Feature Adaptive Co-Segmentation by Complexity Awareness

Fanman Meng, Hongliang Li, *Senior Member, IEEE*, King Ngi Ngan, *Fellow, IEEE*, Liaoyuan Zeng, and Qingbo Wu

***Abstract*— In this paper, we propose a novel feature adaptive co-segmentation method that can learn adaptive features of different image groups for accurate common objects segmentation. We also propose image complexity awareness for adaptive feature learning. In the proposed method, the original images are first ranked according to the image complexities that are measured by superpixel changing cue and object detection cue. Then, the unsupervised segments of the simple images are used to learn the adaptive features, which are achieved using an expectation-minimization algorithm combining *l*1-regularized least squares optimization with the consideration of the confidence of the simple image segmentation accuracies and the fitness of the learned model. The error rate of the final co-segmentation is tested by the experiments on different image groups and verified to be lower than the existing state-of-the-art co-segmentation methods.**

***Index Terms*— Cosegmentation, distance metric learning, image complexity analysis.**

I. Introduction

IN COMPUTER vision area, image segmentation [1]–[8] is a process of segmenting objects from images. The goal of image segmentation is bottom up and unsupervised segmentation of general images. As a key branch of image segmentation, co-segmentation [9]–[24] is to segment common objects from an image group. By assuming a group of images contain common objects, co-segmentation only requires additional images containing the same or similar target objects for accurate segmentation.

The co-segmentation methods are generally developed by adding foreground similarity into single image segmentation models, such as Markov Random Filed (MRF) segmentation [9]–[12], [17], [22], heat diffusion segmentation [19], clustering based segmentation [14], [20], and random walker segmentation [21]. Using the additional foreground similarity constraints guarantees the segmentation of the common objects only, which results in more accurate segmentation than single image segmentation.

TABLE I
THE FEATURES USED IN THE EXISTING CO-SEGMENTATION METHODS

Method	Features	Method	Features
[9]	Color	[17]	SIFT
[10]	Color	[18]	33 features
[11]	Color	[19]	Color
[12]	Color	[20]	SIFT and Color
[14]	Color and SIFT	[21]	Color
[25]	fixed feature	[15]	Color
[22]	Color, LBP, SIFT	[16]	Color, SIFT, Texture
[26]	Color, Shape		

The accuracy of co-segmentation is significantly dependent on the efficiency of the foreground similarities measurement. Many region features, such as color histogram [9]–[13], SIFT [14], [20], contour descriptor [18] and local binary pattern descriptor [22], have been used to evaluate the foreground similarity. Furthermore, $\ell 1$-norm, $\ell 2$-norm, reward strategy and χ^2 distance, were usually used for the feature distance calculation.

However, the existing co-segmentation methods cosegment different image classes using a fixed foreground similarity measurement without change. In general, the fixed features are manually selected or learned from the training data set [18] before co-segmentation. We display the fixed features used in the existing co-segmentation methods in Table I. Using fixed similarity measurement in co-segmentation may lead to some problems in realistic applications. Firstly, since the similar features of the common objects vary in different image groups, the fixed feature can not accurately measure the foreground similarities of different classes, which results in the unsuccessful co-segmentation. Secondly, for the images whose common object varies significantly, a combination of the general features will be required to accurately measure the foreground similarity. However, designing the combinational feature model creates high complexity for the manual selection manner. Thirdly, the training features from the fixed training data may lead to low feature accuracy because the fixed training data cannot accurately represent the similar features

Manuscript received January 4, 2013; revised May 30, 2013 and July 23, 2013; accepted July 31, 2013. Date of publication August 15, 2013; date of current version September 27, 2013. This work was supported in part by the NSFC under Grant 61271289, in part by the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 20110185110002, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2012AA011503, and in part by the Fundamental Research Funds for the Central Universities under Grant E022050205. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Carlo S. Regazzoni.

F. Meng, H. Li, L. Zeng, and Q. Wu are with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: fm-meng@qq.com; hlli@uestc.edu.cn; lyzeng@uestc.edu.cn; wqb.uestc@gmail.com).

K. N. Ngan is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, and also with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (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/TIP.2013.2278461

for a specific class. Hence, to obtain the features that are adaptive to different image classes is necessary to improve the accuracy of the co-segmentation.

To obtain the adaptive features, we note that the common objects in a simple image can be easily extracted by the figure-ground segmentation methods, such as object detection based segmentation and saliency detection based segmentation. It is seen that these simple image segments can be used to learn the features adaptable to each image group and thus increase the accuracy of the co-segmentation. Furthermore, when the original images are collected from various sources, certain images will contain simple background images. Thus, the simple image segments can provide adaptive training data for accurate feature learning.

In this paper, we propose a feature adaptive image co-segmentation method to improve the accuracy of the co-segmentation when the similar features are unknown (first reported in [27]). The simple image segments are used to learn the adaptive features. The proposed method consists of four steps. In the first step, we evaluate the image complexities by the superpixel changing cue and the object detection cue. We then select simple images and segment the initial segments by figure-ground segmentation method. In the third step, we represent the features as a linear combination of the common features, and we learn the linear combination parameters by the EM based algorithm. In the last step, the common objects are segmented according to the learned feature model. We test the performance of the proposed co-segmentation method in terms of error rate in different image groups. The results demonstrate that the lower error rates can be obtained by the proposed method.

The structure of this paper is organized as follows. The related work is discussed in Section II. In Section III and IV, we present the proposed co-segmentation method by demonstrating the image complexity analysis, the adaptive feature learning model and the final co-segmentation achieved by using the learned features. Section V and VI show the experiment of the proposed method and the discussion of the results. Finally, in Section VII, the conclusion is given.

II. Related Work

Co-segmentation is usually modeled as an optimization process with the consideration of the foregrounds similarity constraints added into the single image segmentation models. The MRF based co-segmentation method was first presented by C. Rother *et al.* [9], which segmented common objects through adding foreground similarity constraint into traditional MRF based segmentation methods. $\ell 1$-norm was used to represent the foreground similarity, and the co-segmentation energy was minimized by trust region graph cuts (TRGC) method. Based on Rother's work, several MRF co-segmentation methods deal with the optimization problem using other constraints. In the work of L. Mukherjee *et al.* [10], $\ell 1$-norm was replaced by $\ell 2$-norm and the Pseudo-Boolean optimization was used for the minimization. Instead of penalizing foreground difference, D. S. Hochbaum and V. Singh [11] rewarded the foreground similarity, which can result in the tractable energy function optimization by graph-cuts algorithm. In [12], S. Vicente *et al.* modified Boykov-Jolly model as the foreground similarity measurement, and employed Dual Decomposition to minimize the energy function. Note that in these co-segmentation methods, the common objects are assumed to contain similar colors.

A. Joulin *et al.* [14] segmented common objects using the clustering strategy, in which a classifier produced by spectral clustering technique and positive definite kernel was used as a co-segmentation. The most discriminative classifier was then found as the final co-segmentation by solving a continuous convex searching optimization problem. Both color and SIFT features were used in this work. An interactive co-segmentation method was proposed in the work of D. Batra *et al.* [15], which can segment common objects through user interaction guided by an automatic recommendation system to correct the inconsistent segmentation. In [16], by observing that the rank of the matrix corresponding to the foreground regions still equals to one even if the common objects contain the scale variants, L. Mukherjee *et al.* proposed a scale invariant co-segmentation method which intended to find a matrix comprised of common objects with rank of one. K. Chang *et al.* [17] designed a novel global energy term to represent the foreground similarity and background consistency. Combined with the foreground potentials measured by co-saliency model, the final energy function is submodular which can be minimized by the graph-cut algorithm. S. Vicente *et al.* [18] presented an object co-segmentation method to segment objects of interest. An off-line learning method was used to select the discriminative features from the common features through random forest regressor, which leads to the segmentation of only the interesting common objects. G. Kim *et al.* in [19] used anisotropic heat diffusion segmentation method to segment common objects of multiple classes from a large scale of images group. In Kim's work, the common objects were assumed to contain similar colors, which will result in unsuccessful co-segmentation when the common objects contain other similar features. Y. Chai *et al.* in [25] proposed a Bi-level co-segmentation method (BiCoS) for image classification. Chai's method performs the Grabcuts based segmentation with the initializations of the linear SVM based class models and alternately updates the class models and segmentation until convergence to achieve the image segmentation and classification. Instead of sharing descriptor at the level of individual pixels, Chai's method shares a richer descriptor at the level of superpixels stacked from multiple general sub-descriptors which represent the superpixels' color distribution, SIFT distribution, size, location within the image, and shape. The use of the richer descriptor can improve the co-segmentation accuracy. However, the feature model in the method in [25] reminds a combination of several existing features and the features adaptive to each specific class group is not discussed.

Recently, A. Joulin *et al.* [20] presented a multi-class co-segmentation method which extends the discriminative clustering based co-segmentation [14] to segment the common objects of multiple image classes. Joulin designed a new energy function which consists of spectral-clustering

term and discriminative term. The spectral-clustering term can divide each image into visually and spatially consistent labeled regions, and the discriminative term can maximize the class separability in the image group. The energy function can be finally optimized by using EM algorithm. In Joulin's work, the EM algorithm is to perform image segmentation, where E-step estimates the label of each pixel, and the M-step estimates the parameters of the discriminative classifier. Instead, in our method the EM algorithm is used for the feature learning, where E-step estimates the confidence of the initial segmentation, and the M-step estimates the parameters of the feature model. It is seen that the fixed features are used in the model in [14], while we use adaptive features for more accurate co-segmentation. The co-segmentation method proposed by M. Collins *et al.* [21] adds foreground consistency into the random walker based segmentation method which leads to a tractable energy minimization and speeds up the co-segmentation algorithm compared with the MRF based co-segmentation. J. Rubio *et al.* [22] segmented the common objects by proposing a new graph matching based foreground similarity measurement and alternatively updating the saliency detection and the segmentation, which can enhance the co-segmentation accuracy. In [26], Meng *et al.* used the graph theory to segment the common objects from a large scale image group. A digraph was constructed based on the local region similarity and the co-saliency values. The co-segmentation was then formulated as a shortest path problem, which can be solved by using dynamic programming. In the methods discussed in this paragraph, the fixed features are used to measure the foreground similarity for successful co-segmentation, which will cause unsuccessful co-segmentation when different common objects contain different types of similar features. Hence, in that situation, features that can adapt to different image classes are needed to improve the co-segmentation accuracy.

Another related work is the metric learning [28]–[32], which aims to improve the performance of many applications by learning more accurate distance metric. In general, an objective function representing the consistency between the metric and the training data is first defined. Then, the metric learning is formulated as maximizing the fitness between the metric and the data to obtain the best distance metric according to evaluating the distance parameters. In general, Mahalanobis distance ($d(x, y) = (xy)^T A(xy)$) with parameter A was usually used as the basic distance. Other basic distance representation, such as randomized binary trees, is also employed. The metric learning has been widely used in many computer vision tasks, such as image alignment [29], image classification [31], data clustering, and face recognition. Nguyen *et al.* [29] introduced metric learning in parameterized appearance model based image alignment to overcome the local minima optimization problem. The convex quadratic programming was used for the metric learning. Eric Nowak *et al.* considered the domain specific knowledge in the metric learning for accurate image comparing [30]. This method rewarded the distinct knowledge of the object in the metric learning in terms of a set of randomized binary trees, which resulted in more accurate object comparing. In the work of Nakul Verma *et al.* [31], a hierarchy metric learning model rather than single metric leaning was proposed for the image classification. A set of Mahalanobis distance metrics related to the class taxonomy were trained in a probabilistic nearest-neighbor classification framework. By representing metric in a hierarchal way, accurate distinct distance can be learned. Mensink *et al.* [32] used metric learning to enhance the Large-scale image annotation. The Mahalanobis distance based metric was learned for both k-NN classification and nearest class mean classifier used in the image annotation. To consider the real-time learning in the large-scale datasets, a small fraction of the training data were considered in each iteration by combing stochastic gradient descend (SGD) algorithms and product quantization.

III. The Proposed Co-Segmentation Method

In the proposed method, we learn the adaptive features from the initial segments of simple images. We first select simple images from the image group by image complexity analysis. Then, we use the figure-ground segmentation to extract the initial segments from the simple images, and learn the adaptive feature model based on these segments. The learned feature model is finally used to achieve image co-segmentation. The flowchart of the proposed co-segmentation method is shown in Figure 1, which consists of four steps, i.e., image complexity analysis, simple image segmentation, adaptive feature learning, and co-segmentation.

A. Image Complexity Analysis

In our method, the simple image selection is to simplify the initial object extraction. We can observe that the objects can be easily segmented from the images with simple background, while it is usually difficult to extract the objects from the complex backgrounds. Hence, we define a simple image as the image with homogenous background. On the contrary, an image with complicated background is treated as complex image. In this paper, the image complexity is measured by two cues, i.e., the over-segmentation based image complexity analysis and the object detection based image complexity analysis.

1) Over-Segmentation Based Image Complexity Analysis: It can be observed from the realistic images that the homogenous background contained in a simple image will keep a single local region in the edge based hierarchical over-segmentation, while a complicated background containing many different appearances will be separated into many local regions. We can see that the number of local regions of a simple image is small and stable in the hierarchical over-segmentation results. But a complex image will be assigned a large number of local regions. Motivated by such observation, we use the local regions number in the edge based hierarchical over-segmentation to measure the image complexity. In the measurement, the original image $I_i, i = 1, \cdots, N_i$ is first over-segmented into local regions by the edge based over-segmentation method with different scales. Then, the sum of local region numbers over all scales is counted as the score of the measurement. For I_i, the score of the over-segmentation

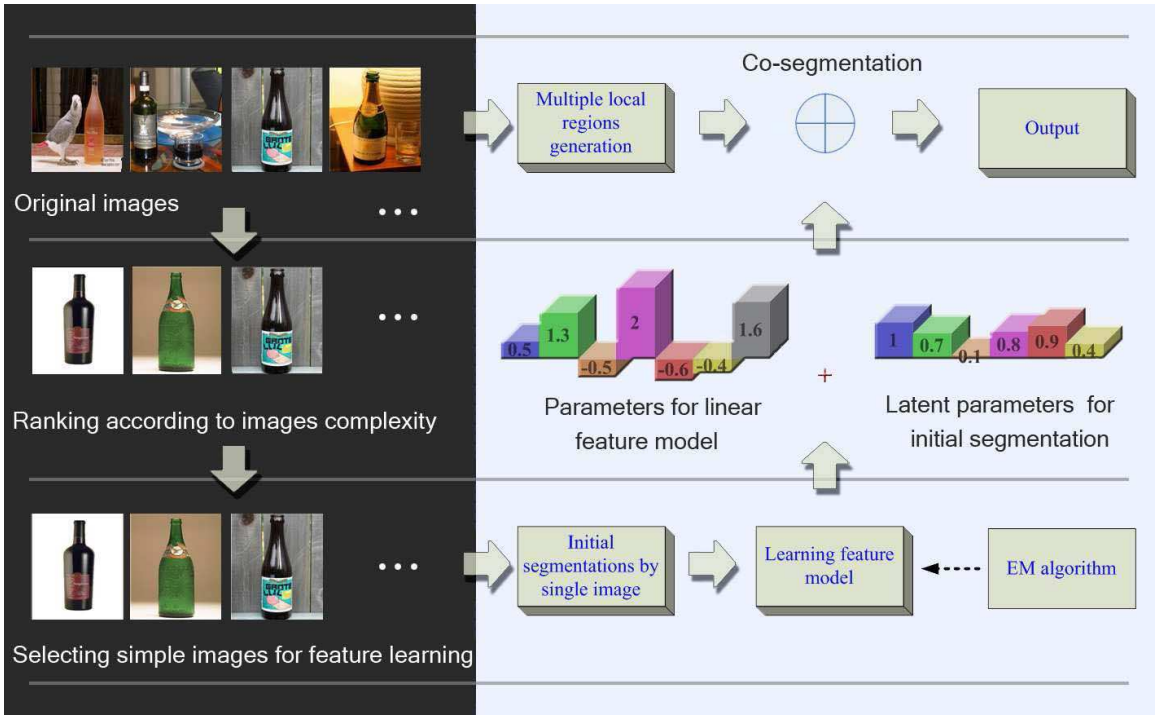


Fig. 1. The flowchart of the proposed method.

based image complexity analysis C_i^1 is calculated by

$$C_i^1 = \sum_{k=1}^{K} n_k^i \tag{1}$$

where n_k^i is the number of the local regions in the k-th scale over-segmentation, K is the number of scales. It is seen that simple image will have small C_i^1. Otherwise, large values will be assigned to the complex images. Based on C_i^1, we sort the image complexity in ascending order and obtain the sorted order ρ_1. Meanwhile, we record I_i by the position (η_1^i) of I_i in ρ_1 and obtain $\eta_1 = \{\eta_1^1, \eta_1^2, \cdots, \eta_1^{N_i}\}$.

We use the method in [33][1] to obtain the hierarchical image over-segmentation. In the method [33], the oriented watershed transform (OWT) is used to form the initial regions. Then, the greedy graph-based region merging algorithm is used to construct the hierarchy of the regions. The hierarchy of the regions is finally treated as an Ultrametric Contour Map (UCM). By setting different thresholds (the scale K) on the UCM, we can obtain a series of over segmentation results. In this paper, we set $K = 40$, 50, 75, 100, 150 and 200 for the hierarchical over-segmentation.

Fig. 2 shows the hierarchical over-segmentation results of three images, where the simple image (the top image) and the complex images (the middle and bottom images) are displayed for comparison. The original images are shown in the first column. The rest columns show the over-segmentation results at different scales. The corresponding scales for the columns are represented above each column. The number of the local region for each over-segmentation result is shown below the over-segmentation result. It is seen that the sum number of the local regions over all scales are 50, 107 and 124 from the top row to the bottom row, respectively. We can see that the number of the simple image is obviously smaller than the number of the complex image, which verifies the validity of the over-segmentation based image complexity analysis. The sorted images by C^1 are shown in the top row of Fig. 4, where the top 12 simple images of *Bottles* are shown. We can see that simple images can be selected by the over-segmentation based image complexity analysis.

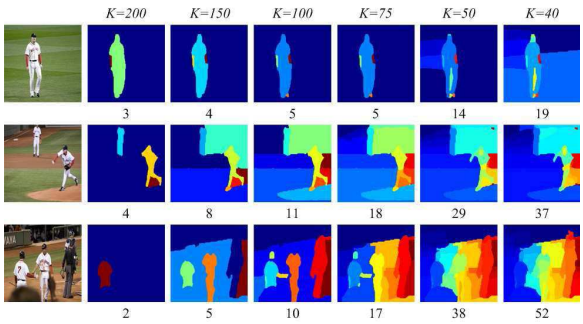


Fig. 2. The over segmentation results by the method in [33]. The first column: Original images. The rest columns: the segmentation results under different K. The C^1 for three images are 50, 107 and 124 from first row to last row, respectively.

2) Object Detection Based Image Complexity Analysis: It is observed that the simple images usually contain single object, while the complex images include many objects, especially the objects in the backgrounds. By performing the object detection method on the simple images, the detected windows will focus on the object, and result in the compact detection. But for complex images, the detections will locate on different objects, and result in dispersive detection. To clearly illustrate this

[1] http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html

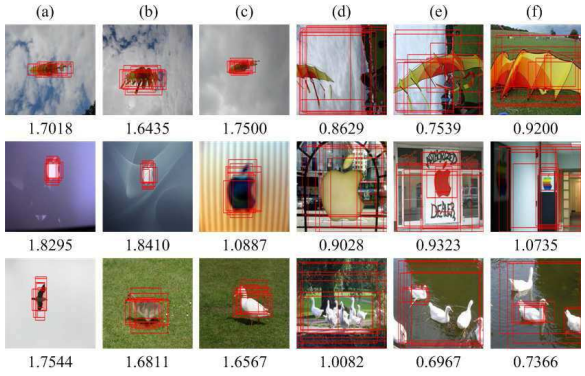


Fig. 3. The detection results of the simple images and the complex images in [34]. (a)-(c): the detection results of simple images. (d)-(f): the detection results of complex images.

observation, we show some detection results in Fig. 3, where the results of the simple images and the complex images are shown in Fig. 3 (a)-(c) and Fig. 3 (d)-(f), respectively. It can be seen that compact windows are obtained from the simple images, such as the apple logo with a simple blue background. Meanwhile, the scatter windows are detected in the complex images, such as ducks under the tree.

We use the scatter degree to evaluate the image complexities. We first perform a sliding window based object detection method in each image I_i. The best N_w windows are selected for the complexity measurement. Then, we represent each window as a binary matrix $M_i, i = 1, \cdots, N_w$, where the size of the matrix is same to the size of the image, and the pixels within the window have value one and zero for the pixels outside the window. Next, we compute M by summing up all binary matrixes, i.e., $M = \sum_{k=1}^{N_w} M_k, 0 \leq M(j,l) \leq N_w$. The complexity of the image I_i is then measured by

$$C_i^2 = \frac{\sum_{(j,l)} \pi(M(j,l), T_w)}{\sum_{(j,l)} \pi(M(j,l), 1)} - \frac{\sum_{(j,l)} \pi(M(j,l), 1)}{\sum_{(j,l)} \pi(M(j,l), 0)} \quad (2)$$

where

$$\pi(a,b) = \begin{cases} 1 & if \quad a \geq b \\ 0 & else \end{cases} \quad (3)$$

It can be seen that there are two terms included in (2). The first term is to evaluate the scatter of the detection by measuring the ratio of the overlapped regions to the whole detected region. It prefers a large value when most overlapped regions focus on an object. In order to avoid the influence of the unsuccessful detections in the complex images, where most of the backgrounds are detected and included in the windows, we introduce the second term by measuring the ratio of the area of the detection region to the whole image region. It is seen that the unsuccessful detections will have low scores by the second term. We sort C_i^2 in descending order and obtain the sorted order ρ_2. We also record each image I_i by the position of the image η_2^i in ρ_2 and obtain $\eta_2 = \{\eta_2^1, \eta_2^2, \cdots, \eta_2^{N_i}\}$.

The method in [34][2] is used as the object detection. We set $N_w = 10$ and $T_w = 8$ for all image groups. In Fig. 3, we also display C_i^2 for each image. The values C_i^2 are shown below each image. It is seen that the simple images have larger C^2 than the values of the complex images, which demonstrates that the object detection based image complexity analysis can describe the complexities of these images. The sorted images based on the object detection based analysis method are shown in the middle row of Fig. 4, which shows the successful selection of the simple images by the object detection based image complexity analysis.

3) Combination of Image Analysis Methods: We combine the above two cues to obtain more accurate image ranking. We believe that the image I_i tends to be a simple image when the values of η_1^i and η_2^i are both small. Thus, we first represent each image I_i by

$$\eta^i = \eta_1^i + \eta_2^i \quad (4)$$

where η^i is the sum of the rankings measured by the two complexity analysis cues. Then, the final sorted order is obtained by sorting η^i in ascending order. The final sorted images of *Bottles* are shown in the bottom row of Fig. 4. Compared with the results in the top row and the middle row, we can see the more accurate sorting by the combined method.

B. Object Extraction from Simple Image

Based on the image complexity analysis, we select the top m simple images, and segment the initial segments $Q = \{Q_1, Q_2, \cdots, Q_m\}$ from these simple images using figure-ground segmentation method. In this paper, we use the saliency extraction based object segmentation method [35][3] to obtain the initial segments.

IV. Feature Learning

After initial segment generation, we next learn the adaptive features of the class. Here, we consider two requirements in the learning. Firstly, some unsuccessful segments may be obtained in the above initial object extraction step, which can interfere the feature learning and result in the inaccurate feature model. We need to avoid these interferences in the learning. Secondly, the learned feature model must fit the initial segment data very well.

A. Feature Model

In our method, the similarity between two initial segments Q_i and Q_j is measured by a linear feature model, i.e., a linear combination of the general region features. Assuming there are n general features, such as the features of color, shape and texture, we evaluate the similarity s_{ij} between two segments Q_i and Q_j by

$$s_{ij} = \omega_1(1 - x_1^{ij}) + \omega_2(1 - x_2^{ij}) + \cdots + \omega_n(1 - x_n^{ij}) \quad (5)$$

where $x_k^{ij} = d(f_k^i, f_k^j)$ is the distance between the k-th features (f_k^i and f_k^j) of the segments Q_i and Q_j, f_k^i denotes the k-th general feature of Q_i, $\omega_1, \cdots, \omega_n$ are the weighting coefficients of the features. In our method, we use five features such as color histogram, inner shape descriptor [36], SIFT descriptor [37], [38], self-similarity descriptor [39] and pHOG descriptor [40] as the general features. Chi-square distance is

[2]http://groups.inf.ed.ac.uk/calvin/objectness/

[3]http://cg.cs.tsinghua.edu.cn/people/~cmm/Saliency/Index.html

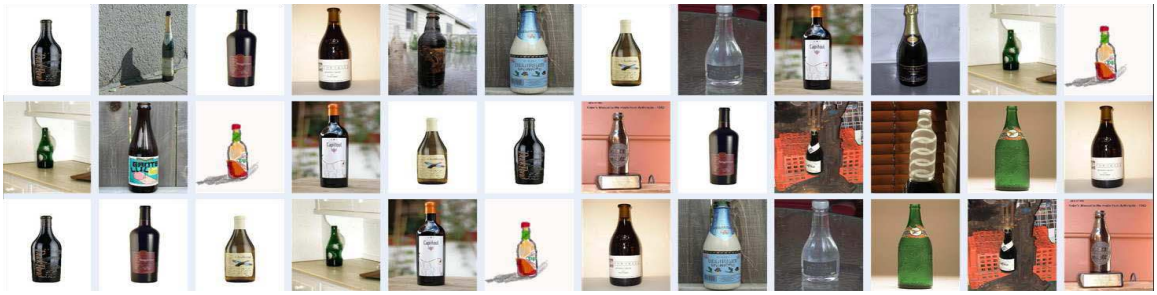


Fig. 4. The ranking of images using the proposed method. The top row: the ranking by the over-segmentation based image complexity analysis. The middle row: the ranking by the object detection based image complexity analysis. The bottom row: the final ranking by the proposed method.

used as the feature similarity evaluation. From (5), we can see that x_k^{ij} is calculated only by the k-th features, i.e., f_k^i and f_k^j. The measurement of the feature distance is only performed by the same feature types. Hence, the model is available although the dimensions of different type of features are not equal to each other.

Setting parameters $\theta = (\omega_1, \omega_2, \cdots, \omega_n)^T$ and

$$X_i = E - \begin{pmatrix} x_1^{i1} & x_2^{i1} & \cdots & x_n^{i1} \\ x_1^{i2} & x_2^{i2} & \cdots & x_n^{im} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{im} & x_2^{im} & \cdots & x_n^{im} \end{pmatrix} \tag{6}$$

$i = 1, \cdots, m$, we obtain

$$\begin{aligned} &S(X_1, \cdots, X_m, \theta) \\ &\quad = (s_{11}, \cdots, s_{1m}, s_{21}, \cdots, s_{2m}, \cdots s_{m1}, \cdots, s_{mm})^T \\ &\quad = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \theta = X\theta \end{aligned} \tag{7}$$

where E is a matrix with all elements 1.

Assuming initial segments are accurately segmented from the simple images, the distance between any pair of the initial segments approximately equals to 0. Hence, the target matrix S' of S is a $m^2 \times 1$ vector with all elements one. However, the feature self-similarities cannot provide useful discriminative information to distinguish the useful features from the other features. Furthermore, the feature learning is based on the feature similarities. The self-similarities of unsuccessful segmentation will interfere the feature learning and result in inaccurate feature model. Hence, we do not consider the self-similarities, and set the values corresponding to the self-similarities to 0 in X and S' such as the i-th row in X_i and the $((i-1)m+i)$-th element in S', $i = 1, \cdots, m$. Then, the parameters θ of the feature model that best fits X can be calculated by

$$\arg\min_{\theta} \|S - S'\|_2^2 + \alpha \cdot \|\theta\|_1 = \arg\min_{\theta} \|X\theta - S'\|_2^2 + \alpha \cdot \|\theta\|_1 \tag{8}$$

where α is the scale factor. However, there may have unsuccessful initial segments. Next, we learn the feature parameters by considering these bad segments.

B. Parameters Learning

Our goal is to find the parameter θ of the feature model that best fits the training data X and also discover the confidences of the initial segments to discard the bad segments. We achieve our goal in the probability framework. We set X_i as the observed data corresponding to the initial segment Q_i. The unknown segmentation confidences for the initial segments are denoted by the unobserved latent variables $Z = \{z_1, z_2, \cdots, z_m\}$ where z_i is the segmentation confidence of the segment Q_i. The complete data set is denoted by $\{X, Z\}$. The goal is to find the maximum posteriori estimation of θ and Z given X, which can be represented by

$$\begin{aligned} \hat{\theta}_{MAP} &= \arg\max_{\theta \in \Omega} p(\theta|X) = \arg\max_{\theta \in \Omega} p(X|\theta) \cdot p(\theta) \\ &= \arg\max_{\theta \in \Omega} \prod_{i=1}^{m} \int p(X_i, z_i|\theta) dz_i \cdot p(\theta) \end{aligned} \tag{9}$$

We solve the problem in (9) by the EM algorithm which seeks to find the MAP iteratively applying the following two steps: *E-step* and *M-step*. In *E-step*, we generate the expectation $\mathcal{Q}(\theta, \theta^{old})$ of the complete-data evaluated using the observed data X and the current parameter θ^{old}, which is represented as

$$\begin{aligned} &\mathcal{Q}(\theta, \theta^{old}) \\ &= \sum_{i=1}^{m} \int p(z_i|X_i, \theta^{old}) \ln p(X_i, z_i|\theta) dz_i + \ln p(\theta) \end{aligned} \tag{10}$$

In *M-step*, the parameter θ^{new} is updated by maximizing the expectation $\mathcal{Q}(\theta, \theta^{old})$, which can be represented by

$$\theta^{new} = \arg\max_{\theta} \mathcal{Q}(\theta, \theta^{old}) \tag{11}$$

The *E-step* and *M-step* are iterated alternately until the convergence of θ and Z. In what follows, we detail the calculation of $p(\theta)$, $p(z_i|X_i, \theta^{old})$, $p(X_i, z_i|\theta)$, respectively.

1) The Distribution of $p(\theta)$: From (8), we can see that the model is designed to be a sparse representation since $\|\theta\|_1$ is minimized. A value of θ with small $\|\theta\|_1$ refers to large probability. Otherwise, a small probability will be given. Hence, we set $p(\theta)$ as

$$p(\theta) = \frac{1}{N_\theta} \exp^{-\alpha\|\theta\|_1} \tag{12}$$

where N_θ is the normalized constant.

2) The Posterior Distribution of $p(z_i|X_i,\theta^{old})$: Given the observed data X_i of the segment Q_i and feature model parameters θ^{old}, the similarities between Q_i and other segments $\tau_i = (\tau_i(1), \cdots, \tau_i(m))^T$ can be obtained by

$$\tau_i = X_i\theta^{old} \tag{13}$$

Since the initial segments are obtained from the simple images, most of the initial segments can be considered as successful segments. It is seen that a successful segmentation will be similar to most of the segments and have large sum of the similarities, i.e., a large value of $\|X_i\theta^{old}\|_1$. Otherwise, unsuccessful segmentation refers to a small value of $\|X_i\theta^{old}\|_1$. We can see that $p(z_i|X_i,\theta^{old})$ is related to $\|X_i\theta^{old}\|_1$, and we set $p(z_i|X_i,\theta^{old})$ as

$$p(z_i|X_i,\theta^{old}) = \mathcal{N}(\frac{\|X_i\theta^{old}\|_1}{N_1}, 1) \tag{14}$$

where N_1 are the normalized constants. We set $N_1 = \max_i \|X_i\theta^{old}\|_1, i = 1, \cdots, m$.

3) The Posterior Distribution of $p(X_i, z_i|\theta)$: In our model, we assume that θ independents to Z. Given a feature model parameter θ, we can measure $p(X_i, z_i|\theta) = p(z_i)p(X_i|\theta, z_i)$ by two terms, i.e., $p(z_i)$ and $p(X_i|\theta, z_i)$.

a) $p(X_i|z_i,\theta)$: In our model, we measure $p(X_i|z_i,\theta)$ by the fitness between the observed data corresponding to X_i and z_i and the target matrix related to z_i and θ. A large $p(X_i|z_i,\theta)$ prefers a good fitness. Otherwise, a small $p(X_i|z_i,\theta)$ will be assigned.

Given z_i, we train our model by only considering the good segment Q_i with large z_i. Two data adjustments are used to select the good segment. The first is to adjust the data X_i according to Z. The j-th data row of X_i with large z_j need to be selected. Otherwise, the data row should be abandoned. We achieve the adjustment by multiplying the values of k-th row of X_i by z_i, i.e.,

$$X_i^{new} = \Lambda_{z_i} X_i \tag{15}$$

and

$$\Lambda_{z_i} = \begin{pmatrix} \min(z_1, z_i) & 0 & \cdots & 0 \\ 0 & \min(z_2, z_i) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \min(z_m, z_i) \end{pmatrix} \tag{16}$$

where X_i^{new} is the adjusted observed data for Q_i. The value $\min(z_k, z_i)$ represents the confidence of a pair of segments (Q_k, Q_i). Here, these $z_k, k \neq i$ are considered as fixed values for X_i. It is seen that the confidence of a pair of segments Q_k and Q_i is represented by the smaller confidence of z_k and z_i, since we believe that the value referring to any bad segment should be abandoned. Hence, by multiplying $\min(z_k, z_i)$, the data in X_i corresponding to the successful segment pairs will be retained, while the date of the unsuccessful segmentation pairs tends to be zeros and to be abandoned.

We next adjust the target S_i' with respect to Z. The original target vector is $m \times 1$ vector $S_i' = (1, 1, \cdots, 0, \cdots, 1)^T$ with only one zero element $S_i'(i) = 0$. Similar to observed data X_i, the target value corresponding to a pair of good segments need to be retained and approximately equal to 1. For unsuccessful segment pairs, the corresponding target value should be close to 0. In our method, we adjust S' by

$$S_i'' = \Lambda_{z_i} S_i' \tag{17}$$

where S_i'' is the adjusted target vector. We can see that the observed data of the successful segment pairs has the value $S_i''(k)$ that is close to one. For the unsuccessful segment, $S_i''(k)$ tends to be zero.

Based on X_i^{new} and S'', the fitness between X_i, θ and latent variable z_i is evaluated by

$$\ell_i(X_i^{new}, \theta, z_i) = \|X_i^{new}\theta - S_i''\|_2^2 = \|\Lambda_{z_i}X_i\theta - \Lambda_{z_i}S_i'\|_2^2 \tag{18}$$

where $\ell_i(X_i^{new}, \theta, z_i)$ (use ℓ_i for short) is the loss function measuring the difference between the similarity matrix $\Lambda_{z_i}X_i\theta$ and the target similarity matrix S_i''. A good fitness prefers small ℓ_i. Based on $\ell_i(X_i^{new}, \theta, z_i)$, we formulate $p(X_i|\theta, z_i)$ as

$$p(X_i|\theta, z_i) = \frac{1}{N_x}\exp(-\ell_i(X_i^{new}, \theta, z_i)) \tag{19}$$

and N_x is the normalized constant.

b) The distribution of $p(z_i)$: Since the initial segments are obtained from the simple images, we believe that most of the initial segments are successfully segmented. Hence, $z_i \approx 1$ for most of segments. In our method, we set $p(z_i)$ as

$$p(z_i) = \frac{1}{N_z}\exp(-\beta|1 - z_i|) \tag{20}$$

with the normalized constant N_z.

c) The distribution of $p(X_i, z_i|\theta)$: Based on the distribution of $p(X_i|\theta, z_i)$ and $p(z_i)$ above, $p(X_i, z_i|\theta)$ can be represented by

$$p(X_i, z_i|\theta) = \frac{1}{N_x N_z}\exp(-\ell(X_i^{new}, \theta, z_i) - \beta|1 - z_i|) \tag{21}$$

4) The Minimization of the Expectation $\mathcal{Q}$: By (14) and (21), $\mathcal{Q}$ in(10) can be represented as

$$\begin{aligned} &\mathcal{Q}(\theta, \theta^{old}) \\ &= \sum_{i=1}^{m}\int p(z_i|X_i,\theta^{old})\ln p(X_i, z_i|\theta)dz_i + \ln p(\theta) \\ &= \sum_{i=1}^{m}[-\ln(N_x N_z) - \int p(z_i|X_i,\theta^{old})\ell_i dz_i \\ &\quad - \int p(z_i|X_i,\theta^{old})\beta(|1 - z_i|)dz_i] - \gamma\|\theta\|_1 \end{aligned} \tag{22}$$

where $\gamma = \frac{\alpha}{N_\theta}$. The derivation of (22) can be found in the appendix. It is seen from (22) that only $\int p(z_i|X_i,\theta^{old})\ell_i dz_i$ and $\gamma\|\theta\|_1$ are related to θ. Hence, maximizing $\mathcal{Q}$ in M-step (11) with respect to θ changes to solve the following

minimization problem, i.e.,

$$\begin{aligned}
&\theta^{new}\\
&= \arg\max_{\theta} \sum_{i=1}^{m} \int -p(z_i|X_i,\theta^{old})\ell_i dz_i - \gamma \|\theta\|_1\\
&= \arg\max_{\theta} \sum_{i=1}^{m} \int -p(z_i|X_i,\theta^{old})\|\Lambda_{z_i}X_i\theta - \Lambda_{z_i}S_i'\|_2^2 dz_i\\
&\quad -\gamma \|\theta\|_1\\
&= \arg\max_{\theta} \sum_{i=1}^{m} \int -p(z_i|X_i,\theta^{old})\|\Lambda_{z_i}(X_i\theta - S_i')\|_2^2 dz_i\\
&\quad -\gamma \|\theta\|_1\\
&= \arg\max_{\theta} \sum_{i=1}^{m} \int -p(z_i|X_i,\theta^{old})(X_i\theta - S_i')^T \Lambda_{z_i}^T\Lambda_{z_i}\\
&\quad (X_i\theta - S_i')dz_i - \gamma \|\theta\|_1\\
&= \arg\max_{\theta} \sum_{i=1}^{m} -(X_i\theta - S_i')^T \int p(z_i|X_i,\theta^{old})\Lambda_{z_i}^T\Lambda_{z_i}dz_i\\
&\quad (X_i\theta - S_i') - \gamma \|\theta\|_1\\
&= \arg\max_{\theta} \sum_{i=1}^{m} -(X_i\theta - S_i')^T \Lambda_{u_i}^T\Lambda_{u_i}(X_i\theta - S_i') - \gamma \|\theta\|_1\\
&= \arg\max_{\theta} -\sum_{i=1}^{m} \|\Lambda_{u_i}X_i\theta - \Lambda_{u_i}S_i')\|_2^2 - \gamma \|\theta\|_1\\
&= \arg\min_{\theta} \|X^{new}\theta - S'')\|_2^2 + \gamma \|\theta\|_1 \qquad (23)
\end{aligned}$$

where $\Lambda_{u_i}^T\Lambda_{u_i} = \int p(z_i|X_i,\theta^{old})\Lambda_{z_i}^T\Lambda_{z_i}dz_i$, $X^{new} = ((X_{u1}^{new})^T,\cdots,(X_{um}^{new})^T)^T$, and $S'' = ((S_{u1}'')^T,\cdots,(S_{um}'')^T)^T$ are the adjusted data of X and S' based on Λ_{u_i}. It is seen that the problem in (23) is a $l1$-Regularized Least Squares problem. We use the method in [41][4] for the minimization.

5) Implementation: The *E-step* and *M-step* are iteratively executed until the convergence. We can see from (23) that Λ_{z_i} consists of m different matrixes over all z_i. Hence, Λ_{u_i} can be calculated by the sum of these piecewise matrixes combined with cumulative distribution function of Gaussian distribution (by $\int p(z_i|X_i,\theta^{old})dz_i$). For simplicity, we directly use $\Lambda_{z_i}, z_i = \frac{\|X_i\theta^{old}\|_1}{N_1}$ to approximate Λ_{u_i}, i.e., $X^{new} = ((X_1^{new})^T,\cdots,(X_m^{new})^T)^T$, and $S'' = ((S_1'')^T,\cdots,(S_m'')^T)^T$ to reduce the computational cost. We set the iteration number (the stop number is 50) as the EM stop condition. We set $m = 10$ for the simple image selection. In (23), $\gamma = 0.01$. Note that these parameters are fixed among different image datasets. The algorithm of the proposed learning method is shown in Algorithm 1.

Algorithm 1 The Algorithm for EM Based Feature Learning Method

Input: Matrix $X_i, i = 1,\cdots,m$ obtained by (6).
Output: The feature model parameters θ and segmentation confidence $Z = \{z_1,\cdots,z_m\}$.
% *Parameters initialization*
$\theta^{old} = (1,1,\cdots,1)_{n\times 1}$.
% *Iteration*
while The stop condition is not satisfied **do**
1) Calculate $\|X_i\theta^{old}\|_1, i = 1,\cdots,m$ based on θ^{old} and X_i in (14) and obtain $N_1 = \max_i(\|X_i\theta^{old}\|_1)$.
2) Obtain $z_i = \frac{\|X_i\theta^{old}\|_1}{N_1}, i = 1,\cdots,m$ and calculate Λ_{z_i} based on $Z = \{z_1,\cdots,z_m\}$ by (16).
3) Calculate $X_i^{new}, i = 1,\cdots,m$ with respect to X_i and Λ_{z_i} by (15), and obtain X^{new} by $X^{new} = ((X_1^{new})^T,\cdots,(X_m^{new})^T)^T$.
4) Calculate target matrix $S_i'', i = 1,\cdots,m$ with respect to S_i' and Λ_{z_i} by (17), and obtain S'' by $S'' = ((S_1'')^T,\cdots,(S_m'')^T)^T$.
5) Obtain θ by solving the $l1$-regularized least squares optimization problem in (23).
6) Update θ: $\theta^{old} \leftarrow \theta$.
end while

C. Co-Segmentation

Based on the learned feature model, we use our previous work in [26] to achieve the co-segmentation task. In the method, the original images are segmented into over-lapping local regions using object detection method, saliency detection method and hierarchy over-segmentation method. Then, the similarities between the local regions are represented by a directed graph structure. The co-segmentation is formulated as a shortest path searching problem and is solved by dynamic programming.

Several improvements are used to achieve adaptive feature learning based co-segmentation. Firstly, in edge weight calculation, we calculate the region term by the learned feature model rather than the original features. Secondly, the initial segments referring to large confidences are used as the co-segmentation results. The co-segmentation result is then treated as the only local region of the related image in the process of the digraph construction.

V. Experimental Results

In this section, we verify the proposed co-segmentation method on many images groups. The subjective and objective assessments of the segmentation results are given.

A. Co-Segmentation Results

1) Test Images Dataset: In the experiments, we collect image groups from well-known image databases such as MSRC database [42],[5] ETHZ shape database [43],[6] and ICoseg database [15][7]. We select 16 classes among the total 20 classes in MSRC dataset and the classes that have more than 20 images in ICoseg dataset for the verification. The total five classes in ETHZ shape database are all used. To completely verify our method, we use all images in each class. We use the ground truth given by [15] and [42] for the ICoseg

[4] http://www.stanford.edu/~boyd/l1_ls/

[5] http://research.microsoft.com/en-us/um/people/antcrim/data_objrec/msrc_objcategimagedatabase_v2.zip

[6] http://www.vision.ee.ethz.ch/~calvin/ethz_shape_classes_v12.tgz

[7] http://chenlab.ece.cornell.edu/projects/touch-coseg/CMU_Cornell_iCoseg_dataset.zip

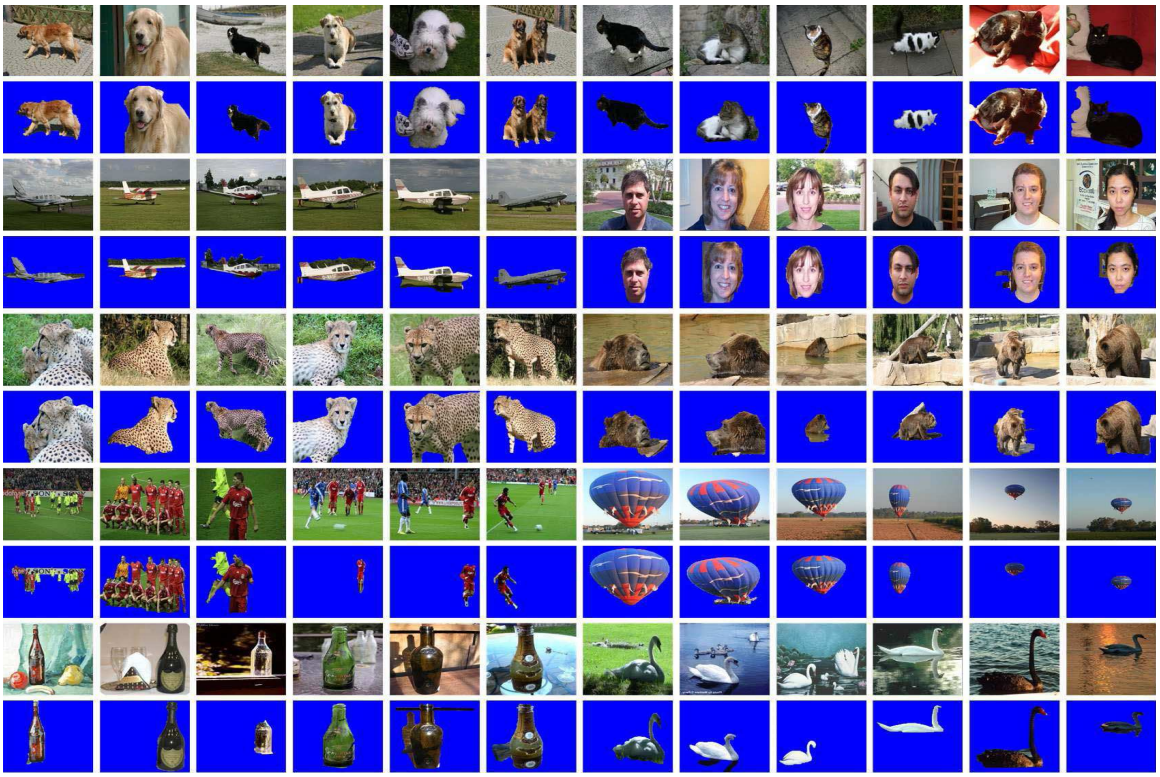


Fig. 5. The segmentation results of the proposed method. From top to bottom: the rows 1, 3, 5, 7 and 9 show the original images. The rows 2, 4, 6, 8 and 10 display the segmentation results.

and the MSRC database, respectively. For ETHZ shape dataset, we obtain the ground truth by the contour based ground truth in [43].

2) The Co-Segmentation Results: The co-segmentation results of ten classes are shown in Fig. 5. For each image class, six original images and the co-segmentation results are presented. From Fig. 5, we can see that the original images have many variations, such as color, shape and texture. It is also seen that the proposed co-segmentation method successfully segments the common objects from these images. For example, the 'cats' in *Cats* vary significantly. The proposed co-segmentation method successfully segments these 'cats', which benefits from the adaptive feature learning.

We also show the results of the feature learning method. The confidences of the initial segments are shown in Fig. 6(a), where the results of six classes are shown. For each class, the original images are shown in the first row. These images are selected by the proposed image complexity analysis method. We can see that simple images can be selected by the proposed method. The initial segments obtained by the unsupervised segmentation method are shown in the second row. It is seen that most of the objects can be successfully segmented from the simple images. Meanwhile, there are a few unsuccessful segments, such as the second image in *Cheetah* and the fifth image in *Mugs*. The learned confidence of each initial segment is shown below the image. It can be seen that the learned confidences fit the human judgments. For example, in *Mugs*, the fifth initial segment is the unsuccessful segmentation. The learned confidence is small (0.0869). Meanwhile, for the first segment which is a successful segmentation, the learned confidence is close to one (0.9634).

Furthermore, the learned feature model corresponding to the classes in Fig. 6(a) are shown in Fig. 6(b). Each feature model is represented by a color-bar, where each color describes a general feature. These colors represent the features of color, shape, SIFT, Self-similarity and pHog from left to right, respectively. The amplitude of each color represents the learned weight coefficient of the corresponding feature. We can see that the learned feature model can represent the similarities between the objects. For example, the class *Mugs* contains similar shape. The weight coefficient of the shape feature is large in the learned model, which indicates that the shape feature plays an important role in the foreground similarity measurement. For the class *Bear*, the weight coefficient of color is large, which fulfills the fact that the 'bears' contain similar colors.

B. Objective Evaluation

We evaluate the proposed co-segmentation method by the error rate which is defined as the ratio of the number of wrongly segmented pixels to the total number of pixels. A small error rate refers to a successful segmentation. The mean error rate over all images is used to evaluate the performance of a class. The error rates of the proposed co-segmentation method are shown in Table II. We can see that the proposed co-segmentation method achieves low error rates in most of the classes. It is also seen that there are unsuccessful segments, such as *Panda* and *Stonehenge*. The unsuccessful segments are caused by the fact that there are no simple images in these classes. The complex images lead to unsuccessful initial segments and further result in inaccurate learning of the feature model.

Fig. 6. (a): The confidences of initial segments. For each block, the first row shows the simple images obtained by complexity analysis. The second row shows the initial segments obtained by method in [35]. The confidences obtained by the proposed learning method are shown under the images. (b): The learned feature models corresponding to the classes in (a). The color in the model represents the features. They are color, shape, SIFT, Self-similarity and pHog from left to right respectively.

In Table II, we also compare our method with the existing co-segmentation methods such as the methods in [14], [19] and [26]. Joulin *et al.* in [14] proposed a co-segmentation model using the discriminative clustering method and the spectral clustering method. In the experiment, the source code given by the authors[8] is used. To improve the co-segmentation results, we adjust the parameter μ for each class. Color feature (for ICoseg dataset and ETHZ dataset) and SIFT feature (for MSRC dataset) suggested by the author are employed. The superpixels are generated by the over-segmentation method in [33] (by setting $k = 100$). The results referring to the method in [14] are shown in the second row of Table II. It is seen that the common objects are successfully segmented from several classes, such as *Liberty* and *Airshows2*. Meanwhile, there are unsuccessful segments, such as *Cheetah* and *Pandas*. The unsuccessful segments are caused by the fact that the classes contain different similar features.

Kim *et al.* in [19] propose multiple class co-segmentation method, which is achieved by the linear anisotropic diffusion based segmentation method. Color feature is used. In the experiment, the code released by the author is used[9]. The intra-image Gaussian weights and the number of segments (K) are adjusted for accurate co-segmentation. The results by the method in [19] are shown in the third row of Table II. We can see that Kim's method can successfully segment common objects in several classes, such as *Liverpool* and *Goose*. Meanwhile, unsuccessful segments are also achieved, such as *Dogs* and *Chairs*. The unsuccessful segmentations are caused by the fact that many classes contain other similar features rather than color.

Meng *et al.* in [26] achieves common objects segmentation by graph theory. The co-segmentation is formulated as the shortest path searching, and the shortest path is found by dynamic programming. In the experiment, we adjust the scaling parameter α for each classes to achieve accurate co-segmentation. We use color feature for ICoseg dataset and MSRC dataset and shape feature for ETHZ dataset. The results by the method in [26] are shown in the fourth row of Table II. We can see that the method in [26] can successfully extract common objects from several images, such as *Soccer* and *Kite1*. Meanwhile, there are unsuccessful segments, such as

[8]www.di.ens.fr/~joulin

[9]http://www.cs.cmu.edu/~gunhee

TABLE II

RESULTS COMPARISON BETWEEN THE PROPOSED CO-SEGMENTATION METHOD AND THE EXISTING METHODS IN TERMS OF ERROR RATE. CLASSES IN ICoseg, MSRC AND ETHZ DATASETS ARE USED

ICoseg									
Method	Air1	Cheetah	Pandas	Kite1	Soccer	Liberty	Kendo	Air2	Liverpool
[14]	0.0832	0.4025	0.4971	0.1185	0.1664	**0.0591**	0.1985	0.0188	0.4488
[19]	0.1538	0.3787	0.4535	0.2740	0.2889	0.1520	0.0335	0.2942	**0.0704**
[26]	0.1182	0.2650	**0.2158**	0.0490	0.0841	0.1046	0.0858	0.1224	0.1533
[35]	**0.0666**	0.3367	0.2808	**0.0366**	**0.0726**	0.0964	**0.0283**	0.0289	0.1056
Ours	0.0985	**0.1065**	0.3262	0.0413	0.0869	0.0864	0.0933	**0.0076**	0.0806
Method	Bear	Goose	Redsox	Stone	Balloons	**Average**			
[14]	0.2214	0.2508	0.4205	0.4584	0.1330	0.2484			
[19]	0.1828	**0.0895**	0.0526	0.3860	0.0808	0.2065			
[26]	**0.1535**	0.1166	**0.0401**	**0.3599**	0.0878	0.1397			
[35]	0.1586	0.2043	0.1506	0.4278	0.1372	0.1522			
Ours	0.1629	0.2014	0.0470	0.3942	**0.0104**	**0.1245**			
MRSC									
Method	Planes	Cows	Faces	Cars	Sheeps	Flowers	Signs	Birds	Chairs
[14]	0.3562	0.2922	0.3564	0.4539	0.3241	0.4858	0.4867	0.2029	0.2672
[19]	0.2183	0.2267	0.3807	0.3742	0.2436	0.2512	0.5104	0.1726	0.3252
[26]	0.2573	0.1167	0.3897	0.3448	0.1257	**0.1621**	0.2631	0.1677	0.3384
[35]	**0.1788**	**0.0832**	0.3573	0.3022	0.1031	0.1633	0.2573	0.1034	**0.2064**
Ours	0.2125	0.1274	**0.3079**	**0.2844**	**0.0646**	0.2860	**0.1282**	**0.0988**	0.3498
Method	Cats	Dogs	Trees	Builds	Bicycles	Books	Humans	**Average**	
[14]	0.3379	0.3125	0.4829	0.3687	**0.2475**	0.3421	0.3784	0.3560	
[19]	0.2391	0.3171	0.3281	0.4104	0.3562	0.4539	0.2249	0.3145	
[26]	0.3141	0.1717	0.2507	0.3401	0.3669	**0.3338**	0.4098	0.2720	
[35]	0.1648	0.1969	0.3048	0.3572	0.3535	0.4707	0.1546	0.2386	
Ours	**0.1432**	**0.1183**	**0.2018**	**0.2149**	0.3971	0.4322	**0.1528**	**0.2200**	
ETHZ									
Method	Bottles	Swans	Logos	Mugs	Giraffes	**Average**			**Average all**
[14]	0.2771	0.1981	0.0952	0.2198	0.2564	0.2093			0.2920
[19]	0.4745	0.1216	0.5824	0.4642	0.5922	0.4470			0.2902
[26]	**0.1113**	**0.0542**	0.0950	0.2049	**0.1428**	**0.1216**			0.1976
[35]	0.1392	0.1441	0.2494	0.1626	0.1301	0.1651			0.1918
Ours	0.1491	0.0764	**0.0728**	**0.1613**	0.1718	0.1263			**0.1684**

Cats and *Cheetah*. These unsuccessful segments are mainly caused by the fact that the given features cannot fully represent the similarities between the common objects.

The comparison results show that the proposed co-segmentation method achieves the lowest error rates for most of the image pairs. For ICoseg dataset, the mean error rates over all classes are 0.2484, 0.2065, 0.1397 and 0.1245 for the methods in [14], [19], [26] and the proposed method, respectively. We can see that the proposed method achieves the smallest error rate. It is also seen that the other comparison methods achieve good performance in the ICoseg dataset, since the common objects contain similar colors in the ICoseg dataset. For MSRC dataset, the mean error rates over all classes are 0.3560, 0.3145, 0.2720 and 0.2200 for the methods in [14], [19], [26] and the proposed method, respectively. It is seen that the error rates are obviously decreased by the proposed method which is caused by the adaptive learning of the feature model. For ETHZ dataset, the mean error rates are 0.2093, 0.4470, 0.1216 and 0.1263 for the method in [14], [19], [26] and the proposed method, respectively. We can see that the method in [26] achieves the smallest error rate in this dataset. The reason is that the shape feature can accurately represent common objects similarity for the classes. By using the shape feature, the method in [26] can achieve accurate co-segmentation. Note that the differences between the method in [26] and the proposed method are small (the difference is 0.0047). Hence, the performance of the proposed method is comparable to the method in [26] in ETHZ dataset. The error rates over all classes are 0.2920, 0.2902, 0.1976 and 0.1684 for the methods in [14], [19], [26] and the proposed method, respectively. It is seen that the proposed method achieves the smallest error rate, which demonstrates the effectiveness of the proposed method.

To further verify our proposed method, we display the results of the initial segment method [35] in Table II. The method in [35] is to first detect the saliency regions by global contrast and then perform grab-cuts to obtain the salient regions. It focuses on the salient regions in each single image instead of the common objects in multiple images. From Table II, we can see that the method in [35] can obtain successful object segmentation in some classes, such as *Air1* and *Planes*. The reason is that the salient objects in these classes are also the common objects. When the images contain other multiple salient regions, these salient regions may be also obtained by the method in [35], such as "Logos" and "Dogs", which results in unsuccessful segmentation. It is also seen from Table II that the mean error rate (0.1684) of the proposed method is smaller than the one (0.1918) in the method [35].

We also show the results of the proposed method by selecting different number of simple images m on the three datasets. We show the results in Fig. 7, where the results of the

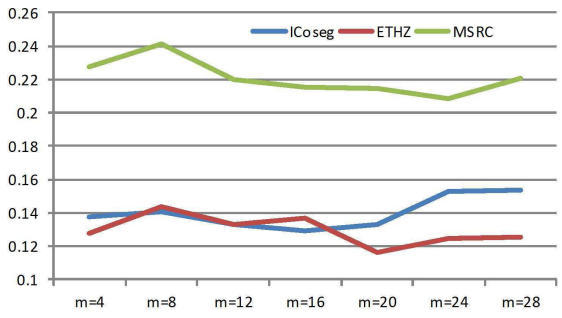


Fig. 7. The results of the proposed with different m. The three datasets (ICoseg, MSRC and ETHZ) and seven m values ($m = 4$, 8, 12, 16, 20, 24 and 28) are shown.

three datasets (ICoseg, MSRC and ETHZ) and seven m values ($m = 4$, 8, 12, 16, 20, 24 and 28) are displayed. We can see that small error rates can be obtained when $m \in [10, 22]$ for the ICoseg and MSRC datasets. Meanwhile, the small and large m can result in the increase of the error rates. The reason is that a small m may not provide enough initial segments for the accurate feature learning, while a large m can introduce more segment noises to interfere the feature learning. Note that the error rates of ETHZ dataset decrease along the considered m. The reason is that the number of the images in the ETHZ classes is large (about average 50 image per class), which leads to a larger m to obtain the small error rates. In our experiments, we set $m = 10$ by considering the small error rate and the low computational cost of the learning.

VI. Discussion

We first discuss the motivation of using simple images to learn the feature model. It is known that the success of a learning scheme is directly associated with an appropriate input data selection [44]. Inaccurate learning will be obtained when the training samples contain many wrong samples. The proposed method adaptively learns the useful features for accurate co-segmentation. In the feature learning, the accurate learning depends on the accuracy of the initial segments. Successful segments can provide useful information to accurately learn the feature model. On the contrary, unsuccessful segments will interfere the feature learning and lead to inaccurate feature model. Hence, it is required to accurately extract the initial objects as much as possible. As we known, extracting the objects from the simple image is much easier than the complex image, which guarantees the requirement of our feature learning. This property motivates us to use the image analysis to select the simple images to achieve the initial segmentation. For the complex images, we believe that the feature learning will be difficult from the complex images because many unsuccessful segments will be generated and used in the feature learning. These incorrect training samples will result in inaccurate learning of the feature model and lead to unsuccessful co-segmentation.

We next discuss the generalization of the proposed model. In order to guarantee the fairness of the comparison, all parameters and the general features used in the feature learning are fixed for different datasets in our experiments. Meanwhile, the original feature pool contains much type of features, such as the color, texture and shape. These features are usually shared by most of the common objects in the realistic images. Hence, the feature learning method can be generalized to other datasets. We verify the generalization of the proposed method on other different image datasets, such as Caltech-UCSD Birds 200-2011 dataset, Stanford Dogs, and Oxford Flowers 102. The segmentation results and the error rates are shown in Fig. 8 and Table III, respectively. It is seen that the proposed method can be successfully generalized to these image classes. Furthermore, we verify the generalization of the learned feature model in Caltech 101 datasets. In the experiments, we use the feature model learned from MSRC or ETHZ to implement the co-segmentation on the same class in the Caltech 101 dataset. The segmentation results and error rates are shown in Fig. 9 and Table III, respectively. The results of *Mugs* and *Aeroplanes* are displayed. We can see that the learned feature model also achieves successfully co-segmentation on a new dataset. The reason is that the images of a class tend to contain the same similar features in different datasets. The feature model learned from a image group can also be used to achieve co-segmentation in the other image groups. The results of the methods in [14], [19] and [26] are also proposed in Table III for comparison. It is seen that the proposed method can also achieve the lowest error rates on most of the classes shown in Table III, which demonstrates that the proposed method can be generalized to other datasets.

In our method, we use the method in [34] to detect the windows, where the initial windows are generated by sliding windows at many scales. Different sizes of windows are generated and are uniformly distributed over the entire image. In the detection, each initial window is first scored based on four cues, such as saliency, color contrast, edge density and superpixels straddling. The best top N_w windows are then selected for the image complexity analysis based on the scores. After windows selection, the overlap regions among the selected windows are extracted using threshold T_w. We can see that the choices of N_w and T_w mainly depend on the scores of the windows instead of the window size.

In our method, we impose the sparsity constrain on the θ as shown in (8). The existing sparsity constrains usually used in the sparse representation, such as $\ell 1$-norm [45], $\ell 2$-norm and elastic net formulation [46] can be used as the constrain. In our model, $\ell 1$-norm is selected based on its natural to obtain both the shrinkage and the variable selection in the regression [46]. Also, the $\ell 1$-norm has been successfully used in many computer vision tasks compared with l2-norm, such as face recognition [47]. Moreover, compared with elastic net formulation, $\ell 1$-norm can sufficiently represent the sparsity here, since we intend to select one or small number of features to represent the foreground similarities.

It is seen from (5) that we use a linear model to learn the adaptive feature. The reason is that linear model is simple and can lead to the easy parameters estimation of the model. The other and also the most important reason is that the linear model is able to capture the foreground similarity consistency. As our method is based on the assumption of

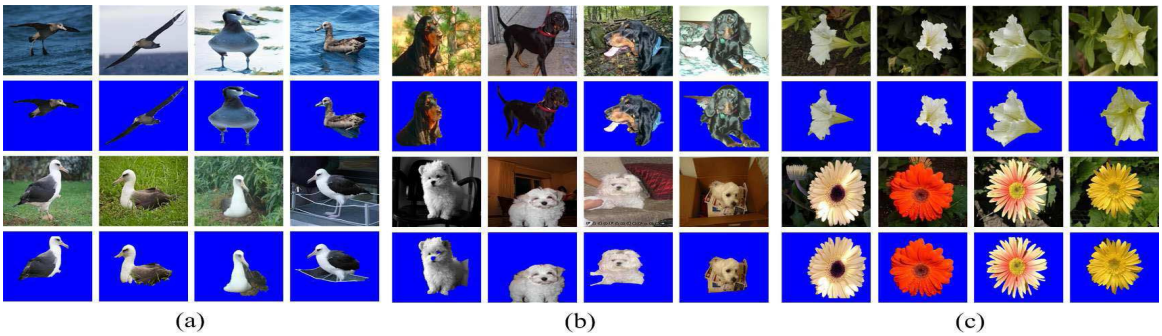


Fig. 8. The results of the proposed method on the other datasets. (a): Caltech-UCSD Birds 200-2011 dataset. (b): Stanford Dogs dataset. (c): Oxford Flowers 102 dataset.

Fig. 9. The segmentation results by generalizing the learned model to the other image datasets. The classes *Aeroplanes* and *Mugs* in Caltech 101 dataset are used. The feature models learned in MSRC datasets (*Aeroplanes*) and ETHZ shape datasets (*Mugs*) are used for the co-segmentation in the Caltech classes.

TABLE III

THE ERROR RATE ON THE OTHER DATASETS WITH THE SAME PARAMETERS, SUCH AS CALTECH-UCSD BIRDS 200-2011 DATASET (*Black* AND *Laysan*), STANFORD DOGS (*Maltese* AND *coonhound*), OXFORD FLOWERS 102 (*Petunia* AND *Barbeton*) AND CALTECH 101 (*Mugs* AND *Aeroplanes*). THE RESULTS OF THE METHODS IN [14], [19] AND [26] ARE ALSO PROPOSED FOR COMPARISON

Datasets	UCSD Birds		Stanford Dogs		Oxford Flowers		Caltech 101	
Classes	Black	Laysan	Maltese	Coonhound	Petunia	Barbeton	**Mugs**	**Aeroplanes**
[14]	0.4606	0.4067	0.3582	0.3537	0.1812	0.4945	0.4739	0.4511
[19]	0.2648	0.2236	**0.2168**	0.1933	0.2515	0.2292	0.2940	0.1016
[26]	**0.0513**	0.1558	0.3087	0.2055	0.0162	0.1186	0.2325	**0.0674**
Ours	0.0665	**0.1123**	0.2211	**0.1905**	**0.0144**	**0.0713**	**0.2075**	0.0927

the sparsity of the features, it is seen that selecting single feature or linearly combing a few of features as used in our linear model is enough to represent the sparsity of the features. Note that non-linear feature selection methods, such as kernel based support vector regression [48] and kernel based logistic LASSO regression [49], can also be used in our feature learning. Since the linear model is a specific case of the no-linear model, the no-linear model may result in better co-segmentation results. Meanwhile, it also leads to more complex analytical and computational properties than the linear model [50]. Hence, linear model is selected in our method.

In our method, the successful feature learning depends on the accuracy of the initial segmentation, as discussion in Section VI. Successful initial segmentation will result in accurate feature learning, while incorrect feature model can be learned from the wrong initial segments. To achieve accurate feature learning, we combine the image complexity evaluation and the saliency based foreground extraction ([35]). It is noted that although it is still difficult to extract the saliency regions from complex scenes, the object foregrounds can fortunately be well extracted from the simple backgrounds by the saliency detection method (such as [35]), which can help the initial segmentation of the common objects.

In the feature learning, we introduce the segmentation confidence to select the success of a segment for the feature learning. The segments with large confidences are used to learn the feature model. Furthermore, we directly use these segments as the co-segmentation results of the corresponding image for simplicity. Hence, some of these results in Fig. 6 are shown as the final results in Fig. 5. Note that these segments will not be cosegmented in the following co-segmentation. For the selected images with small confidences in Fig. 6, they are the bad initial segments and are not used in the feature learning. Hence, we return these images to the rest image group and perform the co-segmentation to obtain the accurate object extraction.

The proposed method and the method in [26] all require the similar features to generate the edge weights in the graph construction. The indeed different is that we automatically

learn the similar features, while the method in [26] manually selects the feature for each class [26]. Compared with manual selecting manner in [26], our model can easily handle more features (5 features) and learn the best feature combinations (by linear model), which results in the improvement of the co-segmentation as shown in Table II. But the improvement is not significant, since the manual selection can also select the similar feature of each class. However, compares with [26], our method is more reasonable due to the automatic feature learning by the computer and the wide applications in realistic computer vision tasks.

VII. Conclusion

In this paper, we proposed a new feature adaptive co-segmentation model to segment common objects from multiple images. We proposed a new image complexity analysis method to rank the images and extract the objects from the simple images by using unsupervised segmentation method. An accurate feature model is learned from the objects by using an EM algorithm combining $l1$-regularized least squares optimization. The feature model is combined with the initial segmentation to extract the common objects. The experiments demonstrate that the error rate of the proposed method is lower than the existing methods when the feature is unknown. In the future, we will extend the proposed feature learning method for images with high complexity and nonlinear model.

Appendix A

The Derivations of the Equation (22)

$$\begin{aligned}\mathcal{Q} &= \sum_{i=1}^{m}\int p(z_i|X_i,\theta^{old})\ln p(X_i,z_i|\theta)dz_i + \ln p(\theta)\\ &= \sum_{i=1}^{m}[\int p(z_i|X_i,\theta^{old})(-\ln(N_xN_z)-\ell_i-\beta|1-z_i|)dz_i]\\ &\quad -\frac{\alpha}{N_\theta}\|\theta\|_1\\ &= \sum_{i=1}^{m}[-\ln(N_xN_z)-\int p(z_i|X_i,\theta^{old})\ell_i dz_i\\ &\quad -\int p(z_i|X_i,\theta^{old})\beta(|1-z_i|)dz_i] - \frac{\alpha}{N_\theta}\|\theta\|_1 \qquad (24)\end{aligned}$$

References

[1] H. Li and K. N. Ngan, "Unsupervised video segmentation with low depth of field," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 12, pp. 1742–1751, Dec. 2007.

[2] N. Jacobson, Y.-L. Lee, V. Mahadevan, N. Vasconcelos, and T. Nguyen, "A novel approach to FRUC using discriminant saliency and frame segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2924–2934, Nov. 2010.

[3] T. Patz and T. Preusser, "Segmentation of stochastic images with a stochastic random walker method," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2424–2433, May 2012.

[4] X. Chen, J. Udupa, U. Bagci, Y. Zhuge, and J. Yao, "Medical image segmentation by combining graph cuts and oriented active appearance models," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2035–2046, Apr. 2012.

[5] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[6] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. Int. Conf. Comput. Vis.*, 2001, pp. 105–112.

[7] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[8] J. Zhang, J. Zheng, and J. Cai, "A diffusion approach to seeded image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2125–2132.

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

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

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

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

[13] D. Batra, D. Parikh, A. Kowdle, T. Chen, and J. Luo, "Seed image selection in interactive cosegmentation," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2393–2396.

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

[15] D. Batra, A. Kowdle, and D. Parikh, "ICoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3169–3176.

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

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

[18] S. Vicente, C. Rother, and V. Kolmogorov, "Object cosegmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2217–2224.

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

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

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

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

[23] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3365–3375, May 2011.

[24] F. Meng, H. Li, G. Liu, and K. N. Ngan, "Image cosegmentation by incorporating color reward strategy and active contour model," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 725–737, Apr. 2013.

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

[26] F. Meng, H. Li, G. Liu, and K. N. Ngan, "Object co-segmentation based on shortest path algorithm and saliency model," *IEEE Trans. Multimedia*, vol. 14, no. 5, pp. 1429–1441, Oct. 2012.

[27] F. Meng and H. Li, "Complexity awareness based feature adaptive co-segmentation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 1–3.

[28] X. Chen, Z. Tong, H. Liu, and D. Cai, "Metric learning with two-dimensional smoothness for visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2533–2538.

[29] M. H. Nguyen and F. de la Torre, "Metric learning for image alignment," *Int. J. Comput. Vis.*, vol. 88, no. 1, pp. 69–84, 2010.

[30] E. Nowak and F. Jurie, "Learning visual similarity measures for comparing never seen objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[31] N. Verma, D. Mahajan, S. Sellamanickam, and V. Nair, "Learning hierarchical similarity metrics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2280–2287.

[32] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Metric learning for large scale image classification: Generalizing to new classes at near-zero cost," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 488–501.

[33] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: An empirical evaluation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2294–2301.

[34] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2010, pp. 73–80.

[35] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 409–416.

[36] H. Ling and D. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007.

[37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[38] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2006, pp. 2169–2178.

[39] T. Deselaers and V. Ferrari, "Global and efficient self-similarity for object classification and detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1633–1640.

[40] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proc. ACM Int. Conf. Image Video Retr.*, 2007, pp. 401–408.

[41] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale $\ell 1$-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[42] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Oct. 2005, pp. 1800–1807.

[43] V. Ferrari, T. Tuytelaars, and L. V. Gool, "Object detection by contour segment networks," in *Proc. Eur. Conf. Comput. Vis.*, Jun. 2006, pp. 14–28.

[44] C. Pedreira, "Learning vector quantization with training data selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 157–162, Jan. 2006.

[45] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

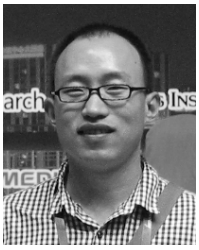
[46] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. B*, vol. 67, no. 2, pp. 301–320, 2005.

[47] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[48] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[49] K. Koh, S.-J. Kim, and S. Boyd, "An interior-point method for large-scale $\ell 1$-regularized logistic regression," *J. Mach. Learn. Res.*, vol. 8, pp. 1519–1555, Jul. 2007.

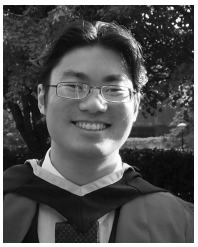
[50] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.



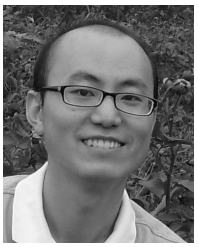
Fanman Meng received the B.Sc. degree in computer science and technology from Shandong Agricultural University, Taian, China, in 2006, and the M.Sc. degree in computer software and theory from Xihua University, Chengdu, China, in 2009. Since September 2009, he has been pursuing the Ph.D. degree with the Intelligent Visual Information Processing and Communication Laboratory, University of Electronic Science and Technology of China, Chengdu. He has been a Visiting Student with the Division of Visual and Interactive Computing, Nanyang Technological University, Singapore, since July 2013. He works in the areas of computer vision and pattern recognition. His work focuses on specific object segmentation and detection. He is currently focusing primarily on co-segmentation and saliency detection.

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 joined the Visual Signal Processing and Communication Laboratory, Chinese University of Hong Kong (CUHK), Hong Kong, as a Research Associate. From 2006 to 2008, he was a Post-Doctoral Fellow with the same laboratory in CUHK. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. His current research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system. He has authored or co-authored numerous technical articles in well known international journals and conferences. He is a co-editor of a Springer book titled *Video Segmentation and its Applications*. He was involved in many professional activities. He is a member of the Editorial Board of the *Journal on Visual Communications and Image Representation*. He served as a TPC member in a number of international conferences, including ICME in 2013, ICME in 2012, ISCAS in 2013, PCM in 2007, PCM in 2009, and VCIP in 2010, and served as a Technical Program Co-Chair in ISPACS in 2009, and a General Co-Chair of the 2010 International Symposium on Intelligent Signal Processing and Communication Systems. He serves as a Local Chair of the 2014 IEEE International Conference on Multimedia and Expo. He was selected for the New Century Excellent Talents in University, Chinese Ministry of Education, China, in 2008.

King Ngi Ngan (F'00) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K. He is currently a Chair Professor with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong. He was a Full Professor with Nanyang Technological University, Singapore, and the University of Western Australia, Crawley, Australia. He holds honorary and visiting professorships with numerous universities in China, Australia, and South East Asia. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal on Visual Communications and Image Representation*, *EURASIP Journal of Signal Processing: Image Communication*, and the *Journal of Applied Signal Processing*. He chaired and co-chaired a number of prestigious international conferences on image and video processing, including the 2010 IEEE International Conference on Image Processing, and served on the advisory and technical committees of numerous professional organizations. He has published three authored books, six edited volumes, and over 300 refereed technical papers, and edited nine special issues in journals. He holds ten patents in the areas of image/video coding and communications. He is a fellow of IET, U.K., and IEAust, Australia, and was an IEEE Distinguished Lecturer from 2006 to 2007.



Liaoyuan Zeng received the B.Eng. degree in telecommunication engineering from Southwest Jiaotong University, Chengdu, China, in 2005, the M.Eng. degree in computer and communication engineering from the University of Limerick, Limerick, Ireland, in 2006, and the Ph.D. degree in electrical engineering from the University of Limerick in 2011. He is currently a Lecturer and Researcher with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu. He was supported by a number of postgraduate scholarships and projects, including those provided by the Government of Ireland Postdoctoral Fellowships in Science, Engineering and Technology and European Cooperation in Science and Technology Action IC0902. He was with the Worcester Polytechnic Institute, Worcester, MA, USA, the Limerick, Ireland, Project Center as a Teaching Assistant in 2009 and 2010, a Teaching Assistant with the Polytechnic University of Catalonia, Catalonia, Spain, and the Limerick Project Center in 2010, and was a Leading Researcher and Representative of the Wireless Access Research Center, University of Limerick in the WUN Cognitive Communications Consortium from 2009 to 2011. He is currently a Researcher/Lecturer with the Intelligent Visual Information Processing and Communications Laboratory, University of Electronic Science and Technology of China. He is an active member of the research community, participating in activities that help facilitate the exchange of ideas between members within the community. He served as a Track Co-Chair of BWCCA in 2012, and he was a member of the Technical Program Committee of VTC in 2012, AICT in 2012, 2011, and 2010, and COCORA in 2012 and 2011.



Qingbo Wu received the B.E. degree in applied electronic technology education from Hebei Normal University, Hebei, China, in 2009. He is currently pursuing the Ph.D. degree from the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. His current research interests include image/video coding, quality evaluation, and perceptual modeling and processing.