On Multiple Image Group Cosegmentation

Fanman Meng[1], Jianfei Cai[2], and Hongliang Li[1]

[1] School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China
[2] School of Computer Engineering, Nanyang Technological University, Singapore,
Correspondence to:asjfcai@ntu.edu.sg.

Abstract. The existing cosegmentation methods use intra-group 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 intra-group information within each image group, but also transfers the inter-group information among different groups so as to more accurate object priors. Particularly, we formulate the multi-group cosegmentation task as an energy minimization problem. Markov random field (MRF) segmentation model and dense correspondence model are used in the model design and the Expectation-Maximization algorithm algorithm (EM) is adapted to solve the optimization. The proposed framework is applied 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 show that the proposed multi-group image cosegmentation framework is able to discover more accurate object priors and significantly outperform state-of-the-art single-group image cosegmentation methods.

1 Introduction

Cosegmentation automatically extracts common objects from multiple images by forcing the segments to be consistent, which can be used in many applications, such as image classification [1] , image retrieval [2] and object recognition [3]. Such a task 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 foreground consistency constraint into traditional segmentation models to achieve the common object extraction, such as graphcut based cosegmentation [2, 4–6], random walker based cosegmentation [7], active contours based cosegmentation [8], discriminative clustering based cosegmentation [9], and heat diffusion based cosegmentation [10].

Although these methods have been successfully used in some scenarios, they mainly focus on the cosegmentation of a single image group, where intra-group information is discovered to achieve the common object extraction. However, in

many other scenarios, multiple image groups with different characteristics but related to the same common object can be formed or already exist. For example, 1) for a given image group with large number of images, we can divide them into several subgroups such as low-complexity 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., a landmark) may be retrieved from several web engines such as Google and Flicker, which naturally results in the generation of several image groups with distinct characteristics according to the searching engines. The common existence of image groups 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?*.

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. The problem with such straightforward methods is that they ignore the subtle prior information among image groups, which could be very helpful for cosegmentation as illustrated in the following examples.

- The in-between group information can provide more accurate object prior and make the model more robust to the background interferences. For example, in the top row of Fig. 1 (a), *Shiny Cowbird* has very smooth texture (just black), which can be easily cosegmented within this group even with complex background. Then, its segmentation results can be used to help the cosegmentation of *Swainson Warbler* group that has complicated texture, as shown in the bottom row of Fig. 1(a).
- Multiple group cosegmentation can simplify the cosegmentation in terms of the 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 group and complex image group, as shown in Fig. 1(b). The object prior can be easily and accurately generated from the simple image group rather than all images. In addition, since the size of the simple image group is smaller than the original one, it will also reduce the time cost of the cosegmentation significantly.
- Multiple group cosegmentation might be able to help on removing noise images. For example, the images of an object retrieved from Google and Flicker are likely to contain independent noise images. By comparing among different groups, we can easily filter out the noise images.

In this paper we propose a framework for multi-group image cosegmentation which utilises the in-between 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: traditional single image segmentation term

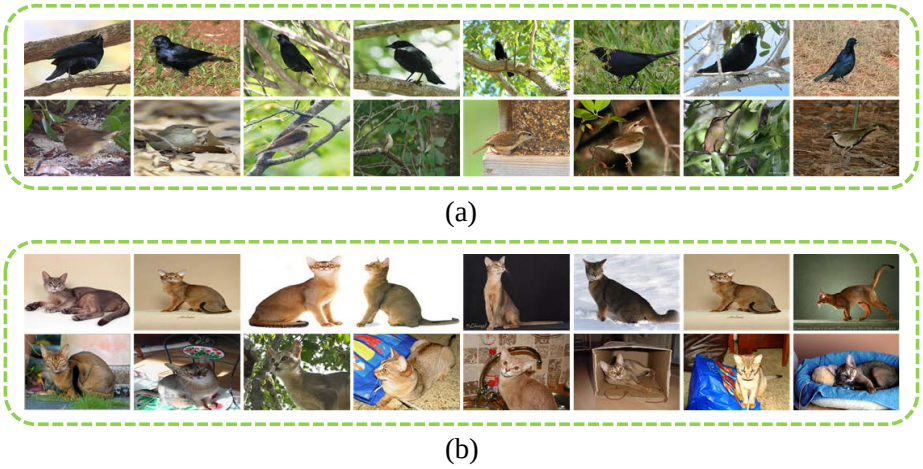

Fig. 1. Examples of the usefulness of inter-group information in cosegmentation. (a): two subspecies groups of bird (with smooth texture and complex texture, respectively). (b): simple background group and complex background group generated from a given image group.

that enforces foreground and background to be smooth and discriminatory, traditional 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. We also introduce hidden variables in the energy function to select useful image pairs within a group and across the groups. The proposed model is finally minimized by the Expectation-Maximization algorithm (EM) algorithm with some adaptations and customizations. Furthermore, we apply our 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 show that our proposed multi-group segmentation significantly outperforms the existing methods in terms of both quantitative intersection-over-union (IOU) values and visual quality.

2 Related Work

The existing cosegmentation methods focus on segmenting common object from a group of images, which is usually designed by adding the foreground consistency constraint into traditional segmentation models, i.e.

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

where E^{image} is the traditional single image segmentation term (single term) to ensure the segment smoothness, and E^{global} is the multiple image term (global term), which is to make the segments consistent with each other. The cosegmentation is then achieved by minimizing (1). Since adding the global term usually makes the energy minimization of (1) difficult, it is critical to design appropriate single term and global term for easy optimization. In the existing methods, several efficient single and global terms have been designed. For example, markov random field segmentation [2, 4–6] , random walker segmentation [7], heat diffusion segmentation [11, 12], and active contours segmentation have been used for E^{image}, while ℓ_1 norm [2], ℓ_2 norm [4] and reward measurement [5] have been proposed for E^{global} to trade off between accurate foreground similarity measurement and simple model minimization. In general, non-linear region similarity measurement is more accurate to measure the foreground consistency, but at cost of difficult energy minimization and local minimum solution. In contrast, linear region similarity measurement can result in simple model optimization, although it is not as accurate as the non-linear region similarity measurement.

Recently, more strategies have been introduced to evaluate the global term E^{global}, such as the region similarity evaluation by clustering output [9, 13], random forest based objectness evaluation model [14], the matric rank for scale invariant objects [15], second order graph matching method [16], co-saliency model [17], graph transduction learning [18] and consistent functional maps [19]. Note that these methods are still based on the model in (1). In other words, they still focus on single image group cosegmentation, where the group level information has not been explored.

There are a few cosegmentation methods that involve multiple image groups, which partially motivated us. In particular, Kim *et al.* [20] proposed a web photo streams based cosegmentation, which tries to extract common objects from multiple web photo streams. Their method focuses on extracting multiple classes of objects from streams by skillfully incorporating the photo storylines, which can improve the classification accuracy via the iteration of segmentation and classification. However, the method is essentially similar to combining the photo streams into a single image group, which does not sufficiently use the group level information in the cosegmentation. Meng *et al.* [21] recently proposed a feature adaptive cosegmentation method, which tries to learn the feature model adaptive to each image group using simple and complicated image subgroups. Since it focuses on the feature learning, its cosegmentation is still within one group.

3 Proposed Framework for Multiple Image Group Cosegmentation

3.1 Problem Formulation

Denoting multiple image groups as $\mathbf{I}^i$, we aim at extracting the common objects ω_j^i from each given image I_j^i, where I_j^i refers to the j-th image in the

i-th image group $\mathbf{I}^i$. Without lose of generalization, let's consider two image groups for simplicity: $\mathbf{I}^0 = \{I_1^0, \cdots, I_{N_0}^0\}$ and $\mathbf{I}^1 = \{I_1^1, \cdots, I_{N_1}^1\}$, where N_i is the number of images in group i, $i \in \{0, 1\}$. Denoting $\mathbf{w}^0 = \{\omega_1^0, \cdots, \omega_{N_0}^0\}$ and $\mathbf{w}^1 = \{\omega_1^1, \cdots, \omega_{N_1}^1\}$ as the set of the common object regions ω_j^i, the goal becomes extract $\mathbf{w}^0$ and $\mathbf{w}^1$ from $\mathbf{I}^0$ and $\mathbf{I}^1$, respectively.

As illustrated in Fig. 2, our basic idea is to combine the single-image consistency, the single-group consistency and the multi-group consistency whenever it is necessary so as to achieve better common object extraction. We formulate the problem as an energy minimization problem with the overall energy function:

$$E = \sum_{i=0}^{1} \alpha_i E_I(\mathbf{w}^i) + \beta_i E_S(\mathbf{w}^i) + \gamma_i E_M(\mathbf{w}^i, \mathbf{w}^{1-i}), \tag{2}$$

where E_I is the **single image segmentation term** that enforces foreground and background to be smooth and discriminatory, E_S is the **single group term** that enforces the consistency between image pairs from the same group, E_M is the **multiple group term** that enforces the consistency between image pairs from different image groups, and α_i, β_i and γ_i are tradeoff factors. In the following, we describe the three terms in detail.

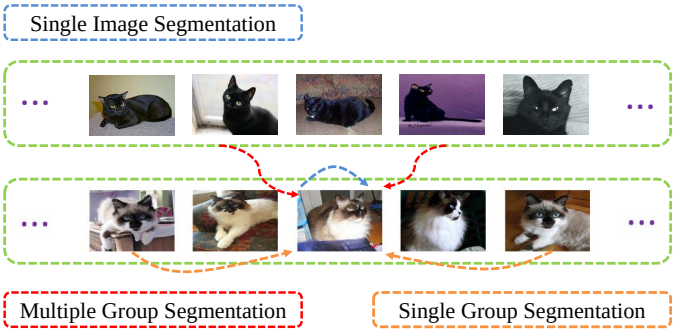

Fig. 2. Illustration of our main idea of combining single image segmentation, single group segmentation and multiple group segmentation.

Single group term $E_S(\mathbf{w}^i)$. Given a foreground set $\mathbf{w}^i$ with N_i foregrounds, $E_S(\mathbf{w}^i)$ is used to evaluate the consistencies between its elements. Particularly, in our model we evaluate the consistency by the sum of the similarities between each pair of images, i.e.

$$E_S(\mathbf{w}^i) = \sum_{(k,l),k \neq l} \mathbf{z}_{sg}(k,l) S(\omega_k^i, \omega_l^i) \tag{3}$$

with the similarity function defined as

$$S(\omega_k^i, \omega_l^i) = \sum_{p \in \omega_k^i} -\log(P(p|F_{\omega_l^i})), \tag{4}$$

where $P(p|F_{\omega_l^i})$ is the probability of pixel p belonging to the Gaussian Mixture Model (GMM) feature model $F_{\omega_l^i}$ of foreground ω_l^i, (k,l) represents foreground pair (ω_k^i, ω_l^i) in set $\mathbf{w}^i$, and $\mathbf{z}_{sg}(k,l)$ is a hidden binary variable to indicate whether ω_k^i and ω_l^i are paired or not with 1 for pairing and 0 for not pairing. Note that (4) is essentially the GMM similarity measurement that has been widely used in MRF models. There are also many other similarity measurements such as $\ell 1$-norm [2], $\ell 2$-norm [4], which could also be adopted here. The reason we choose the GMM similarity measurement is that it is a linear measurement, which leads to simple energy minimization. The introduce of $\mathbf{z}_{sg}(k,l)$ in (3) is to create useful image pairs for consistency enforcement and avoid bringing in bad image pairs that might deteriorate the performance. All these hidden variables together form a matrix $\mathbf{z}_{sg}$ with size $N_i \times N_i$.

Multiple group term $E_M(\mathbf{w}^i, \mathbf{w}^{1-i})$. The multiple group term transfers foreground information among the image groups. Here, we define it as

$$E_M(\mathbf{w}^i, \mathbf{w}^{1-i}) = \sum_{(k,l)} \mathbf{z}_{sm}(k,l) S_m(\omega_k^i, \omega_l^{1-i}), \tag{5}$$

where (k,l) represents a foreground pair of $(\omega_k^i, \omega_l^{1-i})$ from the two different foreground sets, $\mathbf{z}_{sm}(k,l)$ is the hidden binary variable to indicate whether ω_k^i and ω_l^{1-i} from different image groups are paired or not, similar to $\mathbf{z}_{sg}(k,l)$, and S_m is the similarity measurement between foreground pair $(\omega_k^i, \omega_l^{1-i})$. All $\mathbf{z}_{sm}(k,l)$ together form a matrix $\mathbf{z}_{sm}$ with size $N_i \times N_{1-i}$.

Different from the similarity measurement S defined in (4), for the image pair similarity at group level we often want to transfer structure information such as shape from one group (e.g. simple group) to the other (e.g. complex group). Thus, we define the group-level similarity measurement as

$$S_m(\omega_k^i, \omega_l^{1-i}) = \sum_{p \in \omega_k^i} \|f_k^i(p) - f_l^{1-i}(p + \mathbf{v}(p))\|_1, \tag{6}$$

where f_k^i and f_l^{1-i} are the features of image I_k^i and I_l^{1-i}, respectively, $p + \mathbf{v}(p)$ is a pixel in image I_l^{1-i} corresponding to pixel p in image I_k^i, and $\mathbf{v}(p)$ is the flow vector of pixel p. We use the SIFT flow method [22] to obtain the flow vector set $\mathbf{v}$. The feature f_k^i could be SIFT, color, or other features, depending on the applications.

Single image term $E_I(\mathbf{w}^i)$. Single image term is to ensure the smoothness of the segmentation and the distinction of the foreground and the background.

Following common MRF segmentation model, $E_I(\mathbf{w}^i)$ is defined as

$$E_I(\mathbf{w}^i) = \sum_{k=1}^{N_i} S(\omega_k^i, \omega_k^i) + S(\bar{\omega}_k^i, \bar{\omega}_k^i) + V(\omega_k^i) \tag{7}$$

where S is the same as that in (4), $\bar{\omega}_k^i = \{p | p \in \mathbf{\Omega}_k^i, p \notin \omega_k^i\}$ is the background, $\mathbf{\Omega}_k^i$ is the pixel set of image I_k^i, and $V(\omega_k^i)$ is the smoothness term regularizing the segment mask ω_k^i. We select V as the pairwise term in the common MRF segmentation model [2]. The first two terms in (7) are essentially the data terms, respectively measuring how well foreground and background pixels match the foreground and background GMM feature models of the image itself.

3.2 Optimization Solution

We now present our solution to the optimization problem of (2). Considering there are hidden variables in (2), we adapt the EM to find the solution, which consists of two alternatively iterative steps: E-step and M-step. In the E-step, we update the hidden variables $\mathbf{z}_{sg}^i$ and $\mathbf{z}_{sm}^i$ based on the feature consistency of the segments, while in the M-step we refine the segments based on the updated hidden variables. In the following, we describe the two steps in detail.

E-step: updating z. In the E-step, we update $\mathbf{z}$ by the K nearest-neighbor search. Given the segmentation results in the t-th iteration, we represent each segment by a feature such as color or SIFT. Then, for each segment ω, we calculate its K nearest neighbors denoted as $\mathbf{N}(\omega)$. For a segment pair (ω and ω_k), we set the corresponding hidden variable $z = 1$ if $\omega_k \in \mathbf{N}(\omega)$; otherwise, $z = 0$. In this way, we update $\mathbf{z}_{sg}$ and $\mathbf{z}_{sm}$ respectively when the images are in the same group or different groups.

The nearest neighbors are usually searched based on a certain distance metric such as Euclidean distance or Qi-square distance. We observe that these distances may not handle the region interferences very well. For example, in Fig.3(a), the current foreground contains $A + B$, where A is the object region we want while B is the noise region. Directly using those common distance metrics might find the nearest neighbors that contain both A and B such as Fig.3(b) and exclude the ideal neighbors such as Fig.3(c). To avoid such cases, we define the region distance between two foregrounds as

$$D(\omega, \omega_k) = \frac{1}{|\omega_k|} \sum_{q \in \omega_k} \min_{p \in \omega} d(f(p), f(q)), \tag{8}$$

where $f(p)$ is the feature representation of pixel p and d is the Euclidean distance. In this way, the foreground in Fig. 3(c) will have a small distance to that in Fig. 3(a). To speed up the process, we compute the distance in (8) based on the segment (superpixel) obtained by the simple linear iterative clustering (SLIC) superpixel generation method [23] (with the pixel number 300).

(a) (b) (c)

Fig. 3. An example of foreground distance measurement for K nearest-neighbor search.

M-step: refining cosegmentation. In the M-step, we fix $\mathbf{z}_{sg}$ and $\mathbf{z}_{sm}$, and want to refine $\mathbf{w}^0$ and $\mathbf{w}^1$ by minimizing (2). However, directly minimizing (2) is difficult since it involves two image groups and each image group contains multiple images. To make the problem trackable, we propose to solve each image segmentation separately by fixing the foregrounds of other images as constants. In this way, we divide the minimization problem into many sub-minimization problems. The energy function of each sub-minimization problem becomes

$$E_k^i = \alpha_i[S(\omega_k^i, \omega_k^i) + S(\bar{\omega}_k^i, \bar{\omega}_k^i) + V(\omega_k^i)] \\ +\beta_i \sum_{l,l\neq k} \mathbf{z}_{sg}^i(k,l) S(\omega_k^i, \omega_l^i) + \gamma_i \sum_l \mathbf{z}_{sm}^i(k,l) S_m(\omega_k^i, \omega_l^{1-i}). \quad (9)$$

Since the similarity measurements S and S_m are designed as linear measurement, the energy in (9) is submodular. Hence, (9) can be efficiently minimized by the classical graphcut algorithm [24]. By solving the sub-minimization problem in (9) one by one, we then update all ω_j^i.

Overall algorithm. Alg. 1 summarizes the proposed EM based solution. Note that the input includes two 2×2 matrices M_1 and M_2, which are used to specify the propagation relationship and the similarity features used so as to accommodate different application scenarios. Specifically, if we want to use the foreground information of the j-th group for the cosegmentation of the i-th group, we set $M_1(i,j) = 1$; otherwise, we set $M_1(i,j) = 0$. The diagonal elements $M_1(i,i)$ are always set to 1. M_2 is used to specify the features used in the propagation. In this research, we mainly consider color and SIFT features. We set $M_2(i,j)$ to 0 or 1 to respectively indicate color or SIFT feature used in the information transfer from group j to group i. Note that $M_2(i,i)$ specifies the transfer feature used within group i. Based on M_1 and M_2, we can easily design the transfer direction and the corresponding feature used in the transfer.

For the initialization, we set the initial region $\mathbf{w}_0^i, i = 1, 2$ as the rectangles with a fixed distance of $0.1 \times W$ (W is the image width) to the image boundary. $\mathbf{z}_{sg}^0$ is set as zero matrix with one on the diagonal, and $\mathbf{z}_{sm}^0$ is set as zero matrix. The M-step and E-step are run iteratively until the stop condition is met, i.e. reaching the maximum number of iterations N_{stop}. Typically, the EM algorithm converges in four iterations and thus we set $N_{stop} = 4$.

Algorithm 1 Proposed multiple group cosegmentation.

Input:
Two image groups $\mathbf{I}^0$ and $\mathbf{I}^1$
The relationship matrix M_1 and M_2.
Output:
The common foreground region sets $\mathbf{w}^0$ and $\mathbf{w}^1$.
Setting iteration $t = 1$, the initial segments $\mathbf{w}_t^i, i = 0, 1$, $\mathbf{z}_{sg}^t$ and $\mathbf{z}_{sm}^t$;
while $t \leq N_{stop}$ **do**
  // M-step
  for each image I_j^i in $\mathbf{I}^i, i = 0, 1$ **do**
    Based on $\mathbf{w}_t^i, i = 0, 1$, $\mathbf{z}_{sg}^t$ and $\mathbf{z}_{sm}^t$, update ω_j^i for $\mathbf{w}_{t+1}^i$ by minimizing (9);
  end for
  // E-step
  Based on $\mathbf{w}_{t+1}^i, i = 0, 1$, update $\mathbf{z}_{sg}^{t+1}$ and $\mathbf{z}_{sm}^{t+1}$;
  $t = t + 1$;
end while
return $\mathbf{w}_{t+1}^i, i = 0, 1$;

4 Experiments

In this section, we verify the proposed method via three cosegmentation applications: image complexity based group cosegmentation, multiple training group cosegmentation and multiple noise image group cosegmentation. We use four benchmark datasets, including ICoseg [25], Caltech-UCSD Birds 200 [26], Cat-Dog [27] and Noise Image dataset [28].

4.1 Image Complexity Based Group Cosegmentation

Here, we consider the scenario of extracting a common object from a given single image group with large number of images, where some images are of simple background while others have complex background, which are difficult to segment. Following the image complexity analysis in [21], we can divide the given image group into simple image group and complex image group. For simple image group, we can easily extract the object out by using the single image group cosegmentation (setting γ_i in (2) to 0). Then, for the complex image group we perform the multiple image group cosegmetnation using our proposed framework, where the prior information generated from the simple image group is transferred to help the complex image group cosegmentation.

We test this scenario on the ICoseg dataset [25]. Color feature is selected for information transfer between the simple group and the complex group. Fig 4 shows some segmentation results of the images with complex backgrounds from the three classes *cheetah*, *elephant* and *panda2*. We can see that the proposed method can extract the common objects from interfered backgrounds, which is largely due to the accurate object prior provided by the simple group.

We next objectively evaluate the proposed method by IOU value, which is defined as the ration of the intersection area of the segment and the groundtruth

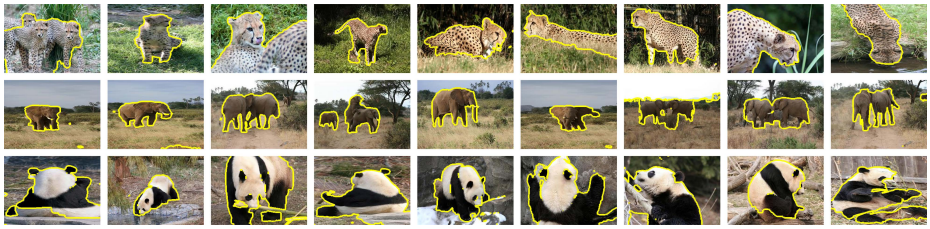

Fig. 4. The segmentation results of the proposed method on ICoseg dataset.

to their union. We use the average value of the IOU results over all the classes of the ICoseg dataset to verify the performance. The average IOU values of the proposed method and the existing methods on ICoseg dataset are shown in the second column of Table 1, where we also compare the methods of our framework without and with the multiple group term, denoted as *ours+s* and *ours+m*, respectively. It can be seen that our proposed method with the multiple group term achieves the best performance with the highest IOU value of 0.7086 on the ICoseg dataset. Note that some image classes in ICoseg only contain small number of images (smaller than ten), which is not suitable for simple and complex group division. Thus, for these small classes, only single image group cosegmentation of the proposed method is performed.

Table 1. The IOU values (Precision value for the Noise Image dataset) of the proposed method and the existing methods on Icoseg, Bird, Cat-Dog and Noise Image dataset.

Method	Icoseg	Bird	Cat	Noise
[9]	0.3947	0.2340	–	0.5270
[11]	0.3927	0.1806	0.4534	0.4695
[13]	0.4264	0.2384	–	0.6168
[28]	0.6763	0.2480	0.3950	0.5892
Ours+s	0.6514	0.3897	0.6235	–
Ours+m	0.7086	0.3957	0.6550	0.8627

Fig. 5 further gives some visual comparison among different methods. We can see that the results of the single group based method often obtain large noise regions, such as the meadow in the *Liverpool* class (the first two columns). This is mainly because these noise regions repeatedly appear in the image group, which are then being considered as part of the foregrounds. Compared with other methods, our proposed group-level cosegmentation method can successfully remove those noise regions due to the nice prior extracted from the simple image group.

Fig. 5. From top to bottom: the original images, the segmentation results of [13], [28], ours+s and ours+m methods.

4.2 Multiple Training Group Cosegmentation

In this subsection, we consider the scenario of given a training collection of a general object such as bird or cat, where there already exists some groupings according to the type of the species. Some subspecies can be easily extracted according to a certain feature while segmenting the others is challenging due to the complicated texture of the object. For such dataset, we apply the single-group image cosegmentation (*ours+s*) using either color or SIFT feature on one selected group that can be easily segmented, and then apply our multi-group image cosegmentation (*ours+m*) on other groups using SIFT feature to transfer the object prior from the easy group to each of the other groups.

For this scenario, we test the proposed method on two classification datasets: Cat-Dog dataset and Caltech-UCSD Bird dataset. The Cat-Dog dataset contains 12 subspecies of cat with about 200 images per class, and we use all the classes. In Bird dataset, there are 200 species of bird with about 30 images per class. We select 13 continuous classes from number 026 (Bronzed Cowbird) to 038 (Great Crested Flycatcher) for verification. Considering some easy group has relatively large number of images, when applying the multi-group cosegmentation, the image matching between groups becomes very time-consuming. In order to reduce the computational cost, only a subset of the images with small number of images is used as the easy group to help cosegment other groups. Specifically, in the Cat-Dog dataset, Bombay cat group with 23 images is used as the easy group, and for UCB-Bird dataset, we select a subset of 029 American Crow with 18 images as the easy group.

Fig. 6 and Fig. 7 show the segmentation results of some difficult groups in the Cat-Dog dataset and the Bird Dataset using our proposed method. Here, we give examples of the images with interfered backgrounds. We can see that the

proposed method can locate the objects from these complicated backgrounds, such as the cat in the indoor scene.

Fig. 6. The segmentation results of some difficult groups in Cat-Dog dataset using our proposed method.

Fig. 7. The segmentation results of some difficult groups in Bird dataset using our proposed method.

The IOU values of ours method *Ours+m*, *Ours+s* and the existing methods on these two datasets are shown in the fourth and fifth columns in Table 1. Again, the proposed multi-group cosegmentation achieves the best performance. Meanwhile, we can see the significant improvement of cosegmentation is mainly caused by our single group version. The reason is that we use several new strategies to improve the single image group cosegmentation performance, such as the dynamic re-neighboring across images, new neighbor selection method and simultaneously considering the segmentation on multiple image and single image. These strategies are able to result in the significant improvement of cosegmentation, especially when the dataset is challenging (such as Bird and Cat with the IOU values of 0.24 and 0.45, respectively). Meanwhile, it can also be seen that our multiple group version can further improve the IOU values over the single group version.

4.3 Noise Image Based Cosegmentation

In this experiment, we intend to demonstrate that our multi-group cosegmentation can help on removing noise or irrelevant images from a given internet

image collection of a general object that for example could be the search results of multiple search engines, such as Google and Bing. We can divide the image collection into multiple groups according to its sources, i.e. where an image is coming from. By assuming the noise images are different from different sources, we can easily remove the noise images in one group by checking whether the noise images appear in another group or not. Such a noise removing method is much simpler than the one proposed in [28].

For demonstration purpose, we construct a noise dataset from the one in [28] to illustrate our idea. Specifically, we add different objects into a common object image set so as to form two different groups. For example, we respectively add a number of face and bird images into the car image set to form two different car groups. Note that for each group, we allow the repetition of the noise images, which cannot be handled by [28].

Some example results of the proposed method are shown in Fig. 8, where the top and bottom rows correspond to the results of the two different groups. We can see that the proposed method can delete the noise images successfully, as evident by no segmentation mask in those noise images. Since it is not meaningful to calculate IOU with empty segmentation mask, here we use the precision value as the evaluation metric, which is defined as the ratio of the number of correctly labelled pixels to the total number of pixels. The precision results of the proposed method are given in the last column of Table 1, which shows the significant improvement by using the proposed multiple group cosegmentation.

Fig. 8. The segmentation results of the proposed method on the Noise image dataset. Note that the noise images are identified in the cosegmentation since they have no segmentation masks.

5 Conclusion

In this paper, we have proposed a multi-group image cosegmentation framework, which is formulated as an energy minimization problem. The proposed energy model consists of three terms: the single image segmentation term, the single group term and the multiple group term, which, together with the hidden variables, can effectively ensure the right consistency to be enforced within an image, within a group and across different groups. The proposed model is minimized by the EM algorithm incorporated with the adopted K-nearest neighbor search

and the graphcut algorithm. The experiments on three practical cosegmentation tasks and four benchmark image datasets have clearly demonstrated the usefulness and powerfulness of utilizing inter-group information.

Acknowledgement. This work was supported in part by the Major State Basic Research Development Program of China (973 Program 2015CB351804), NSFC (No. 61271289), the Singapore National Research Foundation under its IDM Futures Funding Initiative and administered by the Interactive & Digital Media Programme Office, Media Development Authority, the Ph.D. Programs Foundation of Ministry of Education of China (No. 20110185110002), and by The Program for Young Scholars Innovative Research Team of Sichuan Province, China (No. 2014TD0006).

References

1. Chai, Y., Rahu, E., Lempitsky, V., Gool, L.V., Zisserman, A.: Tricos: A tri-level class-discriminative co-segmentation method for image classification. In: European Conference on Computer Vision. (2012)
2. Rother, C., Kolmogorov, V., Minka, T., Blake, A.: Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In: IEEE Conference on Computer Vision and Pattern Recognition. (2006) 993–1000
3. Zhu, H., Lu, J., Cai, J., Zheng, J., Thalmann, N.M.: Multiple foreground recognition and cosegmentation: an object-oriented crf model with robust higher-order potentials. In: IEEE Winter Conference on Applications of Computer Vison. (2014)
4. Mukherjee, L., Singh, V., Dyer, C.R.: Half-integrality based algorithms for cosegmentation of images. In: IEEE Conference on Computer Vision and Pattern Recognition. (2009) 2028–2035
5. Hochbaum, D.S., Singh, V.: An efficient algorithm for co-segmentation. In: International Conference on Computer Vision. (2009) 269–276
6. Vicente, S., Kolmogorov, V., Rother, C.: Cosegmentation revisited: models and optimization. In: European Conference on Computer Vision. (2010) 465–479
7. Collins, M., Xu, J., Grady, L., Singh, V.: Random walks based multi-image cosegmentation: Quasiconvexity results and gpu-based solutions. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 1656–1663
8. Meng, F., Li, H., Liu, G., Ngan, K.N.: Image cosegmentation by incorporating color reward strategy and active contour model. IEEE Transactions on Cybernetics **43** (2013) 725–737
9. Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image cosegmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2010) 1943–1950
10. Chai, Y., Lempitsky, V., Zisserman, A.: Bicos: A bi-level co-segmentation method for image classification. In: International Conference on Computer Vision. (2011) 2579–2586
11. Kim, G., Xing, E.P., Fei-Fei, L., Kanade, T.: Distributed cosegmentation via submodular optimization on anisotropic diffusion. In: International Conference on Computer Vision. (2011) 169–176
12. Kim, G., Xing, E.P.: On multiple foreground cosegmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012)

13. Joulin, A., Bach, F., Ponce, J.: Multi-class cosegmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 542–549
14. Vicente, S., Rother, C., Kolmogorov, V.: Object cosegmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2011) 2217–2224
15. Mukherjee, L., Singh, V., Peng, J.: Scale invariant cosegmentation for image groups. In: IEEE Conference on Computer Vision and Pattern Recognition. (2011) 1881–1888
16. Rubio, J., Serrat, J., López, A., Paragios, N.: Unsupervised co-segmentation through region matching. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 749–756
17. Li, H., Ngan, K.N.: A co-saliency model of image pairs. IEEE Transactions on Image Processing **20** (2011) 3365 –3375
18. Ma, T., Latecki, L.J.: Graph transduction learning with connectivity constraints with application to multiple foreground cosegmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2013)
19. Wang, F., Huang, Q., Guibas, L.J.: Image co-segmentation via consistent functional maps. In: IEEE International Conference on Computer Vision (ICCV), IEEE (2013) 849–856
20. Xing, G.K.E.P.: Jointly aligning and segmenting multiple web photo streams for the inference of collective photo storylines. In: IEEE Conference on Computer Vision and Pattern Recognition. (2013)
21. Meng, F., Li, H., Ngan, K.N., Zeng, L., Wu, Q.: Feature adaptive co-segmentation by complexity awareness. IEEE Transactions on Image Processing **22** (2013) 4809–4824
22. Liu, C., Yuen, J., Torralba, A.: Sift flow: Dense correspondence across scenes and its applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011) 978–994
23. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (2012) 2274–2282
24. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001) 1222–1239
25. Batra, D., Kowdle, A., Parikh, D.: Icoseg: interactive co-segmentation with intelligent scribble guidance. In: IEEE Conference on Computer Vision and Pattern Recognition. (2010) 3169–3176
26. Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., Perona, P.: Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology (2010)
27. Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C.V.: Cats and dogs. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012)
28. Rubinstein, M., Joulin, A., Kopf, J., Liu, C.: Unsupervised joint object discovery and segmentation in internet images. In: IEEE Conference on Computer Vision and Pattern Recognition. (2013)