Topic based Context Model for Object Detection

Jeong-Woo Son, Wonil Jang, Soyoung Park, Keeseong Cho
Broadcasting & Telecommunications Media Research Laboratory, ETRI
E-mail : {jwson, zzango1, bubble, chokis}@etri.re.kr

1. Introduction

Object detection is well-known and has widely studied in the computer vision area. In various techniques, objects are generally detected with two kinds of information: image and context information. Image information is obtained from a specific part of image by extracting particular features like HUG, SIFT, and so on. On the other hand, context information is extracted from the entire image to reflect overall context of the image such as average color, appeared objects, textures, and so on.

In this paper, among various types of contexts, the topic-level context is proposed, and then it is adopted in the Deformable Parts Model (DPM) [1] to improve its performance on object detection. To extract topic-level context for a given image, the proposed method adopts the Latent Dirichlet Allocation (LDA) [2] which has widely used in various tasks to reveal latent topics in documents. By adopting LDA, the proposed method can determine topics of an image that are not explicitly represented in the image. Thus, the topic-level context can contain information distinguishing from image information obtained by traditional image process techniques.

The evaluation of the proposed method is performed with well-known benchmark dataset, PASCAL VOC 2007. The proposed method shows about 6% average improvement over general DPM with respect to average precision (AP). Those improved AP implies the effectiveness of the topic-level context information. The general DPM already adopted context information estimated with a different viewpoint from the topic-level context. Thus, the experimental results also imply that the topic-level context can successfully cooperate with other kinds of context information, because improved AP means the topic-level context represents different information from other contexts.

2. Object Detection with Topic-Level Context

High-conceptual context which is normally determined with possibly appeared objects in an image has widely used in various object detection models. Typically, when DPM was first introduced, the latent SVMs is improved with a context model that determined a probability of an object given other objects in the image. More formally, in this paper, object detection for a specific area *a* for the image *I* is defined as

$$c^* = \underset{c \in C}{\operatorname{argmax}}\, p(c \mid if_a, cf) = \underset{c \in C}{\operatorname{argmax}}\, p(c \mid if_a) \cdot p(c \mid cf)\ , \tag{1}$$

where c is a class of object in $C(|C|=n)$, while if_a and cf denote image and context features. The image feature, if_a, contains image features for a. Thus, it reflect characteristics of a not the image I. On the other hand cf, context feature, contains context features. For example, in DPM, it is represented as a vector of objects. That is, cf, was described as

$$cf = \{ f_{a \in I}(p(c_1 \mid if_a)), f_{a \in I}(p(c_2 \mid if_a), \ldots f_{a \in I}(p(c_n \mid if_a)) \}, \tag{2}$$

where the function f is used to combine all the probabilities of c_i in C for every sub-image a in I. Various operations like product, log-sum, or max can be applied as the function f in this equation. In the case of DPM, the max operation is used as f. Note that, as shown in Equation (1) and (2), $p(c|cf)$ is determined for the image I, not for the given sub-image a.

Equation (2) can have meaningful information when it is guaranteed the simple and general assumption, "an object in an image has relations with other objects in the image". That is, $p(c_1, c_2, \ldots c_n) \neq p(c_1) \cdot p(c_2) \ldots p(c_n)$. Even though the validation of this assumption has been shown by improved performances of DPM in many tasks, the assumption can not be fully reflected with the definition of cf in Equation (2). The reason is that when cf is represented as Equation (2), values in cf is assumed to be independent each other. However, as mentioned in the assumption, objects in an image have dependencies. Thus, Equation (2) can not reflect such dependencies among objects but it only reflects dependencies between a specific object c and others.

To overcome this problem, this paper proposes topic level context representation for object detection. To relax independencies in Equation (2), latent topic $z \in Z$ ($|Z|=m$) is defined as a distribution over objects, and it is assumed that all objects are conditionally independent given topics Z. Then, cf in Equation (2) is redefined as

$$cf' = \{ p(z_1 \mid c_1, c_2, \ldots c_n), p(z_2 \mid c_1, c_2, \ldots c_n), \ldots, p(z_m \mid c_1, c_2, \ldots c_n) \}\ .$$

Since we assumed conditional independence among objects,

$$p(z_1 \mid c_1, c_2, \ldots c_n) = \prod_{i=1}^{n} p(z_1 \mid c_n) \approx \sum_{i=1}^{n} \log(p(z_1 \mid c_n)).$$

In this paper, latent topics Z is determined with the well-known topic model, Latent Dirichlet Allocation. After estimating topics, the probability of the object c given context information, $p(c|cf)$, is defined as a logistic function, $p(c \mid cf) = e^{w_c \cdot cf}$, where $w_c \in \Re^m$ is a parameter to be estimated through learning with labeled data.

3. Experiments

The proposed context representation was evaluated with the well-known dataset, PASCAL VOC 2007. PASCAL VOC dataset is constructed with 9,963 images that contain 24,640 annotated objects. Among them, 2,501 and 2,510 images are used for training and validation respectively. While, remaining 4,952 images are used for testing. In experiments, our topic-level context yields averagely 6% performance improvement of original DPM for 20 objects with respect to the Average Precision (AP). The proposed context representation also showed about 110% of maximum improvement on the class 'Dog'.

There is one thing to note. Original DPM already used context information to improve detection performance. Conceptually, the context information in DPM can be regarded as Equation (2) that estimates context features by using independent assumption. However, even though other context information is used in original DPM, the topic-level context still leads better performance for DPM. That is, the topic-level context represents aspects of images differing from other context information, and moreover, the topic-level context can be harmonized with other context by using just a simple operation. These results prove the efficiency of the topic-level context for object detection.

[Table 1] Experimental results with PASCAL VOC dataset

Class	DPM	DPM+ Topic	Improvement Ratio (%)	Class	DPM	DPM+ Topic	Improvement Ratio (%)
Aeroplane	0.285	0.295	3.2	DiningTable	0.242	0.246	1.7
Bicycle	0.550	0.551	0.0	Dog	0.049	0.109	117.6
Bird	0.006	0.008	28.6	Horse	0.451	0.455	0.6
Boat	0.144	0.161	11.3	Motorbike	0.382	0.403	5.4
Bottle	0.264	0.286	8.0	Person	0.351	0.358	0.2
Bus	0.396	0.410	3.4	PottedPlant	0.089	0.128	42.5
Car	0.500	0.503	0.4	Sheep	0.174	0.183	5.1
Cat	0.164	0.178	7.6	Sofa	0.226	0.230	1.7
Chair	0.163	0.173	6.0	Train	0.340	0.349	2.4
Cow	0.167	0.177	5.6	Tvmonitor	0.382	0.390	2.0

4. Conclusions

This paper proposes a topic-level context for object detection. In the proposed method, an image is represented with a set of objects that are possibly appeared in the image. Then, a set of topics are estimated with LDA based on the set of objects. Estimated topics are finally used to determined object probabilities for the given image by adopting a simple logistic function. Experimental results shows the significant improvement of the DPM in which the topic-level context is cooperated. Those improvements imply the efficiency of the proposed method.

5. ACKNOWLEDGEMENT

This research was funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2014.

6. References

[1] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, 2010.

[2] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3, 2003.