and u_{data} cannot be zero since the result would have no relation to the data if it were. We tried three values 0.5, 1.0, and 2.0 for u_{atlas} and ten values 0.1, 0.2, 0.5, 1.0, 2, 0, 5.0, 10, 20, 50, and 100 for u_{smooth}. We compared the 24 segmentation results with that of the voxel-based results.

To assess the accuracy of the segmentation, we computed the average of the Jaccard Index relative to the ground truth. For the comparison with the result of the voxel-based baseline, we used the weights w_{data}, w_{atlas}, and w_{smooth} that gave the best results, which had the ratio of 100 : 1 : 700.

4.2 Results

Fig. 7 shows the plot of the average Jaccard Index obtained for the 24 test CT images obtained with different relative weights and SLIC *size* parameters, compared with the voxel-based baseline results, which is shown as the horizontal line.

For our supervoxel-based method, on average the best result were obtained when the SLIC *size* parameter was set to 3^3 and the relative weight were $u_{\text{data}} : u_{\text{atlas}} : u_{\text{smooth}} = 1 : 0.5 : 5$, when the average of the Jaccard Index was 0.8069. In contrast the best result the baseline achieved was 0.7612. Thus, our method outperformed the baseline by about 4%.

We show the results from two CT images in Fig. 8. We can see that in these cases our method successfully detected

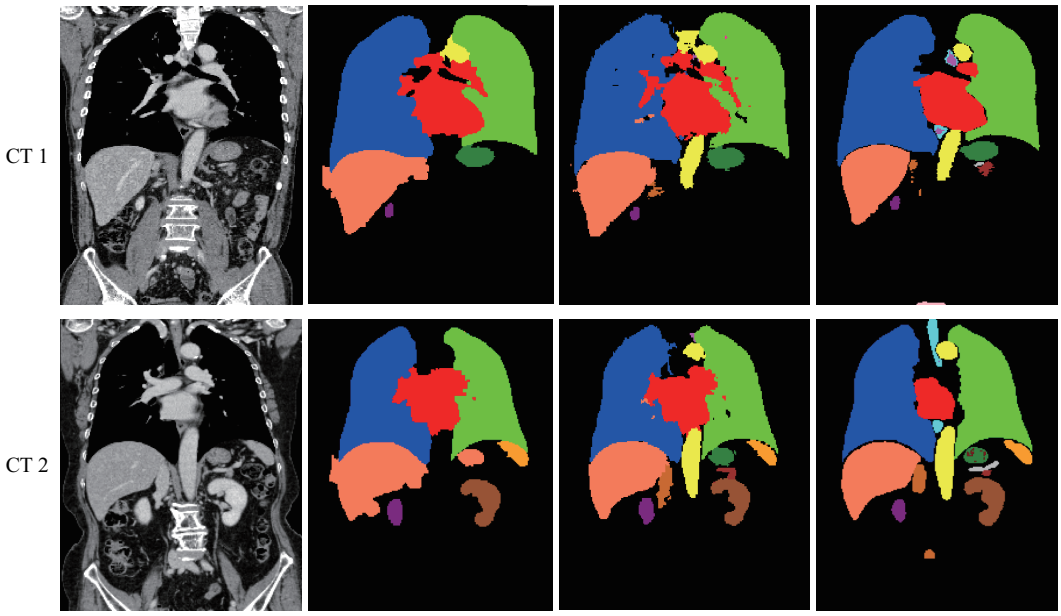

Figure 8. Segmentation results for two CT images. The relative weight in the energy terms were $u_{\text{data}} : u_{\text{atlas}} : u_{\text{smooth}} = 1 : 0.5 : 5$ and the SLIC parameters were *compactness* = 80 and *size* = 3^3.

the aorta that the baseline missed. Also, our method correctly labeled the stomach and the inferior vena cava, which the baseline mislabeled. On the other hand, the result by our method seems a little rougher than those of the baseline, which is the result of labeling isolated supervoxels.

Table 1. Organ labels

Organ	Label color
Background	Black
Right lung	Blue
Left lung	Lime
Heart	Red
Aorta	Yellow
Esophagus	Cyan
Esophageal lumen	Magenta
Liver	Blueviolet
Gallbladder	Deeppink
Stomach or duodenum	Green
Stomach or duodenal lumen	Navy
Stomach or Duodenal contents	Maroon
Spleen	Orange
Right kidney	Purple
Left kidney	Sienna
Inferior vena cava	Chocolate
Hepatic portal vein or Splenic vein or Superior mesenteric artery	Silver
Pancreas	Brown
Bladder	Pink
Prostate	Olive
Uterus	Teal

5 Conclusion

In this paper, we showed that the segmentation accuracy measured by Jaccard Index relative to the ground truth can be improved by using a supervoxel-based method in the energy minimization. Also, some problems with the voxel-based approach, such as missing organs and misshaped segments could be alleviated.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Number 26108003 as well as CREST from JST.

References

[1] C. Rother et al.: "Optimizing Binary MRFs via Extended Roof Duality", Computer Vision and Pattern Recognition, pp.1-8, 2007.

[2] Okagawa. A et al.: "Multi-organ segmentation by minimization of higher-order energy for CT boundary," Machine Vision Applications, 14th IAPR International Conference, Date 18-22. May. 2015.

[3] R. Kéchichian et al.: "Efficient multi-object segmentation of 3D medical images using clustering and graph cuts," IEEE International Conference on Image Processing, 2011.

[4] A. Lucchi et al.: "Supervoxel-Based Segmentation of Mitochondria in EM Image Stacks with Learned Shape Features," IEEE transactions on medical imaging, pp.474-486, Feb. 2012.

[5] S. Z. Li.: "Markov Random Field Models in Computer Vision," European Conference on Computer Vision, pp.361-370, 1994

[6] R. Achanta et al.: "SLIC Superpixels Compared to State-of-the-art Superpixel Methods," IEEE Pattern Analysis and Machine Intelligence, VOL. 34, NO. 11, pp.2274-2282, Nov. 2012.