OBJECTS CO-SEGMENTATION: PROPAGATED FROM SIMPLER IMAGES

Marcus Chen⋆*, Santiago Velasco-Forero*†

⋆ Nanyang Technological University, Singapore,
† CMM, Mines ParisTech, France

Ivor Tsang‡∗*, Tat-Jen Cham*⋆

‡QCIS, University of Technology, Sydney

ABSTRACT

Recent works on image co-segmentation aim to segment common objects among image sets. These methods can co-segment simple images well, but their performance may degrade significantly on more cluttered images. In order to co-segment both simple and complex images well, this paper proposes a novel paradigm to rank images and to propagate the segmentation results from the simple images to more and more complex ones. In the experiments, the proposed paradigm demonstrates its effectiveness in segmenting large image sets with a wide variety in object appearance, sizes, orientations, poses, and multiple objects in one image. It outperformed the current state-of-the-art algorithms significantly, especially in difficult images.

***Index Terms*—** Co-segmentation, image ranking, segmentation propagation, difficult images.

1. INTRODUCTION

Image segmentation is used in many image applications for classification and recognition. Segmentation results often serve as spatial priors for object-based analysis [1] such as in remote sensing [2]. Without a clear definition of subsequent applications, segmentation by itself is not well defined; i.e., the definitions of complete objects vary according to their utilization [3]. For example, if an image contains a pedestrian wearing a hat, a good segmentation for hat recognition would be just the hat, but pedestrian recognition may require both the person and the hat as one segment. This problem is alleviated when a common object exists in a large image set. The common object becomes *a priori* information for segmentation. Image co-segmentation is typically defined as the task of jointly segmenting something similar in an image set.

Images co-segmentation has been actively researched recently [4, 5, 6, 7, 8, 9, 10]. Most of them are unsupervised except [7], which requires interaction with users. They usually leverage on similarities in foregrounds and backgrounds among different images, and integrate pixel classification into the segmentation. In modeling, these methods aim to extract what is common in all images in terms of visual features such as the Scale Invariant Feature Transform (SIFT) [4]. The existing co-segmentation methods face a few challenges. First, the common objects across images may vary substantially in appearance, color, and orientations. It is hard to model object segmentation in feature spaces, especially when image features are high dimensional and samples are few. Cluttered backgrounds may further complicate this problem. Second, including both simple and cluttered images for modeling may result in a non-discriminative model and poor segmentation in simple images as illustrated in Fig. 1. Third, a large image set may contain multiple object classes. Without the class labels, it is challenging for the existing co-segmentation techniques to work well. Lastly, the images may also contain undesired common backgrounds such as leaves, as shown in Fig. 1.

Fig. 1. Examples of co-segmentation results. First row: original images, second row: results from [6].

To meet these challenges, this paper proposes a co-segmentation paradigm to segment images sequentially, from easy to increasingly difficult images. We first propose a novel image ranking measure to rank image segmentation easiness based on a saliency measure. With a saliency prior, single image segmentation is applied to the simplest images to extract complete objects. The complete object masks are then propagated to more complex images, on which the common objects are less salient. This propagation gives a probabilistic estimation of foreground objects in the images, which are then segmented using a graph cut. This process is efficient in computation and memory, and can segment multiple object classes simultaneously without class labels. In experiments,

∗This work is supported by the Australian Research Council Future Fellowship FT130100746.

it achieved better object segmentation than the current state-of-the-art algorithms, especially on more complex images.

2. RECENT WORKS

2.1. Recent Works on Segmentation

There is a recent trend of performing image segmentation in a superpixel representation, which aims to group pixels with good spatial and intensity homogeneity. This representation allows us to process images more efficiently on the pixel group level. Two recent popular superpixel methods are Simple Linear Iterative Clustering (SLIC) [11] and Entropy Rate Superpixel (ERS) [12]. Both methods can handle object boundaries well, and are computationally efficient. SLIC employs k-means clustering in a local manner with a weighted distance measure combining both color and spatial proximity. On the other hand, ERS formulates the superpixel segmentation problem as an optimization problem on graph topology. The ERS method has a parameter on the expected number of superpixels in an image. Usually, superpixels are the intermediate steps for segmentation, as in [12].

Although using superpixels has many benefits, clustering superpixels into a complete object is not an easy task. Some well-known segmentation methods such as mean-shift [13], graph cut [14], and normalized cut methods [15] have been used in an attempt to segment out objects of interest. The mean-shift method recursively shifts the means of regions as they expand to include neighboring pixels. A cluster of pixels converges to a local distribution forming a segment, and small statistically close segments are merged into bigger segments. Both graph cut and normalized cut methods build a graph to represent pixels and their neighborhood relationships. The graph nodes are image pixels and the graph edges model the affinity between pixels. Graph-based methods have proven to be flexible to include multiple desired properties of segmentation, such as known pixels relationships and symmetry.

Unsupervised segmentation methods do not utilize any *a priori* information on the foreground and background. These methods mainly focus on spatial grouping, rather than on foreground segmentation of objects. The resultant images may not give complete foreground objects.

A recent co-segmentation work [6] utilizes the well-known discriminative clustering method for segmentation and reformulates it into an optimization problem. However, it cannot handle object variations well. This is extended in [16] to a multi-class segmentation using both spectral and discriminative clustering for a probabilistic estimation. Another concurrent work [17] formulates the co-segmentation problem using an energy minimization approach for both intra-group information within each image and inter-group information between images. This approach may be used in the propagation step of our proposed method.

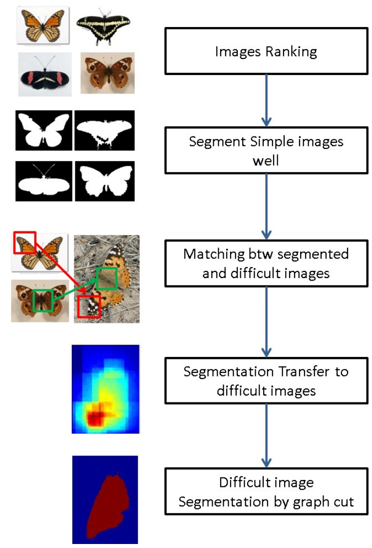

Fig. 2. Overall algorithm framework.

2.2. Segmentation Propagation

There is also a trend towards transferring knowledge to learn a new class from a few training examples by leveraging examples from related classes. Most of these works are intended for object recognition or detection, but not segmentation. The work in [18] proposes a segmentation propagation approach. With some given segmentation masks as training samples, it propagates the masks to the most similar unsegmented images before performing segmentation. These segmented images will form sources for segmentation transfer in the subsequent step. This novel approach maps the segmentation results to the test images based on patch similarity. The underlying assumption is that similar patches share a similar foreground and background segmentation. This approach is supervised as it requires initial labelled images for training.

3. UNSUPERVISED IMAGE SET SEGMENTATION

Building upon the existing unsupervised segmentation methods, this paper proposes a paradigm to co-segment objects in both simple and difficult images. In simple images without cluttered backgrounds, salient objects can quickly capture the user's attention. The common foreground objects in simple images can be segmented out readily and completely. Given sufficient well segmented images, the foreground object masks can be propagated to increasingly difficult images in the manner similar to [18]. The overall algorithm framework is illustrated in Fig. 2 and is explained in more detail as follows.

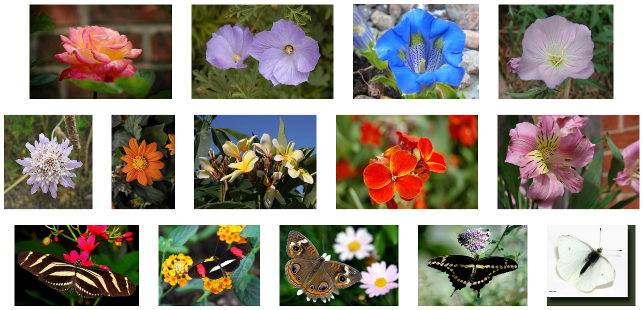

Fig. 3. Ranking of segmentation easiness based on saliency, images in row 1 have higher R_{sal} than in row 2, and are easier to segment. Row 3 images are difficult images in the butterfly dataset.

3.1. Ranking of Segmentation Easiness

An image is easy to segment if the foreground stands out from the homogeneous background. For such images, there should be a clean separation between foreground and background with clear boundaries, and the resultant segments should contain complete foreground objects. This paper proposes a saliency-based continuous measure for segmentation easiness R_{sal} as follows:

$$R_{\text{sal}} = \frac{\sum_{i \in \text{fg}} S(i)}{\sum_i S(i)}, \tag{1}$$

where $\sum_{i \in \text{fg}} S(i)$ is the sum of saliency scores over a foreground region, and $\sum_i S(i)$ is the sum over the whole image. The saliency score of every image region $S(i)$ is estimated via a global contrast saliency score as in [19]. This score is based on the region's color contrast with respect to the whole image, with weighted sum contributions from the neighboring regions. Subsequently, more salient regions are segmented out using a graph cut. Upon segmentation, the saliency ranking R_{sal} is computed. An image with a high R_{sal} score should be easy to segment. Examples of ranking by (1) are presented in Fig. 3.

3.2. Segmentation Propagation

Simple images can be readily segmented to produce good segmentation masks due to a clear separation between foreground and background in these images. The well segmented object masks are then propagated to more difficult images as a segmentation prior. Even in some images that may not be well segmented, the results can be further improved by passing them to the propagation step. The propagation step is elaborated as follows.

Let the image set be $\{I_1, I_2, ..., I_t, I_{t+1}, ...\}$, where I_k is an image according to our ranking R_{sal}. Images in $\{I_1, ..., I_t\}$ have been segmented, and image I_{t+1} is the next to be segmented. The object in image I_{t+1} may not be as salient and the background is more cluttered. This is where the well segmented images can help by propagating the segmentation masks to the similar unsegmented object regions. Since multiple objects could exist in an image, we extract possible object patches from image I_{t+1} for comparison. The image patches are extracted based on *objectness* as defined in [20], and they may overlap. Upon extraction, each patch is then matched to the closest K patches in the segmented set $\{I_1, ..., I_t\}$. The resultant segmentation prior of patch x in image I_{t+1} is defined as follows:

$$P(x) = \frac{1}{K} \sum_{l=1}^{K} exp\left(-d^2(x,l)/2\sigma^2\right), \tag{2}$$

where $P(x)$ is the prior probability of patch x being in the foreground, $d(x,l)$ is the distance between patches x and l, and σ is a parameter to set. The patch distance $d(x,l)$ is computed based on their corresponding GIST features [21]. In this manner, every pixel on the test patch will have a probability of being in the foreground and being in the background.

3.3. Segmentation with Prior Information

After propagating segmentation masks, we then segment image I_{t+1} via a graph cut [22], which solves the following energy minimization problem:

$$E(L) = \sum_i U(L_i) + \sum_{i,j} V(L_i, L_j), \tag{3}$$

where $E(L)$ is the energy to minimize, $U(L_i)$ is the unary potential of pixel i being labelled as L_i, and $V(L_i, L_j)$ is the potentials term modeling the spatial coherence between two neighboring pixels i and j. The unary potentials term is defined as $U(L_i) = -\sum_k \log(P(L_i|C_k)P(C_k))$, where $P(L_i|C_i)$ is the probability of pixel i belonging to class k, $k \in \{0,1\}$, and $P(C_k)$ is the prior probability of class k computed from (2). $P(L_i|C_i)$ is computed based on a Gaussian mixture model. The pair-wise potentials are defined as:

$$V(L_i, L_j) \propto d(i,j)^{-1} \exp\left(-\gamma \sum_{k=R,G,B} |I_i(k) - I_j(k)|_1\right),$$

where $d(i,j)$ is the pixel spatial distance and $|I_i(k) - I_j(k)|_1$ is the intensity difference across RGB channels, and γ is a constant. The minimization and pixels labelling are carried out iteratively until there is no change in pixel labels.

4. EXPERIMENTS

4.1. Data Sets

In the experiments, two data sets are chosen: the Leeds butterfly data set [23] and the flower database from Oxford University [24]. The butterfly data set has 10 categories as shown in

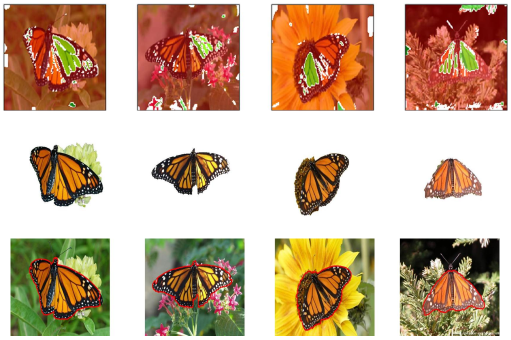

Fig. 4. Examples of co-segmentation results, in the row sequences of DCM [6], SC [19] and our results.

Table 1 and there are 832 images. The flower database consists of 102 classes of common flowers in the United Kingdom, and there are 5198 images chosen. Both data sets have a wide variety of object appearances and poses, and cluttered backgrounds. They have multiple categories of foreground objects. There are both simple and difficult images in these two data sets. The segmentation ground truths are also provided. In this work, we used the average accuracy as the performance measure.

4.2. Results

The discriminative clustering based co-segmentation method (DCM) [6], the Saliency Cut method (SC) [19] (a graph cut based method), and our proposed method are compared using these two data sets. In more complex images shown in Fig. 4, the butterflies are of different orientations, slightly different illuminations, and different sizes. In these cases, the DCM method performed poorly in segmenting out the butterflies. Both SC and our proposed methods could segment out the butterflies well. Since the DCM method could not handle variations in images and multi-class objects well and was computationally expensive to process these two data sets, only the SC method and our method were chosen for the qualitative evaluation.

Table 1 shows the segmentation results on the butterfly data set. Our method significantly outperformed SC, especially on the following categories: Heliconius charitonius (56.4% vs 72.0%), Junonia coenia (61.4% vs 72.7%), Papilio cresphontes (76.7% vs 83.6%), and Pieris rapae (54.7%, 75.2%). These categories of butterflies do not have a clear distinction between foreground and background objects, especially for Pieris rapae (the last image in Fig. 3). Our method could propagate the segmentation results and better handle these difficult images.

For the flower data set, the results are shown in Fig. 5. Our method outperformed the SC method significantly on most of the flower categories with the average score 78.8% as compared to 72.1% by SC. Computationally, it took about 8 hours

S/N	Names	SC [19]	Our Method
1	Danaus plexippus	85.8	**87.7**
2	Heliconius charitonius	56.4	**72.0**
3	Heliconius erato	73.7	**79.7**
4	Junonia coenia	61.4	**72.7**
5	Lycaena phlaeas	**79.2**	79.1
6	Nymphalis antiopa	**87.5**	83.5
7	Papilio cresphontes	76.7	**83.6**
8	Pieris rapae	54.7	**75.2**
9	Vanessa atalanta	**84.1**	83.0
10	Vanessa cardui	79.2	**82.1**
	Mean	73.9	**79.9**

Table 1. Segmentation results on butterfly data set, accuracy in percentage.

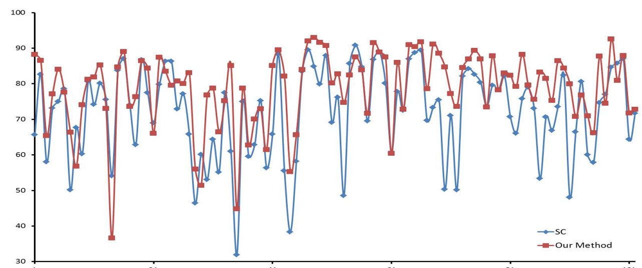

Fig. 5. Segmentation results, accuracy in percentage vs the 102 categories of flowers.

to run the all the steps in the proposed methods (in MATLAB) for 6000 images using a 16GB and 2.13 GHz Xeon machine.

5. CONCLUSION

This paper proposes a novel unsupervised co-segmentation method for large image sets. The proposed method first ranks the image set according to segmentation easiness. Using top level information such as object saliency, it can perform segmentation on simpler images very accurately. This is unsupervised, and no class label information is required. Equipped with the knowledge of both foreground objects and their accurate masks, the proposed method then transfers the segmentation knowledge to more difficult images. This sequential, simple-to-complex manner allows the proposed method to robustly segment complex images, in which objects are not as salient. In the experiments, more than 6,000 images were tested for robustness. The proposed method was compared to some current state-of-the-art algorithms, and outperformed them significantly, especially in complex images.

6. REFERENCES

[1] Tomasz Malisiewicz and Alexei A. Efros, "Improving spatial support for objects via multiple segmentations," in *British Machine Vision Conference (BMVC)*, September 2007.

[2] Thomas Blaschke, "Object based image analysis for remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 65, no. 1, pp. 2–16, 2010.

[3] Sreenath Rao Vantaram and Eli Saber, "Survey of contemporary trends in color image segmentation," *Journal of Electronic Imaging*, vol. 21, no. 4, pp. 040901–1–040901–28, 2012.

[4] Liangliang Cao and Li Fei-Fei, "Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[5] Dorit S Hochbaum and Vikas Singh, "An efficient algorithm for co-segmentation," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 269–276.

[6] Armand Joulin, Francis Bach, and Jean Ponce, "Discriminative clustering for image co-segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1943–1950.

[7] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *CVPR, 2010 IEEE Conference on*. IEEE, 2010, pp. 3169–3176.

[8] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother, "Cosegmentation revisited: Models and optimization," in *Computer Vision–ECCV 2010*, pp. 465–479. Springer, 2010.

[9] Yuning Chai, Victor Lempitsky, and Andrew Zisserman, "Bicos: A bi-level co-segmentation method for image classification," in *Computer Vision, 2011. ICCV 2011. IEEE 13th International Conference on*, 2011.

[10] Yuning Chai, Esa Rahtu, Victor Lempitsky, Luc Van Gool, and Andrew Zisserman, "Tricos: A tri-level class-discriminative co-segmentation method for image classification," in *Computer Vision–ECCV 2012*, pp. 794–807. Springer, 2012.

[11] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *PAMI, IEEE Transactions on*, vol. 34, no. 11, pp. 2274–2282, 2012.

[12] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa, "Entropy rate superpixel segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2097–2104.

[13] Dorin Comaniciu and Peter Meer, "Mean shift: A robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, 2002.

[14] Yuri Boykov and Gareth Funka-Lea, "Graph cuts and efficient nd image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[15] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.

[16] Armand Joulin, Francis Bach, and Jean Ponce, "Multi-class cosegmentation," in *CVPR, 2012 IEEE Conference on*. IEEE, 2012, pp. 542–549.

[17] F. Meng, J. Cai, and H. Li, "On multiple image group cosegmentation," in *ACCV*, 2014, vol. 1, pp. 1–15.

[18] Daniel Kuettel, Matthieu Guillaumin, and Vittorio Ferrari, "Segmentation propagation in imagenet," in *Computer Vision–ECCV 2012*, pp. 459–473. Springer, 2012.

[19] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 409–416.

[20] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari, "Measuring the objectness of image windows," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2189–2202, 2012.

[21] Aude Oliva and Antonio Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *IJCV*, vol. 42, no. 3, pp. 145–175, 2001.

[22] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[23] Josiah Wang, Katja Markert, and Mark Everingham, "Learning models for object recognition from natural language descriptions," in *Proceedings of the British Machine Vision Conference*, 2009.

[24] M-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, Dec. 2008.