Multiple-organ Segmentation Based on Spatially-divided Neighboring Data Energy

Minato Morita[1,2], Asuka Okagawa[1,2], Yuji Oyamada[1,2], Yoshihiko Mochizuki[1,2], and Hiroshi Ishikawa[1,2]

[1]Department of Computer Science and Engineering, Waseda University
[2]JST CREST

Abstract

Medical image segmentation, e.g., Computed Tomography (CT) volume segmentation, is necessary for further medical image analysis and computer aided intervention. In the standard energy minimization scheme for medical image segmentation, three terms exist in the energy: the data term, the Potts smoothing term, and the probabilistic atlas term. In this paper, we propose a novel potential function that extends the data term. The discriminability of the existing data term, which fully depends on how distinctive the objects of interest appear on CT volume, has problem when some of the objects have similar or same CT values. We overcome this limitation by considering the CT values of a pair of neighboring voxels. Increasing the voxel of interest to be evaluated, the data term become more discriminable even if some objects of interest have similar CT values. We also propose to learn the probability of the neighboring data term for each sub-region, not for each voxel. The proposed neighboring data term can be regarded as to combine the standard data term and the probabilistic atlas.

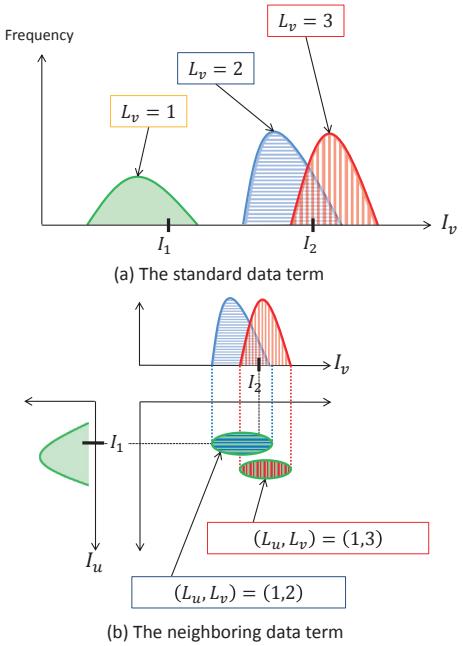

Figure 1. Our contribution. (Top) The standard data term. (Bottom) The proposed neighboring data term.

1 Introduction

Medical image segmentation, *e.g.*, Computed Tomography (CT) volume segmentation, is necessary for further medical image analysis and computer aided intervention. In the standard energy minimization scheme for medical image segmentation, three terms exist in the energy: the data term, the Potts smoothing term, and the probabilistic atlas term. In this paper, we propose a novel potential function that extends the data term. The discriminability of the existing data term, which fully depends on how distinctive the objects of interest appear on CT volume, has problem when some of the objects have similar or same CT values. We overcome this limitation by considering the CT values of a pair of neighboring voxels. Increasing the voxel of interest to be evaluated, the data term become more discriminable even if some objects of interest have similar CT values. We also propose to learn the probability of the neighboring data term for each sub-region, not for each voxel. The proposed neighboring data term can be regarded as to combine the standard data term and the probabilistic atlas.

CT volume is a set of tomographic images observed using X-ray and is used to generate a 3D image of internal objects. Since different objects blocks X-ray beam differently, their CT values are observed differently. A CT volume consists a set of voxels $\mathcal{V}$ and each voxel $v \in \mathcal{V}$ has its CT value I_v.

The task of CT volume segmentation is to decompose an input CT volume into a set of meaningful objects, *e.g.*, internal organs. It can be treated as a labeling problem in which a label representing an internal organ is assigned to each voxel. Let $\mathcal{L}$ be the set of target objects[1]; then the CT volume segmentation$L$ assigns a label $l = L_v \in \mathcal{L}$ to each voxel $v \in \mathcal{V}$.

The problem of choosing the optimal labeling can be formulated as an energy minimization problem. We define an energy function as follows:

$$E(L) = w_{\mathrm{d}} E_{\mathrm{d}}(L) + w_{\mathrm{s}} E_{\mathrm{s}}(L) + w_{\mathrm{a}} E_{\mathrm{a}}(L), \quad (1)$$

[1]Note that $\mathcal{L}$ contains a label indicating background objects that is objects of non-interest. $\mathcal{L} = 2$ for single object segmentation and $\mathcal{L} > 2$ for multiple-object segmentation.

where each term evaluates the labeling L considering different properties of CT volumes and the weights $w_{\mathrm{d}}, w_{\mathrm{s}}$, and w_{a} keep the balance between their contributions. The potentials are defined as

$$E_{\mathrm{d}}(L) = -\sum_{v \in \mathcal{V}} \log P(L_v | I_v), \tag{2}$$

$$E_{\mathrm{s}}(L) = -\sum_{(u,v) \in \mathcal{E}} \log P(L_u, L_v), \tag{3}$$

$$E_{\mathrm{a}}(L) = -\sum_{v \in \mathcal{V}} \log P(L_v), \tag{4}$$

where $\mathcal{E}$ denotes a set of neighboring voxels. The first term E_{d} is called data term that computes the probability of a label L_v given the CT value I_v for a voxel of interest $v \in \mathcal{V}$. In other words, the data term evaluates the likelihood of the organ L_v w.r.t. its CT value I_v. The second term E_{s} is called smoothness term that imposes an assumption that a pair of neighboring voxels $(u, v) \in \mathcal{E}$ tends to have same label. Following the Potts energy [1], its probability is defined as

$$P(L_u, L_v) = \begin{cases} 0 & \text{if } L_u = L_v \\ \dfrac{1}{|I_u - I_v| + \epsilon} & \text{otherwise} \end{cases}, \tag{5}$$

where ϵ is an arbitrarily small quantity for avoiding zero division. The probability is inversely proportional to the CT value difference when the voxels are assigned different labels. The third term E_{d} is called probabilistic atlas term [2] that computes the probability of a label L_v given its position.

There exist several algorithms to minimize Eq. 1 such as Graph Cut [3, 4]. State-of-the-arts used Graph Cut algorithm for either single or multi-organ segmentation [5, 6, 7, 8]

1.1 Our contribution

This paper proposes a novel potential function that extends the existing data term Eq. 2. The discriminability of the existing data term fully depends on how distinctive the objects of interest appear on CT volume. When the objects have clearer difference on their CT values, the data term well-distinguish the objects as expected. The term behaves worse when some of the objects have similar or same CT values. We overcome this limitation by considering the CT values of a pair of neighboring voxels (I_u, I_v). Increasing the voxel of interest to be evaluated, the data term become more discriminable even if some objects of interest have similar CT values. Figure 1 shows a conceptual sketch comparing the data terms. In Fig. 1, we have one distinctive object $L_v = 1$ and two objects, $L_v = 2, 3$, whose CT values are similar. The standard data term Eq. 2 returns higher probability when $I_v = I_1$ but lower probability is returned when $I_v = I_2$ because the objects of label values 2 and 3 have similar CT volume. Hence the probabilities become like $P(2|I_2) \approx P(3|I_2)$. On the other hand, the proposed

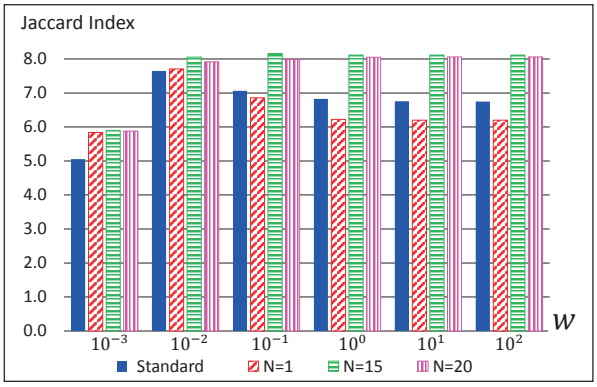

Figure 2. JI w.r.t. the weight w. The standard data term (blue filled), the proposed neighboring data term with N=1 (red diagonal lines), N=15 (green horizontal lines), and N=20 (magenta vertical lines).

neighboring data term evaluates a pair of neighboring voxels (I_u, I_v). The neighboring data term computes the joint probability of the CT values (I_u, I_v) and therefore the proposed term can distinguish the objects of similar CT values as $P(1, 2|I_u, I_v) \gg P(1, 3|I_u, I_v)$ as shown in Fig. 1 (b).

We also propose to learn the probability of the neighboring data term for each sub-region not for each voxel. The proposed neighboring data term is regarded to combine the standard data term and the probabilistic atlas. It means that the probability is trained for each voxel, however voxel based learning has some drawback. To compensate the drawback, we divide a CT volume into a set of sub-regions and then train the probability for each sub-region.

2 The Method

This section describes the proposed neighboring data term that extends the standard data term Eq. 2.

The proposed neighboring data term computes the probability of the labels (L_u, L_v) for each pair of neighboring voxels $(u, v) \in \mathcal{E}$ given their CT values (I_u, I_v). The potential function is formulated as

$$E_{\mathrm{n}}(L) = -\sum_{(u,v) \in \mathcal{E}} \log P(L_u, L_v | I_u, I_v). \tag{6}$$

It requires a lot of training data to learn Eq. 6 because the probability is conditioned by the CT values, which has much more variety than labels do. To learn the probability with less number of data, we redefine the term following Bayes theorem as

$$P(L_u, L_v | I_u, I_v) = \frac{P(I_u, I_v | L_u, L_v) P(L_u, L_v)}{P(I_u, I_v)} \tag{7}$$

$$= \frac{P(I_u, I_v | L_u, L_v) P(L_u | L_v) P(L_v)}{P(I_u, I_v)}. \tag{8}$$

With this definition, we can train the term with less number of training data.

Table 1. Mean value of JI for the case of $w = 10^{-2}$. The abbreviations denote organs: RL(right lung), LL(left lung), H(heart), A(aorta), E(esophagus), EL(esophageal lumen), L(liver), G(gallbladder), St(stomach), SL(stomach lumen), SC(stomach contents), Sp(spleen), RK(right kidney), LK(left kidney), I(inferior vena cava), PV(portal vein), Pa(pancreas), B(bladder), W(womb), and BG(background)

	RL	LL	H	A	E	EL	L	G	St	SL
standard	0.92	0.92	0.68	0.44	0.00	**0.07**	0.82	0.04	**0.34**	**0.05**
N=1	0.91	0.89	0.70	0.47	**0.00**	0.00	0.78	0.05	0.29	0.05
N=15	**0.93**	0.92	0.70	**0.47**	0.00	0.00	**0.82**	**0.07**	0.32	0.05
N=20	0.93	**0.92**	**0.70**	0.46	0.00	0.00	0.81	0.04	0.31	0.05

	SC	Sp	RK	LK	I	PV	Pa	B	W	Total
standard	0.17	**0.67**	0.54	0.55	0.10	0.01	0.03	**0.42**	**0.88**	7.65
N=1	0.17	0.55	0.52	0.58	**0.22**	**0.05**	0.19	**0.42**	**0.88**	7.71
N=15	**0.25**	0.67	**0.58**	0.64	0.16	0.03	0.15	**0.42**	**0.88**	**8.06**
N=20	**0.25**	0.66	0.57	**0.65**	0.17	0.01	0.10	**0.42**	**0.88**	7.92

We learn the probability Eq. 8 for each sub-region not for each voxel. As Eq. 8 shows, the potential is regarded as a combination of the prior probability $P(I_u, I_v|L_u, L_v)$ and the probabilistic atlas $P(L_u|L_v)P(L_v)$ and therefore is expected to be trained for each voxel. It requires more data to learn the probability for each voxel. To compensate this drawback, we divide a CT volume into a set of sub-regions $\mathcal{R}$ and learn the probability for each sub-region $r \in \mathcal{R}$ as

$$P_r(L_u, L_v|I_u, I_v) = \frac{P_r(I_u, I_v|L_u, L_v)P_r(L_u|L_v)P_r(L_v)}{P_r(I_u, I_v)}, \tag{9}$$

$$\approx \frac{P(I_u, I_v|L_u, L_v)P_r(L_u|L_v)P(L_v)}{P_r(I_u, I_v)}, \tag{10}$$

where P_r denotes the probability learned for a sub-region r. The approximation from Eq. 9 to Eq. 10 is derived from the following issues. First, we assume that the CT values conditioned by the label values is constant even in different sub-regions. Second, the probabilistic atlas $P(L_v)$ must be learned for each voxel. Otherwise, the trained probabilistic atlas always prefer to assign background label.

Note that the size of sub-regions may affect the segmentation result and the effect is analyzed in Sec. 3.

Energy minimization

The energy function is formulated as

$$E(L) = w_{\mathrm{nr}} E_{\mathrm{nr}}(L) + w_{\mathrm{s}} E_{\mathrm{s}}(L), \tag{11}$$

$$E_{\mathrm{nr}}(L) = -\sum_{(u,v)\in\mathcal{E}} \log P_r(L_u, L_v|I_u, I_v). \tag{12}$$

The proposed method finds a labeling $\hat{L}$ by minimizing Eq. 11 using Graph cut [4].

3 Experiments

This section validates the proposed method using a set of real CT volumes.

Dataset: For the experiment, 24 CT volumes are used. The size of the volumes is 512×512 pixels, the number of their slices vary from 263 to 538, and the pixel space ranges $0.546 \sim 0.820$ mm. The ground truth for each CT volume was manually generated using a semi-automatic method based on region growing and graph cut method and then a slice-by-slice correction was performed manually by an expert rater. All the CT volumes and their ground truth segmentation are rigidly registered by a simple rigid registration method and the size is set to $209 \times 158 \times 258$ voxels. The total number of organs is 20 including objects of non-interest as background.

Comparison: We compare the proposed neighboring data term Eq. 11 with the standard data term Eq. 1. As mentioned above, we tested three sub-region sizes, $\{1 \times 1 \times 1, 15 \times 15 \times 15, 20 \times 20 \times 20\}$, denoted N=1, N=15, and N=20 respectively. Note that N=1 case means that the learning is executed for each voxel, which is equivalent to sub-region of size $1 \times 1 \times 1$. Since both energy functions Eq. 11 and Eq. 1 share the smoothness term E_{s}, w_{s} is set to 1 and the other weights are varied as follows. For fair comparison, weights w_{d} and w_{a} in Eq. 1 and w_{nd} in Eq. 11 are set to same value and are exponentially proportional to w_{s} from 10^{-3} to 10^{2}.

Evaluation criteria: The segmentation performance was evaluated using the Jaccard index $J \in [0, 1]$,

$$J = \frac{X_{\mathrm{GT}} \cap \hat{X}}{X_{\mathrm{GT}} \cup \hat{X}}, \tag{13}$$

where X_{GT} denotes the ground truth and $\hat{X}$ the segmentation result. Larger Jaccard index (JI) means better segmentation result.

3.1 Results

Leave-one-out cross validation was performed for each test and we evaluate the results based on the mean value of the JI.

Results w.r.t. each weight setting

Figure 2 compares the mean value of the JI w.r.t. each weight value w_{d}, w_{a}, and w_{nd}. Bars filled blue represent JI of the standard data term and bars with colored lines represent the proposed neighboring data

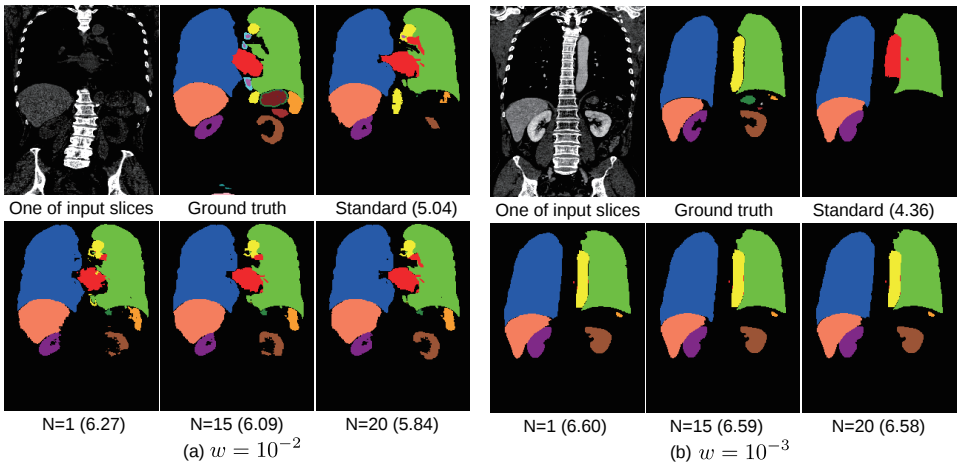

Figure 3. Segmentation results w.r.t. each organ. JI value is written in the parentheses. In the label maps, a color is assigned to each organ: right lung (blue), left lung (lime), heart (red), aorta (yellow), esophagus (cyan), esophageal lumen (magenta), liver (coral), gallbladder (deep pink), stomach (green), stomach lumen (navy), stomach contents (maroon), spleen (orange), right kidney (purple), left kidney (sienna), inferior vena cava (chocolate), portal vein (silver), pancreas (brown), bladder (pink), womb (teal), and background (black).

term: N=1 (red diagonal lines), N=15 (green horizontal lines), and N=20 (magenta vertical lines). All cases are at the peak in terms of JI at $w = 10^{-2}$ or $w = 10^{-1}$ and converge around $w = 10^{0}$. The proposed neighboring data term overcomes the standard data term except N=1 case at $w = 10^{-1}$. The result indicates that the proposed neighboring data term works better than the standard data term with sub-region based learning.

Figure 3 compares label maps, which visualize segmentation, of different CT volumes. In the segmentation maps, a color is assigned to each organ. Fig. 3 (a) shows the result of $w = 10^{-2}$ case. The proposed neighboring data term improved heart and left kidney segmentation. Heart segmented by the standard data term violated into right lung while the proposed data term could avoid such wrong segmentation. The proposed neighboring data term, especially N=15 and N=20 cases, could segment the left kidney better than the standard method. Fig. 3 (b) shows the result of $w = 10^{-3}$ case. The standard data term mis-segment aorta as heart while the proposed data term could segment aorta correctly.

Results w.r.t. each organ

Table 1 compares the mean value of the JI for the case $w = 10^{-2}$. As shown in the table, the proposed neighboring data term achieves better JI comparing to the standard data term.

4 Conclusion

In this paper, we aim at improving the standard data term used in medical image segmentation, especially its discriminability. We propose *neighboring data term* that considers the CT values of a pair of neighboring voxels. By increasing the voxel of interest to be evaluated, we make the data term more discriminable even if some objects of interest have similar CT values. We also propose to learn the probability of the neighboring data term for each sub-region. We validate the neighboring data term by comparing to the standard data term using a set of CT volumes of different persons. As the experimental results show, the proposed neighboring data term outperform the standard one.

Acknowledgement

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

References

[1] R. B. Potts, "Some generalized order-disorder transformations," *Proceedings of the Cambridge Philosophical Society*, vol. 48, no. 1, pp. 106–109, 1952.

[2] H. Park et al., "Construction of an abdominal probabilistic atlas and its application in segmentation," *IEEE TMI*, vol. 22, pp. 483–492, 2003.

[3] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE TPAMI*, vol. 26, no. 9, pp. 1124–1137, 2004.

[4] C. Rother et al., "Optimizing binary mrfs via extended roof duality," in *CVPR*, 2007, pp. 1–8.

[5] A. Afifi and T. Nakaguchi, "Liver segmentation approach using graph cuts and iteratively estimated shape and intensity constrains," in *MICCAI*, 2012, pp. 395–403.

[6] A.Ali et al., "Graph cuts framework for kidney segmentation with prior shape constraints," in *MICCAI*, 2007, pp. 384–392.

[7] B. Ayed et al., "Left ventricle segmentation via graph cut distribution matching," in *MICCAI*, 2009, pp. 901–909.

[8] C. Chu et al., "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3d abdominal ct images," in *MICCAI*, 2013, pp. 165–172.