UNSUPERVISED CO-SEGMENTATION BASED ON A NEW GLOBAL GMM CONSTRAINT IN MRF

Hongkai Yu, Min Xian, and Xiaojun Qi

Department of Computer Science, Utah State University, Logan, UT 84322-4205
hongkai.yu@aggiemail.usu.edu, min.xian@aggiemail.usu.edu, and xiaojun.qi@usu.edu

ABSTRACT

This paper proposes a new Markov Random Fields (MRF) optimization model for co-segmentation. The co-saliency model is incorporated into our model to make it fully unsupervised and work well for images with similar backgrounds. The Gaussian Mixture Model (GMM) based dissimilarity between foregrounds in each image and the common objects in the set is involved as a new global constraint (i.e., energy term) in our model. Finally, we introduce an alternative approximation to represent the energy function, which could be minimized by Graph Cuts iteratively. The experimental results on two datasets show that our algorithm achieves better or comparable accuracy when comparing with state-of-the-art algorithms.

Index Terms— Co-segmentation, Markov Random Fields, Global GMM Constraint, Graph Cuts

1. INTRODUCTION

Co-segmentation known as the task of jointly segmenting the common objects in multiple images attracts many attentions recently. It has a variety of potential applications such as robust image retrieval, 3D model reconstruction, joint object discovery, etc.

Many existing co-segmentation methods model it as an MRF optimization problem. MRF based co-segmentation algorithm was first introduced by Rother *et al.* [1], where they co-segment the common objects in an image pair. L1-norm of dissimilarity between foreground histograms is incorporated as a global constraint (i.e., energy term) in MRF, and Trust Region Graph Cuts is finally used to minimize the energy function [1]. Mukherjee *et al.* [2] replace the global energy term with L2-norm of dissimilarity between foreground histograms and use the Pseudo-Boolean optimization method to minimize the energy function. Hochbaum and Singh [3] modify the global energy term using similarity between foreground histograms and then apply a max-flow solution like Graph Cuts [4] to efficiently optimize the energy function. Chang *et al.* [5] formulate the global energy term by simultaneously considering foreground dissimilarity and dissimilarity between foreground and background. Rubio *et al.* [6] involve an energy term of inter-image region matching in MRF to explore co-segmentation. Rubinstein *et al.* [7] include several energy terms such as saliency, pixel correspondence, and foreground likelihood in MRF for co-segmenting a set of possibly noisy images. All these MRF-based methods are unsupervised. On the other hand, some MRF-based co-segmentation methods are supervised. For example, Cui *et al.* [8] integrate a local color pattern and edge model learned from segmentation of an example image into MRF to segment new images. Batra *et al.* [9] propose an interactive MRF model, which requires human interaction to guide co-segmentation. Vicente *et al.* [10] train a Random Forest regressor from the ground truth segmentation for object co-segmentation.

Besides MRF-based methods, some researchers tackle the co-segmentation problem using other optimization models. Joulin *et al.* [11] build a discriminative clustering framework to treat co-segmentation as a combinatorial optimization problem. Kim *et al.* [12] model co-segmentation as a temperature maximization problem based on anisotropic heat diffusion. Meng *et al.* [13] formulate co-segmentation as the shortest path problem on a digraph.

These existing co-segmentation methods produce impressive co-segmentation results. However, they still have some disadvantages: First, supervised algorithms [8, 9, 10] are limited in their use, so unsupervised algorithms are preferred. Second, some global energy terms in MRF are measured only between an image pair, so they cannot co-segment more than two images [1, 2, 3]. Third, some frameworks are not able to perform co-segmentation in images with similar backgrounds [1, 2, 3]. This paper proposes a new MRF optimization model to overcome these disadvantages. Our contributions are summarized as follows: 1) Building a new MRF optimization model involving the co-saliency model to make our algorithm fully unsupervised and work well for images with similar backgrounds. 2) Proposing a new global energy term into MRF using GMM constraint. This is different from the previous research using histogram constraint as the global energy term. 3) Presenting a new definition of the global energy term by using common objects as an intermediary. This definition makes our model work well for more than two images. 4) Introducing an alternative approximation to optimize the

proposed MRF model.

The rest of the paper is organized as follows: Section 2 presents the proposed MRF optimization model. Section 3 compares our algorithm with state-of-the-art algorithms on two datasets. Section 4 concludes the paper with our contributions and presents the directions for future work.

2. PROPOSED MRF OPTIMIZATION MODEL

We model the co-segmentation problem as a binary labeling problem in MRF. By combining the intra-image and inter-image energy terms, our model assigns optimal labels for a set of N images $I = \{I_i\}_{i=1}^{N}$ containing common objects. To facilitate the discussion later on, we define $\mathbf{x_i} = \{x_i^p\}_{p=1}^{n_i}$ as a set of binary labels in image I_i, where x_i^p is either 0 as background or 1 as foreground for the pth pixel in I_i and n_i is the number of pixels in I_i.

2.1. Co-segmentation energy function

Our model aims to assign optimal labels $\mathbf{X}$ ($\{\mathbf{x_i}\}_{i=1}^{N}$) for the image set by minimizing the following MRF energy function:

$$E(\mathbf{X}) = E^{in}(\mathbf{X}) + E^{Global}(\mathbf{X}) \quad (1)$$

where E^{in} is the intra-image energy term and E^{Global} is the global energy term modeling the inter-image energy under labeling $\mathbf{X}$. E^{in} measures the overall coherence within each image in the set in terms of both co-saliency and local smoothness. E^{Global} measures the overall foreground dissimilarity among all images in the set in terms of GMM constraint.

2.2. Intra-image energy term

The intra-image energy term E^{in} contains data term and smoothness term, where the data term is derived from the co-saliency model. Its definition is as follows:

$$E^{in}(\mathbf{X}) = \sum_{i=1}^{N}[\lambda E^{i,CoSal}(\mathbf{x_i}) + E^{i,Smooth}(\mathbf{x_i})] \quad (2)$$

where $E^{i,CoSal}$ and $E^{i,Smooth}$ respectively represent co-saliency and local feature consistency in image I_i and λ weighs the importance of $E^{i,CoSal}$. By combining these two terms, E^{in} encourages co-salient pixels with local feature consistency to be labeled as foreground.

Co-saliency energy term: The co-saliency model is integrated into our model to explore the common saliency in multiple images. Considering the inter-image correspondence, researchers recently improve the saliency model in single image and propose the co-saliency model for multiple images [5, 14, 15]. Saliency and co-saliency models have been used in some co-segmentation methods [5, 7, 13]. We adopt the co-saliency model due to its following advantages: 1) Co-saliency detection is fully unsupervised, which makes our model unsupervised. 2) Co-saliency considers the correspondence of the common objects in multiple images to strengthen their appearance. 3) Co-saliency is able to highlight the common objects in multiple images to eliminate the effect of similar backgrounds. The cluster-based method [14] combining contrast cue, spatial cue and corresponding cue is used to generate the normalized co-saliency map $\mathbf{M_i}$ (values in [0,1]) for each image I_i in our model. Fig. 1 shows sample saliency and co-saliency maps. It clearly shows that the co-saliency map strengthens the appearance of the common objects and removes background noise. The co-saliency energy term is defined by:

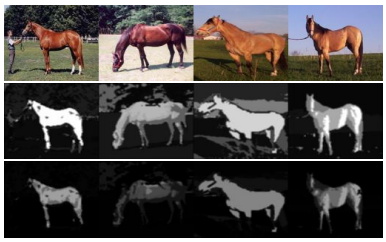

Fig. 1: Comparison of saliency and co-saliency detection results. Top: original images; Middle: single saliency map without considering inter-image correspondence by [14]; Bottom: co-saliency map by [14].

$$E^{i,CoSal}(\mathbf{x_i}) = \sum_{p=1}^{n_i}[M_i^p x_i^p + (1 - M_i^p)(1 - x_i^p)] \quad (3)$$

where M_i^p is the co-saliency value of pixel p in image I_i which measures its probability to be the common objects.

Smoothness energy term: The smoothness energy term $E^{i,Smooth}$ encourages coherence in local regions with similar features in image I_i. It is computed as:

$$E^{i,Smooth}(\mathbf{x_i}) = \sum_{(p,q)\in\mathbf{C_i}} [x_i^p \neq x_i^q]\exp(-\beta\left\|z_i^p - z_i^q\right\|^2) \quad (4)$$

where $[\phi]$ denotes the indicator function taking a value of 1 or 0 for a true or false predicate ϕ, z_i^p is the feature of pixel p in image I_i, $\mathbf{C_i}$ is the set of neighboring pixels in image I_i, and β is a scale computed by $(2\langle(z_i^p - z_i^q)^2\rangle)^{-1}$ with $\langle\cdot\rangle$ denoting expectation over image I_i. $E^{i,Smooth}$ serves as a penalty for discontinuity among the neighboring pixels when their labels are different in the MRF model.

2.3. Global energy term

Many previous studies utilize a global energy term in MRF to model the inter-image information among multiple images. The global energy terms in [1, 2, 3] only model foreground or background correspondence using the histogram constraint between two images. The global term in [5] combines the correspondence of each image pair in the set to co-segment more than two images.

Unlike previous global energy terms modeling foreground or background correspondence between two images, our global energy term measures dissimilarity between foregrounds in each image and the common objects in the set. This is a significantly novel way to define the global energy

term suitable for co-segmenting more than two images. Another novelty is that our global energy term is based on the GMM constraint while other research uses the histogram constraint as the global energy term. GMM offers the following advantages over histogram: 1) GMM is a parametric model of the probability distribution which provides more accurate information. 2) GMM details the distribution of different components of an object. So we choose the GMM constraint to construct the global energy term by:

$$E^{Global}(\mathbf{X}) = \sum_{i=1}^{N} D(\theta_i^f, \theta_{co}) \tag{5}$$

where θ_i^f denotes the foreground GMM of image I_i, θ_{co} denotes the GMM of the common objects constructed from foreground pixels of all images, and $D(,)$ is the L-1 norm dissimilarity function. Let μ_i^k denote the mean of the kth Gaussian component of θ_i^f. π_{co}^t and μ_{co}^t represent the weight and mean of the tth Gaussian component of θ_{co}, respectively. $D(,)$ is defined as:

$$D(\theta_i^f, \theta_{co}) = \sum_{k=1}^{K} \pi_{co}^t |\mu_i^k - \mu_{co}^t| \tag{6}$$

where K is the number of Gaussian components of GMM and t is computed by:

$$t = \underset{t \in \{1,\ldots,K\}}{\operatorname{argmin}} |\mu_i^k - \mu_{co}^t| \tag{7}$$

The computed t ensures that the tth Gaussian component of θ_{co} that is the most similar to the kth Gaussian component of θ_i^f is chosen to compute the dissimilarity function. Minimizing the global energy term would generate the binary labels making foreground in each image consistent with the common objects.

2.4. Optimization

The global energy term makes our energy function extremely hard to minimize. As a result, we introduce an approximate solution to minimize the energy function. Let us define a **global feature energy term** as follows:

$$E^{GF}(\mathbf{X}) = \sum_{i=1}^{N} \sum_{p} -\log[P(z_i^p|\theta_{co})x_i^p + P(z_i^p|\theta_i^b)(1 - x_i^p)] \tag{8}$$

where $P(\cdot)$ is the Gaussian probability distribution and θ_i^b denotes the background GMM of image I_i. E^{GF} is a kind of data term in MRF similar to the data term in GrabCut algorithm [16], which evaluates the overall fit of the labels to the pixel features, given the GMM model (θ_{co} for foreground in all images and θ_i^b for background in image I_i). Assuming foreground and background are not similar in each image, minimizing E^{GF} would segment each image into two parts where the foreground part is close to θ_{co}. In other words, minimizing E^{GF} would force foreground parts of each image to be consistent with the common objects. This result is exactly the effect of minimizing E^{Global}. Therefore, under the above reasonable assumption, minimizing E^{GF} is an approximation to minimize E^{Global}. Then, the co-segmentation energy function E in Eq. 1 could be approximated by:

$$E(\mathbf{X}) \approx E^a(\mathbf{X}) = E^{in}(\mathbf{X}) + E^{GF}(\mathbf{X}) \tag{9}$$

It is obvious that E^a could be minimized by Graph Cuts. To approximately minimize E in Eq. 1, Graph Cuts is applied to minimize E^a iteratively. Given the segmentation result of each image, θ_i^b could be learned after applying K-means to the segmented backgrounds in each image and θ_{co} could be learned after applying K-means to all the segmented foregrounds in the set. θ_i^b of each image and θ_{co} could be initialized from the preliminary segmentation results by minimizing E^{in}. They are then updated from segmentation results by minimizing E^a at each iteration. The approximation process is repeated until convergence (we typically used 5-10 iterations). The procedure of our co-segmentation method is summarized in Algorithm 1. The segmentation results of two sample images at each stage are illustrated in Fig. 2.

Fig. 2: Co-segmentation results at each step. First column: original images; Second column: co-saliency map; Third column: preliminary segmentation by minimizing E^{in}; Columns 4 through 8: segmentation results at iterations 1 through 5.

Algorithm 1 The proposed co-segmentation algorithm

Input: N images containing common objects
Output: segmentation label for each image

1: Generate the co-saliency map for the image set by [14].
2: Apply Graph Cuts [4] to minimize E^{in} (Eq. 2) to get preliminary segmentation results.
3: Initialize θ_i^b $(1 \le i \le N)$ and θ_{co}.
4: **repeat**
5: Apply Graph Cuts to minimize E^a (Eq. 9) to refine the segmentation results.
6: Update θ_i^b $(1 \le i \le N)$ and θ_{co}.
7: **until** convergence

3. EXPERIMENTAL RESULTS

We compare our algorithm with several state-of-the-art unsupervised co-segmentation algorithms in this section. For comprehensive comparison, experiments on two datasets are reported. One is the Pair dataset, a simple set from [3], which contains exactly two images for each class. The foreground in each pair has low variations in pose, size, and viewpoint. Another is the challenging dataset including Weizman horses and MSRC dataset [5, 11], which contains up to 30 images

Table 1: Co-segmentation errors on Pair dataset.

	Stone	Boy	Bear	Dog	Llama	Banana	**Mean**
Ours	1.0%	10.5%	9.1%	6.5%	6.6%	3.3%	6.2%
[11]	0.9%	6.5%	5.5%	6.4%	18.8%	6.5%	7.4%
[3]	1.2%	1.8%	3.9%	3.5%	3.0%	3.1%	2.8%

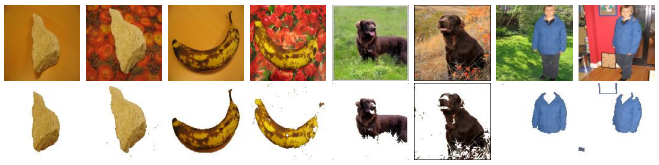

Fig. 3: Sample co-segmentation results on Pair dataset. Original images and corresponding co-segmentation results for Stone, Banana, Dog, and Boy are shown from left to right.

per class. Their foreground exhibits higher variations in pose, size, and viewpoint. The ground truth segmentation of the two datasets are both publicized.

The segmentation accuracy as used in [5, 11] is computed to evaluate the performance. It is calculated as the ratio of the number of correctly labeled pixels to the total number of pixels in an image. The weighting parameter λ, whose value is close to 1, is adjusted heuristically for different classes. The parameter K of GMM is set to be 5. The pixel feature used in experiments is the RGB color feature.

3.1. Experiments with Pair Dataset

Table 1 summarizes the experimental results on the Pair dataset. On average, our algorithm achieves the mean error of 6.2%, which is lower than the mean error of 7.4% obtained by [11] using more complex features (color and Gabor). The smallest mean error of 2.8% is achieved by [3], which uses a priori that the image pair has the same foreground with different backgrounds. Our algorithm is fully unsupervised without this kind of priori.

Fig. 3 shows sample co-segmentation results of our algorithm on the Pair dataset. Our algorithm generates satisfactory results for all classes except for the class Boy due to the heavy edges between the coat and other parts. [11] also has similar error for the class Boy because of the same reason.

3.2. Experiments with Weizman horses and MSRC Dataset

Table 2 summarizes the experimental results on the Weizman horses and MSRC dataset. It demonstrates that our algorithm is effective to co-segment dozens of images with similar backgrounds. Our algorithm achieves the mean accuracy of 83.1% which is significantly higher than the mean accuracy of [11] (79.1%) and [6] (70.8%). Specifically, it yields better accuracy in six out of eight classes than [11] and [6], and obtains the highest accuracy on the two classes (Cow and Horse). Our algorithm yields slightly lower mean accuracy than [5] (86.6%), which involves additional learning for visual vocabulary besides MRF optimization model and more complex features (SIFT). Some classes such as Cat and Cow have foregrounds with different color and texture, which makes [6] and [11] not perform well. However, our MRF optimization

Table 2: Co-segmentation accuracy on the Weizman horses and MSRC dataset. Italic numbers show the classes that our algorithm performs better than one competitor. Bold numbers display the classes in which our algorithm performs better than two competitors. Red bold numbers show the classes where our algorithm performs best.

Class	Images	Ours	[11]	[6]	[5]
Cars(front)	6	*83.6%*	87.7%	65.9%	90.8%
Cars(back)	6	*74.5%*	85.1%	52.4%	85.8%
Face	30	**84.5%**	84.3%	76.3%	87.3%
Cow	30	**91.7%**	81.6%	80.1%	91.4%
Horse	30	**87.6%**	80.1%	74.9%	86.4%
Cat	24	**84.2%**	74.4%	77.1%	86.7%
Plane	30	**85.7%**	75.9%	77.0%	87.7%
Bike	30	**73.2%**	63.3%	62.4%	76.8%
Mean		83.1%	79.1%	70.8%	86.6%

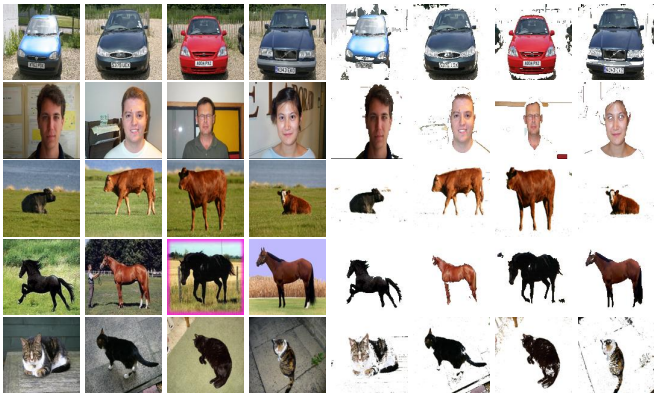

Fig. 4: Sample co-segmentation results on Weizman horses and MSRC dataset. Four original images and corresponding co-segmentation results for Cars(front), Face, Cow, Horse, and Cat are shown from top to bottom.

model combining the co-saliency model and the global GMM constraint reduces this interference and achieves impressive results. All the three algorithms and ours do not perform well on the class Bike due to the special structure of bicycle. Fig. 4 shows sample co-segmentation results by our algorithm on the Weizman horses and MSRC dataset.

4. CONCLUSIONS

We present a novel MRF optimization model for object co-segmentation. Our contributions are: 1) Involving the co-saliency model into our model to make co-segmentation fully unsupervised and work well for images with similar backgrounds. 2) Proposing a new global GMM constraint (i.e., energy term) into MRF. 3) Presenting a new definition for the global energy term by using common objects as an intermediary. 4) Introducing an approximation alternative to optimize our model. Experimental results show that our algorithm using simple color feature achieves better or comparable accuracy when comparing with state-of-the-art algorithms.

In the future, we plan to extend our model to solve multi-class co-segmentation and explore the possibility to employ our model in video co-segmentation.

5. REFERENCES

[1] C. Rother, T. Minka, A. Blake, and V. Kolmogorov, "Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 993–1000.

[2] L. Mukherjee, V. Singh, and C. R. Dyer, "Half-integrality based algorithms for cosegmentation of images," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2028–2035.

[3] D.S. Hochbaum and V. Singh, "An efficient algorithm for co-segmentation," in *Proceedings. IEEE International Conference on Computer Vision(ICCV)*, 2009, pp. 269–276.

[4] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in nd images," in *Proceedings. IEEE International Conference on Computer Vision (ICCV)*, 2001, pp. 105–112.

[5] K. Y. Chang, T. L. Liu, and S. H. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2129–2136.

[6] J. C. Rubio, J. Serrat, A. Lpez, and N. Paragios, "Unsupervised co-segmentation through region matching," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 749–756.

[7] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, "Unsupervised joint object discovery and segmentation in internet images," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1939–1946.

[8] J. Cui, Q. Yang, F. Wen, Q. Wu, C.Zhang, L.V. Gool, and X. Tang, "Transductive object cutout," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[9] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 3169–3176.

[10] S. Vicente, C. Rother, and V. Kolmogorov, "Object cosegmentation," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2217–2224.

[11] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1943–1950.

[12] G. Kim, E.P. Xing, L. Fei-Fei, and T. Kanade, "Distributed cosegmentation via submodular optimization on anisotropic diffusion," in *Proceedings. IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 169–176.

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

[14] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3766–3778, 2013.

[15] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3365–3375, 2011.

[16] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.