Local Shape Transfer for Image Co-segmentation

Wei Teng[1]
tengw@buaa.edu.cn

Yu Zhang[1]
zhangyulb@gmail.com

Xiaowu Chen[†1]
chen@buaa.edu.cn

Jia Li[12]
jiali@buaa.edu.cn

Zhiqiang He[3]
lirong2@lenovo.com

[1] State Key Laboratory of Virtual Reality
Technology and Systems
Beihang University
Beijing, China

[2] International Research Institute for
Multidisciplinary Science
Beihang University
Beijing, China

[3] Lenovo Research

Figure 1: The motivation of this paper. The common objects in these images have different poses, rendering their global shapes inconsistent. However, the local object shapes in different images are highly consistent and provide important cues for co-segmentation.

Image co-segmentation is a challenging computer vision task that aims to segment all pixels of the common objects in an image set. In real-world cases, the common objects often vary greatly in poses, locations and scales, making their global shapes highly inconsistent across images and difficult to be segmented. However their local shapes are often highly consistent (see Fig. 1) and thus transferable. Based on the observation, we propose a novel co-segmentation approach, which transfers patch-level local object shapes and appears more consistently across different images. Given a group of M images,

our framework first estimates coarse initial foreground segmentations by thresholding saliency maps [3]. Meanwhile, we build inter-image connections by constructing a weighted graph on patches sampled from different images using [1], where weights are learned by Locally Linear Embedding [2]. With the patch graph, we refine the initial segmentation in each image by transferring the local shapes among different images. Formally, we minimize the objective

$$\min_{\mathbf{y}} \sum_{i=1}^{M} E_{\text{seg}}(\mathbf{y}^{[i]}) + \alpha \sum_{i=1}^{P} \left\| \vec{y}_i - \sum_{j \in N_i} w_{ij} \vec{y}_j \right\|^2,$$

$$\text{s.t. } \mathbf{y} \in \{0,1\}^{|\mathbf{y}|},$$

where $\mathbf{y}$ concatenates the binary labels of all pixels in the image set, $\mathbf{y}^{[i]}$ is the part from the ith image. The energy E_{seg} implements intra-image foreground/background segmentation, for which we use the popular Markov Random Field energy. The problem is NP-hard and usually large scale as it operates on pixels, which is approximately solved by half quadratic splitting.

We evaluate the proposed approach on two public benchmarks: iCoseg and Fashionista. Experiments show that our approach performs comparably with or better than the state-of-the-arts on iCoseg dataset, while achieving more than 31% relative improvements on Fashionista dataset.

[1] J. Kim, C. Liu, F. Sha, and K. Grauman. Deformable spatial pyramid matching for fast dense correspondences. In *CVPR*, pages 2307–2314, 2013.

[2] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[3] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, and B. Price. Minimum barrier salient object detection at 80 fps. In *ICCV*, pages 1404–1412, 2015.

† correspondence should be addressed to Xiaowu Chen.