RABIT-Model : Unsupervised Scene Understanding via Region and Boundary Integrated Topic Model

Kazuhiko Murasaki Kyoko Sudo Yukinobu Taniguchi
NTT Media Intelligence Laboratories
1-1 Hikari-no-oka Yokosuka, Kanagawa, Japan
{murasaki.kazuhiko, sudo.kyoko, taniguchi.yukinobu}@lab.ntt.co.jp

Abstract

In this paper, we propose a novel topic model that assesses the spatial coherence between neighboring regions; it classifies the boundary attributes as well as the region attributes. Our region and boundary integrated topic model*(RABIT-Model) describes the spatial relationships among superpixels by assigning topics to not only regions but also to the boundaries between them. Conventional models, on the other hand, deal with each superpixel independently and ignore spatial consistency among neighboring regions. RABIT-Model achieves accurate object segmentation and, simultaneously, can classify the attributes of the boundaries of object areas. Image segmentation experiments show that RABIT-Model provides better results than the conventional region-wise topic model. On a geometric context dataset, RABIT-Model achieves layout estimation with over* 80% *accuracy without any supervision in spite of the various appearances present in the dataset.*

1. Introduction

Object recognition and image segmentation are fundamental problems in computer vision. Many recent studies have attempted to solve these two problems in combination to estimate what and where objects are in an image. In particular, interest in topic models as a probabilistic approach for total image understanding has been increasing. Spatial Latent Topic Model, (Spatial-LTM)[3] proposed by Cao et al. assigns topics to over-segmented tiny regions of the image, and achieves concurrent object recognition and image segmentation by an unsupervised scheme that uses topic models. However, in Spatial-LTM, the regions are dealt with independently, though image features between regions or along boundaries are effectively used in other work on image segmentation. Successful approaches include object region extraction using textural variation along object boundary [1]. In our research, we propose a new latent factor for boundaries and evaluate the connectivity of neighboring regions by the use of boundary topics. Adding topics to boundaries enables us not only to classify the attributes of boundaries but also to enhance region-wise object recognition. We assume that the attributes of boundaries are related to the relationship between the neighboring regions; examples include the contour of an object and the boundary within an object. Using semantically-consistent boundaries improves image segmentation accuracy.

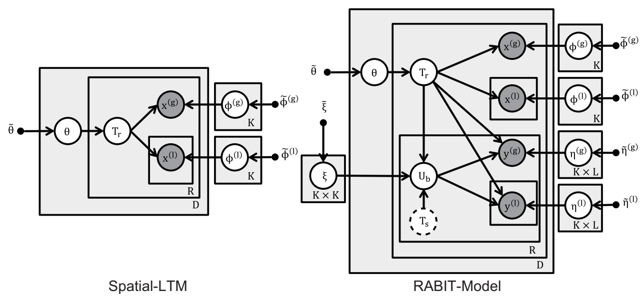

Figure 1. Graphical model of RABIT-Model

2. RABIT-Model

Because Spatial-LTM handles superpixels independently, the topics are estimated without regard to the spatial consistency among neighboring regions. Zhao modified the Spatial-LTM model to include MRF with a cost parameter; the idea was to consider the difference in topics between neighboring superpixels[6], however, their method is likely to collapse small regions because topic differences have constant costs regardless of the boundary. We have proposed a model of latent topics in which topics are related to the boundaries of regions. It is reasonable to assume that attributes can be ascribed to object boundaries such as contours, ridges, and occlusion boundaries. We consider those attributes of the boundaries as topics.

Figure 1 shows graphical models of Spatial-LTM [3] and our RABIT-Model. Parameter T stands for region topic, and parameter U for boundary topic. Boundary topic U is assigned to each neighboring pair of superpixels; it is related to both neighboring region topics. In this figure, neighboring region topic corresponding to each boundary topic is indicated by a dashed line circle. Topic T for the region is generated from $\boldsymbol{\theta}$, the distribution of the topic across the whole image. Global feature $\boldsymbol{x}^{(g)}$ and local feature $\boldsymbol{x}^{(l)}$ obtained from the superpixels are generated from T and $\boldsymbol{\phi}^{(g)}, \boldsymbol{\phi}^{(l)}$. Topic U for the boundary is generated from the combination of topics of the paired regions, and parameter $\boldsymbol{\xi}$. Global feature $\boldsymbol{y}^{(g)}$ and local feature $\boldsymbol{y}^{(l)}$ along the boundary are generated from U, T and $\boldsymbol{\eta}^{(g)}, \boldsymbol{\eta}^{(l)}$. The parameters for the distribution of visual words in the topics represented by $\boldsymbol{\phi}^{(g)}, \boldsymbol{\phi}^{(l)}, \boldsymbol{\eta}^{(g)}$ and $\boldsymbol{\eta}^{(l)}$ are generated from the Dirichlet distributions defined, respectively, by parameters $\tilde{\boldsymbol{\phi}}^{(g)}, \tilde{\boldsymbol{\phi}}^{(l)}, \tilde{\boldsymbol{\eta}}^{(g)}$ and $\tilde{\boldsymbol{\eta}}^{(l)}$.

While Spatial-LTM relates region topics only to region features, RABIT-Model also relates them to boundary topics and boundary features. The topics of a boundary indicate the change from the topics in one region to those in the neighboring region. Estimating the topics of a region and those of its boundary simultaneously improves the spatial consistency between the regions of an image.

To extract image features, we adopt Liu's segmentation method [5] for superpixel division, then extract SIFT and Geometric Blur for local features and extract color and gradient histogram for global features. The parameters of RABIT-Model are inferred by collapsed Gibbs sampling.

3. Experimental result

We conducted two experiments using different datasets to evaluate the effectiveness of RABIT-Model.

For an experiment on extracting the areas of objects and the topics assigned to them, we used the Weizmann Horse Dataset [2], which includes 327 horse images. We extracted topics from all images by RABIT-Model without any supervision, and compared the results to those of some conventional topic models. Figure 2(a) shows some illustrative examples that demonstrate the advances made by RABIT-Model. Unlike spatial-LTM, which deals with superpixels independently and is likely to separate segmented regions, RABIT-Model can extract connected fine regions. Table 1 shows the quantitative result of segmentation accuracy (the ratio of pixels labeled accurately).

Table 1. Segmentation accuracy for horse dataset

	RABIT-Model	Spatial-LTM	TRF
Accuracy	87.9%	79.6%	75.4%

To evaluate the applicability of our model to scene layout understanding, we used the Geometric Context Dataset [4], which consists of 300 landscape images. We extracted 3 topics from each image and confirmed that the 3 region topics matched the expected topics of "support", "vertical" and "sky". Figure 2(b) shows some examples of estimation results. The region labels of "support", "vertical" and "sky" are indicated by red, green and blue, respectively. Table 2 shows the quantitative results of surface layout estimation. Our model achieves clearly better results because it identifies which boundary features are indicative of regions with different topics and which are indicative of regions with the same topic. This yields spatially consistent segmentation even for images with complicated textures.

Table 2. Segmentation accuracy for geometric context dataset

	RABIT-Model	Spatial-LTM	TRF
Accuracy	80.1%	70.4%	50.2%

4. Conclusion

We introduced RABIT-Model, a novel topic model that simultaneously classify regions and boundaries, and demonstrated its superior performance on unsupervised object segmentation and unsupervised scene layout understanding.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. PAMI*, 33(5):898–916, 2011. 1

[2] E. Borenstein and S. Ullman. Learning to Segment. In *Proc. ECCV*, pages 315–328, 2004. 2

[3] L. Cao and L. Fei-Fei. Spatially Coherent Latent Topic Model for Concurrent Object Segmentation and Classification. In *Proc. ICCV*, pages 1–8, 2007. 1

[4] D. Hoiem, A. Efros, and M. Hebert. Recovering Surface Layout from an Image. *Int. J. Comput. Vision*, 75:151–172, 2007. 2

[5] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy Rate Superpixel Segmentation. In *Proc. CVPR*, pages 2097–2104, 2011. 2

[6] B. Zhao, L. FeiFei, and E. Xing. Image Segmentation with Topic Random Field. In *Proc. ECCV*, pages 785–798, 2010. 1

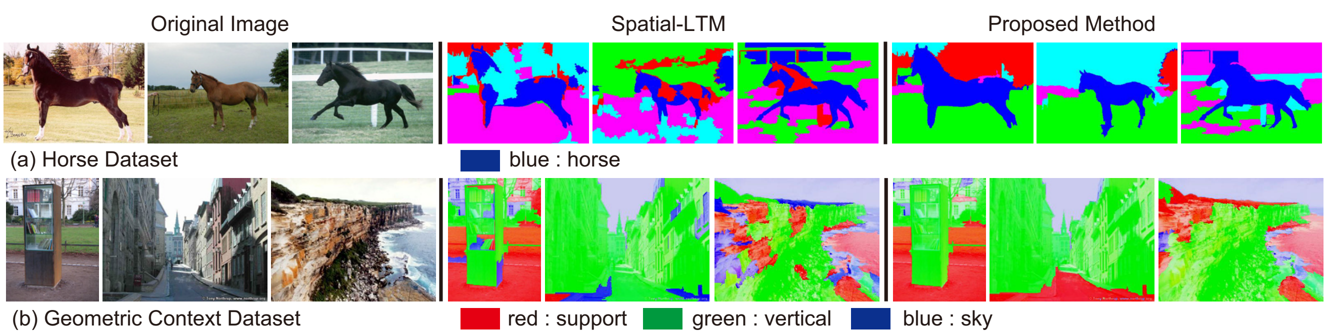

Figure 2. Segmentation results of horse dataset and geometric context dataset