CO-SEGMENTATION OF MULTIPLE IMAGES THROUGH RANDOM WALK ON GRAPHS

Yijie Wang, Byung-Jun Yoon, and Xiaoning Qian

Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX, 77843

ABSTRACT

We present a new image co-segmentation framework to simultaneously segment multiple images by formulating the cosegmentation problem as a multiple graph clustering problem. For each image, we first construct a corresponding segment graph by extracting superpixels as vertices and assigning edge weights between superpixels according to their feature and spatial proximity. To integrate the related information across images, we further compose a similarity graph across all constructed segment graphs, in which edges capture the similarity among superpixels across images. We propose to solve the co-segmentation problem by applying an alternating random walk strategy on both the segment graphs and the similarity graph to borrow strengths across images for better segmentation. The common objects shared in images can be identified by finding low conductance sets based on the transition probability matrix of the alternating random walk on these graphs. Experiments on iCoseg and a sequence of echocardiac images demonstrate that our novel formulation yields promising results and performs better than image segmentation on individual images separately.

Index Terms— Image co-segmentation; Multiple graph clustering; Alternating random walk on graphs

1. INTRODUCTION

Image and video segmentation helps extract useful information in many commercial and biomedical applications but it has been a long-standing problem for the accuracy and automaticity [1-3]. In this paper, we explore the idea of simultaneously analyzing multiple images that share common objects of interest for better image segmentation. Several methods have been recently proposed to study this image cosegmentation problem [4–7]. Typically, co-segmentation has been formulated to identify whether a pixel belongs to the shared common objects without any "external" information about the images or the objects as a way to compensate the lack of supervisory data and the absence of the prior knowledge on the appearance of either images or objects of interest. To the best of our knowledge, the existing methods all pose the problem as a clustering problem in derived feature space for image pixels. For example, a cost function was formulated in [7] for image co-segmentation by considering the common patterns of salient objects in images as well as the consistency of the objects across images to suppress the potential influence from the presence of images with different objects. In [6], a labeling formulation has been proposed to maximize the scoring function in terms of a pairwise similarity learned from a random forest approach. A feature discriminative clustering model has been studied together with a graph partition formulation in [4] to solve image co-segmentation.

Different from the existing formulations, we solve the image co-segmentation by first constructing a multi-partite graph to integrate all the information within and across multiple images and pose the problem as a multiple graph clustering problem. Joint clustering of multiple graphs or networks has been recently investigated in analyzing biological networks to identify cellular functional modules for better understanding of living systems. For example, in [8], we have proposed a joint clustering framework with a novel random walk strategy to simultaneously cluster two proteinprotein interaction networks. With graphical representations derived from multiple images, image co-segmentation can be naturally converted to a joint multiple network clustering problem by intuitively considering each image as an analog of a biological network. We solve this multiple network clustering problem by extending the approach in [8], originally proposed for clustering a pair of networks.

The rest of the paper is organized as follows: In Section 2, we first introduce the way to construct segment graphs and a similarity graph as an integrated multi-partite graph for a given set of images. With the derived graphical representation, we solve the co-segmentation problem by searching for low conductance sets of vertices as desired clusters [8-10] according to the underlying Markov chain of an alternating random walk on the segment graphs and similarity graph. In Section 3, we test our method on the iCoseg benchmark dataset [11] and a sequence of echocardiac images. Compared with image segmentation based on individual images and a feature-based clustering algorithm, our preliminary experimental results show that our method achieves higher segmentation accuracy by integrating information across images. It is promising, considering both segmentation performance and time complexity, to formulate the image co-segmentation problem as a joint multiple network clustering problem.

2. METHODOLOGY

In this section, we formulate the image co-segmentation problem as a multiple network clustering problem. In Section 2.1, we first introduce how to construct the segment graphs by extracting superpixels in multiple images. Then, we present the way to compose the similarity graph with edges revealing the relationships among superpixels across different images in Section 2.2. Finally, in Section 2.3, we derive an alternating



Fig. 1. The first column shows two given images in which a cheetah is the common object of interest. The second column displays the derived superpixels based on image characteristics. The third column provides a schematic illustration of the segment graphs and similarity graph constructed from the derived superpixels across two images. The segment graphs have superpixels represented by red vertices with orange edges capturing the feature and spatial proximity. The similarity graph reflects the superpixel similarity across images represented by blue dash edges. The fourth column illustrates our alternating random walk strategy. The green dash arrows and black dash arrows denote two possible one step walk scheme for our alternating random walk at the yellow vertex. The last column visualizes the results obtained by our graph-based co-segmentation method.

random walk strategy on both segment graphs and similarity graph to solve image co-segmentation.

2.1. Segment Graph Construction

Given a set of images $\mathbf{I} = \{I_1, I_2, ..., I_n\}$ sharing similar objects, we first over-segment each image I_m into k_m superpixels $S_m = \{s_m^1, s_m^2, ..., s_m^{k_m}\}$, where k_m may vary for different images. In the current work, we adopt the VLFeat toolbox [12] for extracting superpixels as similarly done in [4,5]. We note that the quality of the derived superpixels may affect the final segmentation results but it is outside the scope of this paper. With extracted superpixels from each image I_m , we construct a weighted "segment graph" G_m , in which each vertex corresponds to a superpixel. For an arbitrary pair of superpixels, we assign an edge weight based on their spatial and appearance proximity. Intuitively, similar superpixels should have larger weights. With a non-negative adjacency matrix W_m denoting the similarity between superpixels, we compute each edge weight

$$W_m^{ij} = exp(-\lambda \|x_m^i - x_m^j\| - \mu \|d_m^i - d_m^j\|), \quad (1)$$

where x_m^i and x_m^j are the feature descriptors for the superpixels s_m^i and s_m^j and d_m^i and d_m^j are the coordinates of the gravity center for superpixels s_m^i and s_m^j in the image space. In this paper, x_m^i denotes the feature descriptor obtained by computing the color histogram as discussed in [13]. For d_m^i , we simply compute

$$d_m^i = \frac{1}{N_{s_m^i}} \sum_{l \in s_m^i} c(l),$$
 (2)

where $N_{s_m^i}$ indicates the total number of pixels within the superpixel s_m^i and c(l) denotes the image coordinates for the pixel l.

With the adjacency matrix W_m for each given image, we further construct the block adjacency matrix W for an overall segment graph G_w of all n images in I by padding each W_m on the corresponding diagonal block of W as follows:

$$W = \begin{bmatrix} W_1 & 0 & 0 & 0\\ 0 & W_2 & 0 & 0\\ 0 & 0 & \dots & 0\\ 0 & 0 & 0 & W_n \end{bmatrix}_{M \times M},$$
 (3)

where $M = \sum_{m=1,...,n} k_m$ denoting the total number of superpixels across all images. The construction procedure is schematically illustrated in the third column of Fig. 1.

2.2. Similarity Graph Construction

In order to establish the relationships of superpixels in different images for information integration, we further derive a similarity graph G_h based on the superpixel appearance proximity across images. As we mentioned in Section 2.1, for each superpixel s_m^i , we can compute its feature descriptor x_m^i . Hence, the similarity between images I_p and I_q can be similarly computed by the similarity matrix H_{pq} with each matrix element H_{pq}^{ij} estimated as follows:

$$H_{pq}^{ij} = \phi(x_p^i) \times \phi(x_q^j) = \kappa(x_p^i, x_q^j).$$
(4)

In this paper, for simplicity, we use a linear kernel which is the inner product for two feature descriptors. Finally, the overall adjacency matrix of the similarity graph of G_h across all images can be written as:

$$H = \begin{bmatrix} 0 & H_{12} & \dots & H_{1n} \\ H_{12}^T & 0 & \dots & H_{2n} \\ \dots & \dots & 0 & \dots \\ H_{1n}^T & H_{2n}^T & \dots & 0 \end{bmatrix}_{M \times M}$$
(5)

where H is an symmetric matrix. We note that the segment graph G_w and the similarity graph G_h share the same vertex sets but with different types of edges. This graphical representation is a generalization of the multi-partite graph, which motivates us to develop the following random walk strategy to solve the joint graph clustering problem. The third column in Fig. 1 also illustrates the constructed similarity graph.

2.3. An Alternating Random Walk on Graphs

To make use of the information carrying in both the segment graph G_w and the similarity graph G_h , we propose to use the alternating random walk strategy [8], in which we require that the random walker walks on two different types of edges in G_w and G_h in an alternating manner. The transition matrix of the underlying Markov chair for our random walk strategy can be computed as follows:

$$P = \frac{1}{2}P_{w}P_{h} + \frac{1}{2}P_{h}P_{w}$$
(6)

where P_w and P_h are the corresponding transition probability matrices for the random walk on G_w and G_h , respectively. P_w and P_h can be computed by

$$P_w = W D_w^{-1} \text{ and } P_h = H D_h^{-1}, \tag{7}$$

where D_w and D_h are diagonal matrices with the summation of edge weights for each vertex on the corresponding diagonal entry.

The fourth column in Fig. 1 illustrates our alternating random walk strategy. The random walker at the yellow vertex in the fourth column of Fig. 1 can either first walk on the edges in G_w (orange edges) then G_h (blue dash edges) or first walk on edges in G_h (blue dash edges) then G_w (orange edges). These two random walk strategies can be adopted with the equal probability and hence we obtain the corresponding transition probability matrix for the alternating random walk (6).

Based on this formulation, we can solve joint graph clustering by searching for k low conductance sets according to the transition matrix P to identify corresponding k cosegmentations for the given image set [8–10]. Following the derivation in [8,9], searching for k low conductance sets by Pcan be found by solving the following optimization problem:

$$\max \quad trace\left(\frac{Y^T\bar{P}Y}{Y^TD_{\bar{P}}Y}\right) \\ s.t. \quad Y1_k = 1_N, y_{ij} \in \{0,1\},$$

$$(8)$$

where Y is the assignment matrix to assign each superpixel to the corresponding segmentation and $D_{\bar{P}}^{ii} = \sum_{j} \bar{P}^{ij}$. \bar{P} can be computed by

$$\bar{P} = \frac{\pi P + P^T \pi}{2},\tag{9}$$

where π is the stationary distribution of the corresponding Markov chain of the alternating random walk ($P^T \pi = \pi$). We can derive a similar spectral method as in [8, 14] to obtain the approximate solution to the optimization problem (8).

3. EXPERIMENTS

We report experimental results on two datasets: the iCoseg benchmark with the ground truth [11] as well as a sequence of clinical short-axis echocardiac images temporally sampled from the diastole to systole [15].

For the iCoseg dataset, because we know the ground truth, we compute the F-measure to evaluate the performance:



Fig. 2. Visual comparison of segmentation results by different competing methods for the Bear category of images in iCoseg. The first row contains the original images in iCoseg. The second row shows the co-segmentation results obtained by our proposed method. The third row provides the results produced by the feature-based normalized cut based on the overall feature similarity matrix \overline{S} for all derived superpixels across images (SimNCut). The last row displays results based on implementing normalized cut on individual images independently (NCut).

$$F_{measure} = 2 * \frac{precision * recall}{precision + recall}$$
(10)

where $precision = \frac{T \cap R}{T}$ with T denoting the obtained segmentation by segmentation methods and R the ground truth; and $recall = \frac{T \cap R}{R}$. Examples of the segmentation results are also provided for visual inspection. For echocardiac images, only visualization of obtained segmentation results are given due to the lack of the ground truth.

3.1. The iCoseg Dataset

The iCoseg benchmark [11] consists of several categories of images. Each category contains images with the same objects or similar objects of the same kind. It is challenging to segment those images in iCoseg because the objects vary in terms of shape, illumination, and view perspectives. The dataset has been adopted as the benchmark to evaluate different image segmentation methods [11].

In order to demonstrate the strength and potential of our proposed graph-based co-segmentation method, we compare our co-segmentation results with the results obtained from separate segmentation of individual images as well as the results from a feature-based clustering method that considers all derived superpixels from all the given images. For segmenting individual images, we adopt the classical normalized cut algorithm (NCut) [16] as the competing algorithm with the exactly same segment graph construction for each image. For the feature-based clustering method considering all superpixels across images, we apply the normalized cut on the overall similarity matrix \overline{S} , which is similar to the constructed similarity graph with each element reflecting the pairwise similarity based on the feature descriptor of the corresponding pair of superpixels within and across images (SimNCut).

iCoseg	Method	Precision	Recall	F-measure
Cheetah	proposed	0.8317	0.8119	0.8217
	SimNCut	0.7223	0.876	0.7918
	NCut	0.5726	0.6866	0.6245
Bear	proposed	0.8424	0.8536	0.848
	SimNCut	0.8886	0.3777	0.5301
	NCut	0.5776	0.6157	0.5961
Goose	proposed	0.9679	0.9037	0.9347
	SimNCut	0.9227	0.8913	0.9067
	NCut	0.525	0.7617	0.6216
Taj Mahail	proposed	0.4536	0.8969	0.6025
	SimNCut	0.3584	0.7429	0.4835
	NCut	0.3858	0.779	0.516

 Table 1. Segmentation performance comparison for the competing algorithms

Table 1 gives the F-measure together with the precision and recall indices for the obtained segmentation results by all three competing methods for four categories of images in iCoseg. It is clear that our graph-based co-segmentation algorithm outperforms the other two competing algorithms based on the accuracy of segmentation results, which demonstrates that our method can integrate the useful information within and across images based on the constructed multi-partite graph with segment graphs constructed from respective images and the similarity information across images through the similarity graph. By borrowing strengths across images through the proposed alternating random walk, we can achieve better segmentation performance for multiple images. To further illustrate the performance difference of all three competing algorithms, we display all the segmentation results for the category of images with bears in Fig. 2. It is also obvious visually that our method has obtained the most promising segmentation results. Especially, for the image in the third column of Fig. 2, we are able to successfully segment out the bears by borrowing superpixel similarity information from other images compared to other competing methods. The segmentation results by our graph-based co-segmentation for the other three categories of iCoseg images: Cheetah, Goose, and Taj Mahal, are also provided for visual inspection in Fig. 3.1. Overall, our new method based on multiple network clustering yields promising segmentation results in iCoseg.

3.2. Echocardiac Image Sequence

We analyze a sequence of four echocardiac images sampled from the cardiac cycle from diastole to systole. Our co-segmentation results are illustrated in Fig. 4, from which we find that visually, our method extracts endocardiums correctly.

We note that the obtained endocardium boundaries based on the co-segmentation results are not smooth due to the use of superpixels when constructing the graphical representation. We will further explore different strategies to derive appropriate superpixels for more accurate and high-



Fig. 3. Segmentation results for Cheetah, Goose and Taj Mahal categories in the iCoseg benchmark. The first, third, and fifth rows display the original images in iCoseg and the second, fourth, and sixth rows illustrate our segmentation results.

resolution endocardium segmentation in medical image sequences. In addition, segmentation results are also dependent on the feature descriptors as well as adopted feature similarity measures. Depending on the characteristics of given image sets, they may need to be invariant with respect to imaging conditions in different images so that final cosegmentation can provide robust results. We plan to further explore these directions together with our proposed graphbased co-segmentation method to improve the performance in future.



Fig. 4. Co-segmentation results for the echocardiac image sequence. The overlaid images with a white background visualize the identified endocardiums in these four images.

4. CONCLUSIONS

We proposed a new graph-based image co-segmentation framework to simultaneously segment multiple images by viewing each image as a graph of superpixels and integrating information shared among multiple images through a similarity graph. The construction of this multi-partite graph based on superpixels across all images is presented and an alternating random walk strategy is developed to successfully segment similar objects of interest shared in the given images. Our preliminary experiments have demonstrated the superiority of our co-segmentation algorithm over segmentation based on individual images, which motivates us to further investigate improved graph-based information integration for more accurate and robust image and video co-segmentation.

5. REFERENCES

- A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, New York: Academic, 1982.
- [2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [3] T.F. Chan and L.A. Vese, "Active contours with edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [4] A. Joulin, F. Bash, and J. Ponce, "Multi-class cosegmentation," in CVPR, 2012.
- [5] A. Joulin, F. Bash, and J. Ponce, "Discriminative clustering for image co-segmentation," in CVPR, 2010.
- [6] S. Vicente, R. Casten, and K. Vladimir, "Object cosegmentation," in CVPR, 2011.
- [7] M. Rubinsten, A. Joulin, J. Kopf, and C. Liu, "Unsupervised joint object discovery and segmentation in internet images," in CVPR, 2013.
- [8] Y. Wang and X. Qian, "Joint clustering of protein interaction networks through markov random walk," *BMC Systems Biology*, vol. 8, no. suppl 1, pp. S9, 2014.
- [9] Y. Wang and X. Qian, "Functional module identification in protein interaction networks by interaction patterns," *Bioinformatics*, vol. 30, no. 1, pp. 81–93, 2014.
- [10] M. Jerrum and A. Sinclair, "Conductance and the rapid mixing property for Markov chains: The approximation of permanent resolved," in *Proceedings of the 20th annual ACM symposium on Theory of computing (STOC)*, 1988, pp. 235–244.
- [11] D. Batra, A. Kowdle, D. Parikh, and et al, "iCoseg: Interactive co-segmentation with intelligent scribble guidance," in *CVPR*, 2010.
- [12] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision alogrithms," http://www.vlfeat.org, 2008.
- [13] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582– 1596, 2010.
- [14] E. Xing and M. Jordan, "On semidefinite relaxation for normalized k-cut and connections to spectral clustering," Tech. Rep., UC. Berkeley, 2003.

- [15] X. Qian and B. Yoon, "Contour-based hidden markov model to segment 2d ultrasound images," in *ICASSP*, 2011.
- [16] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, 2000.