Graph-based segmentation with homogeneous hue and texture vertices

Lua NGO[1, 2], Jae-Ho HAN[2, 3*]

[1]School of Biomedical Engineering, International University, Vietnam National University,
 Ho Chi Minh City, 720351, Vietnam

[2]Department of Brain and Cognitive Engineering, Korea University,
 145 Anam Rd., Seoul, 02841, South Korea

[3]Department of Artificial Intelligence, Korea University,
 145 Anam Rd., Seoul, 02841, South Korea

[*]Corresponding author: hanjaeho@korea.ac.kr

This work presents an automated segmentation method, based on graph theory, which processes superpixels that exhibit spatially similarities in hue and texture pixel groups, rather than individual pixels. The graph shortest path includes a chain of neighboring superpixels which have minimal intensity changes. This method reduces graphics computational complexity because it provides large decreases in the number of vertices as the superpixel size increases. For the starting vertex prediction, the boundary pixel in first column which is included in this starting vertex is predicted by a trained deep neural network formulated as a regression task. By formulating the problem as a regression scheme, the computational burden is decreased in comparison with classifying each pixel in the entire image. This feasibility approach, when applied as a preliminary study in electron microscopy and optical coherence tomography images, demonstrated high measures of accuracy: 0.9670 for the electron microscopy image and 0.9930 for vitreous/nerve-fiber and inner-segment/outer-segment layer segmentations in the optical coherence tomography image.

Keywords: image segmentation, deep neural network, electron microscopy, optical coherence tomography, pattern recognition.

1. Introduction

Image segmentation techniques have improved in conjunction with advances in imaging modalities and have proved to be crucial in many applications, including semiconductors, medical imaging, object recognition, and traffic control [1, 2]. In the semiconductor manufacturing process, for example, the layer delineation of a wafer is necessary when defect inspections entail the measurement of layer thicknesses for comparison with de-

signed structures. Here, imaging captures sample's epitaxial layer structures with an electron microscope; the captured images are then analyzed to detect possible defects that occur during fabrication [3]. Similar process is utilized in welding inspection of materials, normally visualized by ultrasonic or X-radiography images [4]. In medical imaging, the defect inspections are also applied to recognize the abnormalities appearing in images of human organs. Such images are used to diagnose diseases using optical coherence tomography (OCT) [5–7], to detect lung cancer using positron emission tomography and computed tomography scans [8], and classify melanoma and benign skin lesions in dermoscopic images [9, 10].

The offered benefits and range of major applications suited to image segmentation secures the technique's position in computer vision studies. In addition, fast developments in computer science make automated segmentation an excellent alternative to manual segmentation, which is both laborious and time-consuming. The graph theoretical method plays an important role in automated segmentation. Here, the method maps image pixels to a graph and utilizes graph-based theoretical optimization tools to solve the problem [11, 12].

This study proposes a graph-based segmentation method in which the graph-node unit is represented by superpixels [13, 14] that capture spatially the similarity of pixel groups rather than individual pixels, thus reducing graphics computations. Rather than computing with a pixel count based on *height × width*, the method reduces the computational complexity by taking advantage of the disproportionate and dramatic decrease in the number of vertices encountered with an increase in superpixel size. The calculated graph path portrays a continuous route with the same intensity. The effectiveness of the proposed method is demonstrated by segmenting epitaxial layers in an electron microscopy (EM) image and tissue layers in OCT image.

2. Methods

The method's approach to segmentation is based on graph theory segmentation [12]. However, instead of calculating the graph using pixel intensities, the shortest path is obtained by considering each so-called superpixel as a vertex in the graph. Here, pixels in the same spatial region of the image are grouped into a single large superpixel [13, 14]. The segmentation workflow is illustrated schematically in Fig. 1.
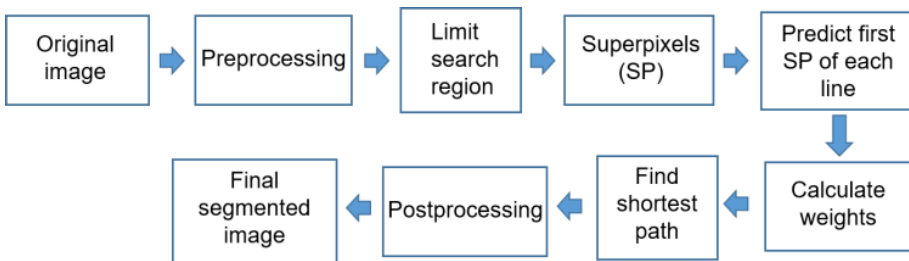

Fig. 1. Algorithm flowchart.

After speckle noise is removed through preprocessing, the regions of interest (ROI) of the images are extracted to save computational resources. Then, the ROI are partitioned into superpixels. In epitaxial wafer EM image or retinal tissue OCT image, each layer normally starts from a specific point in the image, and the intensity of the layer remains invariant or slightly changes because of the noise. Therefore, finding the starting superpixel that includes the starting point is urged to facilitate the graph path computation. The starting pixel is predicted by a deep neural network (DNN) as described in Section 2.1. The shortest path of corresponding layer is then calculated from the starting superpixel by accumulatively searching for the path with minimum total weight. The final layer boundary is obtained after postprocessing procedure.

2.1. First superpixel prediction

The first superpixel is defined by predicting the first pixel of the layer, which is included in the first superpixel. Each layer's first pixel can be distinguished clearly by the sudden changes in intensity and detected by applying a DNN to compare the intensity of each pixel vertically to the accumulated mean intensity of the previous pixels in the first column. In [12], CHIU *et al*. utilized the graph search starting from the first pixel of the images which can cause difficulty in defining the right layer. Hence, the path might go to the wrong position in the image. In addition, in [15], we applied a DNN in a regression task to predict the retinal layer boundary pixels successfully. Therefore, we leverage that characteristic and employ the first pixel prediction in this work to minimize the difficulty in the graph theory to search for the boundary path. The first-pixel prediction is formulated as a regression task in which the training input is the normalized accumulated intensity (NAI) of each column as

$$\text{NAI} = \frac{\text{Intensity-accumulate mean}}{\text{Accumulated standard deviation}} \tag{1}$$

and the training output is the responding vertical coordinate of the marked pixel that belongs to the boundary line.

Deep neural network (DNN) enjoys a wide use in medical applications such as computer-aided diagnosis (CAD) as a classification task [10] or lesion detection as an object recognition task [16, 17] or image segmentation for information interpretation in which each pixel is classified into corresponding class [5]. In the CAD, detail information is not provided to the user, which causes inconvenience in disease screening. On the contrary, the image segmentation provides more helpful details for clinicians. However, classifying each pixel in the image with a deep learning system increases computational complexity that could be limited for almost regular computers. Hence, utilizing the DNN in the regression task for predicting the pixel would leverage the information source without a heavy learning system.

A conditional distribution of the output y, which would be a real-valued x variable, is determined by the regression model as below [18]:

$$p(t\,|\,x, w, b, \beta) = \mathbb{N}(t\,|\,y(x, w, b), \beta^{-1}) \tag{2}$$
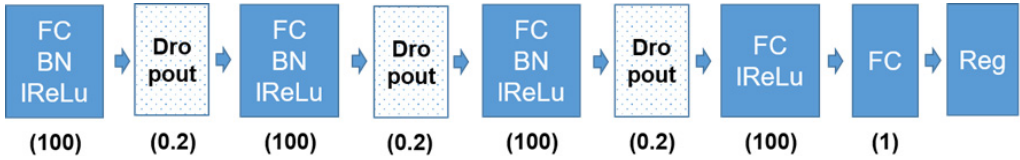
Fig. 2. Regression DNN for first-pixel boundary prediction from NAI.

in which an additive Gaussian noise is presented with zero mean and inverse variance β, w, and b represents the regression parameters. Mean square error (MSE) loss function is utilized to optimize the learning system with n training instances

$$L \;=\; \frac{1}{2n} \sum_{i=1}^{n} \left\| t_i - y_i(x) \right\|^2 \tag{3}$$

Training input and output are extracted from 1024 columns of one training image and one ground-truth image, respectively, for each type (EM image and OCT image). Therefore, 1024 training instances exist for each type of image. NAIs of the first columns in the testing images are calculated and fed to the trained network for predicting the first pixel of each layer. The DNN for regression is presented in Fig. 2.

The network includes three 100-neurons fully connected (FC) layers with batch normalization (BN) and a leakyReLU (lReLU) activation function. These FC convolution layers are interleaved with three dropout layers having dropout probability 0.2, followed by another 100-neurons FC layer with leakyReLU activation and finalized with a one-neuron FC layer without activation function and a regression layer (Reg). Stochastic gradient descent with momentum is used for training optimization; the number of training epochs is set to 100. The whole segmentation process is executed in a Microsoft Windows 10.0 64-bit, Intel Core i7-4790 3.6 GHz CPU 16.0 GB RAM PC using MATLAB R2017a (The MathWorks, Inc., Natick, MA). We utilized the trained network in [15] to predict the first pixel, which is simple for the regression task and is predicted in real-time with approximately 0.001 second using the same computer, without a complicated graphical processing unit (GPU). The training time for this network is about 30 seconds without GPU, showing low computational burden. Further information about computational complexity can be found in [15].

2.2. Graph shortest path calculation

In order to segment the layers separately, the image is partitioned into the superpixels which have similar color and brightness. Each superpixel is represented as a vertex in the graph. The graph is then cut at the path that has minimum total weight, that means the intensity change is limited inside the path. Each pair of neighboring superpixels is assigned a weight that accounts for changes in their average intensity so that the neigh-
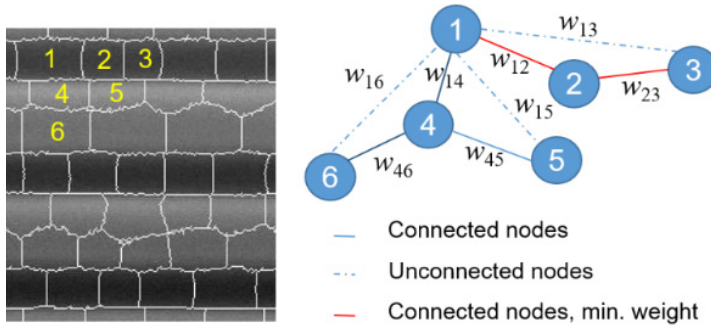
Fig. 3. Illustration of graph-based superpixel-shortest-path calculation.

boring superpixel pair with the smallest intensity change will be considered for addition to the path. The formula for calculating the graph weight is represented as

$$w_{ab} = \begin{cases} |\bar{I}_a - \bar{I}_b| + w_{\min}, & a, b \text{ are neighboring superpixels} \\ \infty, & a, b \text{ are not connected} \end{cases} \quad (4)$$

where $\bar{I}_a$ and $\bar{I}_b$ are the mean intensities of superpixels a and b, respectively, and $w_{\min} = \exp(-5)$ is small to avoid the route weight sum equaling zero. After this routine updates the weights of all superpixels, the graph shortest path is calculated based on the first superpixel of the executing layer. The shortest path of the graph is the one with the minimum total weight. The minimum total weight indicates that the change of the selected superpixels' average intensity in the path is the smallest, where these superpixels are homogeneous in hue and texture.

The shortest path is determined by the minimum-weight-sum route initiated from the predicted first superpixel. This prediction step is explained in next paragraph. The graph path calculation is illustrated in Fig. 3, which departs from the superpixel labeled 1. Direct pathways through the superpixels labeled 3, 5, and 6 do not exist; thus $w_{13}, w_{15}, w_{16} = \infty$. The pathway through superpixel 4 has a greater weight than the pathway through superpixel 2. Thus, $w_{14} > w_{12}$, *i.e.*, the change of intensities between superpixels 1 and 4 is greater than that between superpixels 1 and 2, where $w_{12} \sim w_{\min}$. Hence, the final path in the illustration is the path traversing superpixels 1–2–3. The obtained path is then post-processed with a vertical gradient calculation, thresholding, and smoothing by iterative regression.

3. Results and discussion

A [3 × 3] median filter was implemented for preprocessing to remove speckle noise. ROI extraction was performed. Then, simple linear-iterative clustering (SLIC) was used for partitioning the superpixels. The SLIC superpixel algorithm further adapted

k-means clustering more effectively when assembling the superpixels by considering the pixels' colors and their spatial regions within the target superpixel size [13]. The weight was calculated for each superpixel as in Eq. (1). Then, the shortest path was calculated in the graph from the predicted first superpixels by the breadth-first-search algorithm [19], which traversed through all possible routes in the graph to obtain the least weight-sum path.

After the path was retrieved, the boundary pixels were excerpted in postprocessing by thresholding their vertical gradients. The pixels that lay on the boundary had a greater gradient because they reflected the intensity change between different layers in the image. Then, the segmented boundary was smoothed by iterative regression, in which the adjacent pixels were fitted by smoothing spline regression, iteratively. After regression, the pixels that lay five pixel-units away either horizontally or vertically were removed. Next, the resulting boundary was smoothed again with a smoothing spline. The graph cut computation and postprocessing procedure are described in more details as in the following pseudocode:

```
I ← Image
SP ← Starting superpixel
w_ab ← Weights of all superpixel pairs in the graph
Path ← {SP}
% Calculate graph cut path
For all nodes in graph
    While there are still nodes unvisited
        Path{i} = argmin ∑w_ab; i ← added superpixels; a,b ⊂ graph nodes
    End while
End for
% Iterative regression postprocessing
[~, G] = Gradient(I) % Vertical gradient of image
Bound_G = Path∘G % Element wise multiplication
Bound_ite = Bound_G
For i = 1 : Iteration
    Bound_temp = Bound_ite
    For w = 1 : Image width
        If Bound_ite(w + 1) – Bound_temp(w) > 5
            Remove Bound_ite(w + 1)
            Bound_temp(w + 1) = Bound_temp(w)
        End if
    End for
    Bound_smooth = Spline(Bound_ite) % Smoothing spline regression
    Bound_ite = Bound_smooth
End for
```

The algorithm was verified with an EM image of epitaxial layers which captured the superimposed layers in a cross-sectional manner in the semiconductor materials.
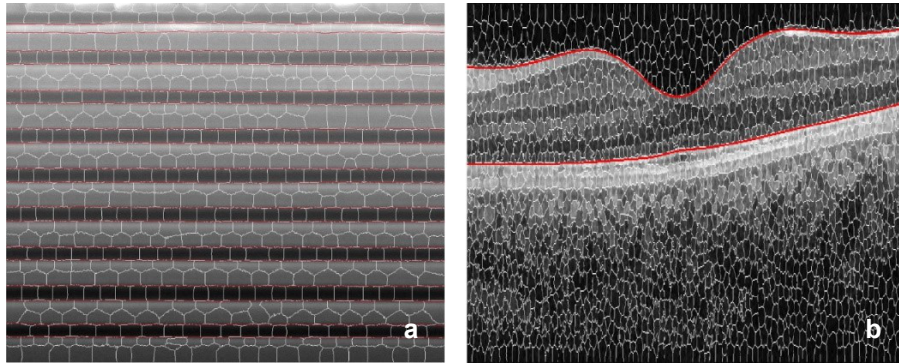
Fig. 4. Superpixel partition and boundary segmentation results (preliminary): (**a**) semiconductor epitaxial EM image with segmented layers, (**b**) retinal-tissue OCT image with segmented layers.

The segmented result of the semiconductor epitaxial-layer structure was compared with the desired design for inspection purposes. The superpixel partition and the segmented result are marked in Fig. 4**a** by solid white and red lines, respectively. In the EM image of 809×1036 dimension, we partition it into 1500 superpixels, which means each super-pixel has approximately 559 pixels, in average. The number of superpixels is a hyper-parameter that is heuristically found to obtain good result for each image. In addition, Fig. 4**b** shows the segmentation result for vitreous/nerve-fiber layer (NFL) and inner-segment/outer-segment (IS/OS) layer boundaries; here, the superpixel partition and final segmented layers appear in solid white and red lines, respectively. Differentiating the IS/OS line plays a critical role in diagnosing epiretinal membranes, which cause poor vision in patients [20]. The anonymous OCT image from Korea University Medical Center is used as a retrospective and feasibility study. 469×1424 pixeled OCT image is partitioned into 2000 superpixels to obtain good result. Therefore, about 334 pixels per each superpixel are employed in this case. Further layer segmentation can be achieved with developing a more time-efficient and accurate line tracking method for more com-plexed shapes of EM cases and OCT retinal images.

The measures of accuracy for the automated and manually segmented EM and OCT images are 0.9670 and 0.9930, respectively. Accuracy is defined as the ratio be-tween similarly labeled pixels over the total number of pixels in the image. Indeed, the accuracy of the OCT segmented result is high because this image has a larger number of background pixels than does the EM image. However, the boundary lines are well delineated in both cases, as Fig. 4 shows. The computational times for segmenting the EM and OCT images are approximately 9 and 13 seconds, respectively. This computa-tional time could be further reduced if the number of superpixels and their division pro-cess could be effectively optimized. In addition, the proposed approach can outperform in real-time using a GPU, given the training time for network in the current system without a GPU that can already be as fast as 30 seconds.

4. Conclusions

In conclusions, a graph-theory-based automated segmentation method that processes superpixels instead of direct individual pixels is proposed in this study. Performing segmentation on the superpixel units reduces the computational burdens of the graph. The calculated graph shortest path traverses through connecting superpixels which have minimal intensity changes, which means these superpixels are homogeneous in hue and texture. Here, there has been no need for data learning process with requesting test images. Instead, the starting pixel of the layer which is included in the starting superpixel is predicted by a DNN regression scheme, where the complexity is less than classifying each pixel in the entire image. As the validation for feasibility with preliminary results, the segmentation delineates the boundary lines effectively in both EM and OCT images with high accuracy: 0.9670 and 0.9930 for the EM image and the vitreous/NFL and IS/OS layer segmentations in the OCT image, respectively. These segmented layers provide crucial information during semiconductor fabrication inspections and epiretinal membranes diagnosis in ophthalmological screening, respectively. Furthermore, the same method could potentially be applied to other imaging modalities such as ultrasound, computerized tomography (CT), and magnetic resonance imaging (MRI) as well.

References

[1] YANG Y., *Image segmentation based on fuzzy clustering with neighborhood information*, Optica Applicata **39**(1), 2009, pp. 135–147.

[2] TANG R., HAN J., ZHANG X., *Efficient iris segmentation method with support vector domain description*, Optica Applicata **39**(2), 2009, pp. 365–374.

[3] NEWELL T., TILLOTSON B., PEARL H., MILLER A., *Detection of electrical defects with SEMVision in semiconductor production mode manufacturing*, [In] *2016 27th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, Saratoga Springs, NY, USA, 2016, DOI: 10.1109/ASMC.2016.7491149.

[4] MALARVEL M., SETHUMADHAVAN G., BHAGI P.C.R., KAR S., THANGAVEL S., *An improved version of Otsu's method for segmentation of weld defects on X-radiography images*, Optik **142**, 2017, pp. 109–118, DOI: 10.1016/j.ijleo.2017.05.066.

[5] ROY A.G., CONJETI S., KARRI S.P.K., SHEET D., KATOUZIAN A., WACHINGER C., NAVAB N., *ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks*, Biomedical Optics Express **8**(8), 2017, pp. 3627–3642, DOI: 10.1364/BOE.8.003627.

[6] KIM H., EOM T.J., KIM J.G., *Vascular morphometric changes during tumor growth and chemotherapy in a murine mammary tumor model using OCT angiography: a preliminary study*, Current Optics and Photonics **3**(1), 2019, pp. 54–65.

[7] HAN J.-H., CHA J., *Intraoperative imaging based on common-path time-domain reflectometry for brain tumor surgery*, Optica Applicata **50**(2), 2020, pp. 223–227, DOI: 10.37190/oa200205.

[8] GROOTJANS W., USMANIJ E.A., OYEN W.J.G., VAN DER HEIJDEN E.H.F.M., VISSER E.P., VISVIKIS D., HATT M., BUSSINK J., DE GEUS-OEI L.-F., *Performance of automatic image segmentation algorithms for calculating total lesion glycolysis for early response monitoring in non-small cell lung cancer patients during concomitant chemoradiotherapy*, Radiotherapy and Oncology **119**(3), 2016, pp. 473 –479, DOI: 10.1016/j.radonc.2016.04.039.

[9] DALILA F., ZOHRA A., REDA K., HOCINE C., *Segmentation and classification of melanoma and benign skin lesions*, Optik **140**, 2017, pp. 749–761, DOI: 10.1016/j.ijleo.2017.04.084.

[10] ESTEVA A., KUPREL B., NOVOA R.A., KO J., SWETTER S.M., BLAU H.M., THRUN S., *Dermatologist -level classification of skin cancer with deep neural networks*, Nature **542**(7639), 2017, pp. 115–118, DOI: 10.1038/nature21056.

[11] PENG B., ZHANG L., ZHANG D., *A survey of graph theoretical approaches to image segmentation*, Pattern Recognition **46**(3), 2013, pp. 1020–1038, DOI: 10.1016/j.patcog.2012.09.015.

[12] CHIU S.J., LI X.T., NICHOLAS P., TOTH C.A., IZATT J.A., FARSIU S., *Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation*, Optics Express **18**(18), 2010, pp. 19413–19428, DOI: 10.1364/OE.18.019413.

[13] ACHANTA R., SHAJI A., SMITH K., LUCCHI A., FUA P., SÜSSTRUNK S., *SLIC superpixels compared to state-of-the-art superpixel methods*, IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(11), 2012, pp. 2274–2282, DOI: 10.1109/TPAMI.2012.120.

[14] ZHANG S., WANG H., HUANG W., YOU Z., *Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG*, Optik **157**, 2018, pp. 866–872, DOI: 10.1016/j.ijleo.2017.11.190.

[15] NGO L., CHA J., HAN J.-H., *Deep neural network regression for automated retinal layer segmentation in optical coherence tomography images*, IEEE Transactions on Image Processing **29**, 2020, pp. 303 –312, DOI: 10.1109/TIP.2019.2931461.

[16] DOU Q., CHEN H., YU L., ZHAO L., QIN J., WANG D., MOK V.C.T., SHI L., HENG P.-A., *Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks*, IEEE Transactions on Medical Imaging **35**(5), 2016, pp. 1182–1195, DOI: 10.1109/TMI.2016.2528129.

[17] PENG H., LI B., LING H., HU W., XIONG W., MAYBANK S.J., *Salient object detection via structured matrix decomposition*, IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 2017, pp. 818–832, DOI: 10.1109/TPAMI.2016.2562626.

[18] BISHOP C.M., *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2016.

[19] YASUI Y., FUJISAWA K., GOTO K., *NUMA-optimized parallel breadth-first search on multicore single -node system*, [In] *2013 IEEE International Conference on Big Data*, Silicon Valley, CA, USA, 2013, DOI: 10.1109/BigData.2013.6691600.

[20] OSTER S.F., MOJANA F., BRAR M., YUSON R.M.S., CHENG L., FREEMAN W.R., *Disruption of the photo-receptor inner segment/outer segment layer on spectral domain-optical coherence tomography is a predictor of poor visual acuity in patients with epiretinal membranes*, Retina **30**(5), 2010, pp. 713 –718, DOI: 10.1097/IAE.0b013e3181c596e3.