J. Vis. Commun. Image R. 31 (2015) 125–137

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Unsupervised regions based segmentation using object discovery ☆

Bai Yang [a], Huimin Yu [a,b,*], Roland Hu [a]

[a] Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[b] The State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 12 August 2014
Accepted 7 June 2015
Available online 16 June 2015

Keywords:
Cosegmentation
Segmentation
Regions based
Object discovery
Saliency
Maximal spanning tree
Structural constraint
Combinatorial optimization

ABSTRACT

We present a new unsupervised algorithm to discovery and segment out common objects from multiple images. Compared with previous cosegmentation methods, our algorithm performs well even when the appearance variations in the foregrounds are more substantial than those in some areas of the backgrounds. Our algorithm mainly includes two parts: the foreground object discovery scheme and the iterative region allocation algorithm. Two terms, a region-saliency prior and a region-repeatness measure, are introduced in the foreground object discovery scheme to detect the foregrounds without any supervisory information. The iterative region allocation algorithm searches the optimal solution for the final segmentation with the constraints from a maximal spanning tree, and an effective color-based model is utilized during this process. The comparative experimental results show that the proposed algorithm matches or outperforms several previous methods on several standard datasets.

1. Introduction

The task of fully unsupervised bottom-up segmentation remains a challenge in computer vision. With the availability of a large amount of online images, it is reasonable to jointly segment multiple images rather than to segment each image independently, because the images often share the same instance or similar objects of the same class. This new approach is termed as cosegmentation, which has recently been a topic of active research [1–8]. The objective of cosegmentation is to simultaneously segment multiple images into segments corresponding to k ($k \geqslant 2$) different classes. When $k = 2$ for a single image, this segmentation process reduces to a case of traditional figure-ground segmentation [9–11], where an image is segmented into foreground and background segments. However, without any supervisory information, it is difficult to classify the segments of a single image into a foreground or background class. Fortunately, the situation changes when prior information is available. For instance, the foreground appearance model is given in an interactive setting, where manually annotated labels are available for different k classes [2,7,12]. In terms of cosegmentation, it is assumed that multiple images contain the same object (or similar objects of the same class) and that these images can be regarded as a form of weak supervisory information to compensate for the lack of prior information in single-image segmentation.

The idea of cosegmentation was first introduced by Rother et al. [1], who proposed a novel energy function that includes global constraints to measure the dissimilarities among the two foreground histograms in addition to the classical Markov random field (MRF) segmentation terms for each image. Since this time, several works [2,3,13] based on these ideas have been presented, and, all of these studies focus on the global energy term. In Ref. [1], the global energy term was defined to use ℓ_1-norm to penalize the dissimilarity between the two foreground histograms, and a new optimization method called the trust region graph cut (TRGC) was utilized to optimize the energy function. Later, Mukherjee et al. [13] replaced the ℓ_1-norm with the ℓ_1-norm. With the ℓ_1-norm, the authors discovered several interesting properties, which enable the use of alternative optimization methods. Then, Hochbaum and Singh [2] proposed another idea by replacing the penalization term with a rewarding term. Vicente et al. [3] compared these models and proposed a new model which is a straightforward extension of the Boykov–Jolly model [14].

However, these existing methods share a commonality, they all utilize MRF-based terms for segmentation and they focus on the optimization algorithms for different forms of the global energy term. At the same time, they have a common limitation. That is, in practice, these techniques work under the assumption that common objects appear in every image, while the backgrounds vary significantly. In fact, when the backgrounds do not change or when

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author at: Institute of Information and Communication Engineering, Department of Information Science and Electronic Engineering, Zhejiang University, No. 38 Zheda Road, Hangzhou 310027, China.

E-mail addresses: yangbai@zju.edu.cn (B. Yang), yhm2005@zju.edu.cn (H. Yu), haoji_hu@zju.edu.cn (R. Hu).

http://dx.doi.org/10.1016/j.jvcir.2015.06.006
1047-3203/

even higher variations appear in the foreground objects, these methods would fail because of the confusion in identifying the foreground objects [15].

To overcome this dilemma, several methods have been proposed. In Ref. [15], Chang et al. introduced a co-saliency prior into the MRF model. Thus, the resulting energy function satisfied the submodular condition and was optimized by the graph-cut algorithm [16]. To the best of our knowledge, this is the first time that visual saliency has been introduced into the cosegmentation. Other methods [17,18] also provide sufficient evidence that saliency can be beneficial to object discovery. Moreover, an interactive scheme [2,19] has been widely used in the segmentation process to provide hints. For example, an interactive scheme using scribbles to select the foreground objects was introduced by Batra et al. [7], and an automatic recommendation system named iCoseg was proposed to intelligently recommend where the user should scribble next.

In this paper, we are motivated to propose a novel fully unsupervised region-based foreground object discovery scheme to address these ambiguity problems for the foreground object. The reasons for choosing the regions as the basic elements of our method are: (1) they encode scale and shape information of objects naturally [20]; (2) each region is represented by a rich set of features such as color, texture and saliency. In this paper we start by using the over-segmentation technique proposed by Arbelaez et al. [21] for each image. Our foreground object discovery scheme is a combination of a region-saliency prior and a region-repeatness measure across the input images. The region-saliency prior, which encodes the saliency feature for each region, is used to detect distinctive regions within each image. The theories of saliency detection [22] assume that the human vision system focuses on parts of an image in detail and that more computational resources should be allocated with the highest priority. The region-repeatness measure is used for selecting the regions that frequently appear in most images. We propose the use of high-dimensional descriptors in the distance measure of any two regions (or combinations of regions) for different images.

For region-based image segmentation and object detection, it remains a challenging problem to obtain the final segmentation result from numerous over-segmented regions. Several methods have been proposed to address this problem. Among them, combinatorial optimization methods play an important role. In [23], Russakovsky and Ng introduced a Steiner tree based approach to select object candidate regions. A maximum-weight connected subgraph technique [24] was proposed to detect non-boxy objects, and the branch-and-bound framework [25] for object localization. For cosegmentation, Kim and Xing [26] proposed a combinatorial optimization method. A combinatorial auction style optimization algorithm is used during an iterative optimization procedure that alternates between foreground modeling and region assignment.

Following these ideas, we propose a tree-constrained iterative region allocation algorithm by incorporating our foreground object discovery scheme mentioned earlier to solve the segmentation problem. The whole system works as follows: First, the initial masks of the foreground objects in different images are generated based on our foreground object discovery scheme. Then, the foreground objects are modeled using a color histogram based approach via their initial masks. Lastly, segmentation is performed via the tree-constrained iterative region allocation algorithm, which alternates between the foreground object modeling and the region allocation procedure. After every step of region allocation, we rebuild the foreground model. An effective color histogram is utilized as the foreground model and as a measurement for distance between regions.

We test our method on the established iCoseg dataset and Caltech UCSD Birds. Our experiments in Section 5 demonstrate that our method provides considerable improvements over previous methods. To summarize, the main contributions of this paper are as follows:

- A fully unsupervised foreground object discovery scheme that combines a region-saliency prior and a region-repeatness measure.
- Combining a tree-constrained iterative region allocation algorithm with the foreground object discovery scheme to solve the segmentation problem automatically.
- An effective color-based histogram serves as the foreground model and as a measurement for distance between regions. As compared in Table 1, our method has some unique characteristics including superior scalability, competitive performance over previous methods, and a desirable ability to address the foreground/background confusion problem.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed foreground object discovery scheme. Section 3 introduces the foreground models. The tree-constrained iterative region allocation algorithm for the combinatorial optimization problem is presented in Section 4. In Section 5, we provide the experimental results, and conclusions are presented in Section 6.

2. Foreground object discovery

Let $I = \{I_1, \ldots, I_N\}$ be the N images for segmentation. The objective is to segment the foreground areas containing the instances of the common objects. Any over-segmentation method can be used, and the technique of [21] is applied to each image to over-segment I_i into m_i regions. Then, our goal is to compute the binary masks $B = \{b_i, \ldots, b_N\}$, where $b_i \in \{0,1\}^{m_i}$, $b_{im} = 1$ indicates that the region m in image I_i belongs to the foreground, and $b_{im} = 0$ indicates the background.

Most previous work [12,27] utilize the central σ percent of the pixels as a Grabcut initialization based on a coarse assumption that the object of interest is likely located at the center of an image. However, in this paper, we propose a more principled way to initialize the foreground binary masks in each image. Because human eyes are easily attracted to unusual things and they focus attention on important parts, we assume that in most of the images, the regions detected as salient areas have a high probability of containing at least parts of the common object [15]. Because the foreground pixels are dissimilar to the other pixels within the images, a high saliency value is assigned by the saliency detection method. In our experiments, we obtain the original saliency maps ϕ from Cheng et al.'s global contrast based saliency [28]. Next, we present the details of our foreground object discovery scheme as follows.

2.1. Region-saliency prior

When given an original saliency map ϕ_i for each image I_i with over-segmented m_i regions, we first compute the mean saliency

Table 1
Comparison with other unsupervised cosegmentation methods. Methods and optimization algorithms are summarized; M and K denote the number of images and number of classes, respectively. Most previous unsupervised methods have focused on small-sized image sets and cannot address the foreground/background confusion.

Methods	Models/algorithms	M	K	FG ambiguity
[1]	MRF + L1 global/Trust Region GC	2	2	–
[2]	MRF + Reward global/Graph Cuts	2	2	–
[8]	MRF + Rank-1 global/Iterative opt.	$\leqslant$20	2	–
[3]	Boykov-Jolly/Dual Decomposition	2	2	–
[4]	Discriminative clustering	$\leqslant$30	2	–
Ours	Combinatorial optimization	$\geqslant$50	2	√