Image Co-Segmentation via Examples Guidance

Rachida Es-Salhi, Imane Daoudi, Hamid El Ouardi
Engineering Research Laboratory,
ENSEM-Hassan II University
Casablanca, Morocco

***Abstract*—Given a collection of images which contains objects from the same category, the co-segmentation methods aim at simultaneously segmenting such common objects in each image. Most of existing co-segmentation approaches rely on computing similarities inter-regions representing foregrounds in these images. However, region similarity measurement is challenging due to the large appearance variations among objects in the same category. In addition, for real-world images which have cluttered backgrounds, the existing co-segmentation approaches miss sufficient robustness to extract the common object from the background. In this paper, we propose a new co-segmentation method which takes advantage of the reliable segmentation of few selected images, in order to guide the segmentation of the remaining images in the collection. A random sample of images is first selected from the image collection. Then, the selected images are segmented using an interactive segmentation method. These segmentation results are used to construct positive/negative samples of the targeted common object and background regions respectively. Finally, these samples are propagated to the remaining images in the collection through computing both local and global consistency. The experiments on the iCoseg and MSRC datasets demonstrate the performance and robustness of the proposed method.**

Keywords*—Co-segmentation; image segmentation; segmentation propagation; MRF based segmentation

I. Introduction

Foreground segmentation is defined as the task of generating pixel level foreground masks for all the objects in a given image or video. Accurate foreground segmentation is very important and basic problem in computer vision field since it has several potential applications like content-based image retrieval [1], image editing [2] and action recognition [3].

In order to highlight the foreground region to be extracted, image segmentation approaches exploit different metrics at the pixel or region level such as saliency, color, texture or shape. However, when dealing with images that have cluttered backgrounds, or images where the foreground has similar attributes as the background, the question of "what to segment out" become more problematic. Considering the limitations of individual image segmentation, in recent years, jointly segmenting multiple images containing a common object has become very popular in a way that the common patterns that exist in a set of similar images can serve as a mean of compensating for the lack of information about visual object foreground. This task of segmenting simultaneously multiple images which contain common or similar objects is known as image co-segmentation.

A. Motivation

Numerous co-segmentation approaches, with various formulations, have been proposed and have proven to be very effective in extracting common objects from a collection of related images. The main idea of all these approaches is to exploit the repeated pattern in the image collection to obtain a form of *a prior* information about the common object to be extracted. On one hand, this weak supervision is an attractive leverage which is not available in the case of single image segmentation, on the other hand the existing co-segmentation models also involve new challenges: 1) Even for images that contain a common object, similarity measurement is challenging due to the large appearance variations among objects in the same category. Also, for images with cluttered background, it could be quite difficult to distinguish the object from the background, and moreover, the image similarity calculation may be useless. 2) Even with a prior information obtained from the related images, the resulting fully automatic segmentation may be imperfect, and in some situations, segmenting images individually performs better, as demonstrated in [4], [5]. Furthermore, in realistic applications, images generally contain similar backgrounds (i.e. similar scenes) such as frames sampled from a video. For these images the co-segmentation process may provide random and insufficiently accurate results. 3) The existing co-segmentation problem is usually formulated using complex models which require a number of parameters to be regulated, especially when dealing with large datasets.

B. Contributions

To deal with the above challenges, the idea of this paper is to use the segmentation of small sample of images to guide the segmentation process in the remaining images. All object/background segments in the sampled set are used as positive/negative samples to be exploited as reliable prior information about the common object in the image collection. Then, the segmentation of a given image is mainly based on similarity between candidate object regions extracted from this image and the positive/negative samples. Particularly, the aim is to transfer the training samples to the unsegmented images by considering simultaneously global and local consistency. The main contributions of this paper are:

- Given the foreground segmentation of only a subset of images, selected randomly, a simple local and global consistency propagation method is proposed to guide the segmentation process of the remaining unsegmented images.

- The proposed method is not limited to segmenting predefined object categories provided in the learning process, which the case of fully supervised methods. As our method is partially interactive, it can segment any object based on the randomly sampled images.
- Instead of propagating only object segmentation samples to the unsegmented images, the proposed method considers both object and background samples in the propagation process. Indeed, this prior information about both the targeted object and the background can better discern the common object in the images, particularly in the case where the image background shares the same features with the foreground.

The rest of this paper is organized as follows: An overview of the related works is presented in Section II. The proposed method is explained in Section III. Experimental results and discussion are given in Section IV, followed by the concluding remarks in Section V.

II. Related work

The co-segmentation problem is a newly explored field of image segmentation. It is defined as the task of jointly segmenting the common region/object from multiple related images. This idea was first introduced in [6]. Since then, numerous formulation of the co-segmentation problem have been proposed ranging from binary-class co-segmentation models(single common foreground object) to multi-class co-segmentation and multi-group co-segmentation. In this study, we are interested in the binary-class co-segmentation model.

In the literature, co-segmentation approaches could be organized into these categories: 1) Markov Random Fields (MRF) based methods [6]–[13], 2) clustering methods [14]–[16] and 3) object proposal selection based methods [17]–[19].

The first family comprises co-segmentation methods based on the Markov Random Field model (MRF). The main idea behind these approaches is to extend the single image segmentation model by adding foreground similarity constraint into the traditional MRF segmentation model. Usually, a new global term is added to the energy function which accounts for foreground similarity. Several foreground similarity measurements are designed, such as L1-norm [6] and L2-norm [11]. In the work of Hochman et al [9], a rewarding similarity measurement is proposed instead of penalizing the foreground difference. This similarity measurement led to a sub-modular energy function which can be easily optimized with graph cuts. Vicente et al [13] compared the three aforementioned MRF-based models and derived a new effective model that was optimized using the dual decomposition method. Later, many works contributed to improve the foreground similarity measure by bringing scale invariance [12, 20]. In the same way, Batra et al [7] have extended the traditional interactive segmentation method by developing an interactive image co-segmentation approach which segmented common objects from the image collection through human interaction. Dong et al. [21] proposed a new interactive co-segmentation method formulated by an unified energy function which encodes the global scribbled energy, inter-image energy and local smooth energy. More recently, [22] introduced the use of higher-order energy to formulate the interactive image co-segmentation problem, where the higher-order term encodes the consistency between the labeled regions and all over-segmentation regions in the image. Instead of relying on the user interaction, other methods used co-saliency, a closely related work to image co-segmentation, to estimate possible foreground locations, then these co-saliency values were exploited to construct the MRF data term. However, adding foreground similarity constraint into the MRF model resulted in non-submodular energy function which is not easy to optimize. So, the focus of all MRF based co-segmentation methods has been on improving approximating solutions which led in most cases to coarse segmentation of the common object.

Other works formulated the co-segmentation problem as a clustering task. In [14], authors handled the segmentation problem in a discriminative framework that combines bottom-up image segmentation with kernel method to assign foreground/background labels jointly to all images. To deal with foreground appearance variations, they used multiple invariant features in the similarity measurement. The discriminative clustering based co-segmentation method [14] was extended in [16] to segment multiple common regions. This method involved a spectral-clustering term and a discriminative term into a new energy function which can be optimized efficiently by using EM algorithm. A large-scale based co-segmentation method was proposed by Kim et al [15], where the joint segmentation task was molded by temperature maximization with finite K heat sources on a linear anisotropic diffusion system. This can be represented as a K-way segmentation that maximizes the segmentation confidence of every pixel in an image. In theory, this temperature function is a sub-modular function, and thus at least a constant approximation of the optimal solution is guaranteed by a greedy algorithm.

MRF based methods and clustering based approaches usually can only provide coarse pixel-level segmentation, thus, large object variations and complicated image backgrounds decrease these methods performance. To this end, methods based on object proposal have been attracting a growing attention [5, 17]–[19, 23]. The main idea behind these methods is to select a subset of the object proposals by evaluating their consistency using region similarity.

These proposals were generated beforehand, and the selected were considered as common targets. In [5], a constraint that the common target has to be an object was added to the co-segmentation framework and an off-line learning method was introduced to retrieve visually similar object proposals among different images. These new aspects contributed to a notable improvement of object co-segmentation performance. In [18], multiple object proposals of all images were represented with a directed graph where similarity between adjacent object proposals were represented by weighted edges. Finally, the common foreground selection was achieved using shortest path algorithm. In the work of [23], additional information such as depth was used to improve proposal based co-segmentation results. These approaches were easily affected by the quality of those generated proposals, and they failed to work well when there were no good proposals in the generated candidates.

All the existing co-segmentation approaches exploited the weak prior information i.e. the same object category contained in collection of images. These co-segmentation approaches constrained correspondence relationship between common objects to better highlight them. For instance, they used additional prior as objectiveness measure [24], or saliency prior or co-saliency measure [8]. By introducing these constraints to object co-segmentation formulation, the common objects could be better segmented even in high appearance variations conditions. Even though, these models still could not obtain robust performance in real-world image collection, where target objects were not salient or shared similar features with the image background.

In this paper, we propose to use the segmentation of few images to guide the segmentation of the remaining images in the collection. In contrast of fully supervised methods which require a large amount of training data from a predefined set of object categories, we demonstrate in this work that the propagation of few images from the image collection can improve considerably the segmentation performance. In such conditions, providing some guidance while segmenting a common object from a complex image collection can improve the segmentation results. Hence, we propose to use the segmentation of few images to guide the segmentation of the remaining images in the collection.

III. The Proposed Method

Given a collection of images all belonging to the same object category, the goal is to extract the common object from all these images. The basic idea of this work is to exploit the segmentation results of randomly selected image samples and use these results to guide the segmentation task of the entire image collection. The work-flow of the proposed method is shown in Fig. 1. First an image sample is randomly selected from the image collection (Fig. 1a), then from each selected image, foreground and background regions are extracted to form a set of positive and negative segments (Fig. 1b) using an interactive segmentation method, in such a way that positive segments are the targeted object instances which we aim to segment out, and the negative segments are representing background regions. Finally, the main step in our proposed approach is to transfer this available information (i.e. positive/negative segments) to the remaining images in the collection (Fig. 1c). To do so, from each remaining image, multiple region candidates are generated. Afterwards, positive/negative segments are transducted to each region candidate by considering both global and local region consistency. The algorithm for the different steps of the proposed co-segmentation method has been detailed in Algorithm 1.

A. Random Image Sample Selection

Consider $\mathfrak{I} = \{I_1, I_2, ..., I_N\}$ a large collection of N images all of which contain instances of the same object category. From the image collection $\mathfrak{I}$, an image subset $\mathfrak{T} = \{I_1, ..., I_M\}$ of M images is randomly selected. In the next step, these selected images will be used to extract the positive and negative samples.

Algorithm 1 Image co-segmentation guided by positive/negative segments

procedure GUIDED-CO-SEGMENTATION
  From the image collection $\mathfrak{I} = \{I_1, I_2, ..., I_N\}$ select a random subset $\mathfrak{T} = \{I_1, ..., I_M\}$ of M images.
  Obtain the segmentation result for each image I_k in $\mathfrak{T} = \{I_1, ..., I_M\}$ using grab-cut algorithm. and construct the positive/negative samples set using these segmentation.
  for each remaining image I_i **do**
    generate a set of candidate regions $\{C_{ij}\}_{j=1}^{R}$
    retrieve a set N_i of most similar images I_k in $\mathfrak{T}$
      Compute the global consistency:
    for each region C_{ij} **do**
      retrieve n_s most similar samples in N_i using equation (5).
      compute the common object estimates $M_{co}(C_{ij})$ of region C_{ij} by equation (6)
      based on $M_{co}(C_{ij})$ of all regions, compute the common object estimates M_i^Gin image I_i
    end for
      Compute local consistency:
    for each image I_k **do**
      from I_i and I_k generate a number $n_r = 10$ of local regions
      for each image r_j^i **do**
        retrieve its most similar local regions in I_k using equation (8).
        compute the local object estimates M_i^L using equation (9)
      end for
    end for
      compute the final common object estimate in I_i using equation 10
      obtain the final segmentation using grab-cut algorithm.
  end for
end procedure

B. Positive/Negative Segments Extraction

In this step, we aim to generate positive and negative segments from the selected image subset $\mathfrak{T} = \{I_1, ..., I_M\}$. For that, we use grabcut method [25] witch is an interactive based segmentation method. Given an image $I_i \in \mathfrak{T}$, the goal is to estimate a label matrix L_i, where $L_i(p) = y_i(p)$ denotes the binary label for the pixel p, and $y_i(p) \in \{0, 1\}$. The label 0 denotes the background and 1 denotes the foreground. The standard grabcut framework [25] involves three steps: initial labeling, learning the appearance model using Gaussian Mixture Model(GMM) and energy minimization.

- Initial labeling: Initially the user provide a bonding box specifying foreground and background regions. Label 1 is assigned to pixels within the foreground region and label 0 for pixels within the background region.

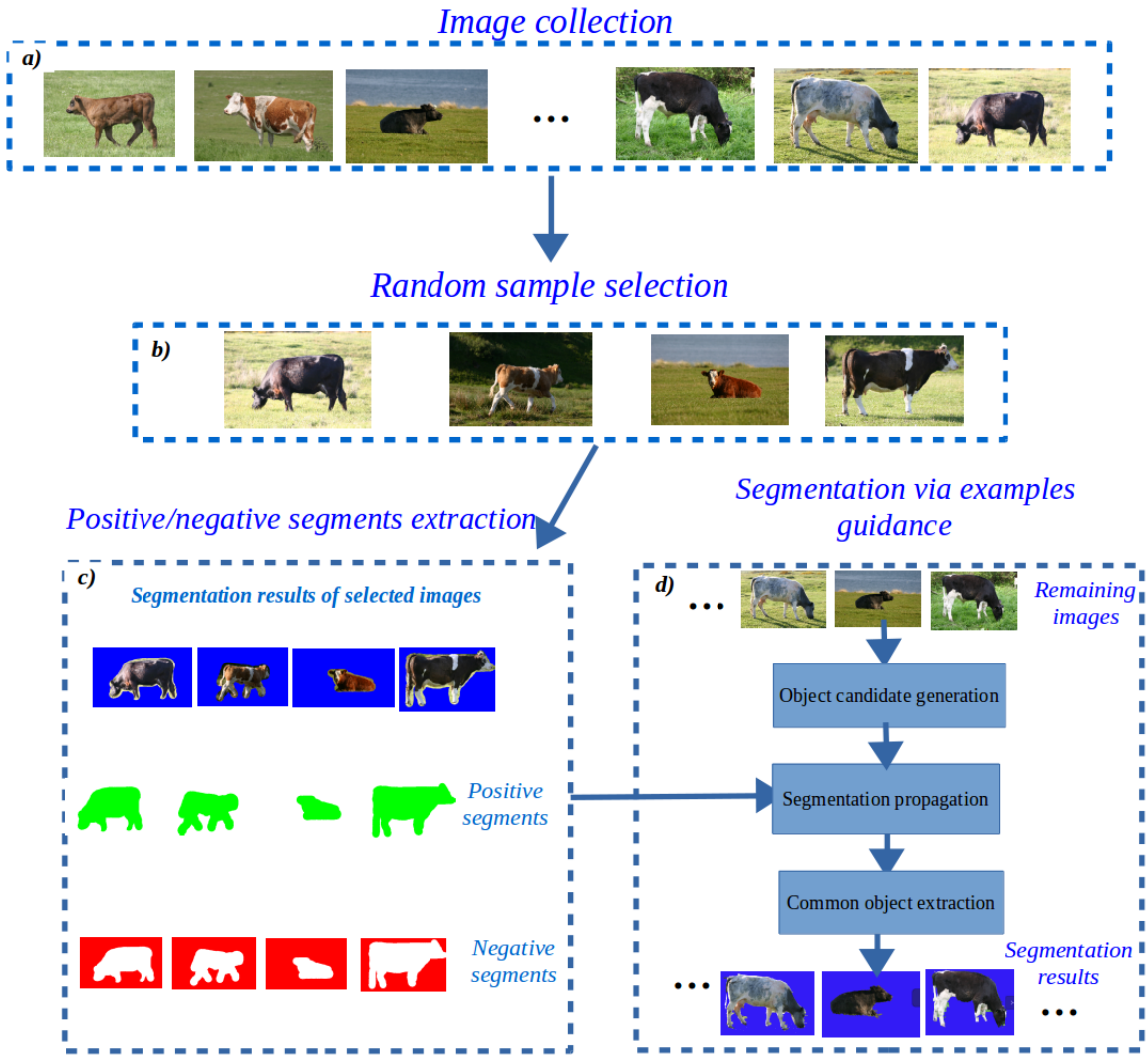

Fig. 1: Flowchart of the proposed co-segmentation method

- Learning the appearance model using GMM: In this step, pixels inside and outside this bounding box are used to learn two Gaussian Mixture Model(GMM) for the foreground and the background in RGB space. Let G_f^i and G_b^i denote those two mixture models. Then, the negative log-likelihood value of a pixel p is computed as follows:

$$\begin{cases} D_i^f(p) = -log(P(z_i(p)|G_f^i)) \\ D_i^b(p) = -log(P(z_i(p)|G_b^i)) \end{cases} \quad (1)$$

Where $z_i(p)$ denotes the RGB color for pixel p in image I_i. This term reflects the cost of assigning a pixel as foreground (or background) according to the GMM models.

- Energy minimization: The object extraction is performed by minimizing the following Gibbs energy function:

$$E(L_i) = U(L_i) + V(L_i) \quad (2)$$

Where $U(L_i)$ is the data term encoding the probability that a pixel p belongs to object or background:

$$U(L_i) = \sum_p [D_i^f(p).L_i(p) + D_i^b(p).(1 - L_i(p))] \quad (3)$$

with $L(p)$ is the label of pixel p that equal to 1 if p belongs to the object and it is equal to 0 if it belongs to the background.
$V(L_i)$ is the smoothness term that penalizes assigning different labels to neighboring pixels with similar color

features. It is defined as follows:

$$V(L_i) = \sum_{(p,q)\in N} [L_i(p) \neq L_i(q)] e^{-\beta d(z_i(p), z_i(q))} \quad (4)$$

where β is a scaling parameter.
The equation 2 is efficiently minimized using grabCut that apply five rounds of iterative refinement, alternating between learning the likelihood values using GMM and obtaining the label estimates.

After obtaining the segmentation results of all images in $\mathfrak{T}$. As shown in Fig. 1c, we extract positive and negative samples; such that all object segmentation results i.e. foreground regions are considered as positive samples and similarly background regions are filed as negative samples. In the next step, all extracted positive and negative regions will be propagated to each remaining image in the collection in order to guide its segmentation.

C. Segmentation via Examples Guidance

In the grabcut based segmentation method, the unary term $U(L_i)$ describes the foreground model which is learned from the user scribbles on the image . In this step, we aim to substitute the user interaction using the pre-segmented image sample. It means that the previously extracted positive/negative samples are used to guide the segmentation of the remaining images. Hence, for each unsegmented image, we aim to define the unary term based on the proposed segmentation propagation process (discussed next), and then perform the grabcut segmentation to extract the object foreground.

1) Object candidate generation: In order to transfer the available positive/negative samples to the unsegmented images, we first extract a number of region candidates which represent object and background regions. To ensure that the common object will be segmented as a local region, the proposed method in [18] is adopted. Namely each image $I_i \in \mathfrak{I} \backslash \mathfrak{T}$ is segmented into R region candidates $\{C_{ij}\}_{j=1}^{R} = \{C1, C2, C3\}$ which comprise three subsets: $C1$ is comprised of super-pixels generated using the over-segmentation method [26], $C2$ contains the segmentation results obtained by saliency detection method [27] and $C3$ includes the segmentation of detected objects in I_i using object detection method [24]. Note that the extracted object regions will form strong match with positive samples, in the same way, particularly for images from similar scenes, background regions are more likely to match the negative samples.

2) Segmentation propagation: After extracting region candidates from each unsegmented image, we propagate the previously constructed positive/negative samples to each region candidate based on region similarities. Furthermore, to deal with object variance among images, we need to propagate the available segmentation samples to the most similar unsegmented images in $\mathfrak{I} \backslash \mathfrak{T}$. Hence, for each image $I_i \in \mathfrak{I} \backslash \mathfrak{T}$, we first retrieve a set N_i of most similar images in $\mathfrak{T}$ and estimate the common object in I_i guided by those images only. In order to account for the foreground region in the similarity measurement, the weighted Gist descriptor [28] is used to represent each image. Basically, given the saliency map S_i of image I_i, a coarse initial foreground/background estimation is computed by thresholding S_i using Otsu method [29], i.e. $\{S_i^f, S_i^b\} = Otsu(S_i)$ and then these pixel estimates are used as a weight of Gist descriptor.
We define the segmentation propagation task using two components, namely, the global consistency and local consistency, so that the global consistency propagates the overall information by considering the whole segment in the similarity measurement. As for the local consistency, and in order to deal with object appearance variations, the local information represented by local patches is propagated to the extracted region candidates.

a) Global consistency: In the global consistency the whole segment information of positive/negative samples is propagated to each unsegmented image. Given an image I_i and the set N_i of its most similar images from the randomly selected images $\mathfrak{T}$. For each object candidate C_{ij} in I_i we first retrieve n most similar samples in N_i, one in each pre-segmented image I_k:

$$l(k) = arg \min_{l} D(C_{ij}, S_{kl}) \quad (5)$$

Where S_{kl} is a positive or negative sample and $D(C_{ij}, S_{kl})$ is the chi-square distance between C_{ij} and S_{kl} features . Then the common object estimates of object candidate C_{ij} is given by the following equation:

$$M_{co}(C_{ij}) = \sum_{k=1}^{n} M(S_{l(k)})(1 - D(C_{ij}, S_{l(k)})) \quad (6)$$

Where $M(S_{l(k)})$ is the object likelihood of the region sample $S_{l(k)}$. Clearly, if regions $S_{l(k)}$ retrieved by equation 5 are positive samples, then their object likelihood $M(S_{l(k)})$ are assigned to 1 as a result, the common object estimates of region C_{ij} is higher and therefore this regions is more likely to belong to the common object. Otherwise, if these regions are negative samples (their object likelihood are assigned to 0), the common object estimates $M_{co}(C_{ij})$ is lower.
Note that object candidate C_{ij} extracted from I_i may be overlapping. So a pixel $I_i(p, q)$ with a location (p, q) may belong to multiple object candidates and will be assigned multiple common object estimates $M_{co}(p, q)$. In this case, the largest one is selected as the common object estimate of the pixel.

$$M_i^G(p, q) = \max_{(p,q)} M_{co}(p, q). \quad (7)$$

b) Local consistency: In real-world conditions the global common object appearance is often inconsistent and difficult to capture due to the large variance of viewpoints, scales and object poses. As a result, considering only the global consistency with the pre-segmented images may not be sufficient to properly estimate the common object in a given image. To handle this problem we look also at local consistency by transferring local regions of positive/negative samples to the unsegmented image. To do so, a set of local patches ,represented by windows, are extracted from both I_k and I_i. Then these local regions are ranked to select $n_r = 10$

relevant patches $\{r_1^k, ...r_{n_r}^k\}$ to represent local information in the pre-segmented image I_k and $\{r_1^i, ...r_{n_r}^i\}$ for image I_i. We also get the local object likelihood $m(r_l^k)$ of a region r_l^k by directly using the object likelihood value of its corresponding positive/negative sample in I_k.
For a local region r_j^i we search for its most similar local regions in I_k based on the distance between feature histograms h_{ij} and h_{kl} of windows regions r_j^i and r_l^k respectively

$$l^* = arg\min_l d(h_{ij}, h_{kl}) \tag{8}$$

Similar to the global consistency computation, we obtain the common object estimates $M_i^L(p,q)$ as follows:

$$M_i^L(p,q) = \sum_{k=1}^{n_s} m(r_{l^*}^k)(1 - d(h_{ij}, h_{kl})) \tag{9}$$

with (p,q) is the pixel location. $m(r_{l^*}^k)$ the object likelihood of positive/negative local region samples (that is equal to 1 if $r_{l^*}^k$ belong to the object and 0 otherwise). As in global consistency computation, a pixel (p,q) may be assigned several local based common object estimates because of the overlapping of detected local regions. In our case the largest one is chosen as the common object estimate.

c) Common object extraction: To obtain the final common object estimates, we combine the global and local consistency maps as follow:

$$M_i^T = \alpha M_i^L + (1-\alpha) M_i^G \tag{10}$$

Where α is a scaling coefficient. From the common object estimates in the image I_i, the object extraction is performed using GrabCut algorithm described in Section III-B. Here the initial label assignment of a pixel p is determined by thresholding M_i^T using the common Otsu's method [29].

$$p \in \begin{cases} F_i & if \quad M_i^T > \tau \\ B_i & if \quad M_i^T < \tau \end{cases} \tag{11}$$

With τ is the global threshold value. Then the final segmentation of image I_i is obtained iteratively through alternating between learning the foreground/background GMM and obtaining the label assignments (equation 1, 2, 4).

IV. Experimental Results

A. Experimental Setting

To demonstrate the efficiency of the proposed method, the experiments are conducted on two publicly available datasets, namely iCoseg [7] and MSRC [30] datasets which have been frequently used in previous co-segmentation studies; MSRC dataset contains 14 categories with 418 images in total. iCoseg dataset contains 38 categories with 643 images in total. Regarding the parameter setting, we set the number of randomly selected images $M = 6$ and the number of nearest neighbors $n_s = 3$. In (10) coefficient α and $1-\alpha$ regulate the importance of the global and local consistency term. We set $\alpha = 0,6$ for all datasets. The color histogram is used for segmentation propagation in ICoseg dataset. For MSRC dataset that exhibits more intra-group variation, color feature for matching the segments will be unreliable. As a result, we used the dense SIFT feature for matching.

Following the literature, two objective measures, Jaccard Similarity (J), and Precision (P) are used for the quantitative results. Denote A_p^f,A_p^b , A_g^f and A_p^b as proposed foreground pixels set, proposed background pixels set, ground-truth foreground pixels set and ground-truth background pixels set, respectively. Here, Jaccard Similarity is defined as the size of intersection divided by the size of union of the proposed and ground truth foreground pixels sets:

$$\frac{\left|A_p^f \cap A_g^f\right|}{\left|A_p^f \cup A_g^f\right|}$$

And Precision [31] is defined as the percentage of pixels that have same labels in both the proposed and ground truth masks:

$$\frac{\left|A_p^f \cap A_g^f\right| + \left|A_p^b \cap A_g^b\right|}{\left|A_g^f \cup A_g^b\right|} * 100$$

The quantitative comparison results between the state-of-the-art algorithms and ours are given in the following subsections.

B. Comparison with the State-of-the-Art

The proposed method is compared with different state-of-the-art object co-segmentation algorithms, including Unsupervised joint object discovery and segmentation in Internet images [4] (named ObjectDiscovery13), Group saliency propagation for large scale and quick image co-segmentation (GSP) [32] and automatic image co-segmentation using geometric mean saliency (GMS) [28]. Image co-segmentation via saliency co-fusion (Kotes16) [33] and a semi-supervised method for image co-segmentation (Es-salhi17) [34].

We note that to compare with the work of Rubinstein et al. [4], the results are reproduced using their publicly available implementation. Moreover, results of [28] and [32] are regenerated by running the codes provided kindly by the authors. For co-segmentation via saliency co-fusion [33], the results reported on the paper are considered.

For iCoseg dataset the precision values obtained by each method on different image groups are depicted in Fig. 2. The precision averages of all groups are shown in the first column. Clearly the proposed method achieves the best result (92.71% accuracy average). Specifically, compared with the work in [34], which transfers the object segmentation of randomly selected images to the unsegmented images, the new proposed method performs better. This demonstrates that transferring both the object segmentation and the background regions to the unsegmented images can accurately extract the common object from interfered or complex background. This is particularly observable for image groups: *bear* (the average accuracy recorded 90,07 %) and *brown bear* (97,85%) where the images share similar foreground and background, this is also the case for *stonehenge* (97,30%), *panda1* (91,03%) , *panda2* (84,29%), *kendo* (97,58%), *kendo2* (98,93%) and *taj mahal* (93,45%) where the proposed method improves considerably

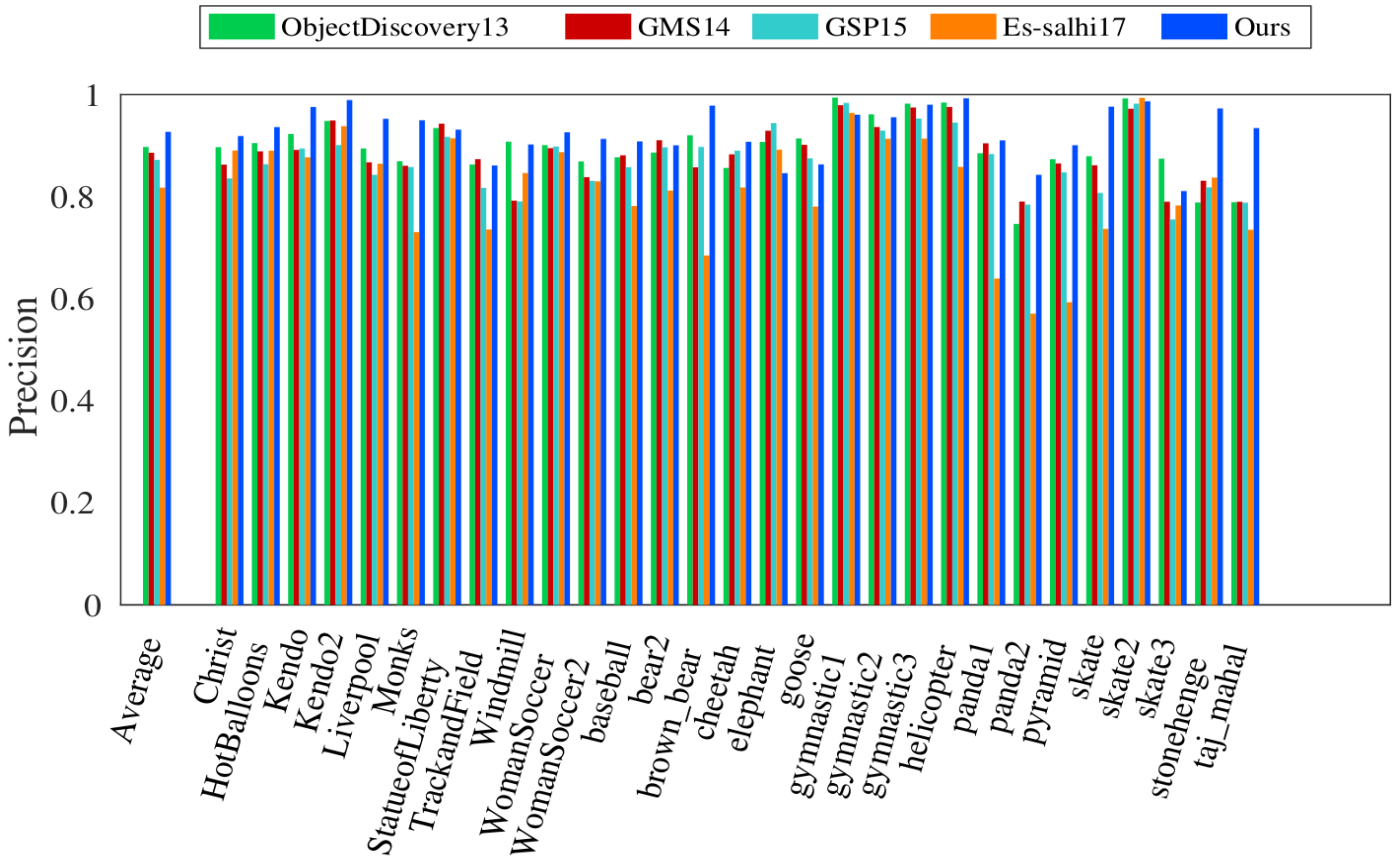

Fig. 2: Comparison between the proposed method and the state-of-the-art methods ObjectDiscovery13 [4], GMS14 [28], GSP [32] and Es-salhi17 [34] on iCoseg dataset.

the segmentation accuracy. Moreover, the proposed method outperforms the other methods on most groups.

We next objectively evaluate the proposed method by the Jaccard similarity metric (J). The results are summarized in Table 2. Obviously, our proposed method outperforms the existing methods on most image groups of the challenging iCoseg dataset. Particularly, the method gives considerably better results than [28] and [4], even if they used dense correspondences to compute consistency between images in the group. This is expected since in a group of related images, the object instances usually appear on similar backgrounds, and consequently computing correspondences between these images can not highlight accurately the common object. However, this is not a crucial issue in our approach, where prior information transferred from positive/negative samples can accurately guide the segmentation of the common object.

Besides, it should be noted that some image groups in iCoseg data set contain small number of images (less than ten images), which is not appropriate for the random image selection step. Thus for these groups, only the segmentation of one randomly selected image is propagated to guide the segmentation of the remaining images. Under these conditions, our method gives appealing results, especially for *brown bear* and *taj mahal* groups.

For overall comparison, Table 1 shows the numeric precision and Jaccard similarity averages on iCoseg dataset compared with the existing methods. Fig. 4 further illustrates visual results of the proposed method on 10 sample groups from iCoseg dataset. The odd columns represent the original images and the even columns display their segmentation results. We can clearly see that the proposed method achieves a smooth segmentation results even when the common object appears in cluttered or similar backgrounds.

TABLE 1: Precision average $\bar{P}$ and Jaccard similarity average $\bar{J}$ on Icoseg dataset.

Method	$\bar{P}$ (%)	$\bar{J}$(%)
[4]	89,74	69,17
[28]	88,61	65,50
[32]	87,21	61,48
[34]	81,77	47,11
Ours	**92,71**	**76,25**

Besides, we evaluate our approach on MSRC dataset, the quantitative results are presented in Fig. 3 where the class-

TABLE 2: Evaluation results comparison between the proposed method and other co-segmentation methods in terms of Jaccard Similarity values. image groups of iCoseg dataset are considered.

	[4]	[28]	[32]	[34]	Ours
Christ	0,770	0.795	0.757	0,692	**0,821**
HotBalloons	0,657	0,763	0,802	0,515	0,692
Kendo	0,778	0,862	0,896	0,663	**0,916**
Kendo2	0,826	0,893	0,921	0,813	**0,962**
Liverpool	0,541	0,512	0,470	0,412	**0,671**
Monks	0,681	0,688	0 ,683	0,446	**0,857**
StatueofLiberty	0,799	0,813	0,863	0,686	0,733
TrackandField	0,519	0,632	0,595	0,313	**0,645**
Windmill	0,492	0,316	0,531	0,220	0,501
WomanSoccer	0,661	0,657	0,699	0,574	**0,728**
WomanSoccer2	0,530	0,538	0,526	0,386	**0,678**
baseball	0,657	0,756	0,703	0,354	0,644
bear2	0,653	0,701	0,675	0,393	0,692
brown bear	0,736	0,662	0,725	0,214	**0,834**
cheetah	0,697	0,754	0,780	0,583	**0,786**
elephant	0,688	0,735	0,799	0,5523	0,706
ferrari	0,724	0,703	0,708	0,566	**0,834**
goose	0,742	0,773	0,503	0,328	0,660
gymnastic1	0,948	0,910	0,976	0,678	0,651
gymnastic2	0,840	0,897	0,831	0,447	0,825
gymnastic3	0 ,896	0.911	0.892	0,508	0,905
helicopter	0,803	0,766	0,803	0,560	**0,904**
panda1	0,759	0,806	0,722	0,253	**0,809**
panda2	0,625	0,718	0,614	0,340	**0,744**
pyramid	0,611	0,686	0,595	0,155	**0,743**
skate	0,735	0,737	0,769	0,376	**0,935**
skate2	0 ,911	0,866	0,900	0,924	0,877
skate3	0,449	0,297	0,491	0,176	**0,528**
stonehenge	0,595	0,714	0,781	0,702	**0,930**
taj mahal	0,460	0,587	0,516	0,396	**0,734**

wise comparison of our method with those of state-of-the art is shown. In this comparison 12 groups are used. It can be seen that our results are very competitive to the best methods [28] and [32]. Particularly our method outperforms other existing methods namely on “cow”, “sheep”, “plane” and “bird” groups.

Furthermore, it is interesting to notice that the proposed method reports good results compared with [34] in almost all image groups. This is expected since this method propagated only the positive segments (regions that contain the targeted object) to other images, while the proposed method is based on both positive and negative segmentation transfer. That allows to have better a performance even when images share similar background or when the common object is depicted in very cluttered image backgrounds.

Fig. 5 shows sample segmentation results from MSRC dataset, we display images from 4 groups to show the performance of our method. First column of each group represents original images and the second column displays the segmented images. By comparing these qualitative results, we can see that the proposed method can distinctly improve the segmentation accuracy.

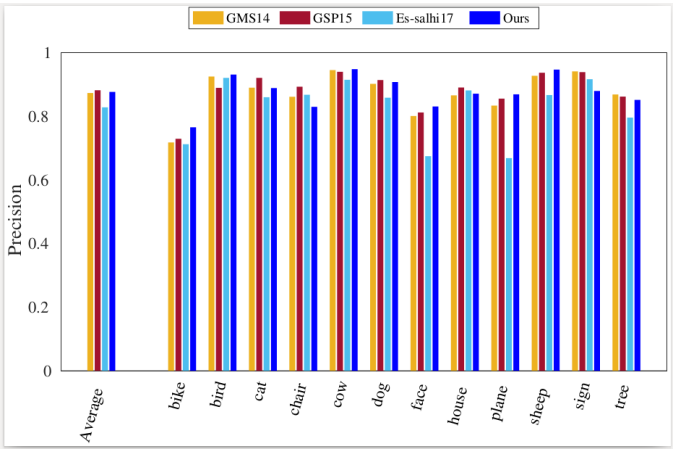

Fig. 3: Comparison between the proposed method and the state-of-the-art methods GMS14 [28], GSP [32] and Es-salhi17 [34] on MSRC data set.

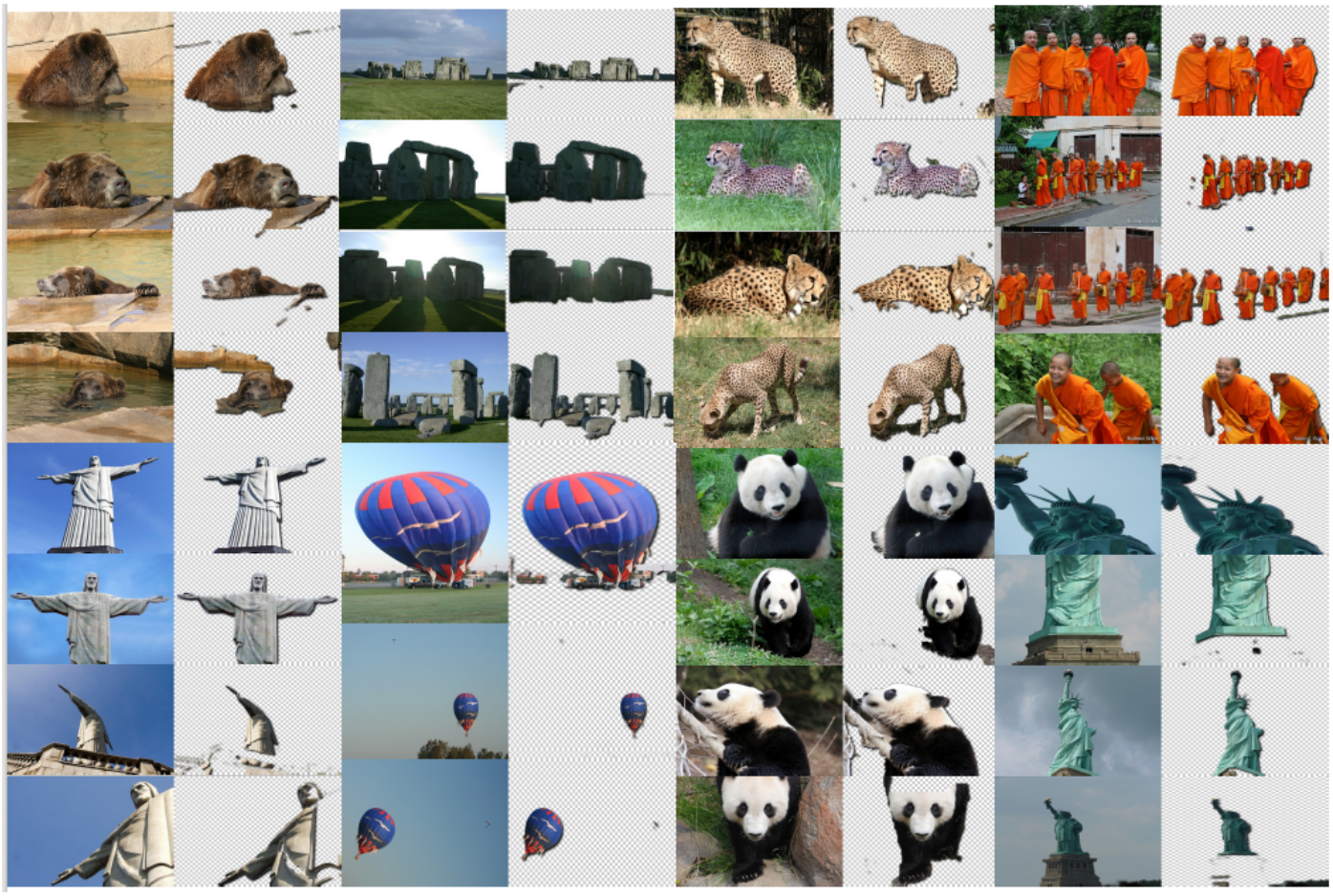

Fig. 4: Sample segmentation results on iCoseg dataset. There are eight groups of images. In each group, the first column represents the original images, and the second column represents the segmentation results.

TABLE 3: Evaluation results comparison between the proposed method and the other co-segmentation methods in terms of Jaccard Similarity values. Classes in MSRC dataset are considered.

	[28]	[32]	[34]	Ours
Bike	0,420	0,424	0,387	0,439
Bird	0,637	0,589	0,662	0,650
Cat	0,668	0,732	0,624	0,637
Chair	0,627	0,671	0,631	0,524
Cow	0,802	0,782	0,745	**0,812**
Dog	0,672	0,682	0,574	0,676
Face	0,577	0,583	0,364	**0,625**
House	0,719	0,755	0,745	0,746
Plane	0,515	0,546	0,391	**0,596**
Sheep	0,781	0,797	0,626	**0,820**
Sign	0,838	0,834	0,779	0,681
Tree	0,760	0,741	0,629	0,739
Average	0,6680	0,6782	0,5965	0.6666

Table 3 lists out the detailed Jaccard similarity results reported for MSRC dataset. The proposed method achieves comparable results with other methods, notably on the following image groups: *cow* (0,812) , *face* (0,625), *plane* (0,596) and *sheep* (0,820).

Fig. 5: Sample segmentation results on MSRC dataset. There are four groups of images. In each group, the first column represents the original images, and the second column represents the segmentation results.

From the experimental results, we can see that propagating both positive and negative prior information constructed by segmenting randomly selected images to the unsegmented images, can guide the segmentation of these images and thus improve the performance of image co-segmentation, especially for the complicated image groups. The proposed global and local complex terms can complement each other in the segmentation propagation step to better handle the object appearance variation among images in the group. In addition, the proposed method does not requires any parameter settings. However, although our approach performs well on benchmark datasets, it is also based on a random image selection step and interactive segmentation of these images, which leads to a semi-supervised approach that may not be suitable for all computer vision applications.

In the context of future work, it is suggested to explore a way to make automatic the positive/negative samples generation step .

V. Conclusion

In this paper, we propose a new method for image co-segmentation. First a random subset of images is selected and segmented using an interactive method, and all region results are used as positive/negative samples to guide the segmentation task. Then for each remaining image, multiple local region generation method are used to segment into a variety of object proposals. All regions in positive/negative set are propagated to all regions of each remaining image by considering both global and local region consistency in the feature space. For each pixel, in the image the maximum foreground estimation value is used to score the foreground estimation as map of the image. Finally, this foreground estimation map is used as a unary term of MRF segmentation based model, and the final segmentation is achieved by graphcut algorithm. The experimental results demonstrate that the proposed method can efficiently segment the common object from a group of images with better precision than many existing co-segmentation methods.

Acknowledgment

The authors wish to acknowledge the anonymous reviewers for their careful readings.

References

[1] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern recognition*, vol. 46, no. 1, pp. 188–198, 2013.

[2] S. Zhang, J. Huang, H. Li, and D. N. Metaxas, "Automatic image annotation and retrieval using group sparsity," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 838–849, 2012.

[3] F. Meng, J. Cai, and H. Li, "Cosegmentation of multiple image groups," *Computer Vision and Image Understanding*, vol. 146, pp. 67–76, 2016.

[4] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, "Unsupervised joint object discovery and segmentation in internet images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13)*, 2013, pp. 1939–1946.

[5] S. Vicente, C. Rother, and V. Kolmogorov, "Object cosegmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*, 2011, pp. 2217–2224.

[6] C. Rother, T. Minka, A. Blake, and V. Kolmogorov, "Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs," in *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR'06)*, 2006, pp. 993–1000.

[7] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR'10)*, 2010, pp. 3169–3176.

[8] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *IEEE conference on Computer vision and pattern recognition, (CVPR'11)*, 2011, pp. 2129–2136.

[9] D. S. Hochbaum and V. Singh, "An efficient algorithm for co-segmentation," in *IEEE12th International Conference on Computer Vision*, 2009, pp. 269–276.

[10] F. Meng, J. Cai, and H. Li, "Cosegmentation of multiple image groups," *Computer Vision and Image Understanding*, vol. 146, no. C, pp. 67–76, 2016.

[11] L. Mukherjee, V. Singh, and C. R. Dyer, "Half-integrality based algorithms for cosegmentation of images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2028–2035.

[12] L. Mukherjee, V. Singh, and J. Peng, "Scale invariant cosegmentation for image groups," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*, 2011, pp. 1881–1888.

[13] S. Vicente, V. Kolmogorov, and C. Rother, "Cosegmentation revisited: Models and optimization," in *European Conference on Computer Vision ECCV' 10*, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Springer Berlin Heidelberg, 2010, pp. 465–479.

[14] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,(CVPR'10)*, 2010, pp. 1943–1950.

[15] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade, "Distributed cosegmentation via submodular optimization on anisotropic diffusion," in *IEEE International Conference on Computer Vision (ICCV'11)*, 2011, pp. 169–176.

[16] A. Joulin, F. Bach, and J. Ponce, "Multi-class cosegmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)*, 2012, pp. 542–549.

[17] K. Li, J. Zhang, and W. Tao, "Unsupervised co-segmentation for indefinite number of common foreground objects," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1898–1909, 2016.

[18] F. Meng, H. Li, G. Liu, and K. N. Ngan, "Object co-segmentation based on shortest path algorithm and saliency model," *IEEE transactions on multimedia*, vol. 14, pp. 1429–1441, 2012.

[19] F. Meng, H. Li, K. N. Ngan, L. Zeng, and Q. Wu, "Feature adaptive co-segmentation by complexity awareness," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4809–4824, 2013.

[20] R. Es-salhi, I. Daoudi, J. Weber, H. El Ouardi, S. Tallal, and H. Medromi, "Multi-scale image co-segmentation," in *Advances in Ubiquitous Networking*. Springer, 2016, pp. 381–390.

[21] X. Dong, J. Shen, L. Shao, and M.-H. Yang, "Interactive cosegmentation using global and local energy optimization," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3966–3977, 2015.

[22] W. Wang and J. Shen, "Higher-order image co-segmentation," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1011–1021, 2016.

[23] H. Fu, D. Xu, S. Lin, and J. Liu, "Object-based rgbd image co-segmentation with mutex constraint," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, 2015.

[24] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, 2010, pp. 73–80.

[25] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM transactions on graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 309–314.

[26] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: An empirical evaluation," in *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR'09)*, 2009, pp. 2294–2301.

[27] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[28] K. R. Jerripothula, J. Cai, F. Meng, and J. Yuan, "Automatic image co-segmentation using geometric mean saliency," in *IEEE International Conference on Image Processing (ICIP'14)*, 2014, pp. 3277–3281.

[29] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models," *Pattern recognition*, vol. 25, no. 10, pp. 1231–1240, 1992.

[30] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *European conference on computer vision*. Springer, 2006, pp. 1–15.

[31] E. Iacona, "Application de l'interférométrie holographique à l'étude de transferts termiques couplés dans un gaz au sein d'une cavité: essai de modélisation," Ph.D. dissertation, Châtenay-Malabry, Ecole centrale de Paris, 2000.

[32] K. R. erripothula, J. Cai, and J. Yuan, "Group saliency propagation for large scale and quick image co-segmentation," in *IEEE International Conference on Image Processing (ICIP'15)*, 2015, pp. 4639–4643.

[33] K. R. Jerripothula, J. Cai, and J. Yuan, "Image co-segmentation via saliency co-fusion," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1896–1909, 2016.

[34] R. Es-salhi, I. Daoudi, and H. El Ouardi, "A new semi-supervised method for image co-segmentation," in *Image Processing Theory, Tools and Applications (IPTA), 2017 Seventh International Conference on*. IEEE, 2017, pp. 1–6.

Rachida Es-salhi received her engineer degree in Computer Sciences from the National Higher School of Electricity and Mechanics- Hassan II university, Casablanca, Morocco, in 2013. She is currently a Ph.D student in image processing and pattern recognition at the Engineering Research Laboratory, Hassan II university, Her research interests include image processing, computer vision and multimedia applications.

Imane Daoudi received her Ph.D degree in Computer Sciences from Mohamed 5 university, Rabat Morocco and the National Institute of Applied Sciences, Lion, France. She is an Associate Professor at the National Higher School of Electricity and Mechanics. Her major research interests include image similarity search, multidimensional indexing and multimedia applications.

Hamid El Ouardi received his first Ph.D degree in Applied Mathematics from Paul Sabatier University, Toulouse, France, and his second Ph.D degree in Applied Mathematics from Chouaïb Doukkali, El Jadidda, Morocco. He is a professor at the National Higher School of Electricity and Mechanics. His research interests include applied mathematics and mathematical modeling.