



A site entropy rate and degree centrality based algorithm for image co-segmentation [☆]



Souradeep Chakraborty ^{a,*}, Pabitra Mitra ^b

^a Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721302, India

^b Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur 721302, India

ARTICLE INFO

Article history:

Received 24 April 2015

Accepted 31 August 2015

Available online 7 September 2015

Keywords:

Site entropy rate

Degree centrality

Co-segmentation

Co-saliency

Markov chain

Object proposal

k-partite graph

Stationary distribution

ABSTRACT

In this paper, we propose a graph based algorithm that efficiently segments common objects from multiple images. We first generate a number of object proposals from each image. Then, an undirected graph is constructed based on proposal similarities and co-saliency maps. Two different methods are followed to extract the proposals containing common objects. They are: (1) degree centrality of nodes obtained after graph thresholding and (2) site entropy rate of nodes calculated on the stationary distribution of Markov chain constructed on the graph. Finally, we obtain the co-segmented image region by selecting the more salient of the object proposals obtained by the two methods, for each image. Multiple instances of the common object are also segmented efficiently. The proposed method has been compared with many existing co-segmentation methods on three standard co-segmentation datasets. Experimental results show its effectiveness in co-segmentation, with larger IoU values as compared to other co-segmentation methods.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Co-segmentation has been an active research topic in the area of image processing. Many practical object segmentation methods are based on generating object priors through human interaction [1,2]. Users are here asked to provide segmentation cues manually [3,4]. When the number of target images is high, users face a huge workload of providing manual segmentation cues. The principle of co-segmentation is to exploit the availability of multiple images that contain instances of the same “object” classes to supplement detailed supervisory information. This reduces the user workload significantly. As opposed to single image segmentation, co-segmentation aggregates information from multiple images (which contain objects with similar features) to improve the segmentation of individual images. Co-segmentation methods which can handle large numbers of images and object classes find potential applications in many fields such as automated image retrieval, object tracking, and object recognition.

Many of the existing methods for co-segmentation are modeled on the Markov random field (MRF)-based optimization procedures [7,9–13] and other graph theoretic methods [19,36]. The MRF

based methods mostly formulate energy functions based on foreground or background consistency constraints and optimize such functions to obtain the segmentations. The graph based methods model image regions as nodes of a graph (which represent object proposals) and then perform graph processing operations such as node clustering and shortest path finding, to extract the strongly connected nodes representing co-segmented image regions. However, these methods do not utilize the essential entropy information furnished from the graphs, such as the rate entropy of the stationary distribution on a Markov chain constructed on the graph, or the edge-weight threshold entropy information. Furthermore, if there are a large number of original images in an image group, the problem becomes more expensive to compute. Development of fast algorithms which avoid time complex optimization procedures as in MRF-based methods, as well as utilize entropy information in constructed graphs, still remains a challenge.

In this paper, we present a simple and effective co-segmentation model, which integrates the notion of degree centrality of nodes, and their site entropy rate information (obtained from the stationary distribution of the constructed Markov chain), to co-segment multiple similar images. The proposed model consists of four main steps. The first step is to segment the original images into a number of local semantic regions, which is achieved by applying object proposal generation algorithm as described in [5]. Out of the several object proposals suggested by the algorithm, we select only those

[☆] This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.

E-mail addresses: sourachakra@gmail.com (S. Chakraborty), pabitra@gmail.com (P. Mitra).

proposals whose mean saliency value is greater than the Otsu threshold calculated on the respective images [6]. In the second step, we construct a graph to represent the local region similarities according to the feature distance and the co-saliency maps generated by using the method in [7]. As the third step, we apply two graph based algorithms on the constructed graph to extract the object proposal (which contains the common object) from each image in the image group. They are: (i) Degree centrality based node selection and (ii) Site entropy rate based node selection. Finally, we select the object proposal with maximum mean saliency value among the two object proposals computed by the two graph based methods, as the co-segmented image region for an image. We evaluate our method on many groups of images. The experimental results demonstrate the effectiveness of our method.

The remainder of this paper is organized as follows. A brief review of the related works in the field of co-segmentation is given in Section 2. We explain the proposed method in Section 3, describing the two sub-methods: site entropy rate based and degree centrality based, in details. Experimental results are provided in Section 4 to support the efficiency of the proposed algorithm. Finally, in Section 5, we draw conclusions with future research issues.

2. Related work

The task of co-segmentation was first introduced by Rother et al. [8], where co-segmentation was modeled as an optimization problem in which a Markov random field (MRF)-based method was proposed to extract the objects from image pairs by adding the constraint of foreground similarity (measured by L1-norm) to traditional MRF-based procedures. Trust region graph cuts (TRGC) method was employed for energy function optimization. Many other methods were proposed [9–12] following this MRF-based optimization framework. Mukherjee et al. replaced the L1-norm by L2-norm in [9]. Pseudo-Boolean optimization method was used for the energy function optimization. Hochbaum and Singh [10] rewarded foreground similarity instead of penalizing foreground difference and that simplified the energy function optimization. In [13], Vicente et al. extended the foreground similarity measurement by employing dual decomposition for the energy function optimization. Chang et al. [12] used the graph-cut algorithm to optimize a global energy term which considered both foreground similarity and background consistency.

Several other co-segmentation models were proposed apart from MRF-based methods. Joulin et al. [14] combined discriminative clustering and spectral clustering methods to perform co-segmentation of multiple classes and for a significantly large number of images. To exploit priors about image more directly, an interactive co-segmentation method was proposed by Batra et al. in [15], which segments common objects through human interaction guided by an automatic recommendation system. Mukherjee et al. [16] put forward a scale-invariant method of co-segmentation with the requirement that the rank of the matrix corresponding to foreground regions should equal one. Vicente et al. [17] proposed a model which emphasizes interesting objects co-segmentation by selecting useful features from a total of 33 features through random forest regressor. In [18], Kimet al. followed a distributed co-segmentation approach via sub-modular optimization on anisotropic diffusion for a highly variable large-scale image collection. Meng et al. [19] designed a digraph to represent the local region similarities according to the feature distance and the saliency map, and formulated the co-segmentation problem as a shortest path problem. In this paper, we adopt a similar graph based approach. Unlike the layered digraph as followed in [19], we construct a k -partite graph and then implement two methods which make use of entropy information: degree centrality based and site

entropy rate based, and then select the object proposal which most accurately segments the common object in the group.

Rubio et al. [20] proposed a multiple-scale multiple-image generative model, which jointly estimated the foreground and background appearance distributions from many images. Meng et al. [21] proposed a model which integrates active contours method and rewarding strategy. They generate a new energy function with two conflicting goals: foreground similarity among the images and background consistency in each image, and then use a mutual evolution approach to minimize the energy function value. In a more recent method of Tao et al. [22], object co-segmentation method based on shape conformability is put forward. It focuses on the shape consistency of the foreground objects in image set. The common shape pattern is extracted if the foreground objects are varied in appearance but share similar shape structures.

There have been many works which have utilized saliency information in the process of segmentation as well as co-segmentation. [23] used saliency to automate the selection of foreground object and background seeds, needed for image segmentation. In a similar work [24], a co-saliency prior has been used as a hint about possible foreground locations for image co-segmentation task. Besides image segmentation, saliency/co-saliency has also been utilized in similar and related applications including image classification [25], ranking [26] and de-blurring [27].

3. Proposed co-segmentation method

The flowchart of the main steps of the proposed method is shown in Fig. 1. It consists of four major steps. In subsequent sections, we describe the object proposal generation method, the approach of graph construction, and the graph based co-segmentation methods used to extract common object from each image.

3.1. Object proposal generation and salient proposal selection

We explain the steps followed to generate object proposals and co-saliency maps for each image, followed by subsequent salient object proposal selection.

Step 1: The object proposal generation procedure in [5] is implemented to segment the original image into a number of local regions by object proposal generation. In this approach, both local and global search procedures are combined in the space of sets of superpixels, to obtain accurate segmentations for all objects of an image. Assume that $I = \{I_1, I_2, \dots, I_m\}$ denotes the original image set, of size m . We first segment each image I_i into a set of overlapping objects proposals $P^i = \{P^i_1, \dots, P^i_{N_i}\}$, where N_i is the number of the object proposals in image I_i . The set of all initial objects proposals is denoted as $P_{initial} = \{P^1_{initial}, \dots, P^m_{initial}\}$.

Step 2: To extract the salient object proposals from the set of proposals obtained (as described in Step 1), we first generate co-saliency maps of all the images in an image set by the self-adaptively weighted co-saliency detection technique recently proposed method by Cao et al. [7]. This method exploits the relationship of multiple saliency cues and obtains the self-adaptive weight to generate the co-saliency maps. In our experiment, we obtain saliency maps from the methods in [28–30], and get co-saliency maps using this self-adaptively weighted co-saliency detection method.

Step 3: After the co-saliency maps are obtained, we select only those object proposals whose mean co-saliency value is greater than a threshold. Following [19], for an object proposal P^i_j from an image I_i , we measure its mean co-saliency value s^i_j as:

$$s^i_j = \left(\frac{\sum_{(k,l) \in P^i_j} S_i(k,l)}{m^i_j} \right) \cdot \left(\frac{m^i_j}{M^i_i} \right), \quad (1)$$

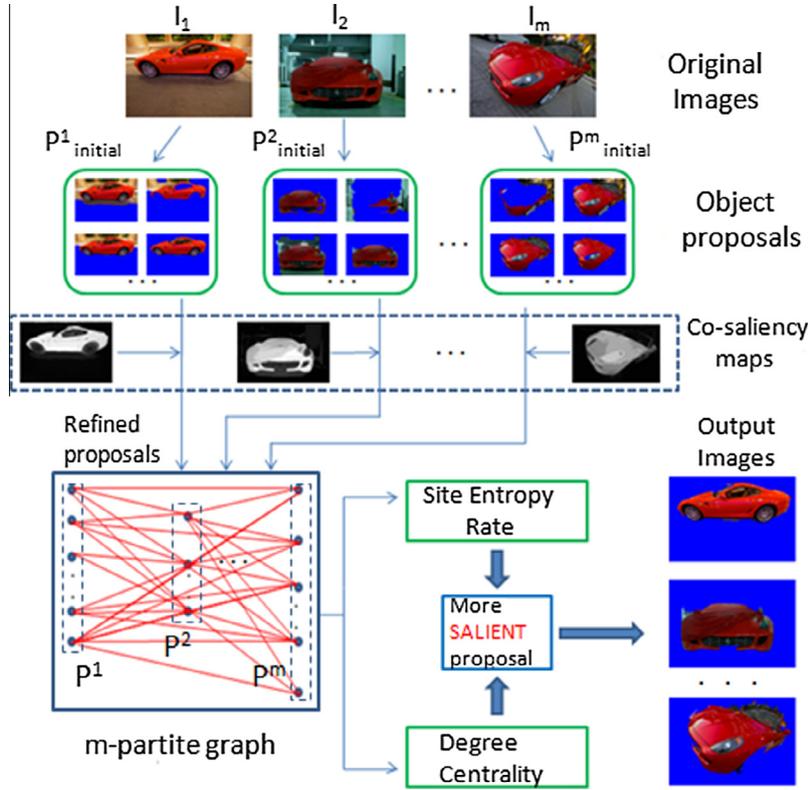


Fig. 1. Flowchart of proposed method for image co-segmentation.

where S_i = co-saliency map for image I_i and m_j^i is the number of the pixels in the proposal P_j^i , m_{ij}^i is the number of the salient pixels in P_j^i , and M_i^i is the number of salient pixels in the image I_i .

We consider salient pixels having saliency value higher than a threshold of 0.5. The first term in Eq. (1) selects local regions with large mean saliency value. Because a small local region may have large mean saliency value, which results in incomplete segmentation, we use the second term to avoid the large value in a small local region. The mean co-saliency threshold value is determined by performing Otsu thresholding [6] on the co-saliency maps obtained by [7]. Let T_{Otsu}^i be the Otsu threshold for an image I_i . We consider only the proposals whose mean co-saliency value (as determined from Eq. (1)), $s_j^i > T_{Otsu}^i$ for graph construction (in Section 3.2). Thus we get a refined set of proposals $P = \{P_1, \dots, P_m\}$.

3.2. Graph construction

Based on the assumption that the common object is segmented as a local region by the multiple local region generation method, the co-segmentation can be achieved by selecting the common objects from the local regions according to their similarities. We represent the similarities between generated proposals, set P by a labeled k -partite graph G . Assuming L_V and L_E denote the set of nodes and edge labels respectively, the graph G is a 3-tuple $G = (V, E, \nu)$, where V is a finite set of nodes, $E \subseteq V \times V$ is the set of edges and ν is the image label of a node. Each node $v_{ij} \in V$ represents an object proposal P_j^i . An undirected edge corresponding to nodes v_1 and v_2 is depicted as $e = E(v_1, v_2) = E(v_2, v_1)$. Furthermore, a weight ω is assigned for each edge. To increase the diversities of a class we do not consider the self-similarities between regions within the same image. We avoid

the self-similarities by introducing two connection constraints: (1) Edge $e = E(v_{ij}, v_{ik}) = 0 \forall i$ and (2) For any pair of edges $e = E(v_{ij}, v_{kl})$ and $e' = E(v_{ij}, v_{k'l'})$ deriving from v_{ij} , we have $k \neq k'$. These constraints generate a k -partite graph.

3.3. Assignment of graph edge weights

We calculate the weight $\omega_{ij,kl}$ of each edge $e = E(v_{ij}, v_{kl})$, by combining factors representing region similarity and mean saliency of nodes corresponding to object proposals.

(1) *Region similarity*: The region term $\omega_{ij,kl}^r$ represents the feature similarity between two local regions and is given by:

$$\omega_{ij,kl}^r = \frac{1}{d(f_{ij}, f_{kl})}, \quad (2)$$

where f_{ij} and f_{kl} are the features vectors of the local regions P_j^i and P_k^l respectively. In this paper, normalized color histogram of a proposal P_j^i is considered as its feature vector f_{ij} . $d(f_{ij}, f_{kl})$ denotes the distance between two feature vectors f_{ij} and f_{kl} . Feature distance d is calculated by χ^2 -distance measure as follows:

$$d(f_{ij}, f_{kl}) = \sum_b \frac{(f_{ij}^b - f_{kl}^b)^2}{(f_{ij}^b + f_{kl}^b)}, \quad (3)$$

where b = number of bins in the normalized color histogram.

(2) *Saliency*: Since the original images have similar background, we introduce the saliency term to distinguish the common objects from the similar backgrounds. As described in Section 3.1, we obtain co-saliency maps S_i for each image I_i in their respective image groups. The saliency term $\omega_{ij,kl}^s$ represents gives more

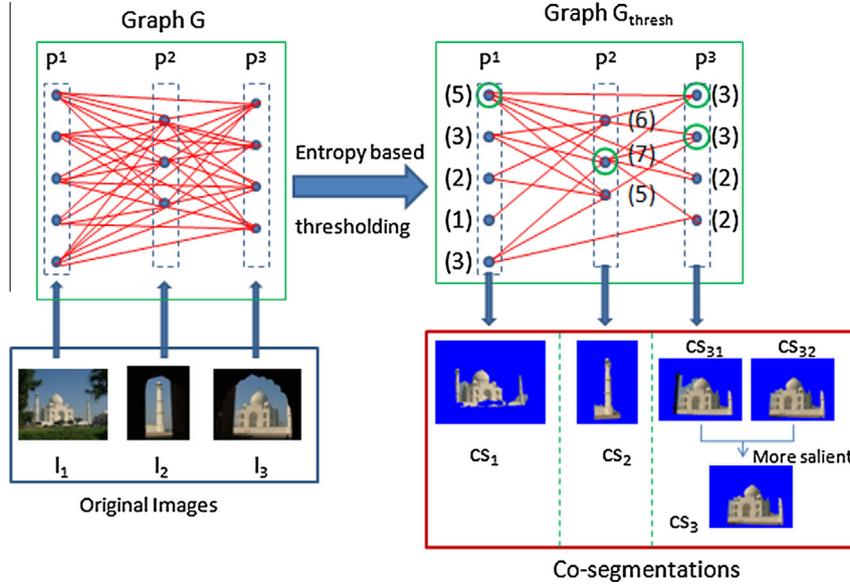
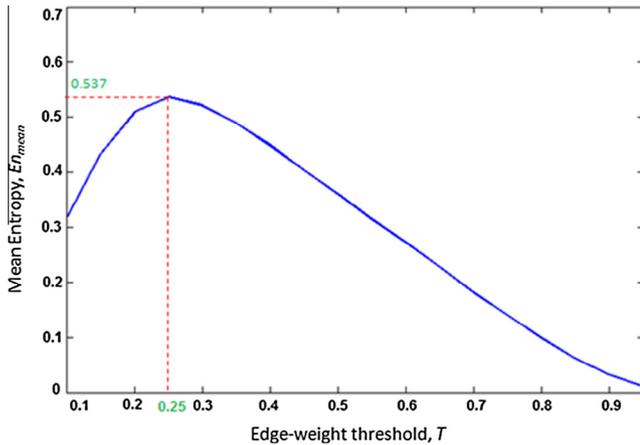


Fig. 2. Degree centrality based co-segmentation.

Fig. 3. Variation of mean edge-weight entropy, En_{mean} with edge-weight threshold, T on iCoseg database [15].

weightage to proposals whose mean co-saliency values (as measured in Section 3.1) are higher than the rest. It is measured as:

$$\omega_{ij,kl}^s = s_{ij} + s_{kl}, \quad (4)$$

where s_{ij} and s_{kl} represent the mean saliency values of the object proposals P_j^i and P_l^k .

Finally, the weight $\omega_{ij,kl}$ (in Eq. (5)) is calculated as the product of the weight corresponding to region similarity (in Eq. (2)) and the weight corresponding to saliency value (in Eq. (4)).

$$\omega_{ij,kl} = \omega_{ij,kl}^r \cdot \omega_{ij,kl}^s \quad (5)$$

3.4. Segmenting common objects from graph

We implement two different methods on the constructed graph to find out the co-segmented regions for each image in the image set with common object/objects. The first method extracts out graph nodes based on their in-degrees on the graph thresholded with entropy based thresholding method as in [31]. The second method finds out nodes whose calculated site entropy rates on the stationary distribution on the constructed Markov chain on the graph are more. We describe below the two methods in details.

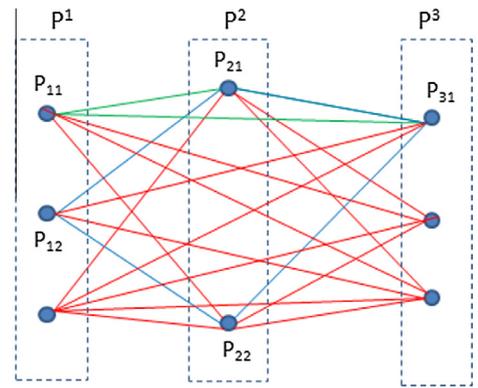


Fig. 4. Aperiodicity in a 3-partite graph A.

3.4.1. Degree centrality based co-segmented object proposal detection

Initially we have a k -partite graph G (k = number of images in an image group) which we threshold to remove connections whose strength fall below a certain threshold. To find an appropriate threshold value, we follow the entropy based thresholding method.

First we select an edge weight threshold T , which is varied between the minimum and the maximum edge weight in the graph G . Next taking this threshold T , we form two sets of edges, one set S_D representing discarded set of edges and the other set, S_S the selected set of edges. Let w_i be the weight of an edge E_i . For a particular threshold T , the ratio of summation of weights for S_D to the weights for the set $S_D \cup S_S$ is calculated as in Eq. (6).

$$r = \frac{\sum_{w_i \leq T} w_i}{\sum_i w_i} \quad (6)$$

The edge-weight entropy En , of the discarded and selected set of edges S_D and S_S respectively is defined as:

$$En = -r \log(r) - (1 - r) \log(1 - r) \quad (7)$$

The edge-weight entropy En is a function of T . The threshold for which edge-weight entropy is maximum is chosen as the edge-weight threshold T . After thresholding, we get a modified thresholded sparse graph G_{thresh} . We intend to compute the in-degrees of the nodes of the graph and select nodes with highest degrees.

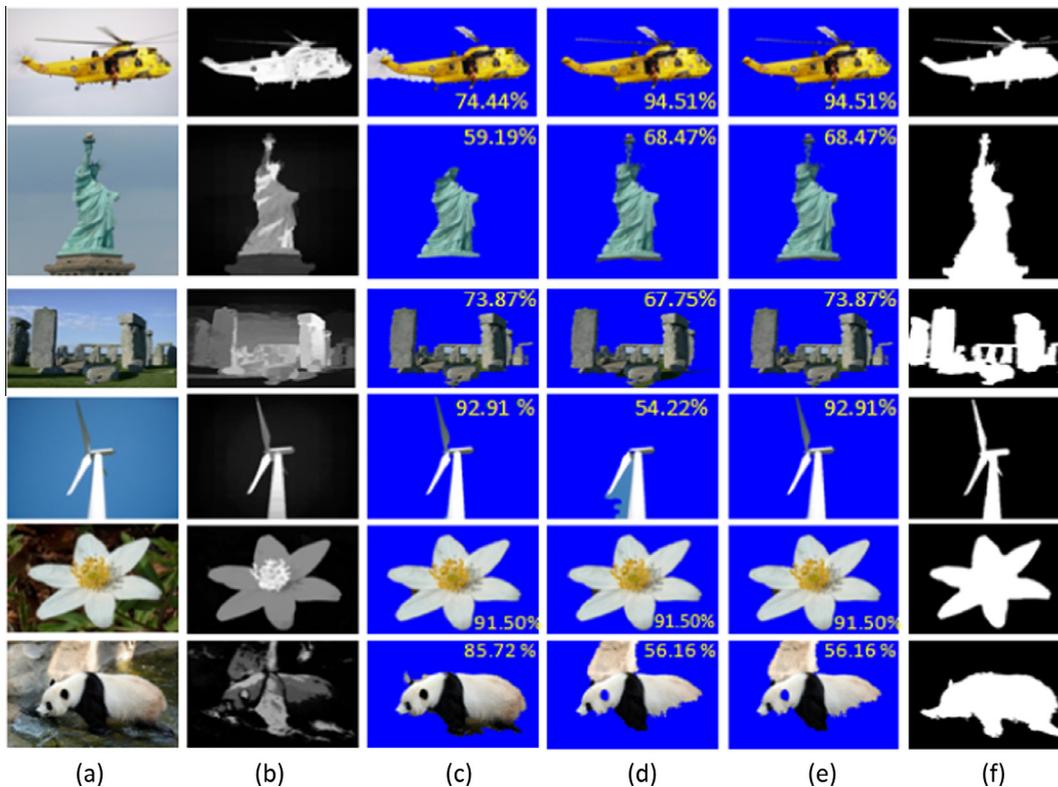


Fig. 5. Comparison of image co-segmentation by DC and SER methods. (a) Original images, (b) co-saliency maps, (c) co-segmentations based on degree centrality, (d) co-segmentations based on site entropy rate, (e) final co-segmented image regions, and (f) ground truth.

However, a fully connected graph has all nodes with the same degree. In such a case, any random node could be a suitable candidate for selection. To circumvent this, we threshold the graph to eliminate weak edges based on the entropy equation (7) and allow edges above a certain threshold value (a threshold that maximizes the entropy value) to participate in the node selection procedure to achieve co-segmentation.

We show the co-segmentation procedure by employing entropy based thresholding on the k -partite graph G in Fig. 2. We get the co-segmented regions for each of the 3 images considered for the *Taj Mahal* object class. The graph G_{thresh} only keeps edges whose edge-weights are greater than the determined threshold T . The numbers aside the nodes (in brackets) denote their respective in-degrees after thresholding. Nodes with the highest in-degrees are marked by green circles. It may be noted that, there are two nodes for image I_3 which have the same highest in-degree value ($= 3$). Therefore, we choose the node corresponding to that object proposal, whose mean saliency value (as calculated in Eq. (1)) is greater. In Fig. 2, the co-segmented proposal CS_{32} has greater mean saliency than proposal CS_{31} and is thus selected as the final co-segmented region CS_3 .

Fig. 3 shows the variation of mean edge-weight entropy, En_{mean} (as in Eq. (7)) with threshold T , on iCoseg dataset used in [15]. In our experiment, we sampled the mean entropy value, En_{mean} at a threshold interval of 0.05, starting from $T_i = 0.10$ and ending with $T_f = 0.95$. A threshold value, $T_{\text{maxEn}} = \text{argmax}_T(En_{\text{mean}}) = 0.25$ was found to yield the highest mean entropy, $En_{\text{mean}} = 0.537$ on the iCoseg dataset [15]. Note here that, $T_{\text{maxEn}} = 0.25$ shown in Fig. 3, indicates the threshold value which yields the highest mean entropy on all image groups in the iCoseg dataset, whereas the threshold T used for an individual image group depends on the maximum entropy value En obtained for that particular group.

Now in the sparse graph G_{thresh} , we compute the in-degrees of nodes. We sort the set N_n of nodes with respect to the calculated in-degrees. It is quite intuitive for nodes with higher in-degrees to have better feature similarities as well as to represent proposals with greater saliency as compared to others with lower in-degrees. Thus we find for every image, the node with highest in-degree, and the object proposals corresponding to those nodes become the co-segmented regions for the respective images. Let this set of object proposals be denoted by set P_{DC} .

3.4.2. Site entropy rate based co-segmented object proposal detection

Next, we consider the original constructed graph G for site entropy rate computation. The stationary/equilibrium distribution of a Markov chain constructed on a graph imparts insight about the proportion of time a random walker would stay on each node in such a distribution. In other words, it is indicative of node prominence in terms of linkage strength with other nodes. Our goal is to find the set of nodes which are most similar to each other, with each node selected from one image in the group. Such a set of nodes (for an image group) can be found when we determine the node with greatest prominence among others, for every image. Thus the use of the stationary distribution of Markov chain on graph G , is quite intuitive in the context of co-segmentation problem.

The stationary distribution of the Markov chain on the constructed k -partite graph G (as formulated in Section 3.2) exists and is unique. For stationary distribution of Markov chain to exist and to be unique, a graph must be ergodic i.e. irreducible and aperiodic. The constructed k -partite graph G is irreducible as there always exists a positive probability path between any of the graph nodes. From Fig. 4, we observe that in a 3-partite graph A , we have two different paths for a random walker who starts from the node

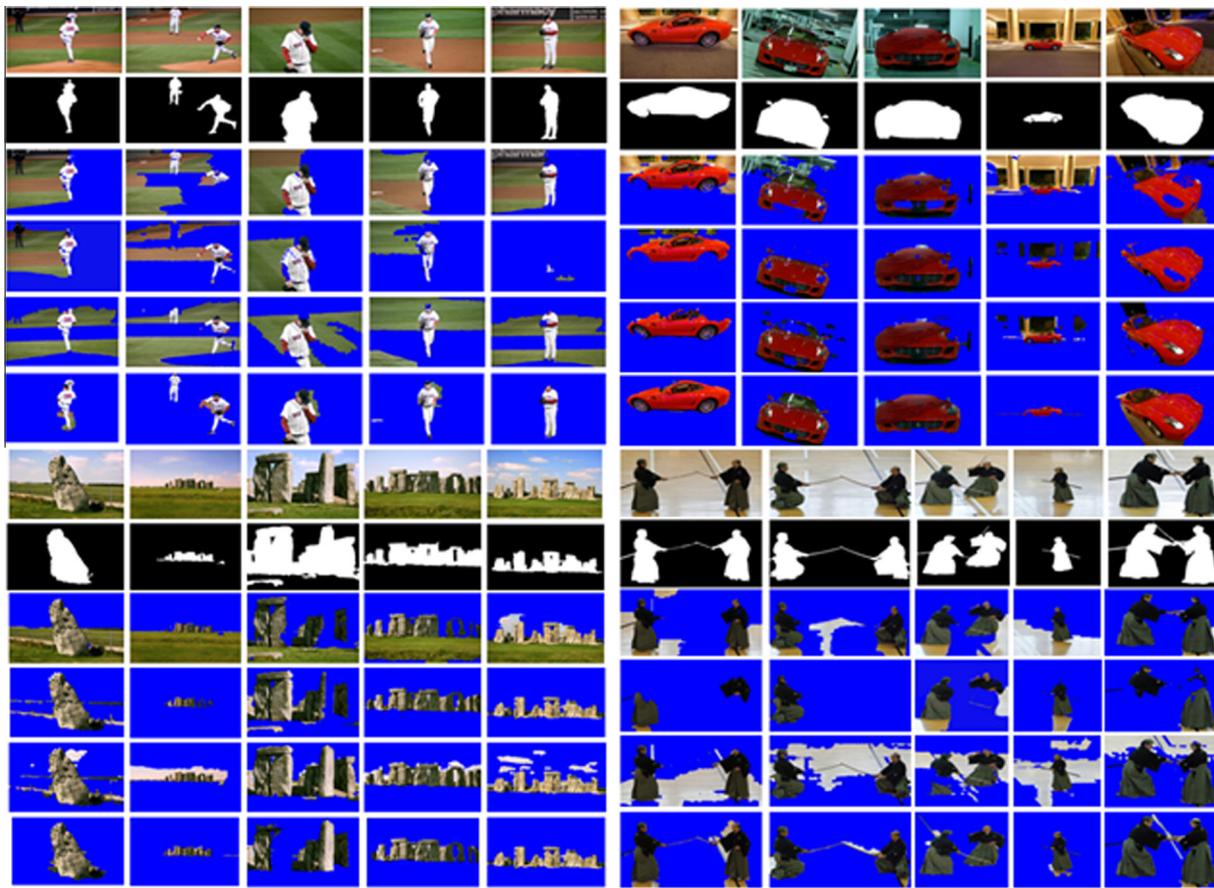


Fig. 6. Co-segmentation results of [14,18,33], and the proposed method on classes of iCoseg dataset. These classes are Redsox players, Ferrari, Stonehenge and Kendo. The rows 1, 7: original images. The rows 2, 8: Ground truth. The rows 3, 9: the results of DM method [14]. The rows 4, 10: the results of AD method [18]. The rows 5, 11: the results of SB method [33]. The rows 6, 12: results of the proposed method.

P_{31} , i.e. path $P_{31} - P_{21} - P_{11} - P_{31}$ (walk length = 3) and $P_{31} - P_{22} - P_{12} - P_{21} - P_{31}$ (walk length = 4). This shows the aperiodic nature of the graph A . Thus, for $k > 2$, we have an ergodic Markov chain on graph G which ensures a unique stationary distribution. For a bipartite graph ($k = 2$) however, the Markov chain stationary distribution will exist, but it may not be unique.

Here, we utilize the site entropy rate (SER) of the equilibrium distribution. SER was first defined by Wang et al. [32] where they adopted it as a visual saliency measure. Inspired by the method, we compute the entropy rates of nodes in the graph and select the nodes with highest SERs for each image label. The selected nodes (from each image) represents the object proposal which most accurately contains the co-segmented image region. Transition probability of the random walk from node i to node j is defined in terms of the normalized edge weights between site i and j as in Eq. (8).

$$TP_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} \quad (8)$$

Next, we compute the stationary distribution π on the Markov chain formed with the transition matrix TP . For a random walk process, the element of π at node i can be simply computed as $\pi_i = \frac{W_i}{2W}$, where $W_i = \sum_j \omega_{ij}$ is the total weight of edges emanating from node i , and $W = \sum_{i,j} \omega_{ij}$ is the sum of the weights of all the edges.

The total information sent from one node to another is decided by two terms: the transmission frequency and the amount of information at each transmission. Site entropy rate measures the information transmission which accounts for the two factors during the

random walk process. Site entropy rate (SER) of a node i is defined as:

$$SER_i = \pi_i \cdot \left(\sum_j -TP_{ij} \cdot \log TP_{ij} \right), \quad (9)$$

The SER can be divided into two parts: the stationary distribution term π_i and the entropy term $\sum_j -TP_{ij} \cdot \log TP_{ij}$. The π_i tells the frequency at which a random walker visits node i . It is also the frequency that node i communicates with the other nodes. The entropy term $\sum_j -TP_{ij} \cdot \log TP_{ij}$ measures the uncertainty of node i jumping to the other nodes at one step. It is related to the amount of information transmitted from node i to the others at one step. Thus we find for every image, the node with highest SER, and the object proposals corresponding to those nodes become the co-segmented regions for the respective images. Let this set of object proposals be denoted by set P_{SER} .

3.4.3. Final co-segmentation

After obtaining the most salient object proposals P_{DC} and P_{SER} from the two methods respectively as described above, we choose the object proposal with the highest mean co-saliency value as the final co-segmented object proposal P_{final}^i for image $I_i \forall i$ as shown in Eq. (10). Fig. 5 shows some instances of co-segmented objects using both degree centrality and site entropy rate based methods and the selection of the more salient object proposal based on co-saliency maps of the respective images. The IoU metrics of the images have also been indicated (in percentages) for better comparison.

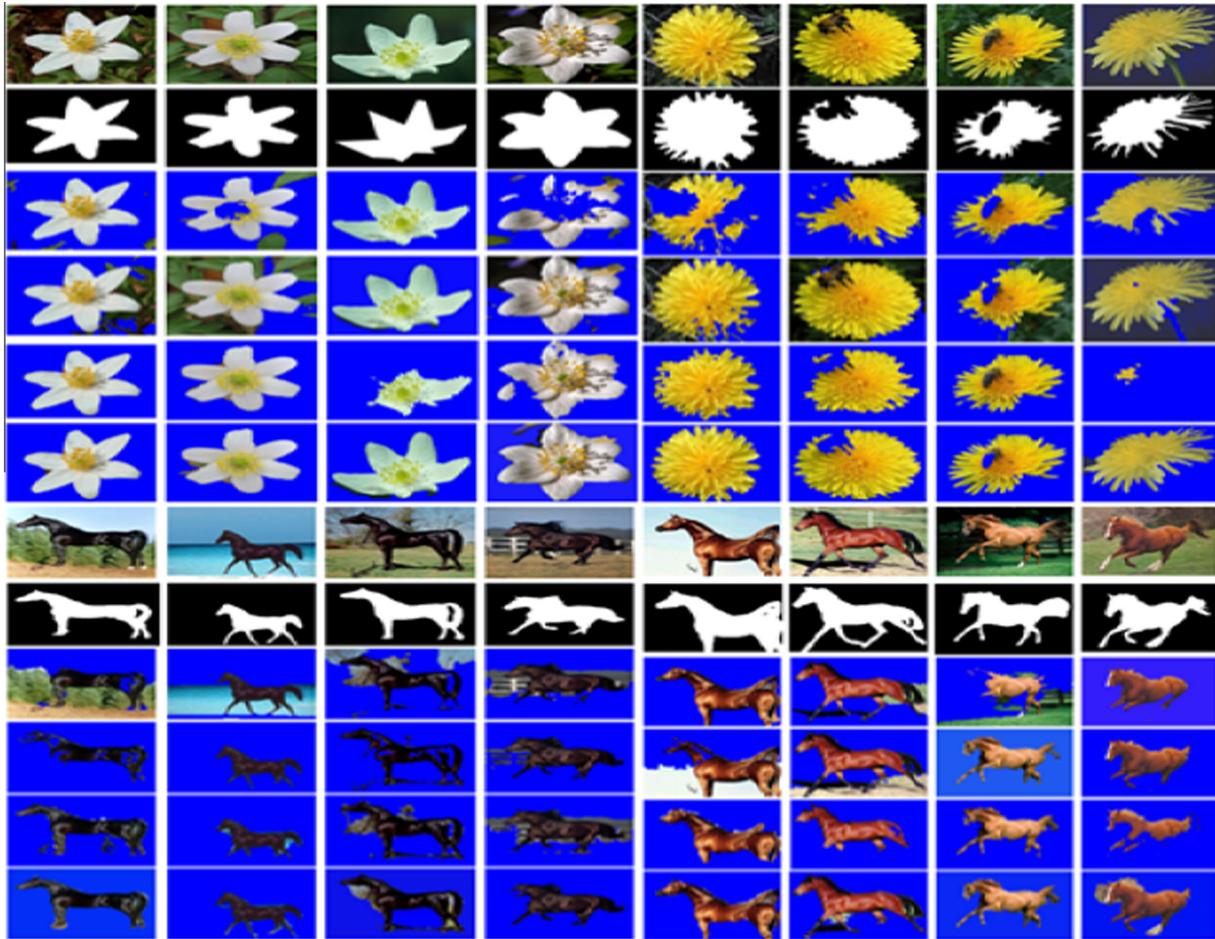


Fig. 7. Co-segmentation results of [14,18,33], and the proposed method on classes of Oxford 17 flowers dataset and Weizmann horses dataset. These classes are *Windflower*, *Dandelion*, *black horses* and *brown horses*. The rows 1, 7: original images. The rows 2, 8: Ground truth. The rows 3, 9: the results of DM method [14]. The rows 4, 10: the results of AD method [18]. The rows 5, 11: the results of SB method [33]. The rows 6, 12: results of the proposed method.

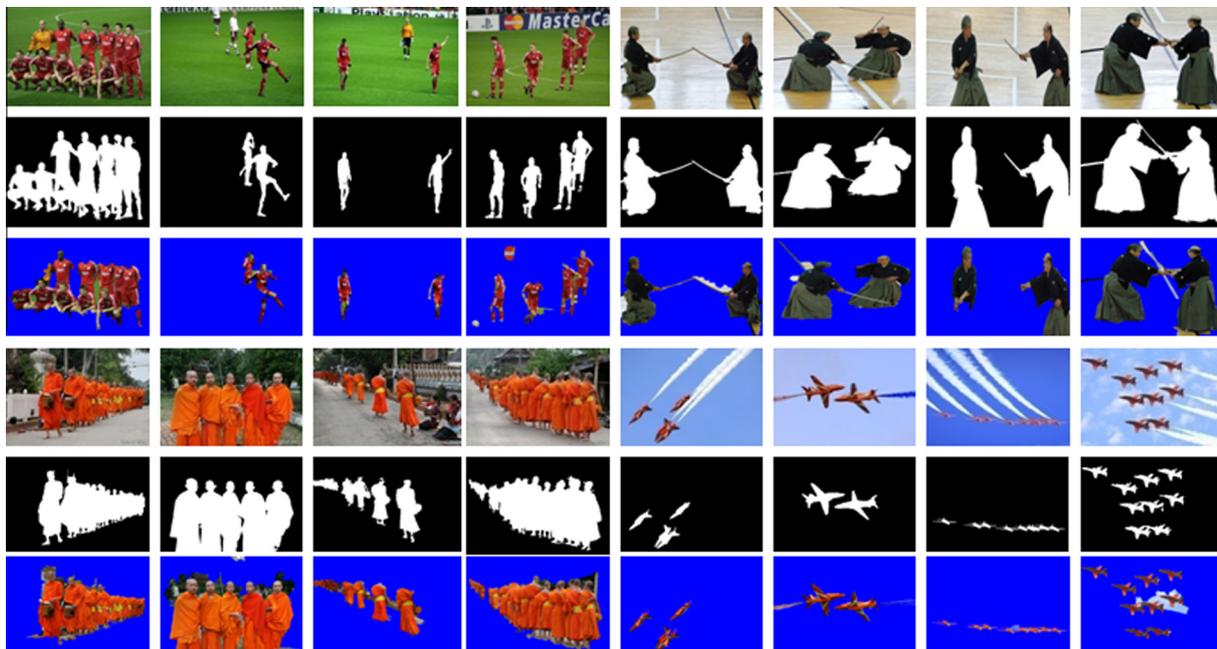


Fig. 8. Co-segmentation of objects with multiple instances in an image by the proposed method.



Fig. 9. Co-segmentation of images with two classes by the proposed method.

Table 1

IoU values of the compared methods (highest values marked in bold) on the object classes of iCoseg dataset.

Class	Method					Class	Method				
	DM [14]	OC [17]	AD [18]	SB [33]	Our		DM [14]	OC [17]	AD [18]	SB [33]	Our
Bear	0.4135	0.5308	0.3078	0.6544	0.7213	Panda	0.3937	0.3606	0.3510	0.6579	0.4788
FC Players	0.2097	0.3817	0.4005	0.6206	0.7819	Kite	0.4894	0.5514	0.3027	0.7267	0.7345
Red Sox	0.5532	0.6896	0.6341	0.5803	0.6944	Gymnastics	0.4212	0.4301	0.3470	0.7435	0.8174
Stonehenge	0.4591	0.4897	0.2964	0.5221	0.5664	Skating	0.4720	0.4460	0.5120	0.7808	0.6825
Liverpool	0.3972	0.4549	0.4535	0.5650	0.6453	Soccer	0.3962	0.4879	0.5023	0.6537	0.4612
Ferrari	0.4814	0.5091	0.6023	0.6518	0.7560	Monk	0.3273	0.7833	0.6836	0.7420	0.8211
Taj Mahal	0.3234	0.3998	0.2176	0.4807	0.5976	Balloon	0.4846	0.5124	0.3643	0.7990	0.8309
Pyramids	0.3925	0.3226	0.4272	0.6058	0.4190	Liberty	0.8792	0.9112	0.2854	0.8572	0.6715
Elephants	0.2542	0.2162	0.4511	0.5679	0.6342	Christ	0.3195	0.5448	0.4651	0.7548	0.7594
Goose	0.4934	0.3444	0.2105	0.6916	0.4507	Speed	0.3314	0.0745	0.2423	0.3598	0.5883
Helicopter	0.6281	0.6886	0.4269	0.8290	0.8769	Track	0.1413	0.4655	0.2754	0.5340	0.4082
Planes	0.2992	0.2730	0.0643	0.4879	0.6465	Windmill	0.2691	0.1219	0.1762	0.3324	0.3894
Cheetah	0.2751	0.4629	0.3779	0.6803	0.7011	Kendo	0.6013	0.5202	0.9034	0.5937	0.7932
						Average	0.4117	0.4605	0.3954	0.6343	0.6511

$$P_{final}^i = \operatorname{argmax}_p(s_{PDC}^i, s_{P_{SER}}^i) \quad (10)$$

In Fig. 5, we observe that for the *Statue of Liberty* and the *Helicopter* classes, site entropy rate (SER) computation yields more accurate segmentation as compared to the degree centrality (DC) based computation. On the other hand, DC method yields more stone area than the SER method for the *Stonehenge* and *Windmill* classes. However both methods co-segment the same region in the *white flower*. In the *Panda* class (last row) however, we observe that Eq. (10) selects the proposal suggested by the SER method, although it has a lesser IoU value than that computed by the DC method. Such erroneous selections may occur when surrounding background regions in an image have greater mean co-saliency values, when compared to a major portion in the

Table 2

IoU values of the compared methods (highest values marked in bold) on the horse classes of Weizmann dataset.

Class	Method			
	DM [14]	AD [18]	SB [33]	Our
White horses	0.6112	0.1113	0.7346	0.8924
Black horses	0.6523	0.8215	0.8426	0.9125
Brown horses	0.5212	0.5834	0.6803	0.8731
Average	0.5949	0.5054	0.7525	0.8927

foreground object region. As observed in the *Panda* class, the entire white fur portion (which constitutes the foreground region) has a lesser mean co-saliency value (as indicated in the co-saliency map) than the surrounding bright patch of sunlight in the

Table 3

IoU values of the compared methods (highest values marked in bold) on the classes of Oxford 17 Flowers dataset.

Class	Method			
	DM [14]	AD [18]	SB [33]	Our
Windflower	0.6043	0.4602	0.7327	0.7218
Sunflower	0.5523	0.4306	0.7317	0.6134
Dandelion	0.5643	0.4126	0.6057	0.7266
Daisy	0.6428	0.4913	0.7863	0.8233
Daffodil	0.6693	0.5812	0.7692	0.7817
Snowdrop	0.1958	0.2351	0.6078	0.6883
Lilyvalley	0.4611	0.4817	0.5126	0.5586
Bluebell	0.3087	0.3591	0.5700	0.5202
Pansy	0.3450	0.2902	0.4663	0.5412
Buttercup	0.4883	0.4213	0.6652	0.6919
Colt's Foot	0.6374	0.5321	0.7576	0.7412
Fritillary	0.6210	0.5810	0.7887	0.7317
Iris	0.6735	0.5238	0.5705	0.7234
Average	0.5203	0.4462	0.6588	0.6818

background. The object proposal obtained by the SER method thus gets selected by Eq. (10), which gives higher selection priority to the proposal with greater mean co-saliency value.

4. Experimental results and evaluation

In this section, we test the proposed co-segmentation algorithm on many groups of images. The qualitative and quantitative assessments of the segmentation results are reported.

4.1. Datasets used

We verify the proposed co-segmentation method on three datasets:

- iCoseg dataset, used in [15].
- Oxford 17 flower dataset, used in [34].
- Weizmann horses dataset, used in [35].

Here we consider all object groups in the iCoseg dataset for quantitative and qualitative evaluation. For the Oxford 17 flower dataset, we test our method on 13 flower groups (30 images considered per group), as the ground truth data for the rest four classes was not available. We group horses in Weizmann horses datasets into three categories based on horse color: white horses, brown horses and black horses (30 images per group) and then use these groups in our evaluation. Pixel ground truth hand annotations provided for all three datasets were used for quantitative evaluation.

4.2. Evaluation

In this subsection, we evaluate our method both from qualitative and quantitative aspects. We compare our method with three well known methods. They are:

- Discriminative clustering method (DM) [14].
- Anisotropic diffusion method (AD) [18].
- Similar background method (SB) [33].

Table 4

Average execution time of each method on iCoseg dataset.

Method	DM [14]	AD [18]	SB [33]	Our
Time (in s)	45.78	13.23	83.07	24.52

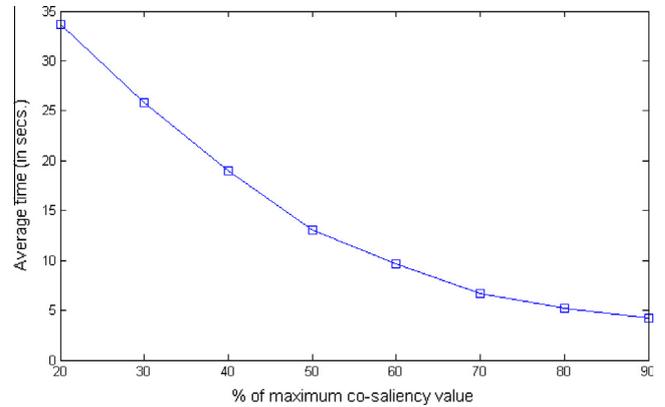


Fig. 10. Average execution time, t_d vs number of object proposals used (in terms of % of maximum co-saliency value).

4.2.1. Qualitative evaluation

Figs. 6 and 7 show the segmentation results of [14,18,33] and the proposed method respectively on iCoseg dataset and Oxford 17 flowers dataset, Weizmann horses datasets respectively. We consider the *Redsox players*, *Ferrari*, *Stonehenge* and *Kendo* image groups in iCoseg dataset for comparison in Fig. 6. The image groups *Windflower* and *Dandelion* from Oxford 17 flowers dataset and *black horses* and *brown horses* from Weizmann horses dataset are used for comparison in Fig. 7. The proposed method gives better segmentations as compared to the compared methods for almost all image groups considered. It might be observed that the compared methods segment undesired image regions along with the common object in most images, which lead to inaccurate segmentations. For some images, such as the 5th image (from left) in the *Ferrari* group and 3rd image (from left) in the *Kendo* group in Fig. 6, the compared methods do not segment out the complete object region. A similar problem is observed in the 4th flower (from left) in the *Sunflower* group in Fig. 7. For the DM method [14], unsuccessful segmentations are obtained as similar local regions of the backgrounds affect the training of the classifier used in the algorithm. Also for the AD method [18], inaccuracies in segmentations are observed because location of seeded points of the common objects and backgrounds becomes difficult when the backgrounds are similar to each other. In the proposed method, we filter out only those object proposals which score well in terms of co-saliency values, for further processing. This step ensures that majority of object proposals containing similar background regions get eliminated before they get considered for the DC and SER methods. Thus our method reduces such inaccuracies to a great extent.

Our method also co-segments images which have more than one instance of the common object. The object proposal generation algorithm [5] we use in our experiment, generates proposals which include those, which have the multiple instances of the object. The accurate segmentation of the multiple instances can be attributed to the edge-weight entropy maximization of the DC method and information transfer entropy maximization of the SER method used in our algorithm. This ensures that only the proposal carrying all the object instances in an image (of an image group) is segmented.

In Fig. 8, we show the segmentation results of the proposed method on images having multiple instances of the common object. Clearly, all possible instances are segmented with good accuracy. In the *Liverpool* class however, we find the red logo portion in the background, segmented along with the red colored dress of the players. This can be due to the dependence of the proposed algorithm on color histograms rather than contour based computation. For most other instances, highly accurate segmentations are obtained.

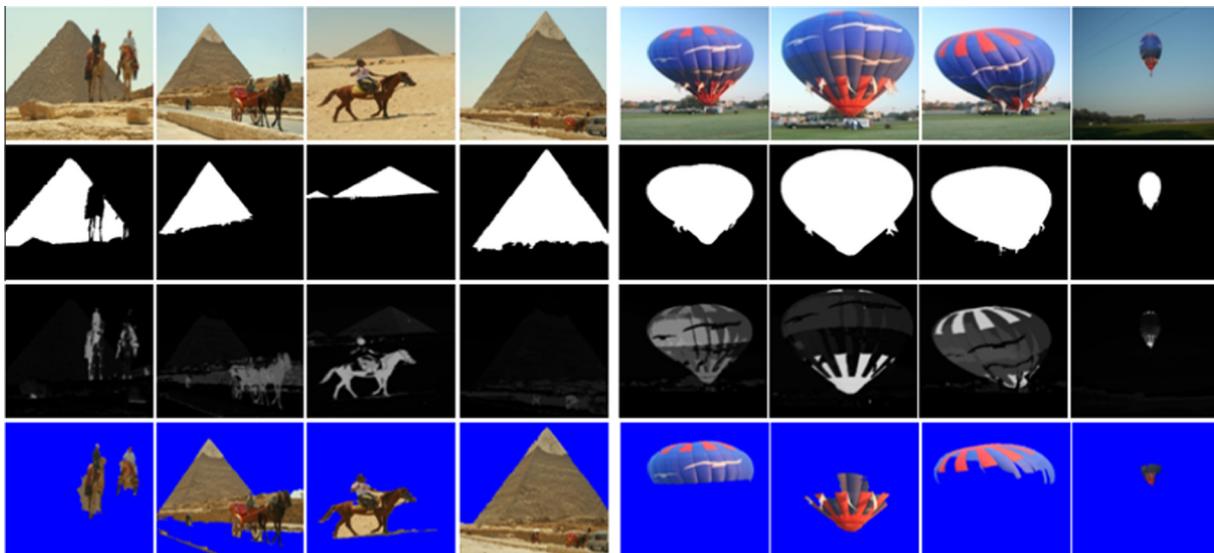


Fig. 11. Failure cases of Pyramid and Hot balloon object classes from iCoseg dataset. 1st row: Original image, 2nd row: Ground truth, 3rd row: Co-saliency map, 4th row: Segmented region.

To investigate the performance of our model on images with multiple classes, we selected images from the iCoseg dataset which contain two classes. Here we consider two distinctly colored objects as belonging to two different classes, as our model uses color histograms to compare object proposals. Fig. 9 depicts the segmentations generated by the proposed algorithm on the images of *Women soccer*, *Airshow planes* and *Liverpool* classes of the iCoseg dataset. Joint co-segmentations of two foreground class instances (e.g. the red and white jersey players in the *Women soccer* class, planes and their smoke trails in the *Airshow planes* class) can be observed. Note here that the segmentation result of the 1st image of the *Airshow planes* group in Fig. 9 is different from the same image in Fig. 8. The white smoke regions are segmented in this case as only selected similar images (containing red planes with white smoke) from the image group are considered for co-segmentation, unlike Fig. 8. The efficient co-segmentation can be attributed to the fact that, the co-saliency maps generated for the images assign high values to image regions corresponding to all the common classes present in the image group. The DC and the SER methods, thus get implemented on the object proposals carrying all class instances. Thus our model successfully co-segments images with more than one class object.

4.2.2. Quantitative evaluation

Here we compare our method with the four methods [14,17,18,33] from quantitative aspect. We objectively evaluate the methods by the segmentation accuracy which is measured by the intersection-over-union metric (IoU). It is defined as, $IoU_i = \frac{1}{|I_i|} \sum_{i \in I_i} \frac{GT_i \cap R_i}{GT_i \cup R_i}$ where GT_i is the ground truth and R_i is the segmentation of image I_i . A large accuracy corresponds to accurate co-segmentation. The mean accuracy over the images in a group is used to evaluate the performance of each group. Tables 1–3 list the IoU values obtained for the image groups in iCoseg dataset, Weizmann horses dataset and Oxford 17 flowers dataset respectively. We also include the proposal based segmentation method in [17] for comparison in Table 1. It may be observed that our method performs better than other methods for most classes. The proposed method has the highest mean IoU on all the three datasets. Of the compared methods, the similar background (SB) method [33] yields IoU values closest to the values obtained for the proposed method. Our method fails to generate segmentations

with good accuracy for some classes such as the *Pyramids*, *Soccer*, *Windmill* and *Track*, as evident from Table 1. The *Soccer* and *Track* classes contain cluttered background regions. The *Pyramids* class has background regions quite similar to the foreground and images of the *Windmill* class all contain sky in the background, a portion of which is also co-segmented along with the windmill in certain images. In such cases, accurate separation of foreground objects from the background becomes a difficult task. This accounts for the low mean IoU values of these image classes.

4.2.3. Computation cost

In addition to the segmentation accuracy, we compare the execution time of different methods. The computational cost of the compared methods on a 2.39 GHz Intel(R) Core i3 CPU with 4 GB RAM, are summarized in Table 4. The software platform used was Matlab R2013a. Table 4 shows the average execution time taken by each method for processing an image on the iCoseg dataset. Our method runs faster than the DM [14] and SB [33] methods. The SB [33] algorithm executes significantly slower than other methods, due to the large number of iterations it needs to obtain the final refined co-segmentations.

In our method, for an image set with n images, the object proposals are generated with complexity $O(n)$, and the two methods considered (DC and SER) take $O(n^2)$ time each. Thus the proposed algorithm has an overall complexity of $O(n^2)$. Other methods [14,18,33] have the same time complexity of $O(n^2)$.

From Fig. 10, we can observe the effect of the number of object proposals considered on the running time of the proposed method (DC and SER methods). Here, running time considered t_d is given by $t_d = t_a - (t_p + t_{cs})$, where, t_a = total average running time of the algorithm, t_p = average time taken for object proposals generation and t_{cs} = average time taken for co-saliency map computation. Note here that, we do not consider all object proposals generated by the method in [5]. We only consider proposals having mean co-saliency value greater than the Otsu threshold, T_{Otsu} of the co-saliency map, as outlined in Section 3.1. However, for the purpose of algorithm execution time study, we vary the threshold, T between 20% and 90% of the maximum co-saliency map value of an image and observe its effect on the execution time of the proposed method. It is observed that the running time falls with increase in the co-saliency threshold value, T . An image dependent

threshold, T_{otsu} has been used in our experiments, as higher values of threshold often leads to consideration of very few object proposals, which results in inadequate segmentation for some object classes.

4.3. Error analysis

In the previous subsections, we made various qualitative and quantitative evaluations of all compared methods. Now we look into some of the instances in image classes where the proposed method fails to yield desired segmentation. Fig. 11 shows such cases. In the 1st and 3rd images (from left) of the *Pyramid* class, we find that horses are the wrongly segmented regions. The 2nd image (from left) segments the horse cart along with the pyramid. These inaccuracies occur due to co-saliency map strength domination over color histogram similarity in the graph weight modeling. In these images, the horses have relatively more co-saliency than the pyramid, thus leading to inaccurate segmentation. Similarly in the *Hot balloon* class, all balloons shown are inadequately segmented. The red stripped regions in the balloons gain more saliency values than the bluish regions on the balloon skin, in the co-saliency map and thus carry more site entropy rate and in-degree values, leading to incomplete segmentations. Note here that, the selection of object proposals (as segmented regions) in our method relies to a large extent on the co-saliency map values of the images. The proposed algorithm only selects from among the object proposals, which carry regions that get more prominence in the co-saliency maps (3rd row in Fig. 11). Both the DC and SER based algorithms are likely to fail in such cases, where the common objects do not get high mean co-saliency values.

5. Conclusion and future work

In this paper, we propose a new co-segmentation model to segment common objects from multiple images. We first generate several possible object proposals on the images. Then, based on the region similarities and co-saliency values, we construct a k -partite graph to represent the relationships between different object proposals. Next, we apply two well known graph based methods: entropy based degree centrality computation and site entropy rate computation to extract object proposals which represent the co-segmented image regions.

In our future work, we will extend the proposed model by considering other features such as texture and contour to model the graph weights, which indicate proposal similarity. Incorporating contour information can highly increase the segmentation accuracy by preserving object boundaries better. We also plan to extend the current graph based co-segmentation model to perform object co-localization.

References

- [1] F. Jing, M. Li, H.J. Zhang, B. Zhang, Relevance feedback in region-based image retrieval, *IEEE Trans. Circuits Syst. Video Technol.* 14 (5) (2004) 672–681.
- [2] H. Li, K.N. Ngan, Q. Liu, FaceSeg: automatic face segmentation for real-time video, *IEEE Trans. Multimedia* 11 (1) (2009) 77–88.
- [3] Y.Y. Boykov, M.P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images, in: Proc. Int. Conf. Computer Vision, vol. 1, pp. 105–112.
- [4] J. Zhang, J. Zheng, J. Cai, A diffusion approach to seeded image segmentation, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2010, pp. 2125–2132.
- [5] P. Rantalankila, J. Kannala, E. Rahtu, Generating object segmentation proposals using global and local search, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2014.
- [6] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* 9 (1) (1979) 62–66.
- [7] X. Cao, Z. Tao, B. Zhang, H. Fu, W. Feng, Self-adaptively weighted co-saliency detection via rank constraint, *IEEE Trans. Image Process. (T-IP)* 23 (9) (2014) 4175–4186.
- [8] C. Rother, V. Kolmogorov, T. Minka, A. Blake, Co-segmentation of image pairs by histogram matching – incorporating a global constraint into MRFs, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, New York, USA, 2006, pp. 993–1000.
- [9] L. Mukherjee, V. Singh, C. Dyer, Half-integrality based algorithms for co-segmentation of images, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, Miami, USA, 2009, pp. 2028–2035.
- [10] D. Hochbaum, V. Singh, An efficient algorithm for co-segmentation, in: Proc. Int. Conf. Computer Vision, Kyoto, Japan, 2009, pp. 269–276.
- [11] Y. Long, Y. Huang, Image based source camera identification using demosaicking, in: Proc. IEEE Eighth Workshop on Multimedia Signal Processing, Victoria, Canada, 2006, pp. 419–424.
- [12] K. Chang, T. Liu, S. Lai, From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011, pp. 2129–2136.
- [13] S. Vicente, V. Kolmogorov, C. Rother, Co-segmentation revisited: models and optimization, in: Proc. Eur. Conf. Computer Vision, 2010, pp. 465–479.
- [14] A. Joulin, F. Bach, J. Ponce, Discriminative clustering for image co-segmentation, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2010, pp. 1943–1950.
- [15] D. Batra, A. Kowdle, D. Parikh, Icoseg: interactive co-segmentation with intelligent scribble guidance, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2010, pp. 3169–3176.
- [16] L. Mukherjee, V. Singh, J. Peng, Scale invariant co-segmentation for image groups, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2011, pp. 1881–1888.
- [17] S. Vicente, C. Rother, V. Kolmogorov, Object co-segmentation, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2011, pp. 2217–2224.
- [18] G. Kim, E.P. Xing, L. Fei-Fei, T. Kanade, Distributed co-segmentation via submodular optimization on anisotropic diffusion, in: Proc. Int. Conf. Computer Vision, November 2011, pp. 169–176.
- [19] F. Meng, H. Li, G. Liu, K.N. Ngan, Object co-segmentation based on shortest path algorithm and saliency model, *IEEE Trans. Multimedia* 14 (5) (2012) 1429–1441.
- [20] J. Rubio, J. Serrat, A. Lopez, N. Paragios, Unsupervised co-segmentation through region matching, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, Providence, USA, 2012, pp. 749–756.
- [21] F. Meng, H. Li, G. Liu, K.N. Ngan, Image co-segmentation by incorporating color reward strategy and active contour model, *IEEE Trans. Cybern.* 43 (2) (2013).
- [22] W. Tao, K. Li, K. Sun, SaCoseg: object co-segmentation by shape conformability, *IEEE Trans. Image Process.* 24 (3) (2015) 943–955.
- [23] Q. Li, Y. Zhou, J. Yang, Saliency based image segmentation, in: International Conference on Multimedia Technology, 2011, pp. 5068–5071.
- [24] K.Y. Chang, T.L. Liu, S.H. Lai, From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model, in: IEEE Int. Conference on Comput. Vision Pattern Recognit., June 2011, pp. 2129–2136.
- [25] G. Sharma, F. Jurie, C. Schmid, Discriminative spatial saliency for image classification, in: IEEE Conf. Computer Vision and Pattern Recognition, June 2012, pp. 3506–3513.
- [26] J. Huang, X. Yang, X. Fang, W. Lin, R. Zhang, Integrating visual saliency and consistency for re-ranking image search results, *IEEE Trans. Multimedia* 13 (4) (2011) 653–661.
- [27] C. Zhang, W. Lin, W. Li, B. Zhou, J. Xie, J. Li, Improved image deblurring based on salient-region segmentation, *Signal Proc.: Image Commun.* 28 (9) (2013) 1171–1186.
- [28] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.
- [29] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2013, pp. 1155–1162.
- [30] H. Fu, X. Cao, Z. Tu, Cluster-based co-saliency detection, in: IEEE Trans. Image Process. vol. 22(10), pp. 3766–3778.
- [31] R. Pal, A. Mukherjee, P. Mitra, J. Mukherjee, Modeling visual saliency using degree centrality, *IET Comput. Vis.* 4 (3) (2010) 218–229.
- [32] W. Wang, Y. Wang, Q. Huang, W. Gao, Measuring visual saliency by site entropy rate, in: Proc. IEEE Int. Conference on Comput. Vision Pattern Recognit., June 2010, pp. 2368–2375.
- [33] F. Meng, H. Li, K. Ngan, B. Zeng, N. Rao, Co-segmentation from Similar Backgrounds, in: IEEE International Symposium on Circuits and Systems (ISCAS), 2014, pp. 353–356.
- [34] M. Nilsback, A. Zisserman, A visual vocabulary for flower classification, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, New York, USA, 2006, pp. 1447–1454.
- [35] E. Borenstein, E. Sharon, S. Ullman, Combining top-down and bottom-up segmentation, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), June 2004.
- [36] F. Meng, B. Luo, C. Huang, Object co-segmentation based on directed graph clustering, in: Conf. Visual Communications and Image Processing (VCIP), November 2013, pp. 1–5.