Computer Vision and Image Understanding 141 (2015) 18–27

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Optimizing the decomposition for multiple foreground cosegmentation ☆

Haw-Shiuan Chang, Yu-Chiang Frank Wang*

Research Center for Information Technology Innovation, Academia Sinica, 128 Academia Road, Sec. 2, Taipei 115, Taiwan

ARTICLE INFO

Article history:
Received 20 November 2014
Accepted 13 June 2015
Available online 18 June 2015

Keywords:
Multiple foreground cosegmentation
Multi-class image cosegmentation
Figure–ground segmentation
Object discovery

ABSTRACT

The goal of multiple foreground cosegmentation (MFC) is to extract a finite number of foreground objects from an input image collection, while only an unknown subset of such objects is presented in each image. In this paper, we propose a novel unsupervised framework for decomposing MFC into three distinct yet mutually related tasks: image segmentation, segment matching, and figure/ground (F/G) assignment. By our decomposition, image segments sharing similar visual appearances will be identified as foreground objects (or their parts), and these segments will be also separated from background regions. To relate the decomposed outputs for discovering high-level object information, we construct foreground object hypotheses, which allows us to determine the foreground objects in each individual image without any user interaction, the use of pre-trained classifiers, or the prior knowledge of foreground object numbers. In our experiments, we first evaluate our proposed decomposition approach on the iCoseg dataset for single foreground cosegmentation. Empirical results on the FlickrMFC dataset will further verify the effectiveness of our approach for MFC problems.

1. Introduction

Aiming at extracting the commonly presented objects, image cosegmentation [1] performs joint segmentation on a set of images sharing overlapping contents. Originally, such cosegmentation is performed on a pair of input images (e.g., [1–4]), later its extension to handling a collection of relevant images attracts more attention from researchers. While supervised cosegmentation methods utilizing user interaction [5] or pre-trained classifiers [6,7] have been presented, some further proposed to observe visual features for performing cosegmentation in an unsupervised setting (e.g., [8,9]), so that the foreground objects can be identified automatically.

Recently, Kim and Xing [10,11] proposed to solve the problem of multiple foreground cosegmentation (MFC), which is to identify multiple foreground objects and the background simultaneously during the cosegmentation process. In MFC, the number of foreground objects in each image is typically unknown. In addition, the background presented across images might be different as well. Therefore, MFC is a very challenging task to address.

In this paper, we propose an unsupervised framework for MFC. As depicted in Fig. 1, we decompose MFC into three distinct computer vision problems: image segmentation, segment matching, and figure/ground (F/G) assignment. While the first task discriminates between image segments, the second task aims at identifying foreground segments across images, and the last task is to separate the foreground segments from background regions. As discussed in Section 3, our decomposition derives and associates solutions of each task in a unified optimization framework. In our experiments, we first evaluate the performance of our method on single foreground object cosegmentation using the iCoseg dataset [5]. The use of the FlickrMFC dataset [10] further verifies the application of our approach for MFC.

1.1. Our contributions

- We propose a novel framework which decomposes MFC into three well-studied computer vision tasks, i.e., image segmentation, segment matching, and figure/ground assignment. By properly associating and updating the outputs from each task, the goal of MFC can be achieved.
- With the proposed decomposition framework, background statistics can be observed across images, and thus background regions can be automatically disregarded. Moreover, the construction of object hypothesis is able to recover foreground objects containing multiple segments, while no prior knowledge on the number of foreground objects is needed.

2. Related works

Markov Random Fields (MRF) have been applied for image cosegmentation, which utilize graph-based optimization for recognizing the common foreground object from a pair of relevant images [1,3,4]. In [3], a variety of MRF models for cosegmentation have been discussed and compared. As noted in [3], dual decomposition

☆ This paper has been recommended for acceptance by M. Pawan Kumar.
* Corresponding author.
E-mail addresses: ken77921@citi.sinica.edu.tw (H.-S. Chang), ycwang@citi.sinica.edu.tw (Y.-C.F. Wang).

http://dx.doi.org/10.1016/j.cviu.2015.06.004
1077-3142/

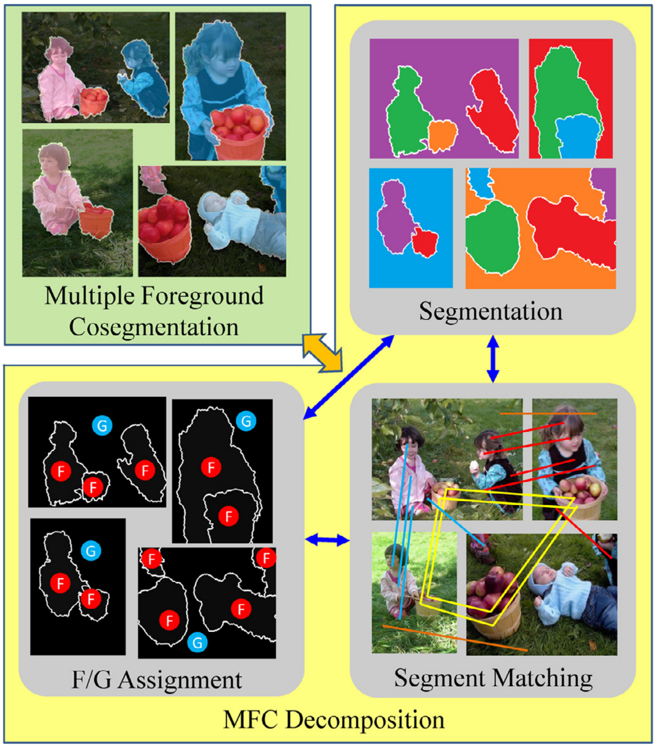

Fig. 1. Illustration of our proposed method, which decomposes MFC into the tasks of segmentation, segment matching, and figure/ground (F/G) assignment.

tackles the cosegmentation problem by advancing alternative optimization, which solves an EM-like optimization task on the smaller sub-problems. We note that, however, existing dual-decomposition based approaches focus on segmenting the single foreground object from a pair of input images with different backgrounds.

If the foreground objects exhibit significant visual appearance variations across multiple input images, more advanced matching techniques will be needed for solving the cosegmentation task (e.g., global descriptor matching [12], random forest regressor [6], graph-based matching [7], and SIFT flow [9]). Since the background regions in the input images are not necessarily distinct, it would be desirable to separate the foreground objects from such regions during cosegmentation. This is known as the figure/ground (F/G) assignment problem, which is typically solved by user interaction [5,13] or pre-trained classifiers [6,7]. Recent unsupervised cosegmentation approaches [8,9,14] derived the background models from each individual image (instead of a set of input images). Therefore, the robustness of their capability in F/G assignment will be limited.

Nevertheless, most of the above cosegmentation approaches focused on extracting a single type of foreground objects from input images. For multi-class cosegmentation methods described in [15,16,20], they did not assign foreground and background labels to their segmentation outputs even if only two classes were of interest.

Recently, Kim and Xing [10,11] proposed a problem called multiple foreground cosegmentation (MFC), which not only segments multiple types of foreground objects from the image collection, but F/G assignment will also be considered. As pointed out in [10], an exhaustive search for proper feature combination for each foreground object would be computationally prohibitive for MFC. Thus, labeled training data are required for F/G assignment in MFC (e.g., MFC-S [10], GTC [17], and MFRC [18]). On the other hand, Wang et al.[19] required the users to provide the exact number of foreground objects in input images. In practice, such user interaction or prior knowledge might not be easy to obtain, especially when the number of input images is large. While CoSand [15] has been applied for MFC in an unsupervised way (i.e., MFC-U in [10]), F/G assignment was not considered.

Table 1
Comparisons of recent cosegmentation methods. The symbol of $\not{X}$ indicates the task is partially addressed.

Methods	Unsupervised	F/G assignment	MFC
MRF	O	O	X
Batra et al. [5]	X	O	X
Vicente et al. [6]	X	O	X
Rubio et al. [7]	O	O	X
Rubinstein et al. [9]	O	O	X
Faktor and Irani [14]	O	O	X
CoSand [15]	O	$\not{X}$	X
Joulin et al. [16]	O	$\not{X}$	X
MFC-S [10]	X	O	O
GTC [17]	X	O	O
MFRC [18]	X	O	O
Wang et al. [19]	$\not{X}$	O	O
Li et al. [20]	O	$\not{X}$	O
MFC-U [10]	O	$\not{X}$	O
Ours	O	O	O

Table 2
The list of notations. $R()$ and $l()$ denote the region and label of interest, respectively.

Notation		Explanation
Region	Label	
$R(C_k)$	C_k	The region of the kth foreground class/part and its label
$R(G_i)$	G_i	The background region in image I_i and its label
$R(\mathcal{F}_r)$	$\mathcal{F}_r$	The region of the rth foreground objects and its label
p_i^j	$l(p_i^j)$	The jth superpixel in image I_i and its label
s_i^n	$l(s_i^n)$	The nth segment in image I_i and its label (i.e., the set of connecting superpixels with the same label)
O_i^m	$l(O_i^m)$	The mth foreground object hypothesis in image I_i and its labels

As highlighted in Table 1, we propose a decomposition framework for MFC in this paper. Our experiments will verify the effectiveness of our approach for both single and multiple foreground cosegmentation. For the ease of understanding, Table 2 summarizes the notations used in this paper.

3. Decomposing MFC

As illustrated in Fig. 2, we propose to decompose MFC into three different tasks, which can be associated with each other for identifying foreground object parts and background regions from the input images $I_1, \ldots, I_N$. For the jth superpixel p_i^j in image I_i, we will determine whether its label $l(p_i^j)$ belongs to one of the foreground class/part C_k or the background regions. Our decomposition can be viewed as solving the following optimization problem:

$$\min \sum_i E(\mathbf{l}_i) \;\; s.t. \begin{cases} l(p_i^j) = \hat{l}(s_i^n), \text{ for } p_i^j \in s_i^n \\ P_{\mathcal{F}}(C_k) > T, \text{ for } l(p_i^j) = C_k, \end{cases} \forall i, j, \tag{1}$$

where E indicates the energy function for segmentation, and $\mathbf{l}_i = [l(p_i^j)]_{j=1,\ldots,N_p}$ is the label vector of image I_i with its length equal to the number of superpixels N_p. The jth element $l(p_i^j)$ in $\mathbf{l}_i$ is the label of superpixel p_i^j in image I_i. We have s_i^n and $\hat{l}(s_i^n)$ as the nth segment and its desirable label in image I_i, respectively. Note that the image segment determined in this work represents the set of connecting superpixels with the same label, and the image segments with similar visual appearances across images will be identified (via segment matching) and be assigned the same label (see more details in Section 3.2). The function $P_{\mathcal{F}}()$ in (1) denotes the foreground