CO-SALIENCY DETECTION VIA SEED PROPAGATION OVER THE INTEGRATED GRAPH WITH A CLUSTER LAYER

Insung Hwang**, Dong-ju Jeong***, Jae Sung Park**,†*, and Nam Ik Cho**

*Department of Electrical and Computer Engineering, INMC, Seoul National University
†Visual Display Division, SAMSUNG Electronics Co. Ltd.

ABSTRACT

This paper presents a method to detect common salient regions in a set of images. Since saliency and co-saliency detection are usually used as a pre-processing step for image processing and vision tasks, it is important to consider the complexity as well as the performance of an algorithm. Thus, we adhere to using low-level features and propose to detect co-salient objects with a new multilayer graph model in a bottom-up manner. Input images are represented by intra- and inter-image graphs composed of superpixels and their clusters, and each of these nodes obtains its initial co-saliency value from several cues. To generate resultant co-saliency values, foreground and background seeds are defined at parts of the unified multilayer graph, over which the seeds are propagated. Our experiments show that the proposed algorithm outperforms comparable methods on widely used public datasets, especially for the images that have various features.

Index Terms— Co-saliency, saliency, seed propagation model, foreground probability, coherence of saliency

1. INTRODUCTION

Recently, co-saliency detection has emerged as an important subtopic of saliency detection, which is to find visually distinct regions and/or objects that commonly appear in a set of images. In other words, the goal of co-saliency detection is to find “common salient” objects while suppressing salient objects/regions that appear only in part of the image group. Thus, it is needed to consider visual coherency among the images besides the cues used in the saliency detection such as contrast [1, 2, 3] and/or boundary priors [3, 4, 5]. The co-saliency detection can be applied to other computer vision tasks, such as co-segmentation [6], video foreground detection [7], image retrieval [8], and weakly supervised localization [9]. It can be utilized to enhance the single-image saliency detection as well [10].

According to [9], the majority of existing methods in the literature perform in a bottom-up manner [8, 11] while several works have presented the ones using high-level features with deep neural networks [6] and the learning-based ones that benefit from online semi-supervised learning [12]. Even though the latter ones have recently shown better performance, they have inherent weaknesses such as expensive computations and need for negative image groups. It is important to consider these aspects because the co-saliency detection is usually used as a pre-processing step for other applications that may not even have negative samples. In addition, fusion-based methods [13] also need several fast sub-algorithms. Thus, the efforts for devising undemanding approaches are crucial in the co-saliency detection area.

In this paper, a co-saliency detection method for multiple images is presented, which uses low-level features and a unified multilayer graph where multiple seeds are propagated. Each image in a group is over-segmented into several regions, which are defined as nodes on each intra-image graph. Then, a clustering step produces inter-image nodes to indirectly connect these intra-image nodes between different images. After the nodes on each image obtain their initial co-saliency values from contrast, center, and coherency priors, several foreground and background seeds are generated from those values and propagated over the unified multilayer graph learned by the semi-supervised learning scheme [14]. Since the seed propagation is performed on the unified graph considering the whole image group, the proposed method effectively detects co-salient objects with simple matrix computation.

2. GRAPH CONSTRUCTION

A set of images is represented by a graph, where over-segmented regions in each image are defined as intra-image nodes, and the images in the set are linked by defining the inter-image nodes. Specifically, a group of M images, $\{I^m\}_{m=1}^{M}$, are represented by intra-image nodes, and inter-image nodes are defined to connect the intra-image nodes of all the images to a unified graph.

To get the intra-image nodes and construct the intra-image graph of each image, it is independently over-segmented with the SLIC algorithm [15]. The produced regions in I^m correspond to its intra-image nodes, which are denoted as $\{v_i^m\}_{i=1}^{N_m}$, where v_i^m and N_m are the i-th intra-image node and the number of intra-image nodes in I^m, respectively. The edge e_{ij}^m between two neighboring nodes v_i^m and v_j^m that share a common boundary of segments connects them with a weight w_{ij}^m, which represents the affinity between them and is calculated using color similarity [16]. If we denote the

Lab color vector of v_i in an image as $\mathbf{x}_i$, the weight w_{ij} is computed as:

$$\begin{aligned} w_{ij} &= \exp\left(-(\mathbf{x}_i - \mathbf{x}_j)^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right) \\ \Sigma &= \frac{1}{N(E)} \sum_{e_{ij} \in E} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \end{aligned} \tag{1}$$

where E is the set of all the edges in the image. Then the affinity matrix for the intra-image graph of I^m is constructed whose (i, j)-th element is the weight between v_i^m and v_j^m:

$$(\mathbf{W}^m)_{i,j} = \begin{cases} w_{ij}^m, & \text{if } j \in Q_i^m, \\ 0, & \text{otherwise,} \end{cases} \quad m = 1, \ldots, M \tag{2}$$

where Q_i^m is an index set of neighbors of the i-th node.

Because $\mathbf{W}^m$ encodes node relations within only its intra-image graph, it is necessary to connect all the nodes of the images in the group for sharing co-saliency information. In [17], a graph matching method finds pairs of the most relevant segments between two images, each of which is connected with its matching score. Though ensuring good matched pairs for similar scenes such as sequential frames of a video that are not severely different from each other, in general, this approach easily fails to find good pairs between the images that have various backgrounds and/or different sizes of objects. Hence, an indirect approach is introduced in this paper to overcome this problem. We basically ignore the connectivity between images, which means that there are no edges that directly connect the intra-image nodes of any two different images. Thus the intra-image graphs are represented just in a form of a block-wise diagonal matrix:

$$\mathbf{W}_I = \begin{bmatrix} \mathbf{W}^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{W}^M \end{bmatrix}. \tag{3}$$

Instead, the proposed method introduces an additional "cluster layer" to consider the interactions between images and indirectly connect the intra-image nodes via the inter-image ones on it.

To define the inter-image nodes, we perform K-means clustering with the descriptor of every intra-image node, composed of its averaged *Lab* color vector, color histogram, and Leung-Malik (LM) filter bank response histogram [18]. In other words, texture features are added to $\mathbf{x}_i$ $(i \in \{1, ..., N\}$, $N = \sum_{m=1}^{M} N_m)$ to treat co-salient objects with different colors. As for the histogram descriptors, we additionally square root each element so that each histogram has unit $L2$-norm and the Euclidean distance (equivalent to the Hellinger distance) can be used for the clustering [19]. Through this procedure, K clusters $\{\mathbf{C}_i\}_{i=1}^K$ and their centroids $\{\mathbf{c}_i\}_{i=1}^K$ are generated, where $\mathbf{c}_i$ is the representative descriptor for $\mathbf{C}_i$ and also defined as an inter-image node. The goal of this step is to construct the affinity matrix of the unified graph including all the intra- and inter-image nodes, so we first connect each $\mathbf{c}_i$ to its elements and compute the weights of the edges using descriptor similarities as:

$$\begin{aligned} w_{ij}^{IC} &= \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{c}_j\|_2}{\sigma_c}\right) \\ (\mathbf{W}_{IC})_{i,j} &= \begin{cases} w_{ij}^{IC}, & \text{if } \mathbf{x}_i \in \mathbf{C}_j \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

where σ_c is a control parameter for the descriptor similarity. In addition, the inter-image nodes are also connected to each other, specifically to their k-nearest neighbors (k-NN), which means that the graph of the cluster layer is as sparse as the intra-image graphs, and its affinity matrix is written as:

$$\begin{aligned} w_{ij}^{C} &= \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2}{\sigma_c}\right) \\ \mathbf{W}_C &= \begin{cases} w_{ij}^{C}, & \text{if } i \in k\text{-NN}(j) \text{ or } j \in k\text{-NN}(i) \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

Finally, the affinity matrix of the unified graph is constructed from $\mathbf{W}_I$, $\mathbf{W}_{IC}$, and $\mathbf{W}_C$, expressed in a block-wise matrix form:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_I & \mathbf{W}_{IC} \\ \mathbf{W}_{IC}^T & \mathbf{W}_C \end{bmatrix}. \tag{6}$$

3. SEED EXTRACTION

As in most of bottom-up methods in the literature, both intra- and inter-image saliency (IrIS and IeIS) maps are needed to extract seeds and propagate them over the unified graph. We adopt our previous work [16] for single-image saliency detection without seed propagation to generate IrIS maps, which utilizes contrast and center priors. The former prior assumes that salient objects have distinctive features so that they show high contrast to other regions while the latter one is based on the observations that those objects are likely to be located in an image center. The resulting IrIS value of each v_i^m is denoted as s_i^m.

Because the IrIS does not consider the correlation among co-salient objects, the IeIS is also needed to measure both the saliency and coherency in the appearance of co-salient objects. Even though the definition of the IeIS is almost the same as that of co-saliency, it can be said that it focuses more on the coherency. To compute the IeIS, salient regions should be defined in advance. The IrIS maps are binarized to 1 or 0 for salient or non-salient regions respectively as:

$$\begin{aligned} b_i^m &= \begin{cases} 1, & s_i^m \geq \tau_m \\ 0, & s_i^m < \tau_m \end{cases} \\ \tau_m &= \max\left(\frac{1}{N_m} \sum_{i=1}^{N_m} s_i^m,\ 0.5\right) \end{aligned} \tag{7}$$

where τ_m is an adaptive threshold defined as the average IrIS value in I^m, and the salient regions of I^m belong to $B_m = \{i | b_i^m = 1\}$. Since a co-saliency map emphasizes common salient objects among an image group, the co-salient regions are expected to have high similarity to $\{B_m\}$. From this, the IeIS value of each intra-image node is defined as the sum of similarities to $\{B_m\}_{m=1}^{M}$ of all the images, which is written as:

$$\begin{gathered} t_i^m = \sum_{p\neq m}\sum_{q\in B_p} \exp\left(-\frac{\left\|\mathbf{x}_i^m - \mathbf{x}_q^p\right\|_2}{\sigma_t}\right) \\ i = 1, ..., N_m,\ m = 1, ..., M \end{gathered} \tag{8}$$

where σ_t is a control parameter for the IeIS penalizing the similarity more in case of smaller σ_t.

Even though an IeIS map itself can highlight co-salient objects in some degree, it is insufficient to suppress background regions. Therefore, initial co-saliency (IC) for seed extraction is obtained by combining the IrIS and IeIS with linear combination as:

$$u_i^m = \eta s_i^m + (1-\eta) t_i^m. \tag{9}$$

Top 10% of co-salient regions with respect to the IC values in each image are extracted as co-saliency seeds.

Meanwhile, the process for background seed extraction is to select boundary nodes of each image as the background seeds, based on the boundary prior. In addition, the ones selected as both the co-saliency and background seeds simultaneously are precluded from both seed sets because those seeds are not reliable. In summary, the co-saliency and background seeds are defined as:

- Co-saliency seeds ($\mathbf{y}_{I,s}$) : high IrIS and IeIS nodes that are not on any image boundaries.
- Background seeds ($\mathbf{y}_{I,b}$) : low IrIS and IeIS nodes on image boundaries.

4. SEED PROPAGATION

From the co-saliency and background seeds, co-saliency values are computed by propagating them to all the (intra-image) nodes in the image group. For this, a graph-based learning method is adopted for effective propagation [14], which makes a full pairwise graph as:

$$\mathbf{W}_L = (1-\alpha)\left(\mathbf{D} - \alpha\mathbf{W}\right)^{-1} = \left[\mathbf{w}_L^1, ..., \mathbf{w}_L^{N+K}\right], \tag{10}$$

where α is a learning balance parameter and $\mathbf{D} = \text{diag}(d_1, ..., d_N)$ is the degree matrix of $\mathbf{W}$ each of whose diagonal entries is the sum of the elements belonging to the same row ($d_i = \sum_j w_{ij}$). As mentioned in section 2, there are no direct inter-image connections between any two intra-image nodes in the graph with the affinity matrix $\mathbf{W}$, so instead the inter-image nodes indirectly connects the pairs of them. However, the learned graph with $\mathbf{W}_L$ has full pairwise relations of all the nodes. In other words, this graph has direct inter-image connections so that it ensures straightforward propagation between images.

For co-saliency detection, the overall affinities to the co-saliency and background seeds are computed respectively, which is written as:

$$\mathbf{f}_s = \mathbf{W}_L\,\mathbf{y}_s = \sum_{i\in S_s} \mathbf{w}_L^i,\ \mathbf{f}_b = \mathbf{W}_L\,\mathbf{y}_b = \sum_{i\in S_b} \mathbf{w}_L^i \tag{11}$$

where $\mathbf{y}_s = [\mathbf{y}_{I,s}; \mathbf{0}]$ and $\mathbf{y}_b = [\mathbf{y}_{I,b}; \mathbf{0}]$ are the co-saliency and background seed vectors respectively each of which is concatenated with a zero vector for the inter-image nodes, and S_s and S_b represent the co-saliency and background seed sets respectively. $\mathbf{f}_s$ and $\mathbf{f}_b$ are decomposed into the vectors for each image and the cluster layer, i.e., $\mathbf{f}_s = \left[\mathbf{f}_s^1; ...; \mathbf{f}_s^M; \mathbf{f}_s^C\right]$ and $\mathbf{f}_b = \left[\mathbf{f}_b^1; ...; \mathbf{f}_b^M; \mathbf{f}_b^C\right]$, and finally the co-saliency map for I^m is computed as:

$$\mathbf{z}^m = \left(\mathbf{f}_s^m - \mathbf{f}_b^m\right) ./ \left(\mathbf{f}_s^m + \mathbf{f}_b^m\right) \tag{12}$$

where ./ is the element-wise division of two vectors. The numerator represents the co-saliency while the denominator maintains the balance among the nodes, and lastly $\mathbf{z}^m$ is normalized to $[0, 1]$.

5. EXPERIMENTAL RESULTS

In our experiments, Image Pair (IP) [20], iCoseg [21], and MSRC [22] datasets are used to evaluate the performance of our algorithm and compare it with others. IP consists of 105 pairs of images, and iCoseg and MSRC are composed of 38 and 8 groups each of which includes 4-42 and 30 images, respectively. The MSRC dataset is used to evaluate the ability to treat co-salient objects that are not consistent in color, where the *grass* group is not used for the evaluation since it has no co-salient objects.

Several parameters are involved in the proposed algorithm: N_m, K, σ_c, k for the graph construction step, and σ_t, η, α for the seed extraction and propagation steps. We consistently set $N_m \simeq 250, K = 100, \sigma_c = 0.25, k = 5, \sigma_t = 0.4, \eta = 0.6, \alpha = 0.99$, and the precise value of N_m is determined by the SLIC algorithm. The relatively low value of σ_c tends to make the edge weights related to the inter-image nodes sparse, by which it is expected to learn a more desired fully pair-wise graph. Meanwhile, σ_t is set relatively high so that we have spatially scattered seeds not to miss co-salient regions.

Table 1. Quantitative comparison with the PR-AUC and ROC-AUC scores.

Meas.	Dataset	CP	CB	HS	MG	Ours
PR-AUC	IP	0.819	0.840	0.902	0.891	**0.914**
	iCoseg	—	0.804	0.839	0.854	**0.861**
	MSRC	—	0.689	0.785	0.690	**0.836**
ROC-AUC	IP	0.926	0.934	0.954	0.950	**0.965**
	iCoseg	—	0.942	0.955	0.957	**0.960**
	MSRC	—	0.798	0.882	0.827	**0.896**

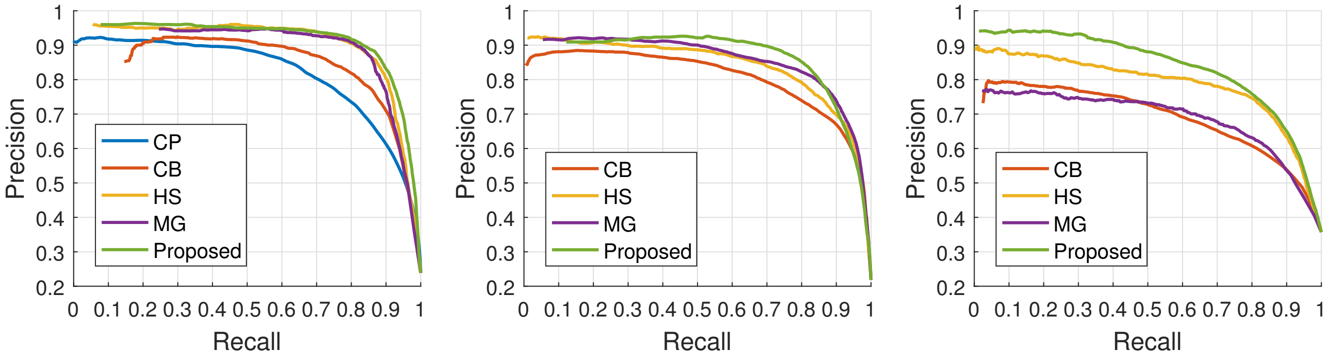

Fig. 1. Comparison of the PR curves on the three datasets (from left to right: the IP, iCoseg, and MSRC datasets). While both ours and several comparable methods have achieved promising performance on IP, the proposed method clearly outperforms the others on the datasets each group of which has a number of images, such as iCoseg and MSRC.

For each dataset, the performance is evaluated with the precision-recall (PR) curve, areas under the PR curve (PR-AUC) and receiver operating characteristic curve (ROC-AUC). Because the PR curve is better than the ROC curve in illustrating the difference between algorithms when there are more true negatives than true positives, we select the PR curve to show the results under each threshold in a form of graph. To this end, the co-saliency maps are normalized to $[0, 255]$ and binarized with a threshold varying from 0 to 255. The measures are calculated under each threshold and averaged over all samples as the standard used in the literature. With these measures, we compare our method with other major algorithms rather than fusion-based or high-level feature-based ones: CP [20] (only for image pairs), CB [8], HS [11], and MG [23].

As can be seen in Fig. 1 showing the PR curves from the three datasets, the proposed method performs better than others in most ranges of the curves. In addition to Fig. 1, Table 1 also shows the effectiveness of our algorithm, especially for the MSRC dataset, which is because we consider the texture features as important as the color ones in spite of the side effects in the cases of IP and iCoseg. Lastly, our co-saliency detection examples are shown in Fig. 2. It is notable that our method is effective in suppressing the objects that are not co-salient but salient in each image, and detects the co-salient regions more uniformly due to the saliency propagation step. The salient objects adjacent to image boundaries can also be detected well since even the regions that are used as the co-saliency/background seeds receive newly propagated saliency values from the seeds.

6. CONCLUSIONS

We have proposed a co-saliency detection method, which finds the regions with high initial co-saliency values from the contrast, center, and coherency priors, and then propagates the seeds from those regions and image boundaries to obtain the final co-saliency values. For this, each image in a group is represented by a graph where each of intra-image nodes is given its intra-image saliency value. Then we obtain the inter-image saliency values from the coherency prior and generate the combined initial co-saliency values. The inter-image nodes indirectly connect the nodes between different images, which makes it possible to propagate the seeds over the unified graph. Experimental results show that the proposed algorithm outperforms the comparable methods on the widely used co-saliency detection datasets with diverse classes.

Acknowledgements

This research was supported in part by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2016-0-00288) supervised by the IITP (Institute for Information & communications Technology Promotion), and in part by the Brain Korea 21 Plus Project in 2017.

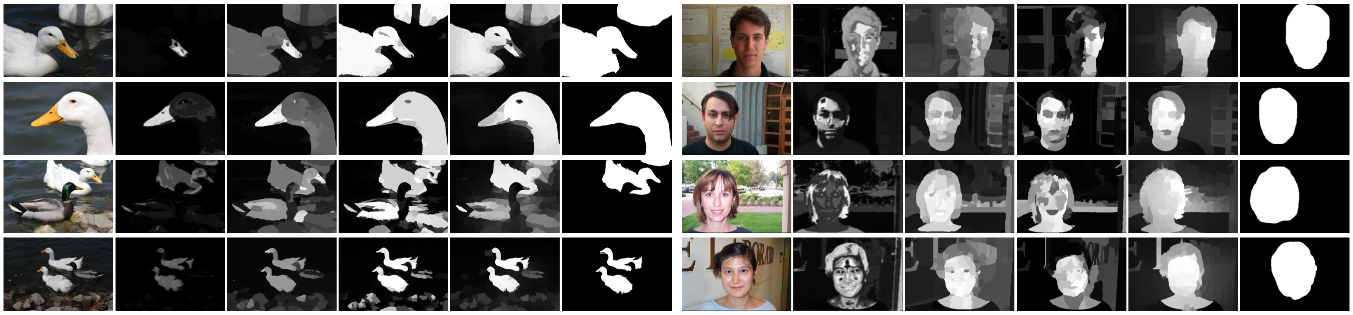

Fig. 2. Visual comparison on the *Goose* set in iCoseg and the *Face* set in MSRC (from left to right: input images, CB, HS, MG, the proposed method, and ground truth images).

7. REFERENCES

[1] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 409–416.

[2] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[3] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 2814–2821.

[4] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. European Conf. Computer Vision*, 2012, pp. 29–42.

[5] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 3166–3173.

[6] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vision*, vol. 120, no. 2, pp. 215–232, Nov. 2016.

[7] H. Fu, D. Xu, B. Zhang, S. Lin, and R. K. Ward, "Object-based multiple foreground video co-segmentation via multi-state selection graph," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3415–3424, Jun. 2015.

[8] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3766–3778, Oct. 2013.

[9] D. Zhang, H. Fu, J. Han, and F. Wu, "A review of co-saliency detection technique: Fundamentals, applications, and challenges," *ArXiv preprint arxiv:1604.07090*, Apr. 2016.

[10] M.-M. Cheng, N. J. Mitra, X. Huang, and S.-M. Hu, "Salientshape: Group saliency in image collections," *Vis. Comput.*, vol. 30, no. 4, pp. 443–453, Apr. 2014.

[11] Z. Liu, W. Zou, L. Li, L. Shen, and O. Le Meur, "Co-saliency detection based on hierarchical segmentation," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 88–92, Jan. 2014.

[12] D. Zhang, D. Meng, C. Li, L. Jiang, Q. Zhao, and J. Han, "A self-paced multiple-instance learning framework for co-saliency detection," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 594–602.

[13] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-adaptively weighted co-saliency detection via rank constraint," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4175–4186, Sep. 2014.

[14] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Advances in Neural Information Processing Systems*, 2004, pp. 321–328.

[15] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[16] I. Hwang, S. H. Lee, J. S. Park, and N. I. Cho, "Saliency detection based on seed propagation in a multilayer graph," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 2111–2129, 2017.

[17] Z. Tan, L. Wan, W. Feng, and C.-M. Pun, "Image co-saliency detection by propagating superpixel affinities," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2013, pp. 2114–2118.

[18] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision*, vol. 43, no. 1, pp. 29–44, Jun. 2001.

[19] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012, pp. 2911–2918.

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

[21] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010, pp. 3169–3176.

[22] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proc. IEEE Int. Conf. Computer Vision*, 2005, pp. 1800–1807.

[23] Y. Li, K. Fu, Z. Liu, and J. Yang, "Efficient saliency-model-guided visual co-saliency detection," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 588–592, May 2015.