Image and Vision Computing 32 (2014) 579–589

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Automatic sub-category partitioning and parts localization for learning a robust object model☆

Linjia Sun, Xiaohui Liang *, Qinping Zhao

State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, BeiHang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 10 March 2014
Accepted 26 June 2014
Available online 3 July 2014

Keywords:
Object model
Topic model
Layout topic
Sub-category partitioning
Part localization

ABSTRACT

This paper introduces a novel topic model for learning a robust object model. In this hierarchical model, the layout topic is used to capture the local relationships among a limited number of parts when the part topic is used to locate the potential part regions. Naturally, an object model is represented as a probability distribution over a set of parts with certain layouts. Rather than a monolithic model, our object model is composed of multiple sub-category models designed to capture the significant variations in appearance and shape of an object category. Given a set of object instances with a bounding box, an iterative learning process is proposed to divide them into several sub-categories and learn the corresponding sub-category models without any supervision. Through an experiment in object detection, the learned object model is examined and the results highlight the advantages of our present method compared with others.

1. Introduction

The object model is one of the indispensable ingredients of many computer vision applications, such as object recognition, object detection and object segmentation. To recognize generic object category rather than just specific object instances, the object model must contain and reflect the largely intra-class variations in visual appearances of different object instances belonging to the same object category.

Many deformable object models have been proposed to capture significant variations in appearance and shape, such as deformable template models [1–5] and part-based models [6–10]. However, any single deformable model is not enough to represent the rich variations [7]. One of the solutions is to partition the generic object category into several specific sub-categories and learn the deformable model for each sub-category, as shown in Fig. 1. Thus, the generic object model is expressed by multiple sub-category models.

The present paper considers the problem of learning such object models from example images by combining the ideas of sub-category partitioning and of part-based modeling. Specifically speaking, given a set of object instances from the same object category, our method will automatically partition them into different sub-categories, locate the meaningful parts in each instance and learn the part-based model for each sub-category. It is very important to study the relationship between the sub-category partitioning and the parts localization in the process of learning object model. For example, when people view cars from various perspectives (e.g. frontal views versus side views), the visually composed parts of cars can be very different; in turn, people can use their visually composed parts to distinguish different viewpoints of car.

To achieve the goal, this paper proposes a novel hierarchical object representation based on the topic model. It is called the Layout-Part Topic Model (*LPTM*). In the *LPTM*, parts are formalized as a group of regions that are spatially clustered. The spatial relationships of parts are formalized as a group of layouts that give a description of the geometric topology for the relative positions of parts. Naturally, an object instance will be represented as a probability distribution over a set of parts with certain layouts, and the object instances belonging to the same sub-category will have similar probability distributions over the layouts and parts. In addition, the layout information is also helpful in sub-category partitioning.

To learn the distribution of object model over the layouts and parts from a given set of object instances, our method divides the *LPTM* into two Spatial Distribution Topic Models (*SDTM*), namely, the layout-based *SDTM* and the part-based *SDTM*. Accordingly, an iterative learning process is proposed based on the two *SDTMs*, i.e. using the layout-based *SDTM* to run the sub-category partitioning and using the part-based *SDTM* to run the parts localization. The sub-category partitioning is to discover several sub-categories from the set of object instances while the parts localization is to locate the possible part regions in each object instance. Each iteration leads to improved consistency between the sub-category partitioning and the parts localization, and the iteration process will stop until the sub-category partitioning no longer changes. Meanwhile, the distribution of the layouts and the parts is also obtained

☆ This paper has been recommended for acceptance by Enrique Dunn, Ph.D.

* Corresponding author at: P.O.Box 6863 , No.37 Xueyuan Road , Haidian District, Beijing 100191,China.

E-mail address: lxh@vrlab.buaa.edu.cn (X. Liang).

http://dx.doi.org/10.1016/j.imavis.2014.06.007
0262-8856/

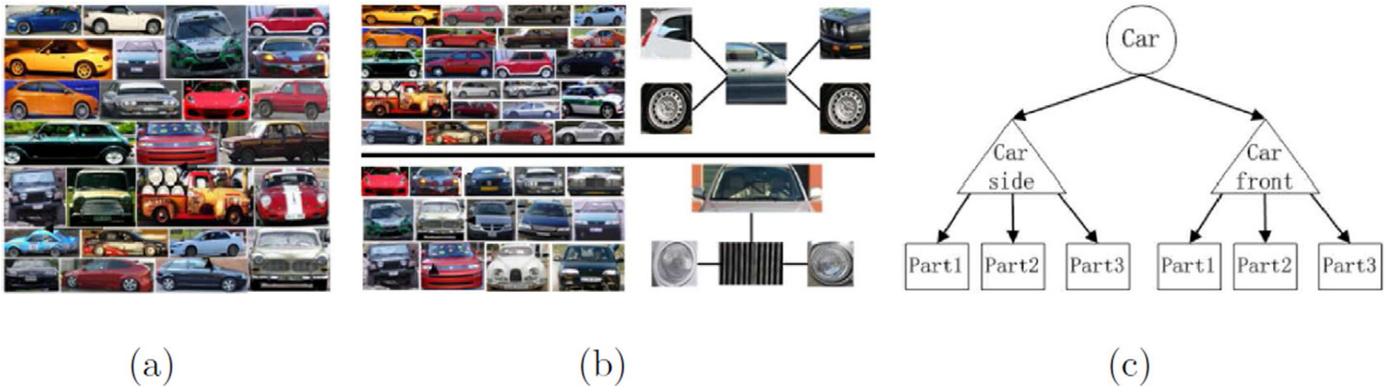

Fig. 1. (a) Given an object category with large intra-class variations, such as the car category, (b) our method will automatically partition it into multiple sub-categories and learn their models. (c) Ultimately, the general car category model is composed of multiple sub-category models.

by using the parameter learning method of the topic model in the iterative process. Finally, our object model is composed of multiple sub-category models, and each sub-category model is a parts-based model.

To sum up, the main features of this paper are: (1) to present a novel hierarchical object representation for learning a generic object model; and (2) to provide an iterative learning process to automatically discover the sub-categories and learn their part-based models from a set of object instances with a bounding box. The new method is implemented to illustrate the capability for learning object models and evaluate them on several public and standard databases through the experiment of object detection. The experimental results show that our method proves to be better than existing methods.

The remainder of this paper is organized as follows: Section 2 is a discussion on the related work. Section 3 introduces the layout-part topic model, and Section 4 describes the learning process in detail. In Section 5, experiments on multiple datasets are conducted by using our method and our method is compared with several others. Conclusions and discussion are presented in Section 6.

2. Related work

Research on learning object models has always been a hot topic in the field of computer vision. Many representative methods have been realized and many public datasets are provided for training and testing the proposed methods. Part-based methods [6] are a powerful representation because of their flexible expressiveness and intuitive interpretation. In a supervised way, each training image is labeled with part locations, and the learning process for a part-based model becomes relatively easy. Nevertheless, researchers have found out that the supervised annotation process is a serious restriction [7,8], i.e. the definition of the part requires high-level prior knowledge about the object category, and the manual annotation process is ambiguous to humans, prone to error, and time-consuming.

Recently, semi-supervised methods have been investigated for learning part-based models. These methods can automatically discover the possible parts from a set of object instances. A better method is proposed by Felzenszwalb [7], who treated the parts as latent variables during training, and the object models were trained by using a discriminative procedure that required only bounding boxes for the objects in a set of images. Similarly, [8] modeled part labels as hidden nodes based on the conditional random field and developed an *EM* algorithm to automatically discover the meaningful object parts from class labels alone. By automatically finding effective parts, the part-based modeling methods can achieve better recognition results. Inspired by these methods, our method also uses latent variables as the parts and introduces the layout information to provide consistent spatial constraints, to improve the correctness of the parts localization.

To model the large variation in visual appearances of different object instances from the same object category, unsupervised sub-category discovery is also proposed, which can simultaneously partition a generic category into multiple specific sub-categories and learn their structure models. Wu et al. [11] generated a cluster boosted tree by iteratively splitting nodes to represent possible sub-categories. Yuan et al. [12] solved the problem with standard *SVM* training, using two kernel functions: one function was used to measure the similarity between the object and the background, and the other function to address the intra-class variation. Todorovic et al. [13] recognized a category by detecting its constituent sub-categories and combined these detection results. Zhu et al. [14] used a recursive compositional model to describe a sub-category, which corresponds to a specific object and viewpoint. Zhang et al. [15] defined a pose-normalized kernel function for sub-category learning that is suitable for kernel-based learning methods. Compared with these studies, our method examines the local parts relationship and the global layouts distribution to discriminate different sub-categories.

Unsupervised object category discovery has also been used to construct object models in recent years. The idea is intuitive that categorizing and modeling should be conducted together without using expensive human annotation. Several methods use a topic model to realize their goal [16–20]. Wang et al. [16] proposed the spatial latent Dirichlet allocation (*SLDA*), which better encoded spatial structures among visual words essential for solving many vision problems. Mario et al. [17] presented a novel method for the discovery and detection of visual object categories that was based on decompositions by using topic models. Erik et al. [18] developed a hierarchical probabilistic model for objects by coupling topic models with spatial transformations. Diane et al. [19] embedded visual vocabulary creation within the object model construction, making it more suitable for object categorization. Marco et al. [20] used a topic model to represent the shape and appearance of each object and could simultaneously discover the objects in each image without supervision from a collection of images. Similarly, our method extends the original topic model by introducing the layout topic, and we focus on the process of automatically finding the possible sub-categories from a collection of instances belonging to the same object category.

3. Layout-part topic model

Usually, an object can be described as a set of parts that have a certain spatial arrangement. This paper uses a topic model to formulate the statistical relationships between the object and its parts. The topic model is originally motivated in the context of text analysis and is a generative model for modeling the statistical relationships among documents, topics and words. The topic model is depicted by the directed graphical model as shown in Fig. 2(a).