Computer Vision and Image Understanding 146 (2016) 67–76

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Cosegmentation of multiple image groups

Fanman Meng[a,*], Jianfei Cai[b], Hongliang Li[a]

[a] School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China
[b] School of Computer Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 29 March 2015
Accepted 6 February 2016

Keywords:
Image segmentation
Cosegmentation
Multiple group cosegmentation

ABSTRACT

The existing cosegmentation methods focus on exploiting inter-image information to extract a common object from a single image group. Observing that in many practical scenarios there often exist multiple image groups with distinct characteristics but related to the same common object, in this paper we propose a multi-group image cosegmentation framework, which not only discoveries inter-image information within each image group, but also transfers inter-group information among different image groups so as to produce more accurate object priors. Particularly, the multi-group cosegmentation task is formulated as an energy minimization problem, where we employ *Markov random field* (MRF) segmentation model and the dense correspondence model in the model design and adapt the *Expectation-Maximization* algorithm (EM) to solve the optimization. We apply the proposed framework on three practical scenarios including image complexity based cosegmentation, multiple training group cosegmentation and multiple noise image group cosegmentation. Experimental results on four benchmark datasets demonstrate that the proposed multi-group image cosegmentation framework is able to discover more accurate object priors and outperform state-of-the-art single-group image cosegmentation methods.

1. Introduction

Image segmentation is a fundamental problem in computer vision. Despite decades of studies, it is still challenging to achieve object-level/semantic segmentation, due to the existence of the semantic gap [1]. The existing segmentation methods can be roughly classified into three categories: unsupervised segmentation [2–4], supervised segmentation [5–7], and weakly supervised segmentation [8,9]. Unsupervised methods typically segment an image into homogeneous regions in an automatic manner, which often leads to over-segmentation such as superpixels instead of semantic regions. Supervised methods learn object priors from manually labelled pixel data, and use them to label new images. Although supervised methods can achieve superior performance for certain object classes, they require huge efforts for constructing pixel-level training labels. On the contrary, weakly supervised methods make use of coarse or partial annotations for object segmentation.

Cosegmentation can be considered as one type of weakly supervised segmentation techniques, which aims at extracting common objects from multiple images [10–12]. Based on the assumption that "*there exist common objects among the images*", cosegmentation accomplishes the segmentation by enforcing the foreground regions from different images to be consistent. Cosegmentation is extremely challenging when dealing with large variations of common objects and the interferences of complex backgrounds. In the past several years, many cosegmentation methods have been proposed, which usually add a certain foreground consistency constraint into conventional segmentation models to achieve the common object extraction, such as graphcut based cosegmentation [11,13–15], random walker based cosegmentation [16], active contours based cosegmentation [17], discriminative clustering based cosegmentation [18], and heat diffusion based cosegmentation [19].

Although the existing cosegmentation methods have been successfully demonstrated in some scenarios, they focus on the cosegmentation of a single image group, where the intra-group consistency is used to extract common objects. However, in many practical scenarios, there often exist multiple image groups with different characteristics but related to the same general object. For example, (1) For a given large-scale image, the images can be classified into several image subgroups, such as simple image group and complex image group. (2) Many training datasets for one general object often contain image groups of multiple classes, such as multiple types of "face" in face recognition and multiple kinds of "bird" species in image classification. (3) The Internet images of an object (e.g., "car") may be retrieved from several web engines such as Google and Flickr, which naturally results in the generation of several image groups with distinct characteristics according to the searching engines. The common existence of image groups

* Corresponding author.
E-mail addresses: fmmeng@uestc.edu.cn (F. Meng), asjfcai@ntu.edu.sg (J. Cai).

http://dx.doi.org/10.1016/j.cviu.2016.02.004

Fig. 1. Examples of the usefulness of inter-group information in cosegmentation. (**Top**): two subspecies groups of birds (with smooth texture and complex texture, respectively). (**Bottom**): simple image subgroup and complex image subgroup generated from a given image group.

naturally brings up the questions: *How to do cosegmentation when there exist multiple image groups with distinct characteristics? How to use the segmentation of one group to help another group?*

Intuitively, there are two straightforward solutions: one is to cosegment each image group independently; the other is to merge all the image groups into one and then use the existing cosegmentation technique to solve it. However, these straightforward solutions ignore the subtle prior information among the image groups, which could be very helpful for the cosegmentation as illustrated in the following examples.

- The in-between group information can provide a more accurate object prior and make the model more robust to the background interferences. For example, in the top row of Fig. 1(**Top**), *Shiny Cowbird* has very smooth texture (just black), which can be easily cosegmented within this group even with complex background. The segmentation results can then be used to help the cosegmentation of *Swainson Warbler* group that has complicated texture, as shown in the bottom row of Fig. 1(**Top**).
- Multiple group cosegmentation can simplify cosegmentation in terms of object prior generation and computational cost. For example, based on some image complexity analysis, we can classify the image group into two subgroups: simple image subgroup and complex image subgroup, as shown in Fig. 1(**Bottom**), where the object prior can be easily and accurately generated from the simple image subgroup instead of from the entire group. Then, the obtained object prior can be used to help cosegment of the complex image subgroup. In addition, since the size of the simple image subgroup is smaller than the original image group, it will also reduce the time cost of the cosegmentation significantly.
- The inter-group information could also help on removing noise images. Here, the noise images are refer to those without containing. For example, the images of an object retrieved from Google and Flickr are likely to contain independent noise images. By comparing among different groups, we can easily figure out the noise images.

In this paper, we propose a framework for multi-group image cosegmentation which utilizes the inter-group information to improve the cosegmentation performance, and can be used in many applications, such as image classification, object detection and object recognition. Particularly, we formulate multi-group image cosegmentation as an energy minimization problem, where our overall energy function consists of three terms: a conventional single image segmentation term that enforces foreground and background to be smooth and discriminatory, a conventional single group term that enforces the consistency between image pairs from the same group, and a novel multiple group term that enforces the consistency between image pairs from different image groups through transferring structure information between image groups. Hidden variables are also introduced to select useful image pairs within a group and across the groups. Finally, the EM algorithm is adapted to solve the energy minimization. We apply the proposed model on three practical scenarios (image complexity based cosegmentation, multiple training group cosegmentation and multiple noise image group cosegmentation) and four benchmark datasets (ICoseg, Caltech-UCSD Birds 200, Cat-Dog and Noise Image dataset). The experimental results show that the proposed method is able to achieve larger IOU values and better precision, compared with the state-of-the-art cosegmentation methods.

The rest of this paper is organized as follows. We introduce the related work in Section 2, and present the proposed framework in Section 3. The experimental results are provided in Section 4. Finally, we draw the conclusions in Section 5.

2. Related work

The existing cosegmentation methods typically extract common objects from a group of images by adding the foreground consistency constraint into the energy function of the conventional segmentation models, which can be generally represented as

$$E = \sum_{i \in \Omega_I} E^{image}(I_i) + \sum_{(i,j) \in \mathcal{C}} E^{global}(I_i, I_j) \tag{1}$$

where I_i is the ith image in an image group, $\Omega_I = \bigcup_{i=1}^{N} I_i$ is the image set, (I_i, I_j) is a pair of images, $\mathcal{C}$ is a cosegmentation relation between images, E^{image} is the conventional single image segmentation term (single term) to ensure the segment smoothness, and E^{global} is the multiple image term (global term) to enforce the foreground consistency. The cosegmentation is then achieved by