Higher-Order Image Co-segmentation

Wenguan Wang and Jianbing Shen, *Senior Member, IEEE*

***Abstract*—A novel interactive image cosegmentation algorithm using likelihood estimation and higher order energy optimization is proposed for extracting common foreground objects from a group of related images. Our approach introduces the higher order clique's, energy into the cosegmentation optimization process successfully. A region-based likelihood estimation procedure is first performed to provide the prior knowledge for our higher order energy function. Then, a new cosegmentation energy function using higher order cliques is developed, which can efficiently cosegment the foreground objects with large appearance variations from a group of images in complex scenes. Both the quantitative and qualitative experimental results on representative datasets demonstrate that the accuracy of our cosegmentation results is much higher than the state-of-the-art cosegmentation methods.**

***Index Terms*—Energy optimization, higher order cliques, image cosegmentation, likelihood estimation.**

I. Introduction

IMAGE co-segmentation is commonly referred as jointly partitioning multiple images into foreground and background components. The idea of co-segmentation is first introduced by Rother *et al.* [5] where they simultaneously segment common foreground objects from a pair of images. The co-segmentation problem has attracted much attention in the last decade, most of the co-segmentation approaches [2], [3], [8], [10], [13], [18], [23], [24] are motivated by traditional Markov Random Field (MRF) based energy functions, which are generally solved by the optimization techniques such as linear programming [8], dual decomposition [18] and network flow model [10]. The main reason may be that the graph-cuts and MRF methods [4], [33] work well for image segmentation and are also widely used to solve the combinatorial optimization problems in multimedia processing. Similar rationale is also adopted by some co-saliency methods [9], [42], [44].

The existing image co-segmentation methods can be roughly classified into two main categories, including unsupervised co-segmentation techniques and interactive co-segmentation approaches. The common idea of the unsupervised techniques [5], [11], [16], [22], [27], [29], [35], [37] formulates image co-segmentation as an energy minimization and binary labeling problem. These approaches usually define the energy function using standard MRF terms and histogram matching term. The former encourages the consistent segmentations in every single image while the later penalizes the differences between the foreground histograms of multiple images.

Inspired by interactive single-image segmentation methods [7], [15], [26], several interactive co-segmentation approaches [17], [19], [21], [28] using user scribbles have been proposed in recent years. The user usually indicates scribbles of foreground or background as additional constraint information to improve the co-segmentation performance. These interactive co-segmentation approaches can handle a group of related images and improve the co-segmentation results by user scribbles. Batra *et al.* [19], [21] proposed an interactive image co-segmentation approach to segment foreground objects with user interactions. They learned foreground/background appearance models using user scribbles. Recently, Collins *et al.* [28] formulated the interactive image co-segmentation problem as the random walk model and added the consistency constraint between the extracted objects from a set of input images. Their method utilized the normalized graph Laplacian matrix and solved the random walk optimization scheme by exploiting its quasi-convexity of foreground objects.

This study formulates the interactive image co-segmentation problem in terms of the higher-order energy optimization, which complements the existing MRF segmentation framework and improves the accuracy of co-segmenting the challenging images with foreground objects that have variations in color and texture only by a few of user seeds. Higher-order energy optimization [12], [14], [20], [25], [31], [32], [34], [41] has been widely used in many fields of computer vision like image denoising [14] and single-image segmentation [12], [34]. We construct higher-order clique as a composed group of three parts: the foreground region, the background region and the over-segmentation region, which considers the correspondence between the over-segmentation region and the labeled region. This strategy makes our framework effective enough in realistic scenarios, instead of a simple foreground/background appearance histogram model. Additionally, our higher-order energy efficiently utilizes the statistical information on a group of pixels by estimating the segmentation quality on higher-order cliques.

As shown in Fig. 1, the interactive co-segmentation results by our higher-order energy achieve more accurate results than previous approaches, especially when foreground and background contain similar colors. The co-segmentation results by random walk co-segmentation (RWCS) method [28] [see Fig. 1(b) and (c)] can not extract the correct foreground objects of planes in this complex scene. Compared to the existing interactive co-segmentation methods using low-order energy function, our high-order energy function optimizes the co-segmentation

Manuscript received September 25, 2015; revised February 24, 2016; accepted March 19, 2016. Date of publication March 22, 2016; date of current version May 13, 2016. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2013CB328805, in part by the National Natural Science Foundation of China under Grant 61272359, and in part by the Fok Ying Tung Education Foundation under Grant 141067 Specialized Fund for Joint Building Program of Beijing Municipal Education Commission. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chengcui Zhang. *(Corresponding author: Jianbing Shen.)*

The authors are with the Beijing Key Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: wangwenguan@bit.edu.cn; shenjianbing@bit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2016.2545409

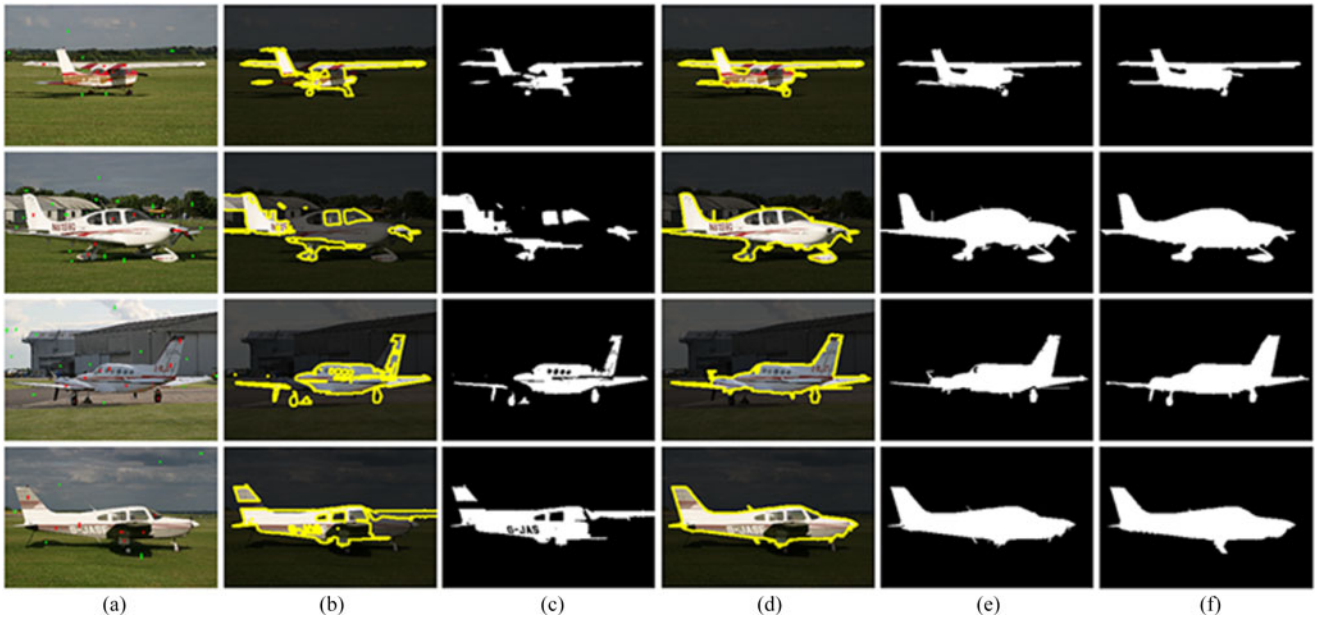

Fig. 1. Co-segmentation comparisons in complex scenes. (a) The input images and user scribbles. (b) and (c) are the co-segmentation results and masks using the interactive RWCS algorithm [28]. (d) and (e) are the co-segmentation results and masks by our algorithm. (f) The ground-truth. Note that our co-segmentation method achieves better performance than the RWCS approach in [28] with the same user scribbles.

process by utilizing richer statistical information of natural images and object relationship by our likelihood estimation. This strategy greatly improves the performance of our co-segmentation results [see Fig. 1(d) and (e)]. Our source code will be available online.[1]

Compared to existing image co-segmentation methods, the proposed approach offers the following contributions.

1) We formulate the interactive image co-segmentation via likelihood estimation and high-order energy optimization, which utilizes the region likelihoods of multiple images and considers the quality of segmentation to achieve promising co-segmentation performance.
2) A novel higher-order clique construction method is proposed using the estimated foreground/background regions and the regions of original images.
3) A new region likelihood estimation method is presented, which provides enough prior information for higher-order energy item for generating final co-segmentation results.

The rest of the paper is organized as follows. Our proposed co-segmentation method with high-order energy term and how to reduce its order is described in detail in Section II. The experimental results are provided in Section III to support the efficiency of our proposed algorithm. Finally, Section IV concludes the paper and gives the future work.

II. Our Approach

A. Overview

Our co-segmentation procedure includes two main steps. The first step is a fast but effective likelihood estimation process, which calculates the probabilities of pixels belonging to foreground/background over entire dataset according to user scribbles. The estimated likelihood offers a rough estimation for foreground /background and is fed into next step as prior knowledge. This process is described in Section II-B. In the second stage, a higher-order energy based co-segmentation function is proposed to obtain final accurate co-segmentation results on a group of images, which is based on higher order cliques. Our higher-order cliques are constructed from a set of foreground and background regions by user scribbles, where all the regions in each image are matched to produce better co-segmentation performance. Additionally, our approach considers the quality of segmentation in higher-order energy to obtain more accurate estimations of foreground/background. We present this part in Section II-C.

B. Likelihood Estimation

Given a group of images $\{I^1, \ldots, I^n\}$ and the user scribbles that indicate foreground or background objects, we first compute pixel likelihood x_k^i for foreground/background in image I^i. The likelihood of pixel x_k^i is denoted by $\pi_{k,l}^i$ where l is a label indicating foreground (1) or background (0) and k is the index value of x_k^i. We compute the likelihoods of regions instead of pixels for computational efficiency. Each input image I^i of the group is divided into regions $r_s^i \in R^i$ using the over-segmentation methods such as mean shift [1] or efficient graph [6] method. For each region r_s^i, the region likelihoods of foreground and background are defined as $z_{s,l}^i$, which is further formulated in a quadratic energy function as follows:

$$\begin{aligned} F_l^i &= F_1 + F_2 \\ &= \lambda^i \sum_{s=1}^{N(R^i)} (z_{s,l}^i - \varepsilon_{s,l}^i)^2 + \sum_{s,s'=1}^{N(R^i)} w_{s,s'}^i (z_{s,l}^i - z_{s',l}^i)^2 \end{aligned} \quad (1)$$

[1][Online]. Available: http://github.com/shenjianbing/hoecoseg

where the first term F_1 defines an unary constraint that each region tends to have the initial likelihood $\varepsilon^i_{s,l}$ estimated through the appearance similarity to foreground/background. The second term F_2 gives the interactive constraint that all regions of the whole image should have same likelihood when their representative colors are similar.

The parameter λ is a positive coefficient for balancing the relative influence between F_1 and F_2. $w^i_{s,s'} = exp(-\|\overline{c}^i_s - \overline{c}^i_{s'}\|)$ is a weighting function that gives a similarity measure for regions r^i_s and $r^i_{s'}$ in color space, and $\overline{c}^i_s$ is the mean color of region r^i_s. $N(R^i)$ is the number of regions of R^i and the parameter $z^i_{s,l}$ indicates the likelihood of region r^i_s. $\varepsilon^i_{s,l}$ defines the initial likelihood for region r^i_s.

Given the user scribbles, we can get the background region set $u_j \in U^{(0)}$ and foreground region set $u_{j'} \in U^{(1)}$. We use the shortest Euclidean distance between region r^i_s and the background/foreground region set (U^0/U^1) in color space to compute the initial likelihood $\varepsilon^i_{s,l}$ for region r^i_s. The initial likelihood $\varepsilon^i_{s,l}$ is formulated as

$$\varepsilon^i_{s,l} = \begin{cases} \dfrac{\min\limits_{u_j \in U^0}(\|\overline{c}^i_s - \overline{c}_j\|)}{\min\limits_{u_j \in U^0}(\|\overline{c}^i_s - \overline{c}_j\|) + \min\limits_{u_{j'} \in U^1}(\|\overline{c}^i_s - \overline{c}_{j'}\|)} & \text{if } l = 1 \\ \dfrac{\min\limits_{u_{j'} \in U^1}(\|\overline{c}^i_s - \overline{c}_{j'}\|)}{\min\limits_{u_j \in U^0}(\|\overline{c}^i_s - \overline{c}_j\|) + \min\limits_{u_{j'} \in U^1}(\|\overline{c}^i_s - \overline{c}_{j'}\|)} & \text{if } l = 0 \end{cases} \quad (2)$$

where $\overline{c}_j$ ($\overline{c}_{j'}$) is the mean color of background region u_j (foreground region $u_{j'}$).

Based on the region likelihoods $\vec{z}^i_l = [z^i_{s,l}]_{N(R^i)\times 1}$ and their initial region likelihoods $\vec{\varepsilon}^i_l = [\varepsilon^i_{s,l}]_{N(R^i)\times 1}$, the quadratic energy function F_l is formulated as the following matrix forms:

$$F^i_l = (\vec{z}^i_l - \vec{\varepsilon}^i_l)^T \Lambda^i (\vec{z}^i_l - \vec{\varepsilon}^i_l) + \vec{z}^{iT}_l (D^i - W^i)\vec{z}^i_l \quad (3)$$

where $W^i = [w^i_{s,s'}]_{N(R^i)\times N(R^i)}$ and $D^i = diag([d^i_1, \ldots, d^i_{N(R^i)}])$. The diagonal elements of the metric D^i are the degrees of the weight matrix W^i: $d^i_s = \sum_{s'=1}^{N(R^i)} w^i_{s,s'}$. The diagonal elements of the metric Λ^i are $diag([\lambda^i, \ldots, \lambda^i])_{N(R^i)\times N(R^i)}$.

(3) is then solved by the following convex optimization:

$$\frac{\partial F^i_l}{\partial \vec{z}^i_l} = \Lambda^i(\vec{z}^i_l - \vec{\varepsilon}^i_l) + (D^i - W^i)\vec{z}^i_l = 0. \quad (4)$$

After solving (4), we finally obtain the region likelihoods $\vec{z}^i_l$ as follows:

$$\vec{z}^i_l = \frac{\Lambda^i \vec{\varepsilon}^i_l}{\Lambda^i + D^i - W^i}. \quad (5)$$

Considering the definition of $\varepsilon^i_{s,l}$ in (2), we have $\varepsilon^i_{s,0} + \varepsilon^i_{s,1} = 1$. According to $\varepsilon^i_{s,0} + \varepsilon^i_{s,1} = 1$ and (5), we have

$$z^i_{s,0} + z^i_{s,1} = 1. \quad (6)$$

We only need to calculate either $\vec{z}^i_0$ or $\vec{z}^i_1$ using (5). (5) is easily computed by least-square and the optimization only takes 0.02 s for 500 over-segmentation regions per image in our tests.

After the region likelihood $\vec{z}^i_l$ is obtained, the pixel likelihood $\pi^i_{k,l}$ is set to the same value as the likelihood of the region that this pixel belongs to

$$\pi^i_{k,l} = z^i_{s^k,l}$$

where s^k indicates the region $r^i_{s^k}$ that pixel x^i_k belongs to.

C. Higher-Order Energy Co-Segmentation

Via our likelihood estimation, we have a fast and rough estimate for foreground/background in each image. For generating more accurate co-segmentation results, we further propose a higher-order energy based co-segmentation function.

In order to simultaneously segment a group of input images $\{I^1, \ldots, I^n\}$ with the labeled images T, we first build a global term $E_{\text{global}}(I^1, \ldots, I^n, T)$ to match all the images with the labeled images T. The proposed energy of our co-segmentation algorithm is expressed as follows:

$$\mathcal{F} = \sum_{i=1}^{n} \left(\epsilon^i_1 E^i_{\text{unary}} + \epsilon^i_2 E^i_{\text{pairwise}}\right) + E_{\text{global}}(I^1, \ldots, I^n, T) \quad (7)$$

where E^i_{unary} and E^i_{pairwise} denote unary term and pairwise term respectively and the global term E_{global} is proposed to match all the input images $\{I^1, \ldots, I^n\}$ with labeled images T. The scalars ϵ weight various terms.

The unary term E^i_{unary} and the pairwise term E^i_{pairwise} for image I^i are defined as follows:

$$E^i_{\text{unary}} = \sum_k -log(\pi^i_{k,1}) \cdot \phi(x^i_k) - log(\pi^i_{k,0}) \cdot (1 - \phi(x^i_k))$$

$$E^i_{\text{pairwise}} = \sum_{k,k' \in \aleph} \|c^i_k - c^i_{k'}\| \cdot |\phi(x^i_k) - \phi(x^i_{k'})| \quad (8)$$

where c^i_k denotes the color value of pixel x^i_k and $\pi^i_{k,l}$ is obtained in our likelihood estimation step. The set $\aleph$ contains all the four-neighbors within one image. $\phi(x^i_k)$ is a binary function indicating the assignment of pixel x^i_k to the background (0) or foreground (1).

The unary term E^i_{unary} is based on the likelihood estimation results and penalizes assignments of pixels with lower likelihood to foreground. The pairwise term E^i_{pairwise} imposes intra-image label smoothness by constraining the segmentation labels to be consistent, which tends to assign the same label to neighboring pixels that have similar color.

The co-segmentation model in (7) is intuitive. Next we discuss how to design the global energy item in the following paragraphs. Previous co-segmentation approaches [5], [10] performed co-segmentation on image pairs and made simple assumption that two input images shared a same/similar foreground object. In contrast, we try to extract common foreground objects that have large variations in color, texture and shape from a group of images with complex background. Rather than building a simple foreground or background appearance model, we collect a region set of foreground/background according to user interaction. The region set $\Im$ of foreground/background

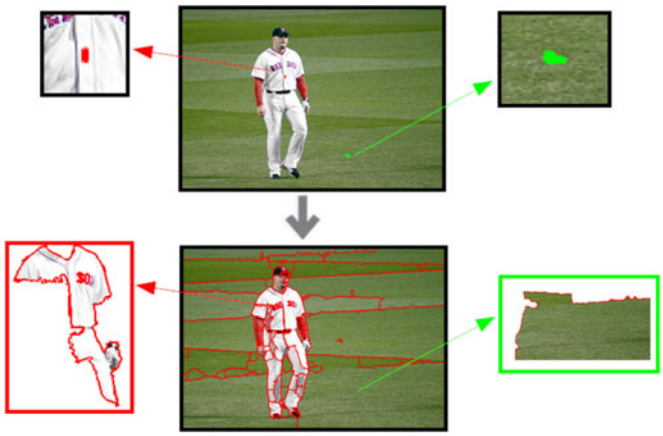

Fig. 2. Illustration of obtaining the region set $\Im$ from user seeds. In the top row, the middle image is one of the labeled images T. The scribble seeds are shown in close-ups, where the red (green) seeds denote the foregrounds (backgrounds). In the bottom row, the middle image denotes the over-segmentation results. Close-ups represent the labeled regions U^l which are extracted from these over-segmentations according to user seeds.

consists of the labeled regions U^l

$$\Im = \{U^0, U^1\}$$

where U^0/U^1 means the background/foreground regions respectively.

The construction process of $\Im$ is accomplished by the previous likelihood estimation step, and all the regions with user scribbles are added into this region set $\Im$. Fig. 2 gives the process of obtaining the region set $\Im$ from user seeds. In order to build the matching relationships between input images $\{I^1, \ldots, I^n\}$ and labeled foreground/background images T, our solution is to make the matching process between the over-segmentation regions R^i of image I^i and the labeled regions from region set $\Im$. Then we define the following higher-order energy item:

$$E_{\text{global}}(I^1, \ldots, I^n, T) = \sum_{i=1}^{n} E_{\text{high}}(R^i, \Im). \tag{9}$$

By associating with (8) and (9), our co-segmentation energy function $\mathcal{F}$ in (7) is then rewritten as

$$\begin{aligned}\mathcal{F} = \sum_{i=1}^{n} \Bigg\{ \sum_{k} \Big(exp^{-\pi^i_{k,1}} \phi(x^i_k) + exp^{-\pi^i_{k,0}} (1 - \phi(x^i_k)) \Big) \\ + \sum_{k,k' \in \aleph} \|c^i_k - c^i_{k'}\| \cdot |\phi(x^i_k) - \phi(x^i_{k'})| \\ + E_{\text{high}}(R^i, \Im) \Bigg\}.\end{aligned} \tag{10}$$

The minimization of unary term and pairwise term in $\mathcal{F}$ (11) can be efficiently solved by the graph cut algorithm. Then we focus on how to design the higher-order term $E_{\text{high}}(R^i, \Im)$ of I^i. We will introduce the higher-order cliques into matching process. The higher-order cliques are composed of three regions: the foreground region, the background region and the over-segmentation region. The co-segmentation process using higher-order energy is shown in Fig. 3. Both the foreground region and the background region are selected to construct our region set $\Im$. Then we build an higher-order energy function $E_{\text{high}}(R^i, \Im)$ on higher-order cliques as follows:

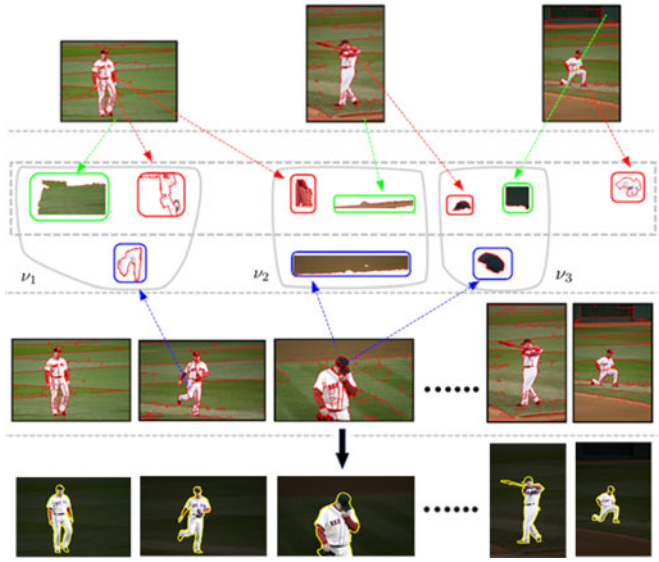

Fig. 3. Illustration of the higher-order cliques. In the second row, higher-order cliques (v_1, v_2, v_3) are constructed by the labeled regions from region set $\Im$ and the over-segmentation region from images. Labeled images T (top row) provide the labeled regions U_1/U_0 to build the region set $\Im$, where the foreground regions U_1 (background regions U_0) are shown in the red (green) rectangles. Input images $\{I^1, \ldots, I^n\}$ (third row) are over-segmented into regions shown in blue rectangles. The final co-segmentation results are given in the bottom row.

$$\nu^i_s = \{r^i_s, u^1(r^i_s), u^0(r^i_s)\} \tag{11}$$

where $u^l(r^i_s) \in U^l$ denotes the most related foreground or background region to r^i_s in Euclidean distance measurement using their mean colors.

For each region r^i_s of image I^i, our algorithm finds the most similar foreground and background region from $\Im$ respectively to make up a higher-order clique. Then the matching energy function using our higher-order cliques is defined as follows:

$$E_{\text{high}}(R^i, \Im) = \sum_{s=1}^{N(R^i)} N(\nu^i_s) \cdot \kappa^i_s \tag{12}$$

where $N(\nu^i_s)$ indicates the number of pixels in clique ν^i_s, which means a large clique will have a large value of weight. κ^i_s defines the matching coefficient which considers both the clique likelihood for foreground/background and the segmentation quality. We define the matching coefficient κ^i_s as follows:

$$\kappa^i_s = \min \left\{ \min_{\mathrm{l}} \left(\mathrm{z}^{\mathrm{l}}(\nu^{\mathrm{i}}_{\mathrm{s}}) \cdot \vartheta^{\mathrm{l}}(\nu^{\mathrm{i}}_{\mathrm{s}}) + \mathrm{z}^{\bar{\mathrm{l}}}(\nu^{\mathrm{i}}_{\mathrm{s}}) \right), 1 \right\} \tag{13}$$

where $z^l(\nu^i_s)$ is the clique likelihood estimated from ν^i_s, which is compute via

$$z^l(\nu^i_s) = \frac{N(r^i_s) z^i_{s,l} + N(u^1(r^i_s)) l + N(u^0(r^i_s))(1 - l)}{N(r^i_s) + N(u^1(r^i_s)) + N(u^0(r^i_s))}. \tag{14}$$

The parameter $\vartheta^l(\nu_s^i)$ in (13) is used to estimate the segmentation quality, which is based on region consistency assumption. The region consistency assumption encourages all pixels belonging to a region to take the same label. We define $\vartheta^l(\nu_s^i)$ as follows:

$$\vartheta^l(\nu_s^i) = \frac{N(r_s^i) - N_l(r_s^i)}{\rho N(r_s^i)}, \qquad 0 < \rho < 1 \tag{15}$$

where $N_l(r_s^i)$ is the number of pixels assigned to foreground/background in region r_s^i. We commonly set $\rho = 0.1$ in our experiments. That means, if more than 90% of the pixels in region r_s^i are classified into foreground, the value of $\vartheta^1(\nu_s^i)$ is less than 1. Similarly, if 50% of the pixels in region r_s^i are classified into foreground, the value of $\vartheta^1(\nu_s^i)$ is set to 5.

According to definition in (15), the more pixels of a region have the same label, the better segmentation quality on this region and the smaller value of $\vartheta^l(\nu_s^i)$ will be. Our higher-order energy function in (13) is a linear truncated function, which means that the higher-order energy function allows some pixels of a region to take different labels.

In clique ν_s^i , we only need to consider the segmentation quality on region r_s^i because pixels in region $u^1(r_s^i)$ or $u^0(r_s^i)$ have been classified into foreground or background. Therefore, the more pixels in region r_s^i are divided into a same label, the higher segmentation quality is obtained and the lower the value of $\vartheta^l(\nu_s^i)$. From (14) and (15), we can find that the number of pixels in foreground/background $N(u^l(r_s^i))$ can influence the clique likelihood and the value of matching coefficient κ_s^i. Then we set $N(u^l(r_s^i)) = 0.5\, N(r_s^i)\varepsilon_{s,l}^i$. If region r_s^i is closer to foreground, $\varepsilon_{s,1}^i$ is larger and $N(u^1(r_s^i))$ will make a greater influence on the matching coefficient κ_s^i.

Higher-order energy function in (13) utilizes clique likelihood $z^l(\nu_s^i)$ for foreground and background. The region consistency in our higher-order clique is also taken into account as an evaluation for segmentation quantity ($\vartheta^l(\nu_s^i)$). In other words, our higher-order energy function considers the similarity between higher-order clique and foreground/background, which encourages all the pixels of a region to take the same label. Next we will introduce our optimization method for higher-order energy in (13). Because the value of $N_l(r_s^i)$ is constant, the problem of minimizing our higher-order energy function can be transformed into a problem of minimizing the matching coefficient κ_s^i, which is defined in the following two important theorems.

Theorem 1: The matching coefficient κ_s^i in (13) can be rewritten as

$$\kappa_s^i = \min\left\{\frac{\sum_{x_s^i \in r_s^i}\phi(x_s^i)}{Q}(1-t)+t, \frac{N(r_s^i) - \sum_{x_s^i \in r_s^i}\phi(x_s^i)}{Q}t + (1-t), 1\right\} \tag{16}$$

where $t = z^1(\nu_s)^i$, $Q = \rho N(r_s^i)$, and pixel x_s^i belongs to region r_s^i. $\phi(\cdot)$ is the binary function indicating the assignment of pixel to background (0) or foreground (1) in (8).

Theorem 2: The matching coefficient κ_s^i in (16) can be transformed into a second-order function by introducing the auxiliary binary variables σ_0 and σ_1

$$\begin{aligned}\kappa_s^i = & \min \Psi(\sigma_0, \sigma_1, \phi(x_s^i)) \\ = & \min_{\sigma_0,\sigma_1} \sigma_0 \frac{\sum_{x_s^i \in r_s^i}\phi(x_s^i)}{Q}(1-t) + (1-\sigma_0)(1-t) \\ & + \sigma_1 \frac{N(r_s^i) - \sum_{x_s^i \in r_s^i}\phi(x_s^i)}{Q}t + (1-\sigma_1)t.\end{aligned} \tag{17}$$

The proofs of Theorem 1 and Theorem 2 are given in detail in Appendix A and Appendix B. Through Theorem 1 and Theorem 2, the matching coefficients κ_s^i in (13) can be rewritten as a second-order function. Therefore, our higher-order energy based co-segmentation function in (7) can be efficiently solved by the conventional graph cut algorithm.

III. Experimental Results

In this section, we first discuss our experiments for evaluating the performance between our algorithm and previous well-known co-segmentation approaches [16], [21], [22], [27], [28], [35]–[38], [40]. Then, we give qualitative and quantitative results obtained by the proposed method with and without the higher-order energy. The experimental evaluations are designed to assess the running time statistics of these algorithms. Then, we give qualitative and quantitative results obtained by the proposed method with and without the higher-order energy. The experimental evaluations are designed to assess the running time statistics of these algorithms. Three parameters λ, ϵ_1 and ϵ_2 are used in our two energy functions (1) and (7). We empirically set $\lambda = 10$, $\epsilon_1 = 1$ and $\epsilon_2 = 30$ for all the test image sets in our experiments.

A. Co-segmentation Results

Our method is first compared with the state-of-the-art interactive co-segmentation methods: intelligent scribble guided co-segmentation (ICOSEG) [21], and RWCS [28] on previous benchmark datasets. To achieve a relatively fair comparison, both the proposed method and other interactive co-segmentation methods [21], [28] use the same scribbles in all experiments.

In the experiments, we collect a variety of image groups from well-known image databases such as iCoseg dataset [21] and Microsoft MSRC database [30]. These two datasets are very popular for image co-segmentation experiments where the ground-truth segmentation masks are also provided. Each group of image collections has a common theme or common foreground object, which makes it challenging to co-segment them with user scribble seeds.

We then quantitatively compare the co-segmentation performance of our algorithm with other eight unsupervised approaches: discriminative clustering co-segmentation (DCCS) [16], multi-class co-segmentation (MCCS) [27], distributed co-segmentation (DCS) [22], region matching based co-segmentation (RMCS) [35], consistent functional maps based co-segmentation (CFCS) [36], joint object discovery and segmentation (JODS) [37], multi-class joint segmentation (MJS) [38], and multiple random walkers based co-segmentation (MRCS) [40]. The experimental results by DCCS, MCCS,

Fig. 4. Comparison results. (a) shows the scribbles that ICOSEG used in [21]; (b) gives the results by ICOSEG method using the scribbles in (a); (c) shows our scribbles; (d) shows co-segmentation results by RWCS approach [28] using the scribble seeds in (c); (e) shows the ground truth masks; and (f) shows co-segmentation results by our algorithm using the scribble seeds in (c). Our algorithm achieves better co-segmentation results than both the ICOSEG [21] and RWCS [28] algorithms.

and DCS are produced by directly running the implementation codes from their websites. And the co-segmentation results of ICOSEG are generated by the implementation code from the authors [21]. The experimental results of JODS are downloaded from their websites. The results by RMCS, CFCS, MJS and MRCS are mainly borrowed from original works, therefore only parts of these results are reported.

Fig. 4 gives a comparison between our algorithm and two well-known interactive co-segmentation approaches: ICOSEG [21] and RWCS [28] for a group of challenging images. The group of Brown bear images is a relatively difficult group in iCoseg dataset. From the co-segmentation results by ICOSEG, we can see that most of regions of the common objects are generally segmented {see Fig. 4(b)]. However, there are still many background regions of which color is similar to foreground are classified falsely [see Fig. 4(b)]. The reason is that the established common appearance model does not work well when the color distributions of foreground and background pixels across the entire dataset have too many overlaps.

Another important interactive co-segmentation method is the RWCS [28] using random walker algorithm [39], [43] as its optimization framework. Based on their appearance model with foreground objects, their algorithm achieves better co-segmentation results [see Fig. 4(d)] than the ICOSEG method [see Fig. 4(b)]. However, the similarity between foreground and background color histograms still influences the performance of RWCS, which may lead to the incorrect segmentation of some foreground regions. Compared with the ground truth

Fig. 5. Comparison co-segmentation results between our method and ICOSEG [21], RWCS [28] approaches. The first row shows the input images. The results in the second and third rows are obtained by ICOSEG [21] and RWCS [28], respectively. The results in the fourth row are generated by our method. The ground truth masks are shown in the bottom row.

masks [see Fig. 4(e)], there are still many regions of bears can not be correctly classified and segmented out by RWCS method [see Fig. 4(d)]. It is clear that our method produces high-quality co-segmentation results of foreground bears, while the results by both ICOSEG and RWCS have more or less lost some important foreground regions. Our approach builds labeled region set instead of using foreground/background appearance model. That makes our method do not rely on the strong assumption that the foreground objects share a common appearance model. Therefore, our algorithm is more applicable and robust in realistic and complicated scenarios, which achieves more satisfying results using fewer scribbles [see Fig. 4(c)] than ICOSEG [see Fig. 4(a)].

In order to further demonstrate the advantage of our algorithm, we present a more challenging co-segmentation example for a group of images where both background and foreground are complex in Fig. 5. There are some overlaps between the color distributions of the foreground and background. For example, the color of the chair in the third column is very similar with the color of background trees. ICOSEG [21] and RWCS [28] cannot handle this situation well (see the second row and the third row of Fig. 5), while our approach successively co-segments the foreground object (see the fourth row of Fig. 5). Moreover, our method preserves more details of co-segmentation objects than the results by the ICOSEG and RWCS approaches, especially the chair regions. Our approach fully considers the region consistency in higher-order cliques and performs the co-segmentation optimization process on both foreground and background region set.

We quantitatively evaluate the co-segmentation results on both the iCoseg dataset and MSRC dataset so as to examine the overall performance of the proposed method. The iCoseg dataset consists of 38 groups with totally 643 images that each group has a common theme or foreground object, and we randomly selected 30 groups of them to perform our experiments. The performance is measured by the proportion of correctly classified labeled pixels (both foreground and background) to the total number of pixels. Fig. 6 summarizes the segmentation accuracy for each class of iCoseg dataset. This figure clearly shows that our algorithm has better performance than other co-segmentation methods. Our overall precision (98.2%) shows significant improvement over other co-segmentation methods. For the per-class precision, we can find that our performance is also better than others. It is also seen that the other methods achieve good performance, since most of the common objects that contain similar color distributions in the iCoseg dataset. Therefore, we further make comparison among our algorithm and those methods on MSRC dataset, which consists of many complex scenes and the foregrounds/backgrounds have high appearance variations. The MSRC dataset consists of 20 groups of images from natural scenes and we randomly select 14 groups of them. Several categories including high variability images lead to significant difficulties while our method still offers substantial improvements in co-segmentation results. We list the precision statistics for each class in Table I. The experiment results in Table I well demonstrate that our method outperforms the state-of-the-art co-segmentation methods on the challenging MSRC dataset. Benefitting from successfully integrating the likelihood estimation and higher-order cliques into our framework, the average precision of our algorithm on MSRC reaches 95.7%, which is much higher than the average precisions by ICOSEG (74.8%) and the ones by RWCS (81.1%).

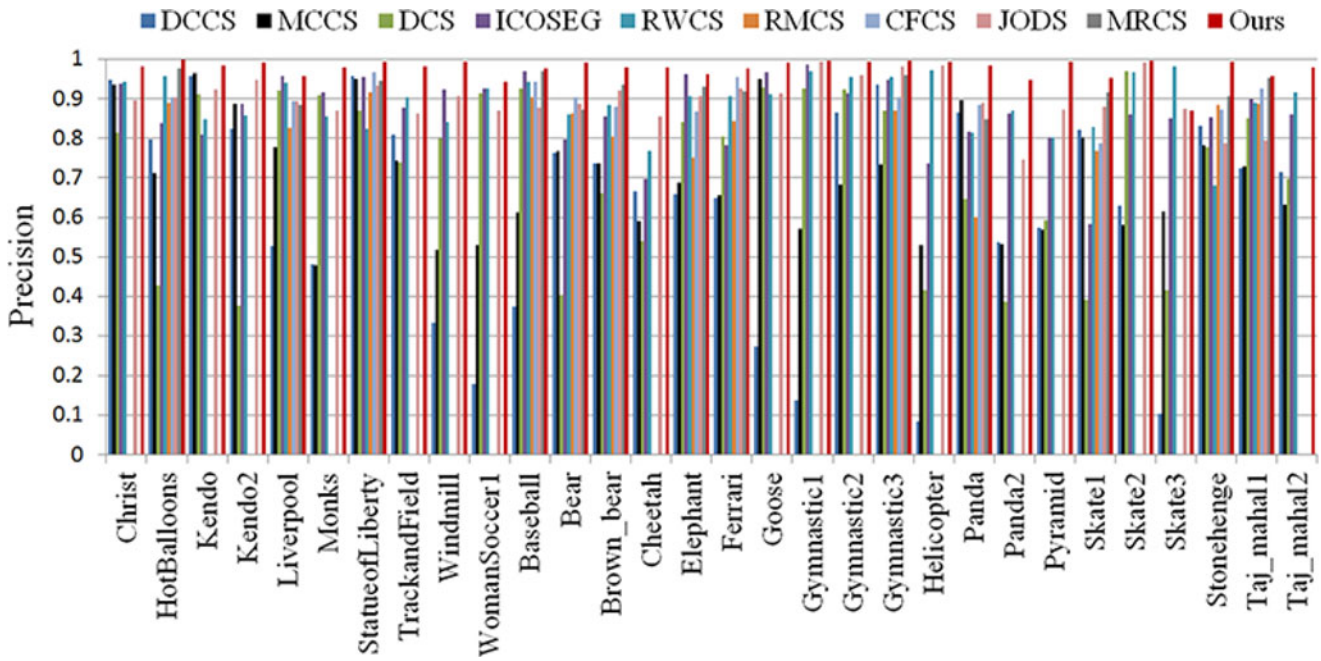

Fig. 6. Comparison results of segmentation accuracy on iCoseg dataset between our co-segmentation method and other co-segmentation approaches (DCCS [16], MCCS [27], and DCS [22], RMCS [35], CFCS [36], JODS [37], and MRCS [40]) and interactive approaches (ICOSEG [21] and RWCS [28]).

TABLE I
SEGMENTATION ACCURACY ON MSRC DATASET BETWEEN OUR CO-SEGMENTATION ALGORITHM AND THE EXISTING STATE-OF-THE-ART CO-SEGMENTATION APPROACHES (DCCS [16], MCCS [27], DCS [22], ICOSEG [21], RWCS [28], RMCS [35], CFCS [36], JODS [37], AND MJS [38])

			supervised				unsupervised						
	class	images	Ours	Ours ($-E_{\text{global}}$)	ICOSEG	RWCS	DCCS	MCCS	DCS	RMCS	CFCS	JODS	MJS
MSRC	bike	30	**90.1**	81.2	74.1	78.1	63.9	68.3	29.9	62.4	-	79.6	51.2
	bird	34	**98.9**	86.3	86.9	85.7	65.8	73.7	25.6	-	95.8	93.2	55.7
	car	30	**95.9**	83.9	72.1	79.9	76.7	79.0	52.9	59.2	83.1	85.6	72.9
	cat	24	**97.8**	89.2	78.3	75.8	63.2	75.2	38.4	77.1	94.5	91.8	65.9
	chair	30	**96.8**	85.3	71.2	66.7	75.2	67.8	72.0	-	-	88.2	46.5
	cow	30	**99.7**	89.1	82.2	75.6	79.8	85.7	83.2	81.6	94.3	97.7	68.4
	dog	30	**98.6**	90.2	75.5	84.0	76.0	75.9	36.5	-	91.3	90.9	55.8
	face	30	88.0	85.5	82.5	85.1	77.4	75.1	56.4	84.3	-	**89.2**	60.9
	flower	32	**98.4**	82.7	83.5	86.3	70.0	68.9	33.8	-	-	88.2	67.2
	house	30	**97.7**	89.5	73.2	82.4	62.6	59.0	56.8	-	-	89.7	56.6
	plane	30	**94.6**	89.5	30.9	92.0	49.4	52.1	56.6	77.0	91.0	87.0	52.2
	sheep	30	**99.2**	93.2	92.5	89.6	89.3	91.5	88.7	-	95.6	94.8	72.2
	sign	30	94.9	89.5	78.3	92.0	80.5	76.7	60.8	-	-	**95.2**	59.1
	tree	30	**89.6**	80.3	65.9	62.1	67.0	85.0	72.6	-	-	87.8	62.0
	Avg.	-	**95.7**	86.8	74.8	81.1	71.2	73.9	54.6	73.6	92.2	89.9	60.4

Higher values are better. The best and the second best results are boldfaced and underlined, respectively.

B. *Effect of Higher-Order Energy*

In this section, we present quantitative and qualitative results for demonstrating the performance improvement of our algorithm after using higher-order cliques energy.

In Table I, we present a baseline, called Ours $(-E_{\text{global}})$, which indicates the co-segmentation results on MSRC dataset without considering higher-order term E_{global} in (7). It is clear that our full approach using higher-order cliques optimization gains higher segmentation accuracy. Fig. 7 gives an intuitive comparison of the co-segmentation results with and without the higher-order energy of our algorithm. Our algorithm can be divided into two stages, likelihood estimation and co-segmentation using higher-order energy function. The initial co-segmentation results by our likelihood estimation stage are shown in Fig. 7(c) where the estimated likelihood maps are also shown in Fig. 7(b). A pixel x_k^i belongs to foreground when its foreground likelihood $\pi_{k,l}^i > 0.5$. As shown in Fig. 7(b) and (c), the initial co-segmentation results by our likelihood estimation only extract the rough foreground objects of panda where the likelihood estimation is not correct in many regions. These initial co-segmentation results are greatly improved and then more

Fig. 7. Improvement for our method through higher-order cliques. (a) The input images; (b) the map of likelihood estimation (in this map, the more white colors that a region has, the higher possibility that the region belongs to foreground); (c) co-segmentation results by our likelihood estimation; (d) co-segmentation results by our method without higher-order cliques energy, which means we only use the unary item and pairwise item without higher-order item of (9); (e) co-segmentation results by our full method with higher-order energy optimization; and (f) the ground truth masks.

accurate co-segmentation results [see Fig. 7(e)] are obtained after we perform our higher-order energy optimization. The estimated likelihood maps are also treated as the prior knowledge in our higher-order optimization process. To see the merit of our higher-order energy item $E_{\text{global}}(I^1, \ldots, I^n, T)$ in (9), we give the comparisons of co-segmentation results with and without our higher-order energy item in Fig. 7(d) and (e). The same unary term E_{unary} and pairwise term E_{pairwise} in (8) are used to produce the comparison results [see Fig. 7(d) and (e)]. The segmentation results consistently demonstrate that our higher-order cliques are helpful in producing high-quality segmentations.

C. Run-Time Statistics

We compare the runtime with ICOSEG [21] and RWCS [28] under the same computer configuration: Intel Xeon E5-2609 @2.50GHz with 64GB RAM. Table II reports the average running time on iCoseg dataset [21] and MSRC dataset [30]. Our approach is faster than the other two methods. We can observe that with an increase in the number of images in the group, the running time of our method only increases linearly. We further analyze the computational complexity of our algorithm.

TABLE II
AVERAGE RUNNING TIME (SECOND) ON ICOSEG (TOP 10 ROWS) AND MSRC (BOTTOM 10 ROWS)

Datasets	Images	Ours	ICOSEG[21]	RWCS[28]
HotBalloons	24	1056.8	2753.2	10 845.3
Kendo	30	1213.8	3087.5	13 361.4
Monks	17	748.2	1047.0	7023.9
StatueofLiberty	41	1745.3	3236.0	23 316.1
Windmill	18	812.1	1165.0	10 530.5
Bear	19	867.4	1391.2	7326.9
Cheetah	33	1211.30	2315.1	14 218.3
Goose	31	1136.4	1865.8	13 350.7
Panda	24	1004.5	1875.7	11 483.6
Stonehenge	18	785.4	1176.8	9971.0
bike	30	312.8	2864.6	3476.3
bird	32	349.2	2671.3	3796.2
car	30	310.7	2130.0	3243.9
cat	24	243.9	1582.6	2932.7
chair	30	363.4	1654.3	3399.6
cow	30	369.5	2129.2	3221.4
dog	30	324.7	2215.5	3648.6
flower	32	383.1	4379.6	3723.1
house	30	370.3	1799.3	3408.1
sheep	30	318.5	1895.7	3038.8

The complexity of our likelihood estimation process is about $O(\sum_{i=1}^{n}[N(R^i)]^2)$, where $N(R^i)$ indicates the number of regions in R^i. Our higher-order energy function can be solved for each image individually and the higher-order cliques are optimized as a second-order function which can be solved by the conventional graph cut algorithm. Therefore, the complexity of higher-order co-segmentation step is about $O(\sum_{i=1}^{n}[N(I^i)]^2)$, where $N(I^i)$ is the number of pixels in I^i. The complexity of our full co-segmentation algorithm is about $O(\sum_{i=1}^{n}[N(I^i)]^2)$, since $N(I^i) \gg N(R^i)$. Therefore, the run-time of our method increases only linearly with additional images.

IV. Conclusion

We have presented a novel interactive co-segmentation approach using the likelihood estimation and high-order energy optimization to extract the complicated foreground objects from a group of related images. A likelihood estimation method is developed to compute the prior knowledge for our higher-order co-segmentation energy function. Our higher-order cliques are built on a set of foreground and background regions obtained by likelihood estimation. Then our co-segmentation process from a group of images is performed at the region level through our higher-order cliques energy optimization. The energy function of our higher-order cliques can be further transformed into a second-order boolean function and thus the traditional graph cuts method can be used to solve them exactly.

The experimental results demonstrated both qualitatively and quantitatively that our method has achieved more accurate co-segmentation results than previous unsupervised and interactive co-segmentation methods, even though the foreground and background have many overlap regions in color distributions or in very complex scenes.

Appendix A
Proofs of (16)

Theorem 1: The matching coefficient κ_s^i in (13) can be written as

$$\kappa_s^i = \min\left\{\frac{\sum_{x_s^i \in r_s^i}\phi(x_s^i)}{Q}(1-t)+t, \frac{N(r_s^i)-\sum_{x_s^i\in r_s^i}\phi(x_s^i)}{Q}t+(1-t), 1\right\} \quad (18)$$

where $t = z^1(\nu_s^i)$, $Q = \rho N(r_s^i)$, and x_s^i indicates the pixels belonging to region r_s^i.

Proof: From (14), we have the property as

$$z^0(\nu_s^i) + z^1(\nu_s^i) = 1. \quad (19)$$

Considering the definition of $N_l(r_s^i)$ in (15), we have

$$\begin{aligned} N_0(r_s^i) + N_1(r_s^i) &= N(r_s^i) \\ N_1(r_s^i) &= \sum_{x_s^i \in r_s^i}\phi(x_s^i) \end{aligned} \quad (20)$$

where $N_l(r_s^i)$ is the number of pixels assigned to foreground/background in region r_s^i.

Considering (19) and (20) and the definition of matching coefficient κ_s^i in (13), we then obtain the solution as

$$\kappa_s^i = \min\left\{\frac{\sum_{x_s^i \in r_s^i}\phi(x_s^i)}{Q}(1-t)+t, \frac{N(r_s^i)-\sum_{x_s^i\in r_s^i}\phi(x_s^i)}{Q}t+(1-t), 1\right\}$$

where $t = z^1(\nu_s^i)$, $Q = \rho N(r_s^i)$, and x_s^i indicates the pixels belonging to region r_s^i.

Appendix B
Proofs of (17)

Theorem 2: The matching coefficient κ_s^i in (16) can be transformed into a second-order function by introducing binary variables σ_0 and σ_1

$$\begin{aligned} \kappa_s^i &= \min \Psi(\sigma_0, \sigma_1, \phi(x_s^i)) \\ &= \min_{\sigma_0,\sigma_1} \sigma_0\frac{\sum_{x_s^i\in r_s^i}\phi(x_s^i)}{Q}(1-t) + (1-\sigma_0)(1-t) \\ &\quad +\sigma_1\frac{N(r_s^i)-\sum_{x_s^i\in r_s^i}\phi(x_s^i)}{Q}t + (1-\sigma_1)t. \end{aligned} \quad (21)$$

Proof: When binary variables σ_0 and σ_1 take all the possible values, there are 2^2 solutions and we can rewrite the above function $\Psi(\sigma_0, \sigma_1, \phi(x_s^i))$ as
$\Psi(\sigma_0, \sigma_1, \phi(x_s^i)) =$

$$\begin{cases} \frac{\sum\phi(x_s^i)}{Q}(1-t)+t & \text{if } \sigma_0=1, \sigma_1=0 \\ \frac{N(r_s^i)-\sum\phi(x_s^i)}{Q}t+(1-t) & \text{if } \sigma_0=0, \sigma_1=1 \\ \frac{\sum\phi(x_s^i)}{Q}(1-t)+\frac{N(r_s^i)-\sum\phi(x_s^i)}{Q}t & \text{if } \sigma_0=1, \sigma_1=1 \\ 1 & \text{if } \sigma_0=0, \sigma_1=0. \end{cases} \quad (22)$$

When $\sum_{x_s^i\in r_s^i}\phi(x_s^i) > Q$, we can get

$$\frac{\sum\phi(x_s^i)}{Q}(1-t)+\frac{N(r_s^i)-\sum\phi(x_s^i)}{Q}t > \frac{\sum\phi(x_s^i)}{Q}(1-t)+t. \quad (23)$$

When $\sum_{x_s^i\in r_s^i}\phi(x_s^i) < N(r_s^i) - Q$, we will have the following property:

$$\frac{\sum\phi(x_s^i)}{Q}(1-t)+\frac{N(r_s^i)-\sum\phi(x_s^i)}{Q}t > \frac{N(r_s^i)-\sum\phi(x_s^i)}{Q}t+(1-t). \quad (24)$$

Combining (22), (23) and (24), the matching coefficient κ_s^i in (16) can be transformed into the following second-order energy

function:

$$\begin{aligned}\kappa_s^i &= \min \Psi(\sigma_0, \sigma_1, \phi(x_s^i)) \\ &= \min_{\sigma_0, \sigma_1} \sigma_0 \frac{\sum_{x_s^i \in r_s^i} \phi(x_s^i)}{Q}(1-t) + (1-\sigma_0)(1-t) \\ &\quad + \sigma_1 \frac{N(r_s^i) - \sum_{x_s^i \in r_s^i} \phi(x_s^i)}{Q} t + (1-\sigma_1)t. \end{aligned} \tag{25}$$

REFERENCES

[1] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[2] Z. Lou and T. Gevers, "Extracting primary objects by video co-segmentation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2110–2117, Dec. 2014.

[3] C. Wang, Y. Guo, J. Zhu, L. Wang, and W. Wang, "Video object co-segmentation via subspace clustering and quadratic pseudo-boolean optimization in an MRF framework," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 903–916, Jun. 2014.

[4] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[5] C. Rother, T. Minka, A. Blake, and V. Kolmogorov, "Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2006, pp. 993–1000.

[6] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.

[7] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, 2006.

[8] L. Mukherjee, V. Singh, and C. R. Dyer, "Half-integrality based algorithms for cosegmentation of images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 2028–2035.

[9] W. Wang, J. Shen, and L. Shao, "Consistent video saliency using local gradient flow optimization and global refinement," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4185–4196, Nov. 2015.

[10] D. S. Hochbaum and V. Singh, "An efficient algorithm for cosegmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep.-Oct. 2009, pp. 269–276.

[11] H. Fu, D. Xu, S. Lin, and J. Liu, "Object-based RGBD image co-segmentation with Mutex constraint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 4428–4436.

[12] P. Kohli, L. Ladicky, and P. Torr, "Robust higher order potentials for enforcing label consistency," *Int. J. Comput. Vis.*, vol. 82, no. 3, pp. 302–324, 2009.

[13] W. Wang, J. Shen, X. Li, and F. Porikli, "Robust video object co-segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3137–3148, Oct. 2015.

[14] I. Hiroshi, "Higher-order clique reduction in binary graph cut," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 2993–3000.

[15] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 3129–3136.

[16] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 1943–1950.

[17] X. Dong, J. Shen, L. Shao, and M. H. Yang, "Interactive co-segmentation using global and local energy optimization," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3966–3977, Nov. 2015.

[18] S. Vicente, V. Kolmogorov, and C. Rother, "Cosegmentation revisited: Models and optimization," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 465–479.

[19] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "iCoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 3169–3176.

[20] T. H. Kim, K. M. Lee, and S. U. Lee. "Nonparametric higher-order learning for interactive segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 3201–3208.

[21] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "Interactively co-segmentating topically related images with intelligent scribble guidance," *Int. J. Comput. Vis.*, vol. 93, pp. 273–292, 2011.

[22] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade, "Distributed cosegmentation via submodular optimization on anisotropic diffusion," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 169–176.

[23] L. Mukherjee, V. Singh, and J. Peng, "Scale invariant cosegmentation for image groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 1881–1888.

[24] K. Chang, T. Liu, and S. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 2129–2136.

[25] A. C. Gallagher, D. Batra, and D. Parikh, "Inference for order reduction in Markov random fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 1857–1864.

[26] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, "Multi-task low-rank affinity pursuit for image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2439–2446.

[27] A. Joulin, F. Bach, and J. Ponce, "Multi-class cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 542–549.

[28] M. D. Collins, J. Xu, L. Grady, and V. Singh, "Random walks based multi-image segmentation: Quasiconvexity results and GPU-based solutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1656–1663.

[29] Y. Chai, V. Lempitsky, and A. Zisserman, "BiCoS: A bi-level co-segmentation method for image classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2579–2586.

[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 *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 1–15.

[31] H. Ishikawa, "Higher-order vlique reduction in binary graph cut," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 2993–3000.

[32] H. Ishikawa, "Transformation of general binary MRF minimization to the first order case," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1234–1249, Jun. 2011.

[33] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep.-Oct. 2009, pp. 731–738.

[34] K. Park and S. Gould, "On learning higher-order consistency potentials for multi-class pixel labeling," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 202–215.

[35] J. Rubio, J. Serrat, A. Lopez, and N. Paragios, "Unsupervised co-segmentation through region matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 749–756.

[36] F. Wang, Q. Huang, and L. Guibas, "Image co-segmentation via consistent functional maps," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 849–856.

[37] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, "Unsupervised joint object discovery and segmentation in internet images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 1939–1946.

[38] F. Wang, Q. Huang, M. Ovsjanikov, and L. J. Guibas, "Unsupervised multi-class joint image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 3142–3149.

[39] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.

[40] C. Lee, W.-D. Jang, J.-Y. Sim, and C.-S. Kim, "Multiple random walkers and their application to image cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 3837–3845.

[41] H. Zhu, J. Lu, J. Cai, J. Zheng, and N. Thalmann, "Multiple foreground recognition and cosegmentation: An object-oriented CRF model with robust higher-order potentials," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2014, pp. 485–492.

[42] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3766–3778, Oct. 2013.

[43] X. Dong, J. Shen, and L. Shao, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.

[44] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-adaptively weighted co-saliency detection via rank constraint," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4175–4186, Sep. 2014.