

Image Cosegmentation by Incorporating Color Reward Strategy and Active Contour Model

Fanman Meng, Hongliang Li, *Senior Member, IEEE*, Guanghui Liu, and King Ngi Ngan, *Fellow, IEEE*

Abstract—The design of robust and efficient cosegmentation algorithms is challenging because of the variety and complexity of the objects and images. In this paper, we propose a new cosegmentation model by incorporating a color reward strategy and an active contour model. A new energy function corresponding to the curve is first generated with two considerations: the foreground similarity between the image pairs and the background consistency in each of the image pair. Furthermore, a new foreground similarity measurement based on the rewarding strategy is proposed. Then, we minimize the energy function value via a mutual procedure which uses dynamic priors to mutually evolve the curves. The proposed method is evaluated on many images from commonly used databases. The experimental results demonstrate that the proposed model can efficiently segment the common objects from the image pairs with generally lower error rate than many existing and conventional cosegmentation methods.

Index Terms—Active contour model, cosegmentation, reward strategy.

I. INTRODUCTION

DURING the past decades, the method of single-image-based segmentation has been developed in a variety of directions. According to the necessity of the object prior, we can classify the directions into two classes, i.e., unsupervised segmentation method and supervised segmentation method. The unsupervised segmentation method aims to segment the image into many uniform and homogeneous regions with respect to texture or color properties [1]–[7]. The supervised segmentation method segments the object using object prior. Compared with the unsupervised segmentation method, the supervised segmentation method can achieve semantic segmentation depending on whether the object prior is precisely modeled or not. To generate precise object prior, several object prior generation methods were proposed in the previous work, such as saliency-detection-based methods [8]–[12], learning-

Manuscript received February 2, 2012; revised August 2, 2012; accepted August 11, 2012. Date of publication September 14, 2012; date of current version April 16, 2013. This work was supported in part by the National Natural Science Foundation of China under Grants 60972109, 61101091 and 61271289, by the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 20110185110002, and by the Fundamental Research Funds for the Central Universities under Grant E022050205. This paper was recommended by Associate Editor I. Bloch.

F. Meng, H. Li, and G. Liu are with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610073, China (e-mail: hlli@uestc.edu.cn).

K. N. Ngan is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: knngan@ee.cuhk.edu.hk).

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/TSMCB.2012.2215316

based methods [13]–[15], and human-interaction-based methods [16]–[18]. Although accurate segmentation is achieved by the supervised segmentation method, it may be hard to generate the corresponding object prior for complex cases. Furthermore, the huge workload will be produced for the user when the number of the target images is high.

Recently, segmenting objects from multiple images has attracted researchers' attention. Generally, the multiple-image-based segmentation method first provides several other images containing the object and then segments the objects by extracting common objects among the images. Compared with the single-image-based segmentation method, the common-object segmentation method is convenient in terms of user interaction and can achieve accurate segmentation results. In this paper, we restrict the focus to segment common objects from a pair of images.

In recent years, extracting common objects from a group of images has become an active research topic. A number of solutions were proposed to solve common-object segmentation problem, such as matching-based method, cosegmentation method, and mutual segmentation method. The matching-based method considers the common-object segmentation as a matching task which generally consists of two steps: 1) locate the common objects among original images and 2) segment the common objects based on the matching results. Since the successful segmentation for the matching-based method depends on the precise location of the common object, the well-known local region descriptors were employed for the locating, such as scale-invariant feature transform (SIFT) descriptor [19], saliency descriptor [20] and Harris-affine interest operator [21]. Other methods, such as conditional random field method and latent Dirichlet allocation (LDA)-based topic discovery method, were also introduced for common-object locating. In [22], SIFT feature was used for the common-object locating. Then, the common objects were segmented by the corresponding growing method. In [23], conditional random field method was used to adapt generic knowledge to the new class. The obtained new knowledge was used to locate and segment the common objects. LDA was used to extract common objects in [24], which performed topic discovery on each image and segmented the common objects based on how well they were explained by the topic. In [25], Harris-affine interest operator and SIFT descriptor were extracted to locate the common objects. Then, the common objects were extracted by the spectral clustering method.

Unlike the aforementioned methods where common objects are segmented based on two steps, cosegmentation [26] simultaneously segments the common objects from a pair of images.

The idea of the cosegmentation is to consider an additional foreground similarity constraint on single-image-based segmentation method to result in the segmentation of the common objects. Several cosegmentation methods were proposed, such as Markov random field (MRF)-based cosegmentation method [26]–[30] and discriminative-clustering-based cosegmentation method [31].

The MRF-based cosegmentation method adds the constraint of the foreground similarity into traditional MRF-based segmentation model to simultaneously consider common-object segmentation and single image segmentation. The key of the MRF-based cosegmentation method is how to measure foreground similarity. Furthermore, since the foreground similarity measurement affects the optimization of the energy function, designing efficient optimization method is another key problem. In the existing MRF-based cosegmentation method, similarity measurements such as $L1$ -norm [26], $L2$ -norm [27], reward model [28], and Boykov–Jolly model [29], [32] were proposed. The corresponding optimization methods, namely, trust-region graph-cut method, quadratic pseudo-Boolean optimization method, maximum flow procedure of graph, and dual decomposition, were also proposed for optimization of the energy function.

Apart from MRF-based cosegmentation method, discriminative clustering and spectral clustering methods were combined to segment common objects in [31], which used spectral clustering technique and positive definite kernels to train a classifier for the common-object classification. In [32], an interactive cosegmentation method was proposed. The authors used human interaction guided by an automatic recommendation system to achieve more accurate common-object segmentation. An object cosegmentation method was proposed in [33]. The authors first used the unsupervised segmentation method to segment the images into a series of local regions. Then, the common objects were selected from the local regions by A^* -search algorithm. Since the common objects can be similar to each other on many features, random forest regressor method was used to select the useful features from a total of 33 features (such as color histogram, textons, SIFT descriptors, boundary curvature, and area). The authors in [34] presented a scale-invariant cosegmentation method by considering the scale variation among the common objects. The fact that the matrix composed of the common objects should have a rank of one was used for modeling the energy function, and the common objects were finally segmented through searching image pixels that made the rank of corresponding matrix be equal to one. The algorithm in [35] solved cosegmentation by generating a novel global energy term which considered the foreground similarity and the background dissimilarity. Meanwhile, a cosaliency map was employed in [35] to construct the unary term of the energy function.

Similar to cosegmentation, the mutual segmentation algorithm given by Riklin-Raviv *et al.* in [36] and [37] simultaneously locates and segments common objects from a pair of images. In mutual segmentation, common objects are extracted by a mutually supporting evolution process, in which the information gained in the evolving segmentation of one image is a dynamic prior for the other. Thus, the segmentation and

the prior information are refined along the process [36]. In [36], the shape feature similarity between the foregrounds was represented by a homograph matrix transform which is robust to the shape variations such as occlusion, noising, and rotation. The significant improvement of the segmentation results was gained.

In this paper, we propose a new cosegmentation model which integrates active contours method and rewarding strategy. We first generate a new energy function with two conflicting goals. One is foreground similarity among the images, which is related to common-object segmentation. The other is background consistency in each image, which corresponds to single image segmentation. We use a color histogram for region representation and employ a rewarding strategy to measure the foreground similarity and the background consistency. Then, we formulate the energy function by level set and use a mutual evolution approach to minimize the energy function value. We evaluate our approach on many images. The experimental results demonstrate the effectiveness of the proposed model.

The contributions of this paper are listed as follows.

- 1) An active-contour-based model is proposed for the cosegmentation problem. The proposed model can segment common objects from a pair of images.
- 2) A new reward-strategy-based foreground similarity measurement is proposed to evaluate foreground similarity, which makes the proposed energy function easy for optimization (be optimized by the method of calculus of variations). In addition, a rewarding strategy considering both the foreground reward and the background reward is proposed.

The rest of this paper is organized as follows. The background including the rewarding strategy and the active contour model is introduced in Section II. In Section III, we present our proposed cosegmentation algorithm. Experimental results are provided in Section IV to validate the efficiency of the proposed model. Finally, Section V gives the conclusion of this paper.

II. BACKGROUND

Our method is related to two previous methods, i.e., the rewarding strategy and the active contour model. In this section, we introduce the background, advantages, and challenges of the two types of methods.

A. Rewarding Strategy

Unlike traditional single image segmentation, cosegmentation [26] aims at simultaneously segmenting common objects from a group of images. This method can segment multiple images jointly instead of segmenting each image independently based on the co-occurrence of objects in the images [38]. It has many potential applications in computer vision, such as image classification [39], object recognition [40], and image retrieval [33].

To achieve common-object segmentation, the cosegmentation is usually formulated as an optimization problem which consists of two terms. One is the single image segmentation term which considers the smoothness of the segmentation in

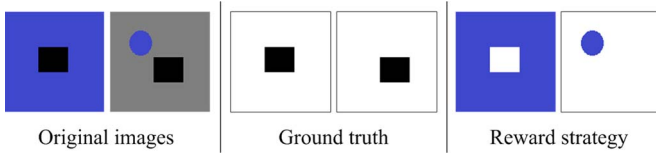


Fig. 1. Example to illustrate the rewarding strategy.

each image. The other is the common-object segmentation term which focuses on extracting similar regions as foregrounds. The key of the cosegmentation model is how to design efficient foreground similarity measurement to form common-object segmentation term.

In the existing cosegmentation approaches, a penalizing strategy is usually used to measure foreground similarity. The penalizing strategy measures foreground similarity by penalizing the differences between the foregrounds, which leads to a small difference between the features of the segmented objects. For example, in Rother's method [26], the similarity between two regions is evaluated by $L1$ -norm $\sum_l |h_1(b) - h_2(b)|$, where h_1 and h_2 are the features of the two regions with the length l . A small $L1$ -norm value corresponds to large similarity. It is seen from the measurement that the similarity measurement is constructed by the differences between the features. A large difference is penalized. Penalizing-strategy-based methods can achieve accurate common-object segmentation results. However, the penalizing-strategy-based foreground similarity measurement leads to nondeterministic-polynomial-time-hard optimization problem.

Instead of penalizing the foreground difference, the rewarding strategy in [28] rewards the similarities between the foregrounds. For example, the similarity between the foregrounds is measured by $\sum_l |h_1(b) \cdot h_2(b)|$, which rewards the similarity instead of penalizing the differences. Compared with the penalizing strategy, the rewarding strategy is advantageous in its effective solution for the energy optimization, since the rewarding-strategy-based foreground similarity measurement leads to simpler optimization problem, such as submodular problem that can be efficiently optimized with graph cuts.

However, the rewarding strategy can lead to unexpected results in the case where the background with uniform color occupies a large area of one image while there are a few regions that are similar to the background in the other image [29]. In this case, the rewarding strategy will classify the background as the common objects, because the reward of background region is larger than the reward of foreground region. As an example, Fig. 1 shows the original images, ground truth masks, and the results of the reward strategy. It can be seen in Fig. 1 that the common objects are the black regions and the blue regions are shared by the two images. Since the blue region takes a big part of the first image, the reward of the blue region will be larger than the reward of the black region. Hence, the final results by the rewarding strategy are the blue regions instead of the black regions.

B. Region-Based Active Contour Model

In active-contour-based segmentation method, the energy function is related to a curve which corresponds to the min-

imum value of the energy function, and the curve serves as the boundary of the object to be segmented. Then, the corresponding curve decreases the energy criterion and converges toward a (local) minimum of the criterion by solving the partial differential equation (PDE) constructed from the energy function. The existing active-contour-based segmentation methods can be classified into two categories, i.e., geometric-based active contour method (geometric-based method) [41]–[45] and region-based active contour method (region-based method) [46]–[51]. The geometric-based methods are based on an energy defined as a curve integral, while region methods have region terms in the energy which are integrals on the regions inside the curves. Compared with the geometric-based method, the region-based method is more robust to the initial curve placement.

One of the well-known region-based methods is the C - V model [47]. The major components of the energy function in the C - V model are shown as

$$E(C) = \int_{\omega^i} (I(x, y) - v_i)^2 dx dy + \int_{\omega^o} (I(x, y) - v_e)^2 dx dy \quad (1)$$

where I is an image, $I(x, y)$ denotes the color value in location (x, y) , $E(C)$ is the energy function that corresponds to the curve C , ω^i and ω^o denote the regions inside and outside the curve C , respectively, and v_i and v_e are the mean color values of regions ω^i and ω^o , respectively.

For the right-hand side of (1), the first term represents the foreground consistency, which can be called interior term. The second term, namely, exterior term, measures the background consistency. Based on the assumption that the textures of foreground and background are uniform, such as medical images, the integrals in interior term and exterior term are equal to zero only if the foreground and background are accurately segmented. Therefore, the object is segmented through searching the curve corresponding to the smallest energy value.

Note that the assumption of the C - V model that the textures of foreground and background are uniform may not be used for natural images, where the foreground and background may be composed of multiple colors. This means that the C - V model may not be suitable to segment objects from natural images. The reason is that the mean value is no longer suitable for representing a region composed of multiple colors. Furthermore, the object information provided by a single image is not sufficient to distinguish foreground and background from an image in the case of complex foreground and background. Thus, segmenting object from the image with complex foreground and background is a challenge for the C - V model.

III. PROPOSED METHOD

In this section, we introduce our proposed model. The energy function is first introduced. Then, the level set formulation and the optimization are illustrated.

A. Proposed Model

Different from the traditional region-based method, two images are considered in the cosegmentation problem. Thus,

the constraint of foregrounds similarity is considered in the proposed model, which is performed at the region level.

1) *Energy Function*: We form the energy function for curve C_k as

$$E_k(C_k) = \mu \cdot \text{Length}(C_k) + \nu \cdot \text{Area}(\omega_k^i) - \lambda_k^i \int_{\omega_k^i} f[I_k(x, y), g(\omega_{1-k}^i)] dx dy - \lambda_k^o \int_{\omega_k^o} f[I_k(x, y), g(\omega_k^o)] dx dy \quad (2)$$

where I_k , $k \in \{0, 1\}$, is the k th image of the image pair, C_k denotes the curve in the image I_k , $E_k(C_k)$ is the energy function corresponding to C_k , ω_k^i and ω_k^o depict the regions inside and outside the curve C_k , $g(\omega)$ is the representation of the region ω , $f(p, g(\omega))$ is used to measure the similarity between pixel p and region ω represented by $g(\omega)$, $\text{Length}(C_k)$ denotes the length of the curve C_k , $\text{Area}(\cdot)$ denotes the area of a region, and ω_{1-k}^i is the region inside the curve in the other image I_{1-k} . $\mu > 0$, $\nu > 0$, $\lambda_k^i > 0$, and $\lambda_k^o > 0$ are parameters.

At the right-hand side of (2), the first term and the second term describe the length and the area of the region inside the curve, respectively. The two terms are related to curve intrinsic properties. The third term, namely, interior term, is composed of the integral of similarities between the foreground pixels of I_k and the region ω_{1-k}^i in I_{1-k} . The interior term describes the foreground similarity. The last term is the exterior term which is the integral of similarities between the background pixels in I_k and the exterior region ω_k^o . The exterior term represents background consistency. It is seen from (2) that the major components of the energy function are composed of an interior term and an exterior term.

2) *Region Representation*: In the proposed model, we use a color histogram h to represent a region ω , i.e., $g(\omega) = h$. For a red (R)–green (G)–blue (B) image, color value is represented by a 3-D vector. A histogram is generated by considering the probability of each color's occurrence in a region. Since a region feature is used for region representation, our method is a region-based method.

3) *Foreground Similarity Measurement*: We set foreground similarity measurement f as $f(p, g(\omega)) = h(p)$ for a pixel value p . A large foreground similarity value will be given if a pixel value is with a large histogram value. Otherwise, a small foreground similarity value is given. In practice, the pixel values are commonly slightly different from each other for a color. Thus, two pixels will have the same color if their color distance is less than β . For a pixel value p , $f(p, g(\omega))$ is then given by

$$f(p, g(\omega)) = \sum_{|p' - p| \leq \beta} h(p') \quad (3)$$

where $|p' - p|$ denotes the distance between two pixel color values p' and p . Furthermore, $|p' - p|$ defines the distance as the largest difference among the differences in R, G, and B color channels.

4) *Analysis of Energy Function*: To analyze the proposed energy function, we assume that only the common objects in two images have the same colors. In addition, most of the areas of the common objects locate within the curves. For a pixel located in the common object in image I_k , we can see that $f[I_k(i, j), g(\omega_{1-k}^i)] > f[I_k(i, j), g(\omega_k^o)]$, where (i, j) denotes the pixel location. On the contrary, we obtain $f[I_k(i, j), g(\omega_{1-k}^i)] < f[I_k(i, j), g(\omega_k^o)]$ for a background pixel. Hence, the curve energy is minimum when the curve accurately separates common objects from images. Otherwise, the curve energy increases. For I_k , the common-object segmentation thus can be depicted as

$$C_k^* = \arg \min_{C_k} E_k(C_k). \quad (4)$$

It is seen from (2) that foreground similarity is designed by the reward strategy, i.e., $\int_{\omega_k^i} f[I_k(x, y), g(\omega_{1-k}^i)] dx dy$. Furthermore, the background is also designed by the reward strategy, i.e., $\int_{\omega_k^o} f[I_k(x, y), g(\omega_k^o)] dx dy$. Hence, rewarding strategy is employed for region similarity measurement. Compared with the reward strategy used in [28], we can see that both foreground similarity and background consistency are rewarded, which improves the traditional rewarding strategy. The improvement is resulted by the fact that, for an image pair where the background with uniform color occupies a large area of one image while there are a few regions that are similar to the background in the other image, a pixel of background with large foreground reward $f[I_k(x, y), g(\omega_{1-k}^i)]$ also has a large background reward $f[I_k(x, y), g(\omega_k^o)]$, which is in contrast with foreground reward. When background reward is larger than foreground reward, i.e., $f[I_k(x, y), g(\omega_k^o)] > f[I_k(x, y), g(\omega_{1-k}^i)]$, the pixel will be considered as background instead of foreground. Thus, the results of traditional rewarding strategy are improved.

B. Level Set Formulation

By manipulating (2), the level-set-based energy function of I_k is expressed as

$$E_k(\phi_k) = \mu \int_{\Omega_k} \delta(\phi_k(x, y)) |\nabla \phi_k(x, y)| dx dy + \nu \int_{\Omega_k} H(\phi_k(x, y)) dx dy - \lambda_k^i \int_{\Omega_k} f[I_k(x, y), g(\omega_{1-k}^i)] H(\phi_k(x, y)) dx dy - \lambda_k^o \int_{\Omega_k} f[I_k(x, y), g(\omega_k^o)] (1 - H(\phi_k(x, y))) dx dy \quad (5)$$

where ϕ_k denotes the zero level set function, $\delta(x)$ is the 1-D Dirac function, $H(x)$ is the Heaviside function, $\delta(\phi_k(x, y))$ represents C_k , $H(\phi_k(x, y))$ denotes the interior region ω_k^i , and $1 - H(\phi_k(x, y))$ means exterior region ω_k^o .

C. Energy Function Minimization

To minimize the proposed energy function, we derive the corresponding Euler–Lagrange equation. For the last term in (5), when $f[I_k(x, y), g(\omega_k^o)]$ is dependent on ϕ_k , the Euler–Lagrange equation is represented as

$$-\lambda_k^o \cdot \frac{\partial f}{\partial \phi_k} (1 - H(\phi_k)) + \lambda_k^o \cdot \delta(\phi_k) f[I_k, g(\omega_k^o)] = 0 \quad (6)$$

where the term $-\lambda_k^o \cdot (\partial f / \partial \phi_k)(1 - H(\phi_k))$ is related to the shape derivative as shown in [44]. In this paper, we use the strategy in [47] to keep $g(\omega_{1-k}^i)$ and $g(\omega_k^o)$ fixed, which means that $f[I_k(x, y), g(\omega_k^o)]$ is independent of ϕ_k . Hence, $\partial f / \partial \phi_k = 0$, and we can rewrite (6) as

$$\lambda_k^o \cdot \delta(\phi_k) f[I_k, g(\omega_k^o)] = 0. \quad (7)$$

Based on (7), the Euler–Lagrange equation of (5) for ϕ_k is shown as

$$\frac{\partial \phi_k}{\partial t} = \delta(\phi_k) \cdot \left\{ \mu \cdot \mathbf{div} \left(\frac{\nabla \phi_k}{|\nabla \phi_k|} \right) - \nu + \lambda_k^i \cdot f[I_k, g(\omega_{1-k}^i)] - \lambda_k^o \cdot f[I_k, g(\omega_k^o)] \right\} = 0 \quad (8)$$

where $t \geq 0$ is an artificial time in $\phi_k(t, x, y)$. Since curvature κ can be calculated by $\kappa = -\mathbf{div}(\nabla \phi / |\nabla \phi|)$ in the differential geometry theory, we may rewrite (8) as

$$\frac{\partial \phi_k}{\partial t} = \delta(\phi_k) \cdot \left\{ -\mu \cdot \kappa - \nu + \lambda_k^i \cdot f[I_k, g(\omega_{1-k}^i)] - \lambda_k^o \cdot f[I_k, g(\omega_k^o)] \right\} = 0. \quad (9)$$

For (9), the corresponding discrete form by explicit scheme can be expressed by

$$\frac{\phi_k^{n+1}(i, j) - \phi_k^n(i, j)}{\Delta t} = \delta(\phi_k^n(i, j)) \cdot \left\{ -\mu \cdot \kappa(i, j) - \nu + \lambda_k^i \cdot f[I_k(i, j), g(\omega_{1-k}^i)] - \lambda_k^o \cdot f[I_k(i, j), g(\omega_k^o)] \right\} \quad (10)$$

where ϕ_k^0 is the initial curve. Setting $\Delta t = 1$, the final evolving form of our model is shown as

$$\begin{aligned} \phi_k^{n+1}(i, j) &= \phi_k^n(i, j) + \delta(\phi_k^n(i, j)) \\ &\cdot \left[-\mu \cdot \kappa(i, j) - \nu + \lambda_k^i \cdot f[I_k(i, j), g(\omega_{1-k}^i)] \right. \\ &\quad \left. - \lambda_k^o \cdot f[I_k(i, j), g(\omega_k^o)] \right]. \end{aligned} \quad (11)$$

Equation (11) is used to minimize the energy function value. In the minimization process, the value $g(\omega_{1-k}^i)$ can be fixed (fixed method) or updated (dynamic method) along the iteration. For the fixed method, $g(\omega_{1-k}^i)$ is calculated using the whole image of I_{1-k} . Thus, $g(\omega_{1-k}^i)$ is a constant value. For the dynamic method, $g(\omega_{1-k}^i)$ is dynamically updated along the evolution of C_{1-k} . Compared with the dynamic method, the fixed method has lower computational cost since there are no additional operators to calculate $g(\omega_{1-k}^i)$ in each iteration. Meanwhile, for the dynamic method, the curves tend to evolve to the boundary of common objects along the iterations of

$\phi_k^n, n \in \mathcal{Z}$, which can provide more accurate common-object information in the minimization process. Hence, the dynamic method usually achieves good performance for the image cosegmentation compared with the fixed method. In this paper, we use the dynamic method for common-object segmentation.

The steps of the proposed algorithm is described in Algorithm (1).

Algorithm 1 The proposed algorithm

- Set initial curves $\phi_k^0, k = 0, 1, n = 0$, and initial parameters.
 - Repeat
 - 1) For each image $I_k, k = 0, 1$, calculate $g(\omega_{1-k}^i)$ and $g(\omega_k^o)$.
 - 2) Based on the obtained $g(\omega_{1-k}^i)$ and $g(\omega_k^o)$, solve PDE in ϕ through (11) to obtain $\phi_k^{n+1}, k = 0, 1$.
 - Until stop condition is satisfied.
 - Extract common objects: The locations with $\phi_k^{n+1} > 0$ are foreground pixels of I_k .
-

IV. EXPERIMENTS

Based on the subjective and objective assessments of the segmentation results, we verify that the proposed cosegmentation algorithm can achieve more accurate segmentation compared with traditional cosegmentation methods.

A. Parameter Setting

In the experiment, we set the initial curve as a rectangle, where each edge has small distance (γ) to the corresponding image edge. We set $\gamma = 5$ for all test image pairs to cover the most area of the common objects. We normalize κ and f in (11) to [0, 1]. Since the sum of histogram g (corresponding to $\beta = 15$) is used for measuring foreground similarity f , we employ the normalized histogram for g to guarantee that $f[I_k(i, j), g(\omega_{1-k}^i)]$ is in the range of [0, 1]. Furthermore, we set $\mu = 0.01$, $\lambda_k^i = \lambda_k^o = 1$, $\beta = 15$, and $\nu = 0.001$ for all image pairs. The number of the iterations is set to $N_i = 1500$, since there are complex foregrounds and backgrounds in some images.

B. Results of the Proposed Method

We collect 30 image pairs from the data sets commonly used in cosegmentation research and other databases such as Microsoft Research Cambridge image database and the Caltech-256 Object Categories database. The image pairs with complex foregrounds and backgrounds and the image pairs with multiple common objects are selected for the experiment. In addition, the image pairs with similar backgrounds are selected to verify our rewarding strategy. Fig. 2 shows the test image pairs and the ground truth masks (binary masks). It can be seen that some images have complex foreground and background, such as *person* (the fourth image pair in the first row), some image pairs

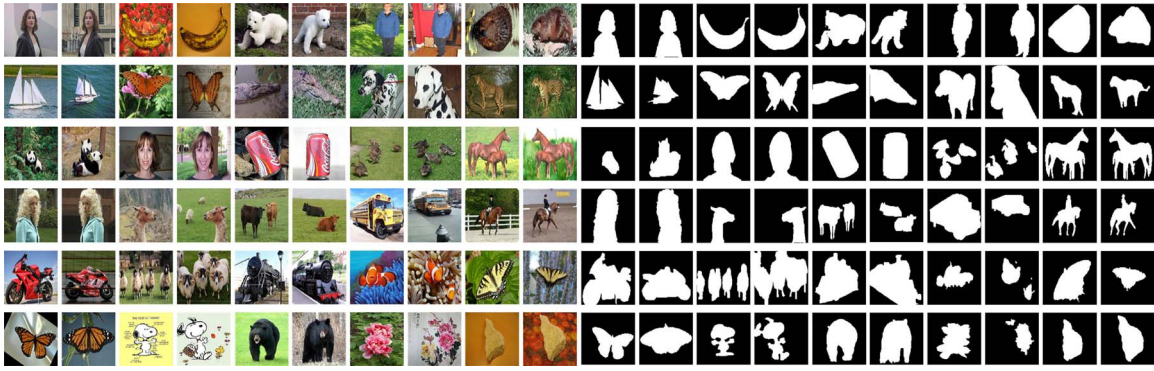


Fig. 2. Original image pairs and the ground truth masks used in this paper.

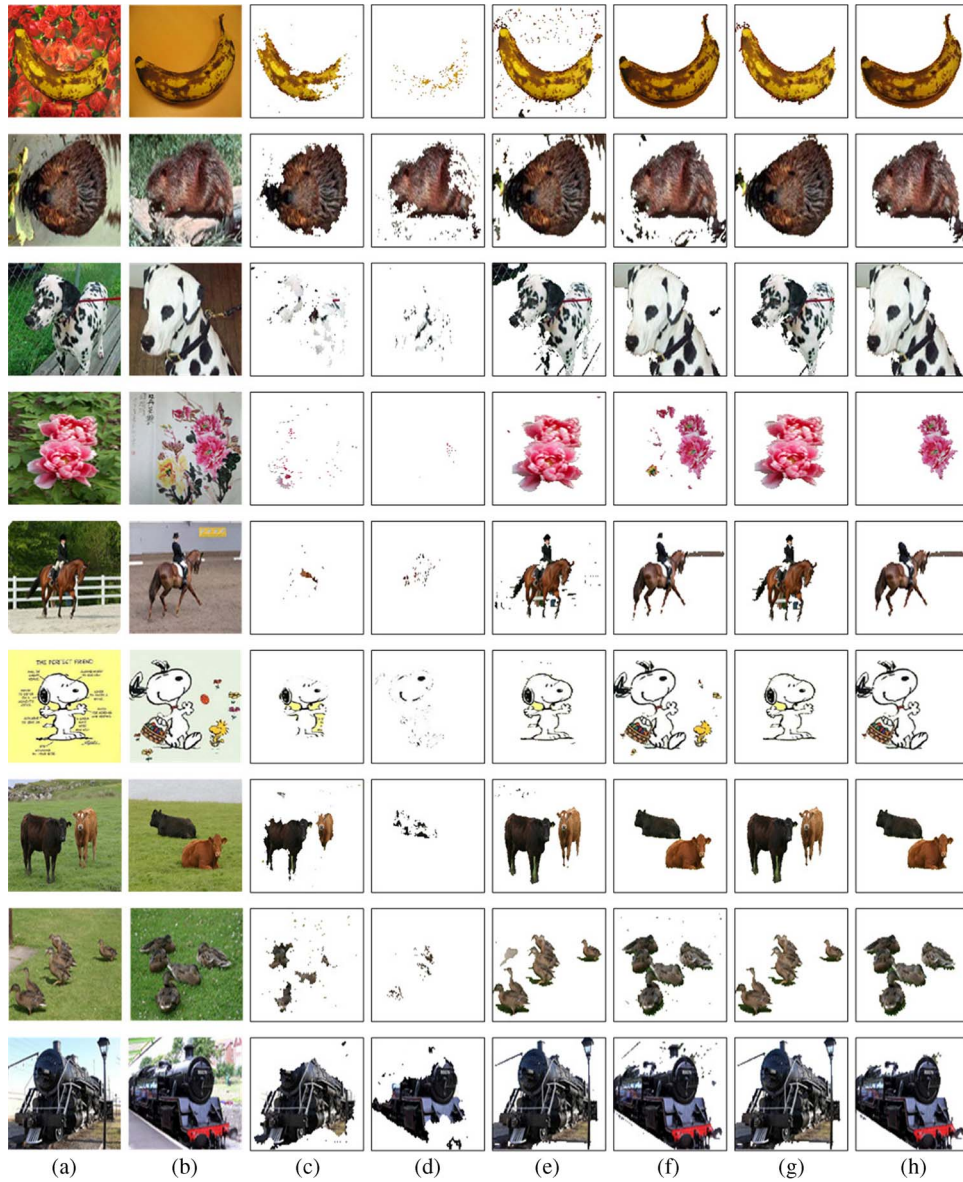


Fig. 3. Segmentation results. (a) and (b) Original image pairs. (c) and (d) Segmentation results by the fixed method. (e) and (f) Segmentation results by the proposed method. (g) and (h) Final segmentation results after postprocessing.

have multiple common objects, such as *duck* (the fourth pair in the third row), and some image pairs have similar backgrounds, such as *mcow* (the third image pair in the fourth row).

The segmentation results for nine image pairs are shown in Fig. 3. The original images are shown in Fig. 3(a) and (b), and the results of our dynamic method are shown in Fig. 3(e) and (f).

TABLE I
RESULTS OF THE FIXED METHOD AND DYNAMIC METHOD AND THE FINAL RESULTS IN TERMS OF ERROR RATE

Name	amira	banana	knut	person	cdbeabera	cdboara	cdbuterflya	cdcora	cddoga	cdleopardb
Fixed Method	0.0570	0.1627	0.3457	0.0792	0.0987	0.1987	0.2265	0.2162	0.4125	0.1046
Dynamic Method	0.1134	0.0425	0.0613	0.1437	0.0820	0.0363	0.0587	0.1821	0.0807	0.1256
Final results	0.0531	0.0374	0.0802	0.0824	0.0492	0.0341	0.0587	0.1690	0.0648	0.1102
Name	cdpandad	cdpersona	coke	duck	horse	kim	llama	mcow	pvoccarb	pvochorseb
Fixed Method	0.1134	0.2929	0.3353	0.1591	0.1245	0.1521	0.0472	0.1191	0.2979	0.1478
Dynamic Method	0.0648	0.1604	0.0224	0.0471	0.0475	0.0212	0.0721	0.0169	0.1235	0.0405
Final Results	0.0290	0.1511	0.0224	0.0405	0.0474	0.0192	0.0316	0.0163	0.0988	0.0383
Name	pvocmotoa	pvocsheepb	pvoctraind	ring008a	ring010a	ring016a	ring017a	ring020a	ring029a	stone
Fixed Method	0.4167	0.3923	0.1920	0.1541	0.0660	0.1817	0.0664	0.3342	0.2045	0.1312
Dynamic Method	0.0665	0.1587	0.1026	0.0261	0.0728	0.0203	0.0591	0.0071	0.0165	0.0137
Final results	0.0654	0.1918	0.1029	0.0265	0.0320	0.0133	0.0426	0.0067	0.0145	0.0106

It can be seen in Fig. 3(e) and (f) that the proposed dynamic method segments common objects from these image pairs. For example, the common objects “dog” in the image pair *dog* with complex texture are segmented from the original image pair (the third image pair in Fig. 3).

To compare with the segmentation results of the proposed dynamic method, Fig. 3(c) and (d) shows the results of the fixed segmentation method with the same parameter setting as in the proposed dynamic method. We can see that the fixed method leads to the loss of some regions of the foreground. For example, the segmentation results of the image pair *banana* (the first image pair) are composed of a number of region fragments. The reason is that the *interior term* of the fixed model is weak. A weak interior term is resulted by the fact that the information of the common objects is interfered by the backgrounds. The proposed dynamic method provides more accurate common-object information than the fixed method, because the curves of the two images evolve to the objects in the energy function minimization process.

From Fig. 3, we can see that the advantages of the proposed method are as follows: 1) A fixed initial curve can segment the common objects from the original image pairs, as shown in Fig. 3, 2) the proposed model can segment multiple common objects from these original images because all these objects have large foreground rewards based on our reward measurement in (2). For example, the two “cows” with different colors are segmented as common objects by using the proposed dynamic method in the seventh image pair.

Note that the image pairs *mcow* (the seventh image pair) and *duck* (the eighth image pair) have similar backgrounds “meadow” in Fig. 3. It is seen that the backgrounds “meadow” have uniform color and occupy most of the area of the two images, which result in a large foreground similarity reward. We can see from the results in Fig. 3(e) and (f) that the proposed method accurately segments the “meadow” as the background due to the reward of the background similarity in our model.

Furthermore, Fig. 3(e) and (f) shows that there are redundant regions or isolated pixels in the segmentation results, such as image pair *banana* (the first image pair). There are two reasons. One is that the isolated pixels have similar colors to the common objects, which results in a large foreground reward of the isolated pixels. For example, the redundant regions in image pair *banana* have colors of black and yellow, which are similar to the colors of the “banana.” The other reason is that the weak smoothness effort [the first term and the second term in (5)] cannot generate enough shrink force to eliminate the redundant regions. The weak smoothness effort is resulted by the small μ ($\mu = 0.01$) and ν ($\nu = 0.001$) that are used in the experiments. Large μ (i.e., $\mu = 0.1$) and ν (i.e., $\nu = 0.1$) can eliminate the redundant regions. However, large μ and ν will also eliminate parts of the common objects. In the proposed method, we use postprocessing to eliminate the redundant regions which are usually small while guaranteeing the common objects to be unaffected, as shown in Fig. 3(g) and (h). In the postprocessing method, all disconnected regions are first identified. Then, the areas of the disconnected regions are calculated $\mathbf{D} = \{D_0, D_1, \dots, D_a\}$, $a \in \mathcal{Z}$, where a is the number of the disconnected regions. To eliminate the redundant regions, we consider the disconnected regions with $D_i > T$ as the regions of the common objects, where T is a threshold. In the experiment, we set $T = 0.2 \cdot \max(\mathbf{D})$ for all the image pairs.

The result of the segmentation is represented by a binary mask, where 0 means background and 1 means foreground. We use the mean error rate for objective evaluation of an image pair, since there are two images in each image pair. The error rate of an image is the ratio of the number of incorrect pixels to the total number of the pixels of the original image. A low error rate indicates an accurate segmentation. Table I shows the resulted error rates of the fixed method, the proposed dynamic method, and the final results (by the postprocessing method). We can see in Table I that the error rates of the dynamic method

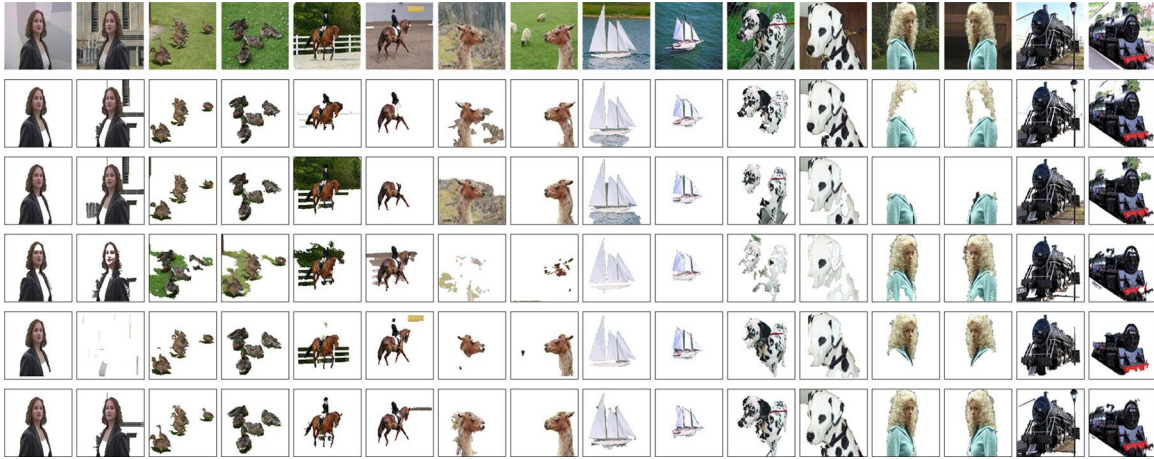


Fig. 4. Results of [28], [31], [36], [52], and the proposed method. The first row displays the original images. The results of [28], [31], [36], [52], and the proposed method are shown from the second row to the last row, respectively.

are lower than those of the fixed method for most of the test image pairs (25 of the 30 image pairs). Furthermore, the final results in Table I show that the error rates are further decreased by the postprocessing method for most of the image pairs (26 of the 30 image pairs). In addition, in some cases, the fixed method gets the lowest error rates, such as *cbleopardb* and *person*. As described in Section III-C, the fixed method uses the other whole image to generate the object prior. When it provides good reference for the common-object prior, the fixed method can also achieve better segmentations.

C. Comparison With Other Methods

We compare our model with the existing work for cosegmentation in [28], [31], and [36]. The comparison is shown in Fig. 4 and Table II. The authors in [28] present a semisupervised MRF-based cosegmentation method which uses a rewarding strategy for cosegmentation. We implement this method using the source code provided by the authors.¹ As the performance of the method in [28] is determined by the initial pixels, we select the most suitable seed pixels for accurate segmentation. In addition, we change parameter λ for more accurate segmentation. The method in [31] aims to classify the common-object regions from all the local regions. Discriminative clustering is employed in [31] for the classification. We use the source code provided by the authors.² Parameter μ is varied to achieve more accurate segmentation. In the implementation of the method in [31], we consider the color descriptor and obtain superpixel using the Ncuts method [1]. The authors in [36] proposed a mutual segmentation approach using a shape feature. We implement the approach with Matlab codes. The parameters for the experiment are set according to [36]. Furthermore, we set the initial curve to be inside the common objects.

In our experiments, the proposed postprocessing method is implemented in [28], [31], and [36] to improve the corresponding segmentation results and compare with the proposed

dynamic method. The related segmentation results are shown in Fig. 4. The original images are shown in the first row of Fig. 4, containing *amira*, *duck*, *pvochorseb*, *llama*, *cdboara*, *cddoga*, *kim*, and *pvoctraind*. By using the postprocessing method, the segmentation results of [28] are shown in the second row of Fig. 4. We can see that the common objects are accurately segmented from the image pairs such as *duck* and *cddoga*. Furthermore, the common objects are not successfully segmented from other image pairs such as *llama* and *kim*, because it is difficult to set precise initial color values of the common objects and backgrounds when the foreground and background have similar colors. The third row of Fig. 4 shows the segmentation results of [31] using the postprocessing method. It can be seen that the common objects are accurately segmented for the image pairs, such as *duck* and *amira*. As the classifier training can be affected by the similarity of the backgrounds, unsuccessful segmentations are obtained for image pairs such as *pvochorseb* and *llama*, as shown in the third row of Fig. 4. The fourth row of Fig. 4 shows the results of [36] applying the postprocessing. We can see that the method in [36] with the postprocessing can achieve successful segmentations for image pairs such as *cdboara* and *kim*. Meanwhile, unsuccessful segmentation is observed for image pairs such as *duck* and *cddoga*, since the precise homograph matrix representing shape similarity is affected by shape variations, such as the variations of pose, scale, and rotation that exist in the data set. For the proposed dynamic method, the segmentation results are shown in the sixth row of Fig. 4. It is illustrated that the common objects are successfully segmented by the proposed cosegmentation method.

Furthermore, we compare our method with the single-image-based segmentation method Grabcut [52] which segments objects through iteratively using graph-cut-based segmentation method. For segmentation result comparison in the experiment, the same initial curve is used in our method and the Grabcut method in [52] with postprocessing. The results of Grabcut are shown in the fifth row of Fig. 4. We can observe that the accurate segmentations are achieved for image pairs such as *duck* and *cdboara*. Furthermore, unsuccessful segmentations are obtained such as *amira* and *kim* due to the initialization sensitivity

¹http://www.biostat.wisc.edu/~vsingh/restricted/myfiles/coseg_vikas_maxflow_distrib.tar.gz

²<http://www.di.ens.fr/~joulain/code/coseg.zip>

TABLE II
RESULTS OF [28], [31], [36], [52], AND THE PROPOSED METHOD IN TERMS OF ERROR RATE

Name	amira	banana	knut	person	cdbeabera	cdboara	cdbuterflya	cdcora	cddoga	cdleopardb
Method of [28]	0.0814	0.0525	0.0355	0.0399	0.0650	0.0746	0.1031	0.3539	0.1183	0.4593
Method of [31]	0.0970	0.0645	0.0335	0.3732	0.2722	0.0945	0.5417	0.2934	0.1868	0.2913
Method of [36]	0.0856	0.1401	0.1031	0.2271	0.0739	0.0559	0.2425	0.2145	0.2231	0.1300
Method of [52]	0.1652	0.0379	0.0572	0.1616	0.0381	0.0241	0.0473	0.1354	0.0873	0.0647
Our results	0.0531	0.0374	0.0802	0.0824	0.0492	0.0341	0.0587	0.1690	0.0648	0.1102
Name	cdpandad	cdpersona	coke	duck	horse	kim	llama	mcow	pvoccarb	pvochorseb
Method of [28]	0.0865	0.3513	0.1910	0.0792	0.0249	0.1153	0.0401	0.1466	0.2630	0.0839
Method of [31]	0.5769	0.3882	0.5032	0.0885	0.0985	0.2115	0.2115	0.2030	0.5565	0.2903
Method of [36]	0.1153	0.2222	0.1177	0.1932	0.2674	0.0469	0.1741	0.1094	0.2636	0.1474
Method of [52]	0.0139	0.1698	0.0242	0.0472	0.1977	0.1585	0.0633	0.0190	0.1137	0.0864
Our results	0.0290	0.1511	0.0224	0.0405	0.0474	0.0192	0.0316	0.0163	0.0988	0.0383
Name	pvocmota	pvocsheepb	pvoctraind	ring008a	ring010a	ring016a	ring017a	ring020a	ring029a	stone
Method of [28]	0.1538	0.3270	0.1674	0.0681	0.0883	0.0956	0.0609	0.0311	0.0497	0.0092
Method of [31]	0.5180	0.4949	0.1660	0.4584	0.2721	0.1770	0.3957	0.0433	0.4178	0.0093
Method of [36]	0.2202	0.2848	0.1284	0.1999	0.1761	0.0638	0.2043	0.0292	0.1656	0.0701
Method of [52]	0.1002	0.3098	0.1065	0.0217	0.1029	0.0126	0.0880	0.0123	0.1285	0.0094
Our results	0.0654	0.1918	0.1029	0.0265	0.0320	0.0133	0.0426	0.0067	0.0145	0.0106

TABLE III
COMPUTATIONAL COMPLEXITIES FOR THE COMPARISON METHODS AND OUR METHOD

Methods	[28]	[31]	[36]	[52]	Ours
Computational complexity	$O(n \log n)$	$O(n^2)$	$O(n^2)$	$O(n \log n)$	$O(n^2)$

of the Grabcut method to the initial curve placement. The initial curve placement determines the background which defines the corresponding Gaussian mixture model and affects the segmentation results. Note that the enhanced segmentation results can be obtained when the initial curve is set manually in [52].

For all the images, the error rates of the methods analyzed in Section IV-C are summarized in Table II. It shows that the proposed method segments the common objects with the minimum error rates for most of the image pairs. Furthermore, we calculate the mean error rate by taking the average of over 30 image pairs for comparison. The mean error rate of our method (5.80% error rate) is about 2.88% less than the method in [52] (8.68% error rate) which obtains the minimum mean error rate among the other methods. In some cases, the existing method achieves the lowest error rates compared with our method, such as *cdcora* and *knut*. The reason is that our method uses fixed parameters for all image pairs in order to perform fair comparison. More accurate segmentation can be obtained by adjusting these parameters for each image pair.

We compare the time complexity of the proposed method with those of the methods in [28], [31], [36], and [52]. In our method, the number of iterations is related to the image size n . Also, the pixel number and the calculation of the histogram for the inner region at each iteration correspond to n . Thus, the time complexity of our method is proportional to $n \times (n + n) = 2n^2$, which means that the time complexity is proportional to $O(n^2)$. Table III shows the time complexities of the proposed method and the other methods. In Table III, it can be seen that the methods in [28] and [52] have lower time complexities, i.e., $O(n \log n)$, since the graph-cut algorithm is used for the optimization. The proposed method has the same time complexity with [31] and [36].

D. Discussion

The proposed cosegmentation model considers the initializations of two curves, since (11) is related to two curves C_k and C_{1-k} for $k \in \{0, 1\}$. For image I_k , we assume that P_1 and P_2

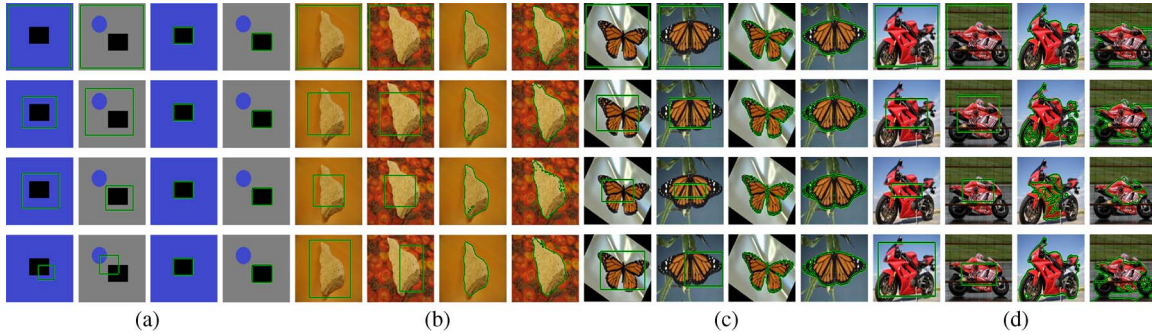


Fig. 5. Segmentation results under various initial curves.

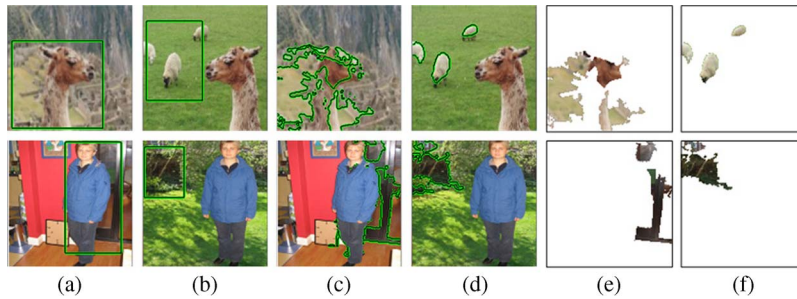


Fig. 6. Unsuccessful segmentation results under specific initializations.

denote the sets of the common-object pixels and the background pixels located on the curve C_k . To facilitate the selection of the initial curve in the proposed method, the requirements of the initial curves are described as follows.

- 1) For a pixel $I_k(i, j) \in P_1$ with location (i, j) , the initial curves should make the value of $f[I_k(i, j), g(\omega_{1-k}^i)]$ greater than $f[I_k(i, j), g(\omega_k^o)]$, i.e., $f[I_k(i, j), g(\omega_{1-k}^i)] > f[I_k(i, j), g(\omega_k^o)]$.
- 2) On the contrary, for a pixel $I_k(i, j) \in P_2$ with location (i, j) , the initial curves should guarantee that $f[I_k(i, j), g(\omega_{1-k}^i)] \leq f[I_k(i, j), g(\omega_k^o)]$.

Based on the assumption that the original images have different backgrounds, we can obtain $f[I_k(i, j), g(\omega_{1-k}^i)] = 0$ for $I_k(i, j) \in P_2$, which means that requirement 2) can always be satisfied.

To fulfill requirement 1), we assume that the initial curves cover most of the regions of the objects and set the initial curve as a rectangle in our model. Each edge of the rectangle has small distance to the corresponding image edge. It is seen that our curve initialization satisfies requirement 1). The reason is that $f[I_k(i, j), g(\omega_k^o)] \approx 0$ for $I_k(i, j) \in P_1$ and $f[I_k(i, j), g(\omega_{1-k}^i)] > 0$.

Fig. 5 shows the results of our method with various initial curves. Both artificial image pairs (the first image pair) and natural image pairs (the remaining three image pairs) are considered. For each image pair, initial curves partially or fully covering the common objects are considered. They are shown in the first and second columns for each image pair block in Fig. 5(a)–(d). We can see that some initial curves cover the whole regions of the common objects, and the others partially cover the common objects. For example, the first initial curve for *stone* covers all the region of the common

object. For the third initial curve in the *butterfly*, only a small region of the common object locates inside. We can see from the results that the curves (except the third initial curve in *motor*) evolve to the boundary of the common objects. For example, the fourth initial curves in Fig. 5(a) successfully segment the common objects from the original images, although the initial curves both partially cover the regions of the common objects. The successful segmentations are contributed by the fulfillment of requirement 1) by the initial curves. For the incorrect segmentation in *motor*, only the red regions are segmented as common objects, because the initial curves cover only the red regions of the common objects, which violates requirement 1).

Furthermore, we show several unsuccessful segmentations by the specific initializations. These results are shown in Fig. 6, where the initial curves are shown in Fig. 6(a) and (b). It can be seen that the initial curve does not cover the common objects in one of the images. The initial segmentation results (without postprocessing) are shown in Fig. 6(c) and (d). The final segmentation results are shown in Fig. 6(e) and (f). We can see from Fig. 6(c)–(f) that incorrect initialization could lead to wrong segmentation results, since the curve dose not include the common objects. Compared with the results in Fig. 5, we can see that good performance can be achieved if the initial curves contain enough foreground information. This condition can be easily satisfied when the initial curve is set as the rectangle with small distance to the image edge.

We discuss the curve evolution process along the iterations. Fig. 7 shows the results containing four image pairs. For each image pair, five curve evolution states corresponding to iteration numbers 1, 101, 201, 301, and 401 are displayed. Fig. 7 demonstrates that the curves for the images evolve to the boundary of the common objects. For example, the curves evolve to the

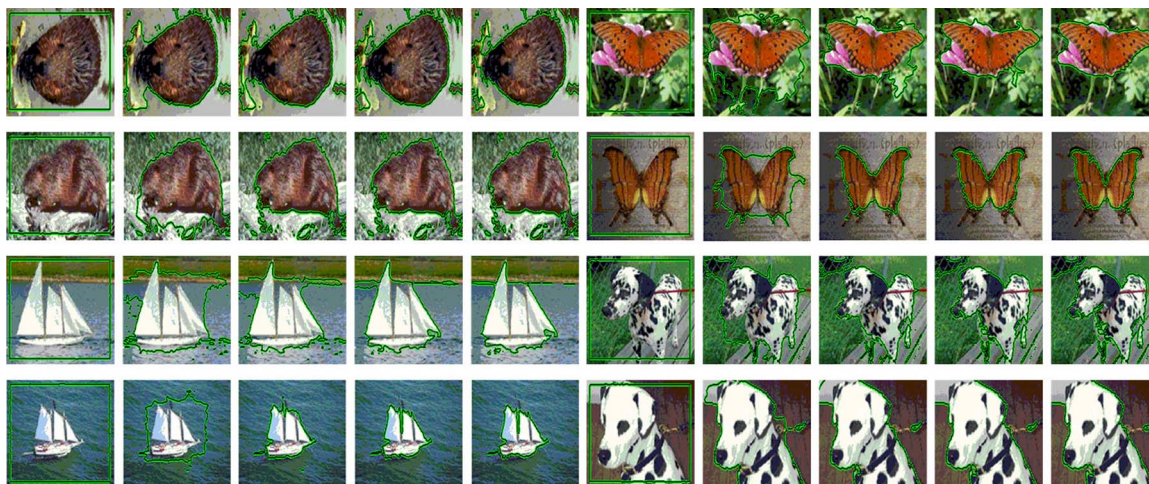


Fig. 7. Curve evolution processes under different iterations. There are five curve states for each original image corresponding to iteration numbers 1, 101, 201, 301, and 401.

boundary of the objects at the 300th iteration in the first image pair.

Compared with the well-known $C-V$ model, the proposed model has two advantages. First, we use color histogram instead of mean value for region representation, which is more suitable to represent a region with multiple colors. Second, for an image pair, we introduce the color information of one image into the energy function of the other image, which increases the robustness to the initial curve replacement. The reason is that, in the $C-V$ model, the foreground information is modeled by the interior region. Thus, the foreground information is affected by the initial curve placement. A careful initial curve setting is necessary for obtaining the precise foreground information. In our model, the foreground information of an image I_k is provided by the other image I_{1-k} in the image pair. Based on the assumption that most areas of the common objects locate within the curve in I_{1-k} , precise foreground information can be provided to I_k . Note that it is simple to set the initial curve as a rectangle to cover most areas of I_{1-k} .

There are small regions in the initial segmentation. These small regions are caused by the topology changes of the level sets. Limiting the topology changes can remove these small regions and avoid the need of the postprocessing. In our model, we do not limit the topology changes of the level set because we use the topology changes to cope with the variations among the common objects, such as the variations of the shape and the pose. Limiting the topology changes may also limit the proposed method to handle these variations. Meanwhile, these small regions can be removed by a simple postprocessing. Note that limiting the topology change is suitable to solve the cosegmentation with similar shapes. In the future, we will extend the model to segment common objects with similar shapes by introducing the limitation of the topology changes.

In this paper, we use level sets to minimize the energy function, which searches the minimum in a local manner. Note that, in previous works, some authors use graph techniques to search for a global optimization, such as [28] and [35]. In these methods, the successful segmentation depends on the accurate object prior to build the unary term of the energy function,

which usually needs user's interaction to initialize the object location or appearance [28]. Our method is statical-region-based method where the contour is not evolved by fitting to local gradient information (as in the Snakes) but rather by fitting statistical models to intensity, color, texture, or motion within each of the separated regions [53]. It has been studied that the cost functionals of the statical-region-based method tend to have less local minima for most realistic images [53]. Hence, the level sets are suitable to solve the proposed energy function in realistic images. The experimental results also demonstrate the effectiveness of the level-set-based minimization.

V. CONCLUSION

In this paper, we have proposed a novel cosegmentation method to segment the common objects from image pairs. In the proposed model, the energy function is modeled by considering the foreground similarity and the background consistency. A color histogram is used to represent a region, and the rewarding strategy is used to evaluate the foreground similarity. The background reward is used to improve the classical reward strategy. Furthermore, the level set is used to represent the curve. Then, a mutual optimization procedure is used for the minimization of the energy function value. The experimental results demonstrate that the proposed model can efficiently segment the common objects from the image pairs with generally lower error rates than many existing and conventional cosegmentation methods. In the future, we will select other region features to obtain more accurate segmentation, and we will modify the proposed method to further reduce the time complexity.

ACKNOWLEDGMENT

The authors would like to thank all the anonymous reviewers for the constructive comments and useful suggestions that led to improvements in the quality, presentation, and organization of this paper. The authors would also like to thank T. Riklin-Raviv for her kindness, advice, generosity, and great help.

REFERENCES

- [1] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [2] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 326–333.
- [3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [4] 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*, 2009, pp. 2294–2301.
- [5] L. Chen, C. Chen, and M. Lu, "A multiple-kernel fuzzy c-means algorithm for image segmentation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 5, pp. 1263–1274, Oct. 2011.
- [6] S.-Y. Chien, Y.-W. Huang, and L.-G. Chen, "Predictive watershed: A fast watershed algorithm for video segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 5, pp. 453–461, May 2003.
- [7] M. Segundo, L. Silva, O. Bellon, and C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 5, pp. 1319–1330, Oct. 2010.
- [8] H. Li and K. N. Ngan, "Unsupervised video segmentation with low depth of field," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1742–1751, Dec. 2007.
- [9] N. Jacobson, Y. L. Lee, V. Mahadevan, N. Vasconcelos, and T. Q. Nguyen, "A novel approach to FRUC using discriminant saliency and frame segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2924–2934, Nov. 2010.
- [10] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, Jan. 2006.
- [11] S. Goferman, L. Z. Manor, and A. Tal, "Context-aware saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2376–2383.
- [12] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, Jan. 2010.
- [13] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Fast pedestrian detection using a cascade of boosted covariance features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1140–1151, Aug. 2008.
- [14] F. Jing, M. Li, H. J. Zhang, and B. Zhang, "Relevance feedback in region-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 672–681, May 2004.
- [15] H. Li, K. N. Ngan, and Q. Liu, "Faceseg: Automatic face segmentation for real-time video," *IEEE Transactions on Multimedia*, vol. 11, no. 1, pp. 77–88, Jan. 2009.
- [16] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *International Conference on Computer Vision*, 2001, pp. 105–112.
- [17] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [18] J. Zhang, J. Zheng, and J. Cai, "A diffusion approach to seeded image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2125–2132.
- [19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [20] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.
- [21] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [22] M. Cho, Y. M. Shin, and K. M. Lee, "Unsupervised detection and segmentation of identical objects," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1617–1624.
- [23] T. Deselaers, B. Alexe, and V. Ferrari, "Localizing objects while learning their appearance," in *European Conference on Computer Vision*, 2010, pp. 452–466.
- [24] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, "Using multiple segmentations to discover objects and their extent in image collections," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1605–1614.
- [25] K. Grauman and T. Darrell, "Unsupervised learning of categories from sets of partially matching image features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 19–25.
- [26] C. Rother, V. Kolmogorov, T. Minka, and A. Blake, "Cosegmentation of image pairs by histogram matching—Incorporating a global constraint into MRFs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 993–1000.
- [27] 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.
- [28] D. S. Hochbaum and V. Singh, "An efficient algorithm for cosegmentation," in *International Conference on Computer Vision*, 2009, pp. 269–276.
- [29] S. Vicente, V. Kolmogorov, and C. Rother, "Cosegmentation revisited: Models and optimization," in *European Conference on Computer Vision*, 2010, pp. 465–479.
- [30] D. Batra, D. Parikh, A. Kowdle, T. Chen, and J. Luo, "Seed image selection in interactive cosegmentation," in *IEEE International Conference on Image Processing*, 2009, pp. 2393–2396.
- [31] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image cosegmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1943–1950.
- [32] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "Icoseg: Interactive cosegmentation with intelligent scribble guidance," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3169–3176.
- [33] S. Vicente, C. Rother, and V. Kolmogorov, "Object cosegmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2217–2224.
- [34] L. Mukherjee, V. Singh, and J. Peng, "Scale invariant cosegmentation for image groups," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1881–1888.
- [35] K. Chang, T. Liu, and S. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2129–2136.
- [36] T. Riklin-Raviv, N. Sochen, and N. Kiryati, "Shape-based mutual segmentation," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 231–245, 2008.
- [37] T. Riklin-Raviv, N. Sochen, and N. Kiryati, "Mutual segmentation with level-sets," in *IEEE CVPR Workshop on Perceptual Organization in Computer Vision*, 2006.
- [38] G. Kim and E. P. Xing, "On multiple foreground cosegmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- [39] Y. Chai, V. Lempitsky, and A. Zisserman, "Bicos: A bi-level cosegmentation method for image classification," in *IEEE International Conference on Computer Vision*, 2011, pp. 2579–2586.
- [40] A. C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [41] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1987.
- [42] L. D. Cohen, "On active contour models and balloons," *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 211–218, 1991.
- [43] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.
- [44] S. J. Besson, M. Barlaud, and G. Aubert, "Shape gradients for histogram segmentation using active contours," in *International Conference on Computer Vision*, 2003, pp. 408–415.
- [45] X. Gao, B. Wang, D. Tao, and X. Li, "A relay level set method for automatic image segmentation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 2, pp. 518–525, Apr. 2011.
- [46] L. D. Cohen, "Avoiding local minima for deformable curves in image analysis," in *Curves and Surfaces With Applications in CAGD*, A. Le Méhauté, C. Rabut, and L. L. Schumaker, Eds, 1997, pp. 77–84.
- [47] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [48] T. F. Chan, B. Y. Sandberg, and L. A. Vese, "Active contours without edge for vector-valued images," *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 130–141, Jun. 2000.
- [49] N. Paragios and R. Deriche, "Geodesic active regions: A new framework to deal with frame partition problems in computer vision," *Journal of Visual Communication and Image Representation*, vol. 13, no. 1/2, pp. 249–268, Mar. 2002.

- [50] Q. Chen, Q. Sun, P. A. Heng, and D. Xia, "Two-stage object tracking method based on kernel and active contour," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 605–609, Apr. 2010.
- [51] B. Wang, X. Gao, D. Tao, and X. Li, "A unified tensor level set for image segmentation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 857–867, Jun. 2010.
- [52] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, Aug. 2004.
- [53] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape," *International Journal of Computer Vision*, vol. 72, no. 2, pp. 195–215, 2007.



Fanman Meng received the B.Sc. degree in computer science and technology in Shandong Agricultural University, Taian, China, 2006, and the M.Sc. degree in computer software and theory in Xihua University, Chengdu, China, 2009, respectively. Since 2009, he has been working toward the Ph.D. degree in the Intelligent Visual Information Processing and Communication Laboratory, University of Electronic Science and Technology of China, Chengdu, China.

His research interests include image segmentation, object detection, and visual attention.



Hongliang Li (SM'12) received the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, Xi'an, China, in 2005.

From 2005 to 2006, he was a Research Associate with the Visual Signal Processing and Communication Laboratory, The Chinese University of Hong Kong, Shatin, Hong Kong, where he was a Postdoctoral Fellow from 2006 to 2008. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. He is a Coeditor of a

Springer book titled "Video segmentation and its applications." He is a Member of the Editorial Board of the *Journal on Visual Communications and Image Representation*. He has authored or coauthored numerous technical articles in well-known international journals and conferences. His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system.

Dr. Li was involved in many professional activities. He served as a Technical Program Committee member in a number of international conferences, e.g., the 2005 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), PCM 2007, Pacific Rim Conference on Multimedia 2009, and Visual Communications and Image Processing 2010; as a Technical Program Cochair in ISPACS2009; and as a General Cochair of the 2010 International Symposium on Intelligent Signal Processing and Communication Systems. He will serve as a Local Chair of the 2014 IEEE International Conference on Multimedia and Expo. He was selected as the New Century Excellent Talents in University, Chinese Ministry of Education, China, in 2008.



Guanghui Liu received the M.Sc. and Ph.D. degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2002 and 2005 respectively.

In 2005, he joined Samsung Electronics, Suwon, South Korea, as a Senior Engineer. Since 2009, he has been an Associate Professor with the School of Electronic Engineering, UESTC. His general research interests include digital signal processing and telecommunications, with emphasis on digital video transmission, and orthogonal frequency-division multiplexing techniques. In these areas, he has published tens of papers in refereed journals or conferences and received more than ten patents (four U.S. granted patents).

Dr. Liu served as the Publication Chair of the 2010 International Symposium on Intelligent Signal Processing and Communication Systems.



King Ngi Ngan (F'00) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K.

He was a Full Professor with Nanyang Technological University, Singapore, and The University of Western Australia, Perth, Australia. He is currently a Chair Professor with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. He holds honorary and visiting professorships in numerous universities in China, Australia, and southeast Asia. He served as an Associate Editor of the *Journal on Visual Communications and Image Representation*, *European Association for Signal Processing Journal of Signal Processing: Image Communication*, and *Journal of Applied Signal Processing*. He has published extensively, including three authored books, six edited volumes, over 300 refereed technical papers, and edited nine special issues in journals. In addition, he is the holder of ten patents in the areas of image/video coding and communications.

Prof. Ngan is a fellow of the Institution of Engineering and Technology (U.K.) and The Institution of Engineers Australia (Australia) and an IEEE Distinguished Lecturer in 2006–2007. He served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He chaired a number of prestigious international conferences on video signal processing and communications and served on the advisory and technical committees of numerous professional organizations. He cochaired the IEEE International Conference on Image Processing held in Hong Kong in September 2010.

Prof. Ngan is a fellow of the Institution of Engineering and Technology (U.K.) and The Institution of Engineers Australia (Australia) and an IEEE Distinguished Lecturer in 2006–2007. He served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He chaired a number of prestigious international conferences on video signal processing and communications and served on the advisory and technical committees of numerous professional organizations. He cochaired the IEEE International Conference on Image Processing held in Hong Kong in September 2010.