

RESEARCH

Open Access

A mutual GrabCut method to solve co-segmentation

Zhisheng Gao¹, Peng Shi^{2,3}, Hamid Reza Karimi^{4*} and Zheng Pei¹

Abstract

Co-segmentation aims at segmenting common objects from a group of images. Markov random field (MRF) has been widely used to solve co-segmentation, which introduces a global constraint to make the foreground similar to each other. However, it is difficult to minimize the new model. In this paper, we propose a new Markov random field-based co-segmentation model to solve co-segmentation problem without minimization problem. In our model, foreground similarity constraint is added into the unary term of MRF model rather than the global term, which can be minimized by graph cut method. In the model, a new energy function is designed by considering both the foreground similarity and the background consistency. Then, a mutual optimization approach is used to minimize the energy function. We test the proposed method on many pairs of images. The experimental results demonstrate the effectiveness of the proposed method.

Keywords: Co-segmentation, GrabCut, Graph cut algorithm, Markov random field

1 Introduction

Image segmentation is a fundamental problem for many computer vision tasks, such as object recognition [1,2], image understanding [3], and retrieval [4]. Due to variations of the objects, image segmentation remains a challenging problem. Recently, co-segmentation [5-15] has attracted much attention from the community. The goal of co-segmentation is to segment common objects from a group of images. Unlike traditional single-image segmentation, the co-segmentation method can segment multiple images jointly rather than independently segmenting each image based on the co-occurrence of objects in the images [16]. Several examples can be found in Figure 1, where six image pairs are shown. In each image pair, the co-segmentation aims to extract the common objects from the image pair, such as the 'plane' and 'banana' in the first two image pairs. Compared with traditional segmentation methods, co-segmentation can accurately segment objects from images by several related images and requires less user workload [17]. It has many potential applications in computer vision, such as image classification, object

recognition, and image retrieval. This paper focuses on the co-segmentation problem.

The existing co-segmentation models address co-segmentation as an optimization problem, which achieves common objects by adding foreground similarity into segmentation models. Both the local smoothness in each image and the foreground similarity among the images are considered. Many traditional segmentation methods have been improved to solve co-segmentation method, such as the Markov random field (MRF)-based segmentation methods [5-8], random walker-based segmentation method [18], and discriminative clustering-based segmentation method [10,19]. Analyzing these methods, these co-segmentation methods can be concluded as the extensions of the interactive-based segmentation methods since it is natural to replace the initial seeds manually given in the traditional method by searching the local similar regions shared by images.

Several well-known interactive-based segmentation methods have been extended to solve co-segmentation problem. MRF-based segmentation method was first extended for co-segmentation task by Rother [5], which introduced a global term representing foreground similarity into the MRF-based image segmentation model. Kim et al. [15] extended heat diffusion-based interactive segmentation method to solve multi-class co-segmentation

*Correspondence: hamid.r.karimi@uia.no

⁴Department of Engineering, Faculty of Technology and Science, University of Agder, 4898 Grimstad, Norway

Full list of author information is available at the end of the article



Figure 1 Some examples of co-segmentation showing six image pairs.

problem. The heat diffusion-based method spreads the heat from the source seeds to the other pixels by pixel similarity. To solve multiple images in co-segmentation, the heat was diffused among the common objects by foreground similarity. The random walker-based interactive segmentation method was extended to solve co-segmentation problem in the work of Collins et al. [18], which introduces foreground similarity constraint into the random walker-based method. In the work of Meng et al. [16,20], the active contour-based model (Chan-Vese model) was extended to fit co-segmentation task by considering both foreground similarity constraint and background consistency.

Among these methods, MRF-based co-segmentation methods attract much attentions since the success of the MRF-based segmentation method on single-image segmentation. Several MRF-based co-segmentation methods have been proposed [5-8]. Their differences focus on the formulation of foreground similarity constraints. Several foreground similarity constraints have been added, such as L1-norm [5], L2-norm [6], and reward strategy [7]. However, it remains challenging to minimize the MRF-based co-segmentation energy function although many global terms have been introduced. To cope with the minimization problem, the existing methods search approximate solutions [5,6,8] or require user to provide foreground appearance and locations [7]. Other methods use saliency map [13] to obtain initial object appearance model. For these methods, the results depend on the accuracies of the initial appearance models.

GrabCut [21] is an important MRF-based co-segmentation method, which segments the objects from a manual rectangle setting by graph cut algorithm. The main advantage of GrabCut is that the energy function can be efficiently minimized by mutually applying graph cut algorithm in polynomial time. Hence, it can be used in many real-time applications. Furthermore, it models

the foreground and background appearance priors by a simple rectangle setting, which is convenient compared with the other interactive-based segmentation methods. It is seen that performing co-segmentation based on the GrabCut model can result in efficient optimization and prior model generation. Meanwhile, the GrabCut model can also benefit from co-segmentation task. The GrabCut model will be more robust to initial curve setting. The reason is that the prior provided by a pair of images in co-segmentation is more sufficient compared with a single image. Hence, automatically segmenting objects by GrabCut (without manual curve setting) can be achieved in co-segmentation task.

In this paper, we propose a new MRF-based co-segmentation method namely mutual GrabCut (MGrabCut) for common object segmentation, which extends GrabCut [21] to solve co-segmentation. In the method, the region outside each initial rectangle is treated as background region. Meanwhile, the regions inside initial rectangles are used to model unary potential of the foreground. To segment similar foregrounds, we introduce the foreground model of the other image in the unary term of the current image. The final co-segmentation results are achieved by graph cuts with iteratively updating unary term of the foreground appearance model and background appearance model. The main advantage of the proposed method is that compared with existing MRF-based co-segmentation methods, we consider foreground similarity into unary term rather than global term, which results in easier minimization. Hence, the proposed model is efficient and real time. Secondly, the proposed method is robust to initial curve setting because the common objects can be more accurately located by the constraint of foreground similarity. A fixed initial curve can be used for all pairs of images. Thirdly, since the foreground model is dynamically updated along the iteration, a more accurate appearance model is obtained by the proposed method.

We test the proposed method on many pairs of images. The experimental results demonstrate the effectiveness of the proposed method.

The contributions of the proposed method are listed as follows:

1. A novel MRF-based co-segmentation model is designed. In the model, the foreground similarity constraint is added into unary term rather than global term, which results in the efficient minimization by graph cut algorithm.
2. Compared with traditional GrabCut model, the proposed model is more robust to initial curve setting and can segment objects with fixed initial curves. The benefit is caused by considering a pair of images instead of a single image.
3. A mutual graph cut-based minimization method is developed to minimize the energy pairs.

2 Related work

In image segmentation, many minimization techniques have been used to achieve accurate object segmentation. Boykov et al. in [22] used graph cut algorithm to minimize the energy in MRF-based segmentation model. In the work of Meng et al. [16], the active contour-based energy function was minimized by level set techniques and the method of calculus of variations. In [17], the shortest path algorithm achieved by dynamic programming method was used for object segmentation. In the work of Zeng et al. [23], a hybrid extended Kalman filter and switching particle swarm optimization algorithm were proposed for model parameter estimation. In [24], a new particle filter was developed to simultaneously identify both the states and parameters of the model. In [25], Zineddin et al. presented a new image reconstruction algorithm using the cellular neural network that solves the Navier-Stokes equation, which offered a robust method for estimating the background signal within the gene spot region.

In the existing co-segmentation methods, co-segmentation is commonly modeled as an optimization problem, which introduces foreground similarity to fit common object segmentation. For MRF-based co-segmentation model, the energy function is usually defined as

$$E = U_{\text{pixel}} + V_{\text{pair}} + G_{\text{global}} \quad (1)$$

where U_{pixel} is the data term which evaluates the potential of the pixel to the foreground or background. V_{pair} is the smoothness term to measure the smoothness of local pixels. These two terms are single-image segmentation-based term. The term G_{global} is the global term evaluates the similarity between the foregrounds. By minimizing the energy function, only common objects are extracted.

Although the global term makes the foreground similar, it also results in difficult minimization since searching the regions with similar appearance is challenging. The existing methods employ various global terms to cope with the minimizations. Rother et al. [5] used L1-norm to measure foreground similarity. The trust region graph cut method was proposed for energy optimization. Mukherjee et al. [6] replaced L1-norm with L2-norm. Pseudo-Boolean optimization was used for optimization. Instead of penalizing foreground difference, Hochbaum and Singh [7] rewarded foreground similarity. Vicente et al. in [8] modified the Boykov-Jolly model for foreground similarity measurement. Dual decomposition was employed for minimization.

Other methods have also been used for co-segmentation task. Joulin et al. [10] segmented common objects by clustering strategy. The main idea was that the common objects can be classified into the same class since they have similar features. Hence, by searching a classifier based on spectral clustering technique and positive definite kernels that best classified the common objects, co-segmentation was achieved. In the work of Batra et al. [11], an interactive co-segmentation method which segmented common objects through human interaction guided by an automatic recommendation system was proposed. Mukherjee et al. [12] proposed a scale-invariant co-segmentation method to segment common objects through the fact that the rank of the matrix corresponding to foreground regions should be equal to 1. The algorithm of Chang et al. [13] solved co-segmentation by a novel global energy term which used the co-saliency model to measure foreground potentials. The energy function considering both foreground similarity and background consistency was submodular and can be efficiently minimized by graph cut algorithm. Vicente et al. [14] focused on interesting object co-segmentation. A useful feature to distinguish the common objects was trained from a total of 33 features through random forest regression. The common objects were segmented by loop belief propagation on a full connected graph. Kim et al. in [15] solved multiple-class-based co-segmentation problem by anisotropic heat diffusion. By combining clustering method and random walk segmentation method, multiple classes can be successfully labeled from a large number of images. Recently, Joulin et al. in [19] focused on multi-class co-segmentation, which considers discriminative clustering and multi-class co-segmentations into account. More accurate segmentation results were obtained. Collins et al. in [18] solved co-segmentation by random walker-based segmentation method which added foreground consistency into traditional random walker-based method. Compared with MRF-based co-segmentation, the random walker-based co-segmentation method was efficient. Rubio et al. in [26] segmented common objects by modifying the wrongly

segmented from the other successful segmentations. A co-segmentation framework was formulated by MRF, and a new global term based on graph matching was proposed. In the work of Meng [17], co-segmentation from a large number of original images with similar backgrounds was considered. A digraph was constructed by foreground similarity and saliency values. The co-segmentation problem was formulated as the shortest path problem and was solved by dynamic programming method.

3 The proposed model

In this section, we first introduce the GrabCut method. Then, the proposed method is illustrated.

3.1 GrabCut segmentation

GrabCut is an interactive image segmentation method. It has been widely used in many computer vision tasks. In GrabCut, the image segmentation is a label problem which assigns a label $\alpha_i \in \{0, 1\}$, $i = 1, \dots, N$ to each image pixels z_i , $i = 1, \dots, N$ with $\alpha_i = 1$ for foreground and 0 for background. N is the number of pixels. The label problem is then set as an optimization problem by minimizing the energy function

$$E(\alpha, \mathbf{k}, \theta, \mathbf{z}) = U(\alpha, \mathbf{k}, \theta, \mathbf{z}) + V(\alpha, \mathbf{z}) \quad (2)$$

where $\alpha = (\alpha_1, \dots, \alpha_N)$, $\mathbf{z} = (z_1, \dots, z_N)$, and θ describes image foreground and background appearance model which is represented as

$$\theta = \{h(z; \alpha), \alpha = \{0, 1\}\} \quad (3)$$

where $\alpha = 0$ for the background model and $\alpha = 1$ for the foreground model. h is the appearance model, which is represented as a Gaussian mixture model. In the model, a full-covariance Gaussian mixture with K components is considered for the construction. With a Gaussian mixture model (GMM) for the foreground or the background, each pixel z_i is assigned a unique GMM component k_i either from the background or the foreground model according to $\alpha = 0$ or 1, where $\mathbf{k} = (k_1, \dots, k_N)$, $\theta = \{\pi(\alpha, k_i), \mu(\alpha, k_i), \Sigma(\alpha, k_i), \alpha = 0, 1, i = 1, \dots, N\}$. Here, $\pi(\cdot)$ are mixture weighting coefficients, and $\mu(\cdot)$ and $\Sigma(\cdot)$ are means and covariances of the distribution $p(\cdot)$.

The data term $U(\alpha, \mathbf{k}, \theta, \mathbf{z})$ in Equation 2 evaluates the fit of the label α to the data \mathbf{z} with θ and \mathbf{k} and is represented as

$$U(\alpha, \mathbf{k}, \theta, \mathbf{z}) = \sum_n D(\alpha_n, k_n, \theta, z_n) \quad (4)$$

where n is the number of pixels and

$$D(\alpha_n, k_n, \theta, z_n) = -\log p(z_n | \alpha_n, k_n, \theta) - \log \pi(\alpha_n, k_n). \quad (5)$$

The smoothness term $V(\alpha, \mathbf{z})$ in Equation 2 encourages coherence in local regions and is defined as

$$V(\alpha, \mathbf{z}) = \gamma \sum_{(m,n) \in \mathbf{C}} [\alpha_n \neq \alpha_m] \exp -\beta \|z_m - z_n\|^2 \quad (6)$$

where $[\cdot]$ denotes the indicator function taking values 0,1 for a predicate \cdot . β is constant. \mathbf{C} is the set of pairs of neighboring pixels. The pixels are neighbors if they are adjacent either horizontally/vertically or diagonally.

Based on Equation 2, the segmentation is obtained by minimizing Equation 2 represented as

$$\hat{\alpha} = \arg \min_{\alpha} E(\alpha, \mathbf{k}, \theta, \mathbf{z}) \quad (7)$$

By fixing \mathbf{k} and θ , the problem in Equation 2 is solved by minimum cut algorithm (graph cut algorithm). In GrabCut, the energy minimization scheme works iteratively, which updates \mathbf{k} and θ by current segmentation and uses new \mathbf{k} and θ to obtain new segmentation by solving the problem in Equation 2. The algorithm starts from an initial curve setting manually. The iteration stops when convergence criterion is satisfied.

3.2 The proposed method

Unlike single-image-based GrabCut method, a pair of images \mathbf{z}^l , $l = 0, 1$ is considered in the proposed model. Set z_i^l is the i th pixel in the l th image and $\mathbf{z}^l = (z_1^l, \dots, z_N^l)$. The label for image \mathbf{z}^l , $l = 0, 1$ is α^l , $l = 0, 1$. The proposed method sets co-segmentation as a label problem that assigns 1 for pixels on the common objects and 0 otherwise. To segment common objects, we design a new unary term by considering foreground similarity, which guarantees that only common objects are considered. In the method, the unary term is defined as

$$U(\alpha^l, \mathbf{k}^l, \theta^l, \theta^{1-l}, \mathbf{z}^l) = \begin{cases} \sum_n (\lambda \cdot D^1(\alpha_n^l, k_n^l, \theta^l, z_n^l) + (1 - \lambda) \\ \times D^2(\alpha_n^l, k_n^{1-l}, \theta^{1-l}, z_n^l)), & \text{if } \alpha_n^l = 1; \\ D^1(\alpha_n^l, k_n^l, \theta^l, z_n^l), & \text{else} \end{cases} \quad (8)$$

where θ^l and \mathbf{k}^l are the parameter sets of GMM representation of \mathbf{z}^l , which is similar to the definition in GrabCut. λ is the scale factor to balance the impacts of the foregrounds in the current image and the other image. D^1 evaluates the fit of the label α^l to the data \mathbf{z}^l with θ^l and \mathbf{k}^l in the current image and is represented as

$$D^1(\alpha_n^l, k_n^l, \theta^l, z_n^l) = -\log p(z_n^l | \alpha_n^l, k_n^l, \theta^l) - \log \pi(\alpha_n^l, k_n^l). \quad (9)$$

The foreground similarity term D^2 evaluates the similarity between the foregrounds and is defined as

$$D^2(\alpha_n^l, k_n^{1-l}, \theta^{1-l}, z_n^l) = -\log p(z_n^l | \alpha_n^l, k_n^{1-l}, \theta^{1-l}) - \log \pi(\alpha_n^l, k_n^{1-l}). \quad (10)$$

We use the smoothness term in GrabCut shown in Equation 6 to form the smoothness term of the proposed method. Then, the co-segmentation is set to minimize the energy function represented as

$$\hat{\alpha}^l = \arg \min_{\alpha^l} E(\alpha^l, \mathbf{k}^l, \theta^l, \mathbf{k}^{1-l}, \theta^{1-l}, \mathbf{z}^l). \quad (11)$$

We can see from Equation 10 that D^2 evaluates the fit of the pixels with $\alpha_n^l = 1$ in the current image to the foreground model θ^{1-l} in the next image. The pixels on common objects have small D^2 since they are similar to the common objects in the next image. Hence, it intends to be assigned 1. For other pixels, a larger D^2 will be obtained. Hence, it intends to be a background pixel.

By keeping \mathbf{k}^l , θ^l , \mathbf{k}^{1-l} , and θ^{1-l} fixed, the energy function is minimized by minimum cut method. Similar to GrabCut, we iteratively update the foreground model and the background model to accurately segment the common objects. The main difference is that there are two images in our model. Hence, we improve the iteration method by simultaneously updating the foreground model and the background model of two images. In the optimization method, the initial curve is first set to each image. The initial segmentations are obtained by treating the pixels inside the curve as the foreground and the pixels outside the curve as the background. Then, based on the initial segmentation, we model the foreground model and background model θ_k^l and \mathbf{k}^l for each image which are then used to obtain the foreground potential and background potential for each image according to Equations 9 and 10. Finally, we optimize the two energies by Equation 11 to obtain segmentation results. The segmentation results are used as the new initial segmentations for the next iteration. The algorithm stops when stop condition is satisfied.

We analyze next the proposed model compared with the GrabCut. Their difference can be found in Figure 2, where Figure 2a shows the model of the GrabCut, which is related to a single image. There is an initial curve C^0 which separates the image Z^0 into two regions, i.e., the region inside the curve and the region outside the curve. The GrabCut considers the region inside the curve as

the foreground and the region outside the curve as the background. Then, the GMM of the foreground and the background are determined based on the two regions. The GMM is represented as k^0 and θ^0 . For a pixel (the blue points), there are two influences in the GrabCut model. One is the foreground model represented by the green lines. The other is the background model represented by the yellow lines. Based on the two aspects, the point will be given a label. We can see that GrabCut is sensitive to the initial curve setting because the change of initial curve will also change the parameters of the foreground model and background model, which results in different segmentations. Hence, for GrabCut, manually selecting the initial curve is used for the segmentation.

The proposed model is represented in Figure 2b, where there are two images, Z^0 and Z^1 , rather than a single image. For each image, there is a curve. The curve also segments the image into two regions: the region inside the curve and the region outside the curve. Like the analysis of the GrabCut in Figure 2a, we consider the blue points in Z^0 . We can see that there are three terms in our model. The first two are the foreground model (the green line in Z^0) and the background model (the yellow line) in the current image Z^0 . These two terms are similar to the two in GrabCut. The third is the foreground model in Z^1 . For the third influence, since only the common objects share similar colors, the pixels on the objects will have large response of the third term. While for a background pixel, it has a small response, which results in the label of background. Hence, the pixels on the common objects will be considered as foreground.

Comparing our model with GrabCut, the difference is that we introduce the third term in our model, which results in the segmentation of the common objects. We can see that the third term also results in the robustness to initial curve setting. The reason is that the initial curve setting of the current image may change the foreground model. However, the next image can provide the accurate foreground model when the curve C^1 covers most of the area of the image pairs. The appearance model of the third model can improve the label of the pixels and result in successful segmentation. Here, we have to guarantee that

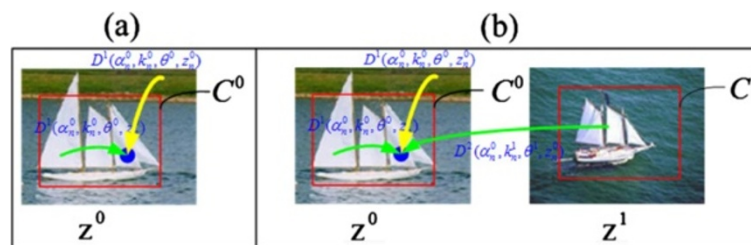


Figure 2 Difference between GrabCut and the proposed model. (a) The segmentation model of GrabCut. (b) The segmentation model of the proposed method.



Figure 3 The initial curve setting used in this paper. Two image pairs are shown. v is the distance between the curve and the image edge.

the curve in the next image covers the most area on the common objects. This can be simply satisfied by setting the initial curve as the rectangle with small distance to the image edge. We can see that this initial curve setting can be used for all image pairs, which means that the proposed method does not need to manually set the initial curve.

Note that other initial curve settings, such as the saliency map-based initial curve setting or manual setting, can also be used as the initial curve setting.

In this paper, we set the initial curve as a rectangle with small distance ($v = 5$) to the image edge; some examples are shown in Figure 3. The iteration stops when the

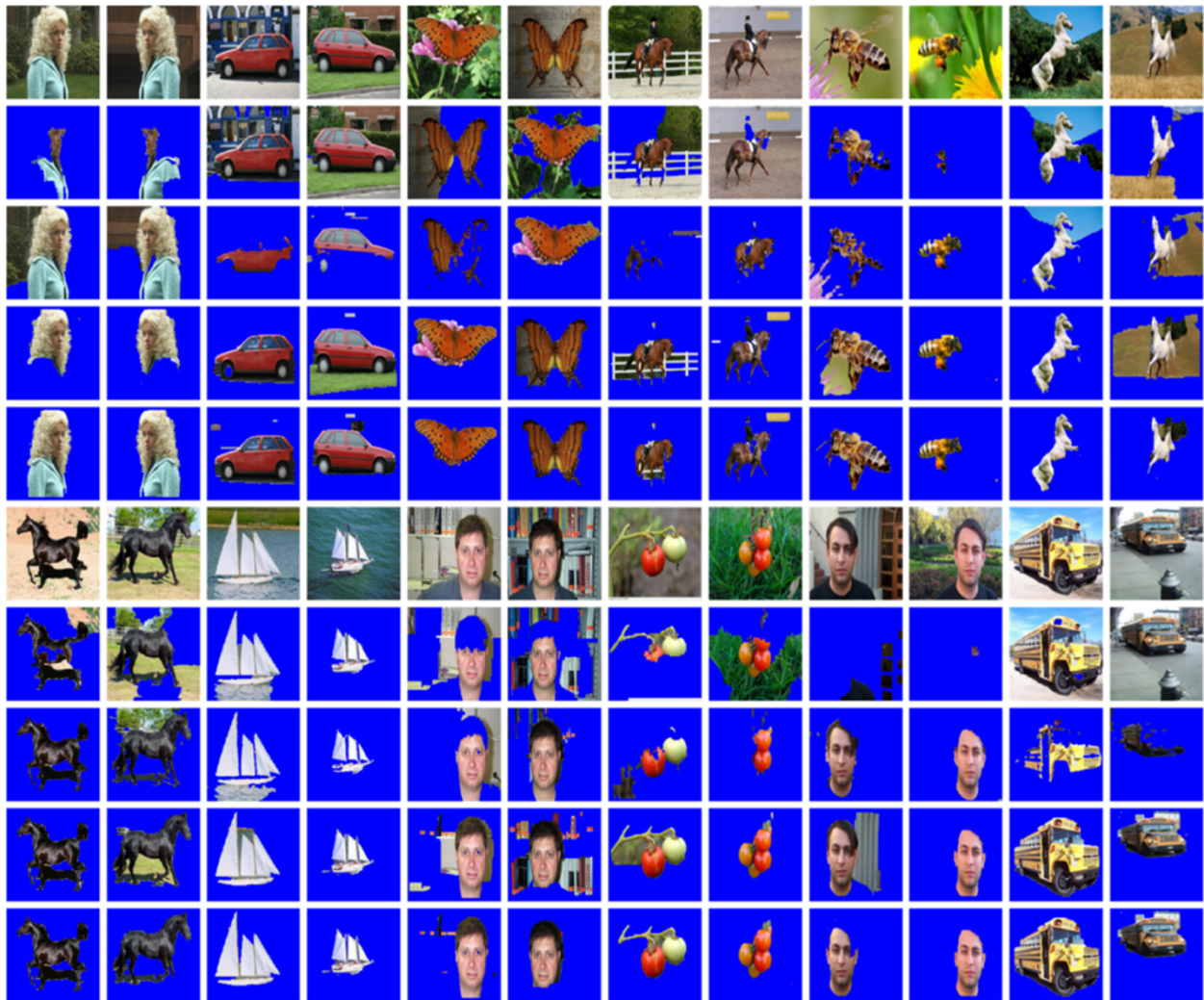


Figure 4 The results by the proposed method. For each block, the first row shows the original images. The results by the methods in [10,15,21] and the proposed method are shown in the second row to the last row, respectively.

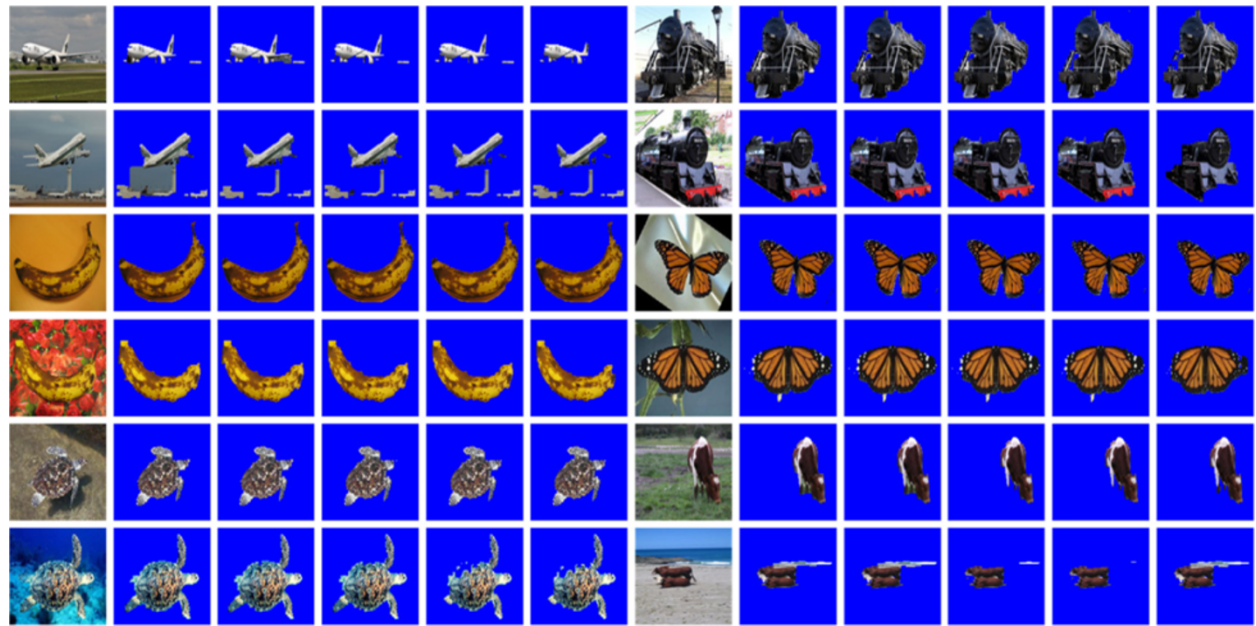


Figure 5 The segmentation results under various λ . For each image block, the original images are shown in the first column. The results with $\lambda = 0.1, 0.2, 0.3, 0.4$, and 0.5 are shown in the second-to-the-last column, respectively.

difference between the old segmentation and new segmentation is less than a threshold T_s . The algorithm of the proposed method is shown in Algorithm 1.

Algorithm 1 The algorithm for MGrabCut

Input: A pair of images, $z^l, l = 0, 1$.

Output: The co-segmentation labels: $\hat{\alpha}^l, l = 0, 1$.

% Parameter initialization

$K = 5$ for GMM, $\nu = 5, T_s = 0.1$.

% Iteration

while The stop condition is not satisfied **do**

(1) Obtain initial segmentation by setting the pixels inside the curve as the foreground and the other pixels as background.

(2) Obtain θ^l, \mathbf{k}^l by GMM for $l = 0, 1$.

(3) Obtain U according to Equation 8.

(4) Obtain V according to Equation 6.

(5) Obtain $\hat{\alpha}^l, l = 0, 1$ by solving Equation 11.

(6) Return to (2).

end while

Output $\hat{\alpha}^l, l = 0, 1$.

4 Experimental results

In this section, we introduce the experimental results. The subjective results and objective results are illustrated.

4.1 Datasets

We use the co-saliency database given in [27]. The co-saliency database contains 105 image pairs which

are collected from several well-known datasets, such as the Microsoft Research Cambridge image database, the Caltech-256 Object Categories database, and PASCAL VOC dataset. Each image pair contains a common object. All image pairs are considered in our method. Due to the complexities of the backgrounds and the changes of the foregrounds, the co-saliency dataset is challenging for co-segmentation task.

4.2 Results of the proposed method

We first introduce the parameter setting. In Equation 8, $\lambda = 0.2$. For GMM, we set the number of Gaussian distribution $N = 5$ for the foreground model and $N = 3$ for the background model. The stop condition of the iteration is set as the number of the iteration for simplicity. We set the stop number as 9.

The results of the proposed method are shown in Figure 4, where the first row for each image block shows the original images. The segmentation results by the proposed method are shown in the fifth row. We can see that the original images have complex backgrounds. Meanwhile, the proposed method successfully segments the common objects from these images. For example, the ‘bus’ in the last image pair *schoolbus* are segmented from the original images although the backgrounds are complex.

We also compare our method with GrabCut [21] and several existing co-segmentations such as [10,15]. Joulin et al. in [10] proposed co-segmentation model using discriminative clustering and spectral clustering method. In the method, a supervised classifier trained from a label of



Figure 6 The segmentation results under different initial curve settings.

Table 1 The comparison results compared with the existing methods in terms of the error rate

Name	Method of [10]	Method of [15]	Method of [21]	Our results
amira	0.0970	0.2633	0.1562	0.1289
banan	0.0645	0.1419	0.0327	0.0295
cdcartoona	0.1684	0.0351	0.0671	0.0186
cdbeabeara	0.2722	0.2634	0.0462	0.0207
cdboara	0.0945	0.0877	0.0391	0.0189
cdboata	0.1424	0.1727	0.1761	0.0398
cdcora	0.2934	0.3749	0.3249	0.2253
cddoga	0.1868	0.5232	0.1630	0.2000
cdleopardb	0.2913	0.1696	0.0589	0.0861
cdflamingoa	0.2510	0.1839	0.0501	0.0159
cdpandad	0.5769	0.1322	0.0202	0.0269
cdpersona	0.3882	0.2899	0.3578	0.1958
duck	0.0885	0.0634	0.0766	0.0400
horse	0.0985	0.3317	0.0427	0.0314
kim	0.2115	0.1418	0.0396	0.0432
llama	0.2115	0.2005	0.0430	0.1126
mcow	0.2030	0.1319	0.0159	0.0243
pvoccarb	0.5565	0.2323	0.0688	0.0595
pvochorseb	0.2903	0.1074	0.0561	0.0651
pvoccata	0.2184	0.1033	0.0743	0.0137
pvocmotoa	0.5180	0.2718	0.0184	0.0498
pvocsheepb	0.4949	0.3688	0.5120	0.3354
pvoctrain	0.1660	0.3507	0.1237	0.0575
ring008a	0.4584	0.1339	0.0269	0.0077
ring010a	0.2721	0.0929	0.0225	0.1047
ring014a	0.1770	0.2859	0.0199	0.0232
ring017a	0.3957	0.0471	0.6477	0.2988
ring023a	0.0523	0.0113	0.0047	0.0036
ring029a	0.4178	0.0945	0.2015	0.0104
stone	0.0093	0.1612	0.0146	0.0069

the images corresponds to a separation. The label leading to the maximal separation of the two classes is the co-segmentation result. The searching problem is solved by relaxing to a continuous convex optimization problem. Superpixels are generated by the method in Ncuts [28]. The results by the method in [10] are shown in the second row of each image block in Figure 4. It is seen that the common objects are successfully segmented from original images by [10], such as the 'boats' in *boats*. Meanwhile, there are unsuccessful segmentations, such as first image pairs *kim*. These unsuccessful segmentations are caused by the complexity and similarity of the background.

The method in [15] focuses on segmenting multiple common objects, which uses color information to label the similar objects. By using linear anisotropic diffusion method into co-segmentation, the co-segmentation is molded as a K -way segmentation problem that maximizes the temperature on anisotropic heat diffusion. Greedy algorithm is employed for optimization. In the experiment, the code released by the author is used. The intra-image Gaussian weights and the number of segments (K) are adjusted to obtain more accurate results. The results by the method in [15] are shown in the third row of each image block in Figure 4. We can see that the method achieves successful segmentation on several classes, such as *boats* and *faces2*. Unsuccessful results are also obtained,

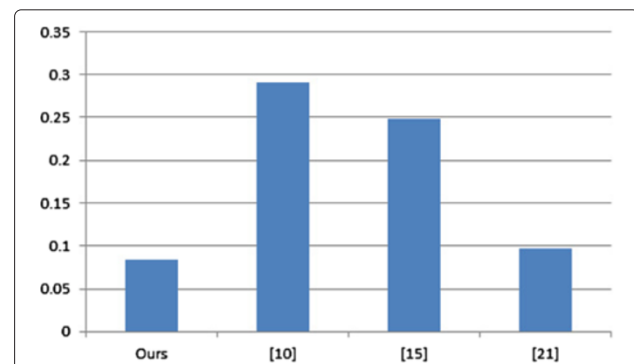
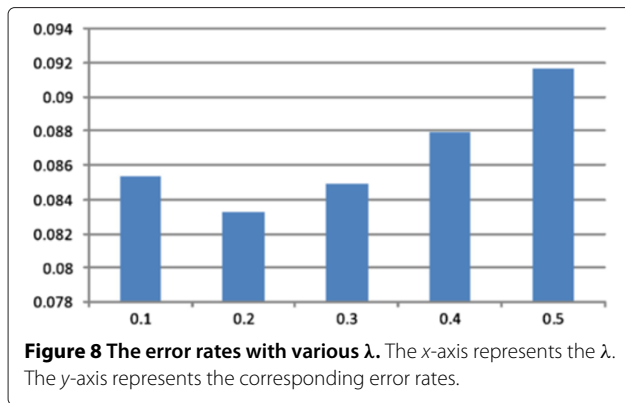


Figure 7 The error 1.0 rate by the methods in [10,15,21], and the proposed method.



such as *kim* and *schoolbus*. The reason is mainly caused by the fact that the complex background interferes with the common object extraction.

For GrabCut, we use the same initial curve for fair comparison. The results by the GrabCut-based method are shown in the fourth row in each image block of Figure 4. It is seen that GrabCut successfully segments the common objects from the original images, such as the ‘car’ in the first image of *car*. There are also unsuccessful segmentations, such as the ‘butterfly’ in the first image of *butterfly* where the red flower is also segmented as the foreground. The unsuccessful segmentations are caused by the fact that it is not enough to distinguish the objects from the background by only considering a single image. For example, the red flower is located inside the initial curve. Hence, GrabCut segments the red flower as the foreground. For MGrabCut, the red flower is segmented as the background since there are no similar regions in the next images.

Furthermore, we show the segmentation results under different scale λ which balances the foreground potential that is similar to the foregrounds in the current image or the other image. The results by various λ are shown in Figure 5, where the original images are shown in the first column of each image block. The results with $\lambda = 0.1, 0.2, 0.3, 0.4$, and 0.5 are shown in the second-to-the-last column, respectively. Six image pairs are shown. We can see that the proposed method is robust to λ . Meanwhile, slight differences are obtained by adjusting λ . A small λ results in segmentation similar to single-image segmentation, which may contain redundant regions, such as the segmentation of *plane*. While a large λ induces to the segmentation of common objects. However, several regions may be lost, such as the segmentation of *train*. Hence, we set $\lambda = 0.2$ for the trade-off

between single-image segmentation and common object segmentation.

Figure 6 displays some segmentation results under different initial curve settings. Three image pairs are shown. For each image pair, we segment the common objects by three initial curve settings, i.e., the initial curves that cover most parts of the common objects, the initial curves that partially cover the common objects, and the initial curves that cover only one of most parts of the common objects. The results of the three initial curve settings are shown in each row. From Figure 4, we can see that the proposed method can achieve successful segmentation in these image pairs with various initial curve settings, which demonstrates that the proposed model is robust to the initial curve setting.

4.3 Objective results

We introduce next the objective evaluation. We evaluate the segmentation performance based on the error rate which is defined as the ratio of the number of wrongly segmented pixels to the total number of pixels. The error rate is small when the object is accurately segmented. Since there are 105 image pairs, we only show the error rates of the 30 image pairs here. The error rates are shown in Table 1. We can see that the proposed method successfully segments the common objects in most of the image pairs. Meanwhile, there are several unsuccessful segmentations, such as ‘cdcora’ and ‘pvocsheepb’. The reason for the unsuccessful segmentation is that the common objects have color variations, which does not fulfill our assumption that the common objects have similar colors.

The error rates of the existing method such as the methods in [10,15,21] are also shown for comparison. From the results, we can see that the proposed method achieves the lowest error rates in most of the image pairs. We also calculate the average error rate of all image pairs for comparison. The error rates by the existing methods and the proposed method are shown in Figure 7. We can see that the proposed method obtains the smallest mean error rate, which demonstrates the effectiveness of the proposed method. Compared with the original GrabCut method [21], we can see that the MGrabCut achieves lower error rates. The improvements are a benefit from considering foreground similarity.

The error rates with various λ are shown in Figure 8, where the error rate is shown in the y -axis. The x -axis displays different λ . We can see that the error rate is smallest when $\lambda = 0.2$, which means that

Table 2 The computational complexities for the comparison methods and our method

Methods	[7]	[10]	[29]	[21]	[16]	Ours
Computational complexity	$O(n \log n)$	$O(n^2)$	$O(n^2)$	$O(n \log n)$	$O(n^2)$	$O(n \log n)$

considering the foregrounds of both the current image and the other image will result in a more accurate co-segmentation.

4.4 Computational complexity analysis

In the proposed method, the minimization is achieved by graph cut algorithm. Since there are pairs of images, the computational complexity of the proposed method is two times that of the graph cut algorithm $O(n \log n)$, which equals to $O(n \log n)$. Hence, the computational complexity of the proposed method is $O(n \log n)$, which has the same computational complexity with the existing graph cut-based segmentation methods such as [7,21]. Meanwhile, because of the efficiency of the graph cut minimization, the computational complexity of the proposed method is lower than the computational complexities of the other co-segmentation methods such as [10,16,29], as shown in Table 2. It is seen that the computational complexity of the proposed method is low compared with the existing methods.

5 Conclusions

This paper proposes a new co-segmentation model by extending GrabCut to MGrabCut. To consider common object segmentation, we introduce the foreground appearance model of the other image to construct the unary term of current images. Both the foreground similarity and background consistency are considered to design our model. The common objects are finally segmented by mutually updating the foreground model and the background model of two images. The experimental results demonstrate the effectiveness of the proposed method. In the future, we will extend the proposed model to solve images with more than two images. Furthermore, other local features will be considered for more accurate segmentation.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Center for Radio Administration & Technology Development, Xihua University Chengdu, Sichuan, 610039, China. ²College of Engineering and Science, Victoria University, Melbourne, Victoria, 8001, Australia. ³School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, South Australia, 5005, Australia. ⁴Department of Engineering, Faculty of Technology and Science, University of Agder, 4898 Grimstad, Norway.

Acknowledgements

This work is partially supported by the research fund of Sichuan Key Laboratory of Intelligent Network Information Processing (SGXZD1002-10), Sichuan Key Technology Research and Development Program (2012GZ0019, 2013GZX0155), and Xihua University Key Laboratory Development Program (szjj2011-021).

Received: 20 November 2012 Accepted: 14 March 2013

Published: 20 April 2013

References

1. G Mori, X Ren, A Efros, J Malik, in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2. Recovering human body configurations: combining segmentation and recognition (IEEE Computer Society, Washington, DC, 27 June–2 July 2004), pp. 326–333

2. C Gu, JJ Lim, P Arbeláez, J Malik, in *IEEE Conference on Computer Vision and Pattern Recognition*. Recognition using regions (IEEE Computer Society, Florida, June 2009), pp. 20–25
3. N Jacobson, YL Lee, V Mahadevan, N Vasconcelos, T Nguyen, A novel approach to fruc using discriminant saliency and frame segmentation, *IEEE Trans. Image Process.* **19**(11), 2924–2934 (2010)
4. F Jing, M Li, HJ Zhang, B Zhang, Relevance feedback in region-based image retrieval. *IEEE Trans. Circuits Syst. for Video Technol.* **14**(5), 672–681 (2004)
5. C Rother, V Kolmogorov, T Minka, A Blake, in *IEEE Conference on Computer Vision and Pattern Recognition*. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs (IEEE Computer Society, New York, June 2006), pp. 17–22
6. L Mukherjee, V Singh, CR Dyer, in *IEEE Conference on Computer Vision and Pattern Recognition*. Half-integrality based algorithms for cosegmentation of images (IEEE Computer Society, Florida, June 2009), pp. 20–25
7. DS Hochbaum, V Singh, in *International Conference on Computer Vision*. An efficient algorithm for co-segmentation (IEEE Inc., Kyoto, 29 September–2 October, 2009)
8. S Vicente, V Kolmogorov, C Rother, in *European Conference on Computer Vision*. Cosegmentation revisited: models and optimization (Springer-Verlag, Grete, September 2010), pp. 5–11
9. D Batra, D Parikh, A Kowdle, T Chen, J Luo, in *IEEE International Conference on Image Processing*. Seed image selection in interactive cosegmentation (IEEE Computer Society, Cairo, November 2009), pp. 7–10
10. A Joulin, F Bach, J Ponce, in *IEEE Conference on Computer Vision and Pattern Recognition*. Discriminative clustering for image co-segmentation (IEEE Computer Society, San Francisco, June 2010), pp. 13–18
11. D Batra, A Kowdle, D Parikh, in *IEEE Conference on Computer Vision and Pattern Recognition*. Icoseg: interactive co-segmentation with intelligent scribble guidance (IEEE Computer Society, San Francisco, June 2010), pp. 13–18
12. L Mukherjee, V Singh, J Peng, in *IEEE Conference on Computer Vision and Pattern Recognition*. Scale invariant cosegmentation for image groups (IEEE Computer Society, Colorado Springs, June 2011), pp. 20–25
13. K Chang, T Liu, S Lai, in *IEEE Conference on Computer Vision and Pattern Recognition*. From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model (IEEE Computer Society, Colorado Springs, June 2011), pp. 20–25
14. S Vicente, C Rother, V Kolmogorov, in *IEEE Conference on Computer Vision and Pattern Recognition*. Object cosegmentation (IEEE Computer Society, Colorado Springs, June 2011), pp. 20–25
15. G Kim, EP Xing, L Fei-Fei, T Kanade, in *International Conference on Computer Vision*. Distributed cosegmentation via submodular optimization on anisotropic diffusion (IEEE Inc., Barcelona, November 2011), pp. 6–13
16. F Meng, H Li, G Liu, KN Ngan, Image cosegmentation by incorporating color reward strategy and active contour model. *IEEE Trans. on Cybern.* **43**(2), 725–737 (2013)
17. F Meng, H Li, G Liu, KN Ngan, Object co-segmentation based on shortest path algorithm and saliency model. *IEEE Trans. Multimedia.* **14**(5), 1429–1441 (2012)
18. M Collins, J Xu, L Grady, V Singh, in *IEEE Conference on Computer Vision and Pattern Recognition*. Random walks for multi-image cosegmentation: quasiconvexity results and gpu-based solutions (IEEE Computer Society, Rhode Island, June 2012), pp. 16–21
19. A Joulin, F Bach, J Ponce, in *IEEE Conference on Computer Vision and Pattern Recognition*. Multi-class cosegmentation (IEEE Computer Society, Rhode Island, June 2012), pp. 16–21
20. F Meng, H Li, G Liu, in *IEEE International Symposium on Circuits and Systems*. Image co-segmentation via active contours (IEEE Inc., Seoul, May 2012), pp. 20–23
21. C Rother, V Kolmogorov, A Blake, Grabcut: interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **23**(3), 309–314 (2004)
22. YY Boykov, MP Jolly, in *International Conference on Computer Vision*. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images (IEEE Inc., Vancouver, July 2001), pp. 7–14

23. N Zeng, Z Wang, Y Li, M Du, X Liu, A hybrid ekf and switching pso algorithm for joint state and parameter estimation of lateral flow immunoassay models. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **9**(2), 321–329 (2012)
24. N Zeng, Z Wang, Y Li, M Du, X Liu, Identification of nonlinear lateral flow immunoassay state-space models via particle filter approach. *IEEE Trans. Nanotechnol.* **11**(2), 321–327 (2012)
25. B Zineddin, Z Wang, X Liu, Cellular neural networks, Navier-Stokes equation and Trans, microarray image reconstruction. *IEEE. Image Process.* **20**(11), 3296–3301 (2011)
26. J Rubio, J Serrat, A López, N Paragios, in *IEEE Conference on Computer Vision and Pattern Recognition*. Unsupervised co-segmentation through region matching (IEEE Computer Society, Rhode Island, June 2012), pp. 16–21
27. H Li, KN Ngan, A co-saliency model of image pairs. *IEEE Trans. Image Process.* **20**(12), 3365–3375 (2011)
28. J Shi, J Malik, Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000)
29. TR Raviv, N Socheny, N Kiryati, Shape-based mutual segmentation. *Int. J. Comput. Vis.* **79**, 231–245 (2008)

doi:10.1186/1687-5281-2013-20

Cite this article as: Gao et al.: A mutual GrabCut method to solve co-segmentation. *EURASIP Journal on Image and Video Processing* 2013 **2013**:20.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com