SEMI-SUPERVISED LEARNING WITH STRUCTURED KNOWLEDGE FOR BODY HAIR DETECTION IN PHOTOACOUSTIC IMAGE

Ryo Kikkawa[1], *Hiroyuki Sekiguchi*[2], *Itaru Tsuge*[2], *Susumu Saito*[2] *and Ryoma Bise*[1,2]

[1]Kyushu University, [2]Kyoto University Hospital

ABSTRACT

Photoacoustic (PA) imaging is a promising new imaging technology for non-invasively visualizing blood vessels inside biological tissues. In addition to blood vessels, body hairs are also visualized in PA imaging, and the body hair signals degrade the visibility of blood vessels. For learning a body hair classifier, the amount of real training and test data is limited, because PA imaging is a new modality. To address this problem, we propose a novel semi-supervised learning (SSL) method for extracting body hairs. The method effectively learns the discriminative model from small labeled training data and small unlabeled test data by introducing prior knowledge, of the orientation similarity among adjacent body hairs, into SSL. Experimental results using real PA data demonstrate that the proposed approach is effective for extracting body hairs as compared with several baseline methods.

1. INTRODUCTION

Photoacoustic (PA) imaging can noninvasively visualize the 3D structure of blood vessels in vivo, which is useful in early clinical diagnosis of cancer and many other diseases. In the PA imaging process, the imaged tissues absorb laser energy and convert it into thermoelastic expansion, thereafter emitting ultrasonic waves. The 3D structures of the sound sources can then be reconstructed by sensing the PA waves [1]. In addition to blood vessels, body hairs also have a PA characteristic. Even with body hairs shaved before imaging, hair roots under skin surfaces are visualized in the PA image, as shown in Fig. 1. The body hair signals degrade the visibility of blood vessels. In addition, body hair metrics, such as the number and thickness of hairs are useful for research on alopecia and cosmetic surgery. In this paper, we propose a body hair detection method for PA imaging.

To extract body hair regions, conventional image processing approaches, such as the combination of thresholding and morphological operations, have limitations, because the intensity features of blood vessels and body hairs are similar and they often touch each other along with PA artifacts, as shown in Fig. 1. Supervised machine learning methods for 3D segmentation, such as V-net [2], usually require a large number of labeled samples. It takes a high annotation cost, however, to obtain such large annotation data from 3D volume images, and the amount of data captured from human bodies is limited, because PA imaging is a new technology.

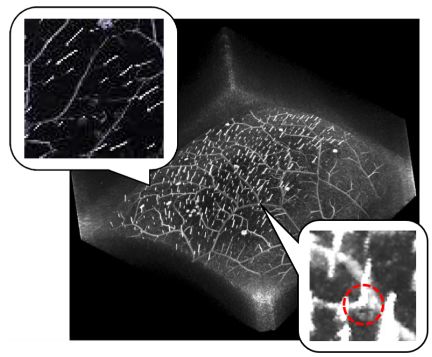

Fig. 1. Example of 3D volume captured by PA imaging. Orientations in a local area tend to be oriented in the same direction. Body hairs touch with vessels, where a red circle indicates the touching area.

To effectively learn a model from a small training dataset, semi-supervised learning (SSL) methods have been proposed [3]. To improve classification performance, SSL exploits the distribution information of large unlabeled data in the feature space. Because the amount of unlabeled data captured by new PA devices is also limited, the current SSL approaches are not sufficient to address our problem.

On the other hand, in medical images, prior knowledge such as the structure of an entire image is often useful for detection and classification. In our case, body hairs under the skin surface in a local area tend to be oriented in the same direction, as seen in Fig. 1. In conventional SSL, the feature space is extracted from a single target object but not from a global spatial structure, such as combination features of only positive samples. In our case, the orientation distribution differs among different test images, and the positive samples are unknown in the test data. Therefore, it is difficult to directly incorporate the orientation features of only positive samples in an SSL framework, because the candidate objects also include many negative samples and their labels are unknown. It is not trivial to introduce such combination features of only positive samples into an SSL framework.

Contribution : Hence, the main contribution of this work is to propose a novel SSL method for extracting body hairs in PA imaging. The method effectively learns the discriminative model from small labeled training data and small unlabeled test data by introducing prior knowledge (i.e., the orientation similarity among adjacent body hairs) into SSL. To achieve this goal, we first investigated the characteristics of body hair regions in PA images and designed object-level features that are effective for identifying body hairs. In addition, to introduce the structured knowledge that is combination features of only positive samples, the method alternatively learns the regressor for the feature space extracted from each sample and that for the feature space expressing the orientation similarity among adjacent body hairs, such that each regressor estimates the likelihood that a candidate region is a body hair. Our method requires only small labeled data, and each label for a body hair only requires to annotating the two endpoints of the hair. This annotation process is much less costly than pixel-level annotation in a 3D volume. To the best of our knowledge, this is the first attempt to automatically extract body hair regions in PA images. Experimental results using real PA data demonstrate that the proposed approach is effective for extracting body hairs as compared with several baseline methods.

Related works : Many SSL methods have been proposed to effectively use a large amount of unlabeled data [3][8]. For example, the self-training approach first trains with a small amount of labeled data, and then confident unlabeled data is added to the training data and the classifier is trained iteratively [4]. Graph-based SSL expresses the input distribution by a graph in which the nodes are labeled and unlabeled samples and edges reflect the similarities of examples. Based on the assumption that similar data points have similar labels, graph-based SSL then propagates labels from labeled data nodes to unlabeled data nodes [5][6][7]. These methods usually use only object-level features extracted from each sample rather than globally structured features. Co-training methods [9] iteratively use the predictions made by models trained on different views of the same data to label the unlabeled set and update the model with the predicted labels. The co-training approach is based on the principle of maximizing the consensus among multiple independent hypotheses. Brefeld *et.al.* [10] introduced this principle into a semi-supervised support vector learning algorithm for joint input-output spaces in the field of natural language processing (NLP). None of these methods consider how to handle features extracted from the spatial structure, such as combination features of only positive samples, in the SSL framework.

2. SSL USING STRUCTURE KNOWLEDGE

Figure 2 shows an overview of our SSL method. The proposed method first detects a set of redundant candidate regions. Then, it identifies whether a candidate region is a body hair. To effectively use prior knowledge (the orientation similarity among adjacent body hairs), the method alternately trains a regressor for the object-level features and a regressor for the features expressing the distribution of orientation

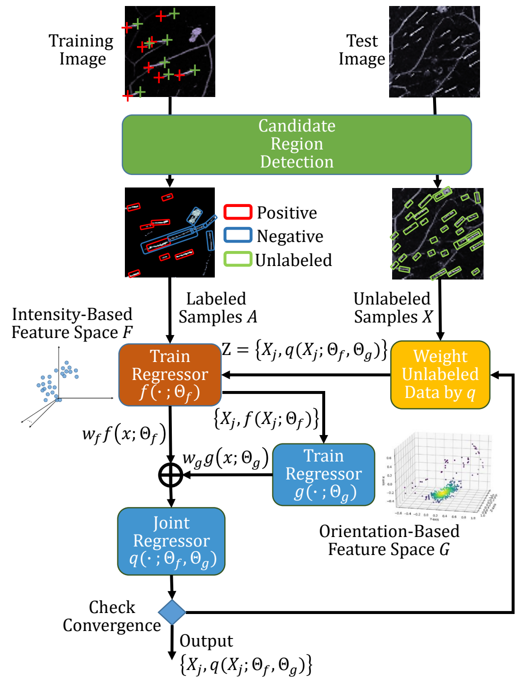

Fig. 2. Illustration of the overview of our SSL method.

similarity among adjacent body hairs. In each iteration, unlabeled data is weighted by the current regressors, and both the labeled training data and the weighted unlabeled data are used for updating both regressors. The following sections explain the details of each step.

2.1. Candidate region detection

The goal in this step is to produce a set of candidate regions that may include many false positives but very few false negatives. This would indicate that most true positives are included in the candidate set. As shown in Fig. 1, body hairs often touch blood vessels in PA imaging, because of radial artifacts in which the intensities of the regions between the body hairs and blood vessels are higher than the background intensity. Because the intensity of an artifact is usually slightly less than that of a sound source (i.e., a vessel or hair), candidate regions are identified by segmenting all regions through multiple-level thresholding. We set K thresholds that are equally spaced, and each threshold is used to segment images at a particular level of intensity. The detected regions constitute a tree structure with the candidates as nodes. To reduce the number of candidates, if a node has only one child, the child node is pruned [11]. In the training data set, if a detected candidate region includes two annotation points placed at the endpoints of a body hair, it is labeled as a positive sample; otherwise, it is labeled as negative. We denote the set of labeled candidate regions as $\mathbf{A} = \{A_i, l_i\}_{i=1}^{N_l}$, where N_l is the number of regions, and l_i is the i-th region's label. In a test image, all candidate regions are used for unlabeled data denoted by $\mathbf{X} = \{X_i\}_{i=1}^{N_u}$.

2.2. Regressor in object-level feature space

In this section, we explain the object-level features that are extracted from each candidate region. To design the object-level features to enable identification of body hair or other regions, we first investigated the characteristics of PA imaging. In general, a PA spectrum of a sample can be recorded by using different wavelengths of light. This spectrum can be used to identify the absorbing components of the sample. In such multi-spectrum images, we can observe the following characteristics: (1) the intensities of a body hair region tend to be larger than those of a blood vessel region; (2) the difference between signals captured using two different wavelengths of light (e.g., 756 and 797nm) tends to be larger in a body hair region than in other regions; and (3) the major axis length tends to be larger for a body hair region than for other regions.

Given these characteristics, we use the intensity histogram, the intensity gradient histogram, the intensity histogram of the difference between images captured by two different wavelengths, and the major axis length of the candidate region as object-level features. Then, we train the regressor $f(\cdot\,;\Theta_f)$ in the object-level feature space F by using logistic regression, in which the regressor estimates the likelihood that a candidate region is a body hair. The regressor trained with the labeled data A is used as an initial regressor in the subsequent step. In the inference for a test image, this regressor is iteratively trained and updated using the labeled data and weighted unlabeled data in the test image.

2.3. Regressor in structured-prior-based feature space

We here define the feature space G that expresses the orientation similarity among adjacent body hairs, and the regressor g for this space. To express the orientation similarity in a local area, we use a sliding window method to separate the entire volume into M local regions denoted as $L = \{L_1, ..., L_M\}$, where the window size is 1/16 of the entire volume, and the sliding distance is half the window size. This indicates that G and g consist of M spaces, $G = \{G_1, ..., G_M\}$, and M regressors, $g = \{g_1, ..., g_M\}$, respectively. In the test image, when the center of L_m is the nearest to X_j, each unlabeled candidate X_j is assigned to the corresponding space G_m.

We next explain each feature space G_i and the corresponding regressor g_i. The feature space G_i expresses the orientation distribution of candidate regions in the corresponding local window L_i. To compute the orientation of the j-th candidate region, we use the normalized first principal component (x_j, y_j, z_j) of the region as the orientation vector, and then all orientation vectors in L_i are mapped to space G_i, as illustrated in Fig. 3. Because the orientation vector is normalized to have a norm of 1, the points are distributed on the surface of a sphere of radius 1. To measure distance on the sphere's surface, we map the 3D space to a 2D space by using stereographic projection as shown in the figure. In this distribution, because the orientations among adjacent body

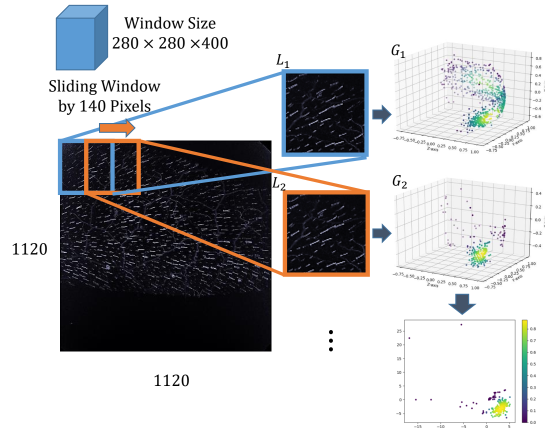

Fig. 3. Examples of local windows and their corresponding feature spaces, G. Left column: maximum intensity projection (MIP) of entire volume. Middle column: MIPs of local windows. Right column: normalized orientation distributions and their stereographic projection.

hairs tend to be similar, we assume that the standard deviation of positive samples (body hairs) becomes smaller than that of negative samples (other regions), given approximation by a Gaussian distribution. From this assumption, if the label of each sample is known, then we can simply estimate the posterior $P(hair|X_j)$ with a Bayesian classifier:

$$\begin{aligned} g(X_j;\Theta_g) &= P(hair|X_j) \\ &= \frac{p(X_j|hair)P(hair)}{P(hair)p(X_j|hair) + P(other)p(X_j|other)} \end{aligned} \tag{1}$$

where $p(X_j|hair)$ and $p(X_j|other)$ are estimated from Gaussian distributions of the positive and negative samples, respectively, and $P(hair)$ and $P(other)$ are the proportions of the samples. Θ_g is a set of parameters for modeling the distributions.

In the test step, however, the label of each sample is unknown. Instead of labels, we thus use the likelihood obtained from the object-level feature regressor f in the first iteration step. Each candidate region is weighted using $f(X_j)$, and then the Bayesian classifier is iteratively updated with the weighted unlabeled data. Note here that the orientation distribution of body hairs in the training (labeled) data may differ from that in the test (unlabeled) data. Thus, we only use the weighted unlabeled data in the inference for a test image.

2.4. Alternative semi-supervised-learning in inference

Here, we give an overview of our method. In the pre-training step, the method trains the object-level regressor $f(\cdot\,;\Theta_f)$ by using the labeled data $\mathbf{A}$. In the inference step, the method first detects candidate regions as unlabeled samples $\mathbf{X}$ from the test data. Given $\mathbf{A}$ and $\mathbf{X}$, it infers the likelihood that an unlabeled sample is positive (body hair), for all unlabeled samples $\mathbf{X}$ by using the regressor $f(\cdot\,;\Theta_f)$ pre-trained with $\mathbf{A}$, and the likelihood $f(X_j\cdot\,;\Theta_f)$ is used for the weights to

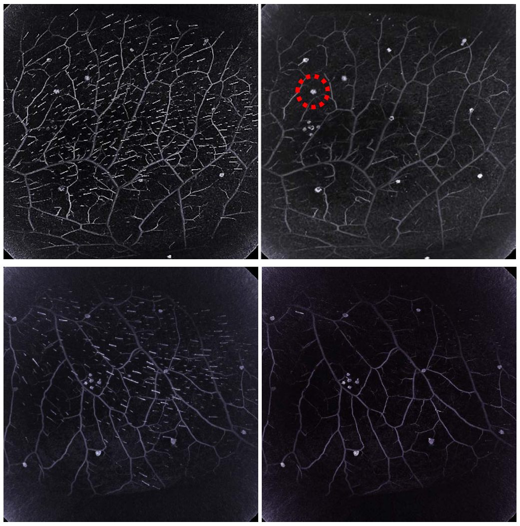

Fig. 4. Examples of virtual hair removal results, with the image contrast manually adjusted for visualization purposes. Left column: original PA images. Right column: results after hair removal.

train $g(\,\cdot\,;\Theta_g)$. Then, each unlabeled sample X_j is weighted with the joint likelihood $q(X_j\,;\Theta_f,\Theta_g) = w_f f(X_j;\Theta_f) + w_g g(X_j;\Theta_g)$, where w_f and w_g are parameters for the user setting. The set of weighted unlabeled data is denoted as $\mathbf{Z} = \{X_i, q(X_j)\}_{j=1}^{N_u}$. Using both the labeled data $\mathbf{A}$ and the weighted unlabeled data $\mathbf{Z}$, the method re-trains the regressors $f(\,\cdot\,;\Theta_f)$ and $g(\,\cdot\,;\Theta_g)$. This procedure is iterated done until it converges or reaches maximum number of iterations. Empirically, it converged in less than ten iterations in all our experiments.

3. EXPERIMENTAL RESULTS

We first quantitatively evaluated our method on a real 3D PA volume dataset captured from multiple patients, in which each volume had dimensions of $1120 \times 1120 \times 400$. To generate the ground truth, we annotated the endpoints of body hairs for three real 3D PA volumes. We set the parameter K as 20, which did not affect the result but just changed the computational cost. Then, the candidate regions and annotated points were matched to generate labeled training data. In all experiments, we simply set $w_f = w_g = 0.5$.

Because there is no existing method for automatically extracting body hair regions in PA images, we also developed two baseline methods to show the effectiveness of the proposed method. These two baselines were also based on the same candidate detection and classification approach as the proposed method. Baseline 1 was a supervised method that trained the logistic regressor in the proposed intensity-based feature space by using only the training data A; it then classified the unlabeled samples by using the trained regressor. Baseline 2 was a semi-supervised method that first pre-trained the regressor by using the training data and then iteratively self-trained the regressor by adding confident data to the training data in the inference step. As this method has the confidence threshold as a user-set parameter, we applied several thresholds (0.6, 0.7, 0.8, 0.9) and then selected the best case (0.6) and worst case (0.9) for comparison.

Table 1. Accuracy of compared methods.

	Precision	Recall	F1 score
Supervised	0.908	0.958	0.930
Semi-supervised (0.9)	0.915	0.961	0.935
Semi-supervised (0.6)	0.917	0.966	0.940
Proposed	**0.923**	**0.985**	**0.952**

To evaluate these methods, we performed cross-validation leaving one volume out and computed the average precision, recall, and F1 score. Table 1 lists the results. In this evaluation, the proposed method achieved the best performance by all metrics. The results show the effectiveness of the proposed method for training the classifier from small labeled training data and small unlabeled test data.

In addition to the above quantitative evaluation, we applied our method to virtual hair removal to improve the visibility of blood vessels in PA imaging. We first classified body hair regions by the proposed method. Then, if a candidate region was classified as a body hair, we filled the region with the average intensity of the periphery of the region. This procedure was applied to eight real PA images that were unlabeled. Figure 4 shows examples of the virtual hair removal results. The left column shows maximum intensity projection (MIP) images of the original 3D PA volume. The right column shows the resulting images after body hair removal. Note that the high-intensity points are not body hairs but markers for identifying position, as shown by the red circle in the upper-right image. These points were correctly classified as negatives by our method. In these experiments, the proposed method worked well for removing body hairs, even though the test data included diverse conditions, such as different numbers of body hairs and different intensity distributions.

4. CONCLUSION

In this paper, we proposed a novel SSL method for extracting body hair regions in PA imaging. The method can effectively learn a discriminative model from small labeled training data and small unlabeled test data by introducing prior knowledge, consisting of the orientation similarity among adjacent body hairs, into SSL. Experimental results using real PA data demonstrated that the proposed approach is effective for extracting body hairs as compared with several baseline methods.

Acknowledgments.

This work was supported by ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan) and JSPS KAKENHI Grant Number JP18025450.

5. REFERENCES

[1] Li, C., and Wang, L.V.; Photoacoustic tomography and sensing in biomedicine, Phys Med. Biol. 54 (19), pp. 59-97 (2009).

[2] Milletari, F., Navab, N., and Ahmadi, S.A.; V-Net: Fully convolutional neural networks for volumetric medical image segmentation, 3DV, pp. 565-571 (2016)

[3] Zhu, X.; Semi-supervised learning literature survey, Computer Sciences TR 1530, University of Wisconsin - Madison (2007).

[4] Rosenberg, C., Hebert, M., and Schneiderman, H.; Semi-supervised self-training of object detection models, IEEE WACV, vol. 1, pp. 29-36 (2005).

[5] Blum, A., Lafferty, J., Rwebangira, M. R., and Reddy, R.; Semi-supervised learning using randomized mincuts, ICML, pp. 13-20 (2004).

[6] Subramanya, A., Petrov, S., and Pereira, F.; Efficient graph-based semi-supervised learning of structured tagging models, in EMNLP, pp. 167-176 (2010).

[7] Kingma, D.P., Rezende, D.J., Mohamed, M., and Welling, M.; Semi-supervised learning with deep generative models, NIPS, pp. 1-9 (2014).

[8] Li, Y., and Zemel, R.; High order regularization for semi-supervised learning of structured output problems, ICML vol. 32, pp. 1368-1376 (2014).

[9] Blum, A., and Mitchell, T.; Combining labeled and unlabeled data with co-training, COLT (1998).

[10] Brefeld, U., and Scheffer, T.; Semi-supervised learning for structured output variables, ICML (2006).

[11] Bise, R., and Sato, Y.; Cell detection from redundant candidate regions under non-overlapping constraints, IEEE TMI, vol. 34, pp. 1417-1427 (2015).