Object-based RGBD Image Co-segmentation with Mutex Constraint

Huazhu Fu[1], Dong Xu[1], Stephen Lin[2], Jiang Liu[3]
[1]School of Computer Engineering, Nanyang Technological University. [2]Microsoft Research, Beijing, China. [3]I2R, A*STAR, Singapore.

The goal of co-segmentation is to extract similar foreground objects from among a set of related images. In contrast to single-image segmentation, co-segmentation makes use of the information in multiple images to infer the objects to extract. However, most existing methods operate on RGB images and utilize color-based features. They cannot distinguish a foreground from a similarly colored background, and are sensitive to illumination differences among images. In this paper, we present an object-based RGBD image co-segmentation method based on *RGBD co-saliency maps*, which capitalize on depth cues to enhance identification of common foreground objects among images, as shown in Fig. 1. Depth is also utilized to provide additional local features for region comparison and to improve selection of object-like regions. Objectness has been used in co-segmentation to overcome limitations of low-level features in separating complex foregrounds and backgrounds [3, 6], but such methods have been formulated with an assumption that *exactly one* common object exists in *all* of the images. If the common foreground is missing in an image, an irrelevant region will be extracted instead. In our work, we additionally address this issue through a fully-connected graph formulation that enables the option of selecting no regions or more than one region in an image.

Given a set of images and their depth maps, we first generate a foreground candidate pool for each image using [4]. We denote the set of object candidates by $\{x_1,...,x_M\}$, and introduce a binary label variable u_i for each object candidate x_i, which takes either the foreground label $u_i = 1$, or the background label $u_i = 0$. The task of object-based co-segmentation is formulated as a binary labeling problem in a weighted graph. Moreover, a mutex constraint modeled as a binary matrix $\mathbf{M} \in \{0,1\}^{M\times M}$ is added to connect the nodes in the graph, where M is the node number. If $M(i,j) = 1$ then the two nodes u_i, u_j cannot belong to the same label. For all nodes u_i, we set $M(i,i) = 0$. The goal of our RGBD co-segmentation is to find a labeling $\mathbf{u} = [u_1,...,u_M]^T$ that minimizes the following objective function:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} \frac{1}{2}\alpha\mathbf{u}^T\mathbf{A}\mathbf{u} - \mathbf{b}^T\mathbf{u}, \qquad (1)$$
$$\text{s.t. } \mathbf{u}^T\mathbf{M}\mathbf{u} = 0, \text{ and } \forall i \in \mathcal{V} : u_i \in \{0,1\},$$

where $\mathbf{A}$ is the pairwise matrix, and $\mathbf{b}$ associates a positive unary term b_i to each node u_i.

Unary term b measures the likelihood that the candidate belongs to the foreground, and it is defined as

$$b_i = Obj(u_i)\cdot Sal(u_i). \qquad (2)$$

The first term is the objectness score $Obj(u_i)$ computed from [4], which reflects the confidence that a region contains a generic object in the RGB-D image. The second term is the RGBD co-saliency score $Sal(u_i)$. We combine the depth and co-saliency cues from multiple RGBD images by integrating them within the saliency map fusion framework of [2]. The RGBD co-saliency score $Sal(u_i)$ for node u_i is computed as

$$Sal(u_i) = S(u_i) + \log\left(\frac{S(u_i)}{S(\bar{u}_i)} + 1\right), \qquad (3)$$

where $S(u_i)$ denotes the mean RGBD co-saliency map value among the pixels of object candidate u_i, and $\bar{u}_i$ denotes the pixels outside of candidate u_i but within the minimum bounding box enclosing the candidate.

Pairwise matrix A measures the similarity between two object candidates, and is defined by using L2-norm distance of combined color, shape and depth features.

Mutex constraint M is used to control the selection of overlapping candidates within the same image. It is a necessary component for our graph structure in which all candidates from all images are connected to each other. Since significantly overlapping candidates within an image often have a low pairwise distance, they may be selected together in the co-segmentation result if a mutex constraint is not applied between them. In this paper, we measure the overlap between two candidates as

$$Overlap(i,j) = \frac{R(u_i)\cap R(u_j)}{\min\left(R(u_i),R(u_j)\right)}, \qquad (4)$$

where $R(u_i)$ denotes the area of candidate u_i. Based on this overlap measure, we define the mutex constraint matrix as

$$M(i,j) = \begin{cases} 1, & \text{if } Overlap(i,j) \geq \tau \text{ and } i \neq j \\ 0, & \text{otherwise.} \end{cases} \qquad (5)$$

where τ is an overlap threshold (we set $\tau = 0.2$ in our experiments).

To infer the co-segmentation solution, we formulate Eq. (1) as an integer quadratic program (IQP). We combine the mutex constraints into the pairwise matrix to obtain the following objective function:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} \frac{1}{2}\mathbf{u}^T(\alpha\mathbf{A} + \gamma\mathbf{M})\mathbf{u} - \mathbf{b}^T\mathbf{u}, \qquad (6)$$
$$\text{s.t. } \forall i \in \mathcal{V} : u_i \in \{0,1\},$$

Similar to [1], we solve the IQP problem by employing the fixed-point iteration method.

Experiments show that our object-based RGBD co-segmentation with mutex constraints outperforms related techniques on an RGBD co-segmentation dataset. With a fully-connected graph structure and mutex constraints, our method is able to properly deal with image sets that contain outliers. Moreover, we also show that our method provides performance comparable to state-of-the-art RGB co-segmentation techniques on regular RGB images with depth maps estimated by using [5].

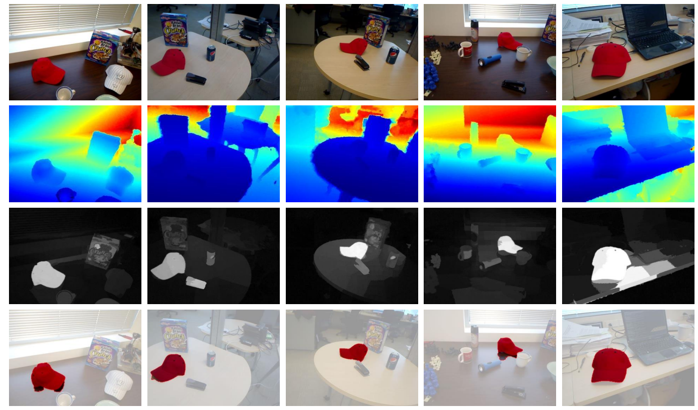

Figure 1: Given a set of RGB images (first row) with the corresponding depth maps (second row), our method employs a RGBD co-saliency map to detect common regions (third row), and then uses a co-segmentation graph to segment the common foregrounds (last row).

[1] W. Brendel and S. Todorovic. Segmentation as maximum-weight independent set. In *NIPS*, 2010.

[2] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng. Self-adaptively weighted co-saliency detection via rank constraint. *TIP*, 23(9):4175–4186, 2014.

[3] H. Fu, D. Xu, B. Zhang, and S. Lin. Object-based multiple foreground video co-segmentation. In *CVPR*, 2014.

[4] S. Gupta, R. Girshick, and P. Arbel. Learning Rich Features from RGB-D Images for Object Detection and Segmentation. In *ECCV*, 2014.

[5] K. Karsch, C. Liu, and S.B. Kang. Depth transfer: Depth extraction from video using non-parametric sampling. *TPAMI*, 36(11):2144–2158, 2014.

[6] S. Vicente, C. Rother, and V. Kolmogorov. Object cosegmentation. In *CVPR*, 2011.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.