Efficient Saliency-Model-Guided Visual Co-Saliency Detection

Yijun Li, Keren Fu, Zhi Liu, *Member, IEEE*, and Jie Yang

***Abstract*—This letter proposes a novel framework to detect common salient objects in a group of images automatically and efficiently. Different from most existing co-saliency models which directly redesign algorithms for multiple images, the saliency model for a single image is fully exploited under the proposed framework to guide the co-saliency detection. Given single image saliency maps, a two-stage guided detection pipeline led by queries is proposed to obtain the guided saliency maps of the image set through a ranking scheme. Then the guided saliency maps generated by different queries are fused in a way that takes advantages of both averaging and multiplication. The proposed model makes existing saliency models work well in co-saliency scenarios. Experimental results on two benchmark databases demonstrate that the proposed framework outperforms the state-of-the-art models in terms of both accuracy and efficiency.**

***Index Terms*—Co-saliency detection, efficient manifold ranking, fusion, saliency model.**

I. Introduction

THE era of Big Data and ubiquity of Internet use result in an explosion of information medium and challenge us to deal with a great amount of subjects instead of a single one when facing a specific task. Co-saliency detection is exactly such a task that it aims at highlighting the common salient objects in a group of images to simulate human visual attention mechanism. It derives from saliency detection which targets on popping out the salient object in a single image and has become a booming research issue in recent years. When the attention is shifted from the single to multiple subjects, a salient object within one image could be no longer salient anymore among the image corpus. Detection results of a co-saliency model are a set of co-saliency maps, which are useful in the subsequent processing such as object co-segmentation [1] and co-recognition [2].

The co-saliency detection begins with some proposed models for a pair of images. In an early work [3] where two images are captured under highly similar background, the local structure changes induced by salient objects are exploited to train a co-saliency model. However, as different background could be a more general case, the following models no longer require any constraints on background. In [4], the joint information provided by an image pair is used under a pre-attentive scheme. In [5], co-saliency maps are generated via the combination of single saliency maps and an inter-image saliency map that is based on a co-multilayer graph.

When it comes to a collection of images, several co-saliency models that abide by a fundamental diagram have emerged. They are basically formulated as follows:

$$Co-saliency = Saliency \times Repetitiveness \quad (1)$$

where $Saliency$ is either generated by a saliency model [1] or redesigned as the intra-image saliency [6]–[8] for each image, and $Repetitiveness$ (correspondence) describes how frequently the object recurs across the image collection [1]. However, such a pipeline has several limitations:

(i) If ever the common salient parts fail to be detected in $Saliency$ by a saliency model, the multiplication operation cannot re-pop them out even they have high $Repetitiveness$ values.

(ii) The redesign of $Saliency$ algorithm often has a high computational cost because comparisons between pixels/regions are no longer within one image, but across the image collection, which restricts the practical use of the model.

(iii) The case with irrelevant images which do not contain common salient objects is rarely considered.

One most related work [9] generates the co-saliency map by fusing saliency maps generated by different saliency models based on rank-one constraint. However, it is complex and time-consuming to run multiple saliency models. Therefore we propose an efficient co-saliency model by fully exploiting one saliency model to solve aforementioned problems. The main contributions of this letter lie in the following two aspects:

(i) A saliency-model-guided scheme is proposed to make existing saliency models work well in co-saliency scenarios. Not only non-common parts are suppressed well, but also the common salient object which originally fails to be detected is successfully recovered.

(ii) With the model accuracy guaranteed, the model efficiency is significantly enhanced as well.

The remainder of this letter is organized as follows. Section II describes the proposed model in details. Section III conducts experiments to compare the proposed model with existing counterparts and results are also analyzed. The conclusion is given in Section IV.

Manuscript received August 04, 2014; revised October 09, 2014; accepted October 20, 2014. Date of publication October 23, 2014; date of current version November 04, 2014. This work was supported by the National Natural Science Foundation of China under Grants 61273258 and 61171144, and by the 973 Plan of China under Grant 2015CB856004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rui Liao.

Y. Li, K. Fu, and J. Yang are with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: leexiaoju@sjtu.edu.cn; fkrsuper@sjtu.edu.cn; jieyang@sjtu.edu.cn).

Z. Liu is with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China (e-mail: liuzhisjtu@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2364896

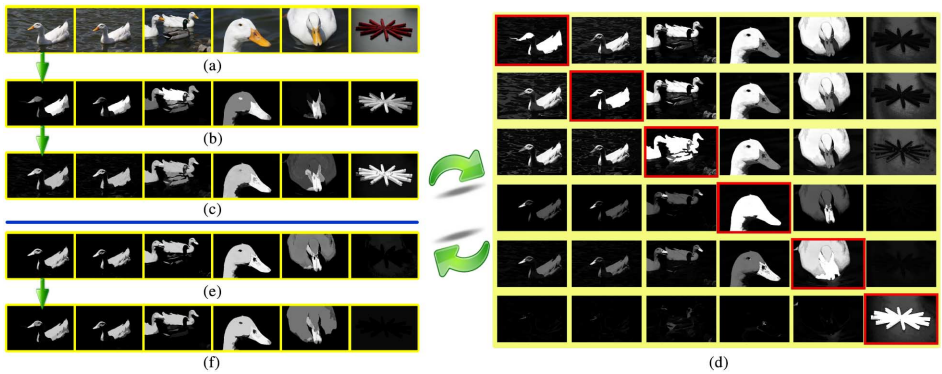

Fig. 1. The framework of our co-saliency model. The red box in (d) represents the query image and each row presents corresponding ranking results. (a) Input image set (b) Single image saliency map (c) First stage guided detection results (d) Second stage guided detection results (red: query) (e) Fusion results (f) Final co-saliency maps.

II. Methodology

The framework of our model is shown in Fig 1. It consists of three main steps, i.e. single image saliency map, guided co-saliency detection and fusion. Details are given in the following subsections.

A. Single Image Saliency Map

Instead of specifically redesign intra-image or inter-image saliency detection algorithm arduously, our approach directly takes advantage of saliency maps generated by state-of-the-art saliency models. Given a collection of images $\{Im_i\}_{i=1}^{M}$, our previous validated work [10] is selected to generate the corresponding saliency maps $\{SM_i\}_{i=1}^{M}$. However, other saliency models are also feasible(see the results in Section III-A). In [10], the boundary prior is mainly utilized to regard four borders of the image as the potential background. With four virtual nodes set to connect four borders respectively under a graph structure, the geodesic saliency measure is used to obtain four saliency maps. Finally four maps are combined to render the single image saliency map (Fig. 1(b)).

B. Guided Co-Saliency Detection

Single image saliency maps could be spotty but they provide us with sufficient information of the common salient object for the subsequent co-saliency detection. In this section we propose a two-stage query-based detection scheme under a ranking framework. The co-saliency value is measured based on their relevances to the given query. Naturally, the ranking scheme and query selection are two important issues to consider in this step.

For the ranking scheme, we adopt the efficient manifold ranking (EMR) [11] which has been proved accurate and efficient in image retrieval. The EMR learns a ranking function by exploiting the intrinsic manifold structure of data and an anchor graph to accelerate the computation. For our co-saliency model, data points $\chi = \{x_1, x_2, \ldots, x_n\}$ involved in ranking are LAB color vectors of all pixels in the image collection. We cluster all data points to d classes using k-means and select clustering centers as anchors $U = \{u_1, u_2, \ldots, u_d\}$. Each data point x_i is connected to its $s(s < d)$ nearest anchors and the connection is assigned with the weight z_{ki} defined in Eq. (2) to obtain a weight matrix $Z_{d\times n}$:

$$z_{ki} = \frac{K(\frac{|x_i - u_k|}{\lambda_i})}{\sum_{l=1}^{d} K(\frac{|x_i - u_l|}{\lambda_i})}, K(t) = \begin{cases} \frac{3}{4}(1-t^2), & if|t| \le 1 \\ 0, & otherwise \end{cases} \tag{2}$$

where $K(\cdot)$ is the *Epanechnikov* quadratic kernel and $\lambda_i = |x_i - u_{[s]}|$ where $u_{[s]}$ is the sth closest anchor of x_i. The final adjacency matrix $W = Z^T Z$. Based on the graph, the optimal ranking function r^* can be written as:

$$r^* = \left(I_n - H^T\left(HH^T - \frac{1}{\alpha_1} I_d\right)^{-1} H\right) y \tag{3}$$

where y is a binarized vector with '1' for queries and '0' for the unlabelled. $H = ZD^{-\frac{1}{2}}$ where D is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} w_{ij}$. I_n and I_d are the identity matrix with the dimension of n and d. α_1 is a weight parameter, and set as 0.5 in our experiment. Please refer to [11] for more details. The ranking scores in the vector r^* are then directly assigned to all pixels as the their new saliency values.

Then two stages of the query selection are conducted to form our two-stage detection scheme. In the first stage, we reshape SM_i into a vector y_i and binarize it as $\hat{y}_i$ by a threshold $th_i = mean(y_i)$. The query vector y is defined as:

$$y = [\hat{y}_1; \hat{y}_2; \ldots; \hat{y}_M], \tag{4}$$

and thus queries of the image collection are utilized together to obtain new saliency maps $\{SM_i^1\}_{i=1}^{M}$ with Eq. (3). This step is critical since it recovers co-salient parts missed in the single image saliency map. To put all queries with confidence of '1' in a holistic way means that we desire to maximally pop out object parts that are similar to queries. As shown in the 5th column of Fig. 1(c), the body of the white goose is successfully recovered from that of Fig. 1(b) because white parts are detected in top four images and hence Eq. (3) guides similar white parts in the 5th image to rank high. Unfortunately, as queries are not all precise, results are still not satisfying since non-common parts are also likely to be recovered and falsely highlighted background regions need to be suppressed.

In the second stage, we reshape SM_i^1 into a vector y_i^1. Different from the first stage, queries in each image take turns to be the query. Specifically, in each round, y_i^1 is regarded as the query

vector and other $y_j^1 (j \neq i)$ are set to zero as unlabelled to form y. Therefore for each image Im_i, M saliency maps $\{SM_{ji}^2\}_{j=1}^M$ are generated. Fig. 1(d) shows an exemplary result of the second stage guided ranking scheme where the query is bounded in the red box in each round.

The key insight of the proposed two-stage guided process illustrates the essence of $Repetitiveness$. The first stage is exploited to promote similar pixels and prevailing saliency values through the image collection to effectively recover and suppress some regions. The effect of the second stage is more obvious. As the common salient object exists in most images, ever since it is successfully detected in some image Im_i, similar parts in other images will be popped out with relatively high ranking score when SM_i^1 is set as the query. As for non-common parts, they are expected to get low saliency values when the query image doesn't contain these non-common parts. The most representative one is the 5th row of Fig. 1(d) where the 5th map serves as the query. The query with high probability of being salient object is the body and the beak. Therefore similar parts in other five images are expected to be detected and dissimilar parts like the dark goose and red pens in the 3rd and 6th image are suppressed. Next a fusion scheme is required to reinforce the saliency value of common salient parts and alleviate that of non-common parts.

C. Fusion

Given $\{SM_{ji}^2\}_{j=1}^M$ for Im_i, we propose a fusion strategy to obtain the initially guided co-saliency map SM_i^*. Popular ways of combination of multiple saliency maps are averaging [12] and multiplication [13]. We take advantages of both to better highlight the co-salient object and suppress the background. In each SM_{ji}^2, we set a threshold by employing $OTSU$ algorithm [14] as a criterion to distinguish the co-salient object and the background. A recorder $Count$ is defined to count how many times a pixel is classified as being co-salient among M maps. Finally the fusion strategy is defined as follows:

$$SM_i^*(k) = \begin{cases} \frac{1}{M} \sum_{j=1}^{M} SM_{ji}^2(k), & if\ Count(k) > \frac{M}{2} \\ \prod_{j=1}^{M} SM_{ji}^2(k), & otherwise \end{cases} \quad (5)$$

where k is the pixel index in Im_i and M is the number of images in the collection.

Since all pixel values in maps are normalized to the range of [0,1], the multiplication of values will be much less than any of them. It works well as what we expect for background saliency, i.e., the smaller, the better. However in such a way of simple multiplication, co-salient parts will also be affected, namely being over-suppressed. To avoid over-suppression and maximally stretch out the difference on co-saliency values between co-salient object and background, we additionally exploit the average strategy. For a pixel k in Im_i, if it appears to be salient in most of maps in $\{SM_{ji}^2\}_{j=1}^M$, where "most" means more than a half of images here, it will be regarded as being co-salient and its co-saliency is obtained by averaging the corresponding M values. Fig. 1(e) verifies the effectiveness of our fusion strategy. For the common salient white geese, though they are not detected in the 6th row of Fig. 1(d), the averaging operation keeps their high saliency values. For non-common parts like the dark goose and red pens, though they gain high values when regarded as the query, other five detection results guided by other queries determine that the multiplication should be conducted to suppress them.

Fig. 2. Examples of co-saliency detection on CP database.

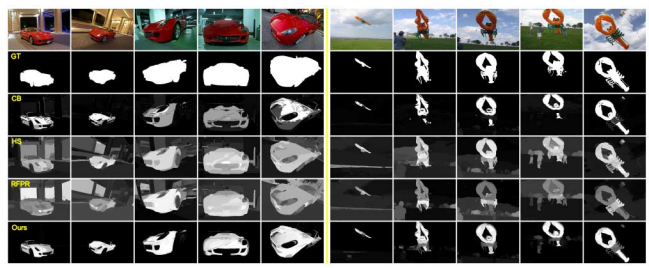

Fig. 3. Examples of co-saliency detection on iCoseg database.

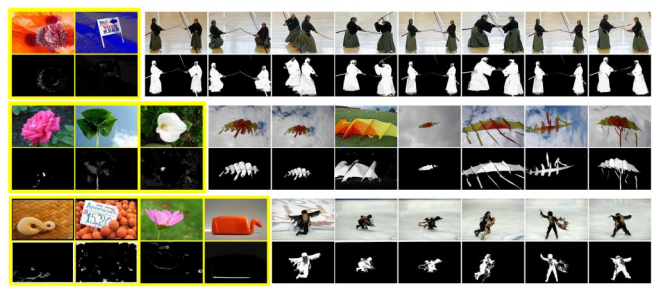

Fig. 4. Examples of irrelevant images involved (yellow: irrelevant images).

To eliminate unnecessary artifacts like some isolated salient pixels or small holes within a co-salient region, a simple smooth post-processing is performed to obtain final co-saliency maps. Each image is segmented by mean shift segmentation [15], and the saliency of each region is measured as the average saliency values of pixels within itself.

III. Experiments and Comparisons

The proposed co-saliency model has been tested on two benchmark databases, i.e., the CP database [5] which contains 105 image pairs, and a subset of the iCoseg database [16] which contains 185 images from 37 classes, each of which has 5 images. All parameters are set to default values in their original works [10][11] except for the number of anchors

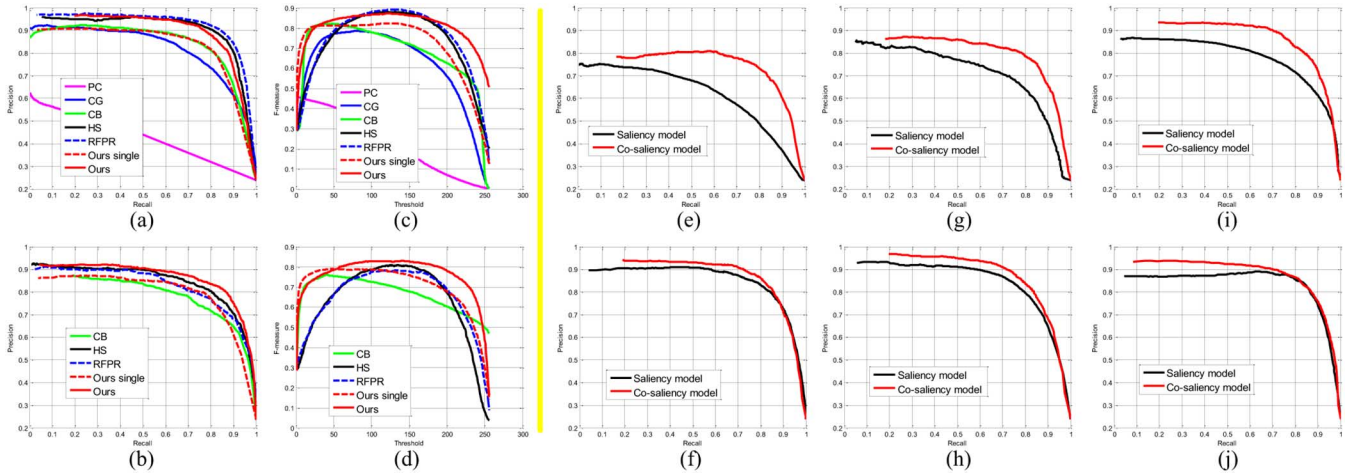

Fig. 5. Precision-Recall curves comparisons: (a)–(d) Comparisons between different co-saliency models on different databases; (e)–(j) Comparisons between the proposed co-saliency model guided by different saliency models on CP database. (a) PR curves on CP database (b) PR curves on iCoseg database (c) F-measure curves on CP database (d) F-measure curves on iCoseg database (e) FT (f) MR (g) RC (h) GC (i) PCA (j) ST.

$d = min\{10M, 50\}$ where M is the number of images, the number of nearest anchors to connect $s = 4$. It is justified in [11] that the performance is not sensitive to both d and s. For the mean shift segmentation in the last step, we set its parameter of allowable minimum region area to 0.2% of the image area.

A. Performance Comparisons

We evaluate our model against several mainstream co-saliency detection models, including PC [4], CG [5], CB [6], HS [7] and RFPR [8]. Visually, Fig. 2 and Fig. 3 demonstrate the validity and efficacy of our model both for the image pair and image collection. It can be seen that our method gains higher quality saliency maps compared with others. Then we extend the original 5-image set in the iCoseg database [16] to a 10-image set with several relevant and irrelevant images involved. Fig. 4 presents exemplary detection results. The co-saliency maps of irrelevant images are almost black although they contain salient objects, indicating a wider applicable scope of our model.

Similar as [10], we also implement quantitative evaluation by employing two criteria: the precision recall (PR) curve and the F-measure curve. $F - measure = \frac{(1+\beta^2)\times Precision \times Recall}{\beta^2 \times Precision + Recall}$ where β^2 is set as 0.3 to emphasize the precision [13][17]. Intuitively, these curves represents the robustness of the algorithm. Fig. 5(a)–(d) present the PR and F-measure curves of all models and it can be seen that our model outperforms most existing co-saliency models. In Fig. 5(a), the PR curve of our model is a little faded against HS [7] and RFPR [8] due to the image pair. The information for guiding is limited as there are only two images. As the number of images increases in each image class of the iCoseg database, Fig. 5(b) shows that our model surpasses HS and RFPR. Besides, the F-measure curves shown in Fig. 5(c)–(d) tell that our model is stable in a wider range of threshold than any other model.

Then we replace the saliency model in the first step with several other saliency models including FT [18], RC [19], PCA [20], MR [13], GC [17] and ST [21]. The guided effect of each input saliency model under our framework is shown in Fig. 5(e)–(j). The remarkable improvement of curves illustrates the following two advantages of our model:

TABLE I
AVERAGE RUNNING TIME (SECONDS PER IMAGE)

Methods	PC[4]	CG[5]	CB[6]	HS[7]	RFPR[8]	Ours
CP	7.16	596.15	1.85	23.20	25.30	1.34
iCoseg	N/A	N/A	10.55	144.04	162.55	8.65

(i) Our co-saliency model does improve the detection results of the saliency model without any attenuated performance. As shown in Fig. 5(b), even our previous saliency model [10] (Ours single) outperforms the co-saliency model CB [6]. It clearly implies that starting from existing proven saliency models is a feasible and better way for co-saliency detection.

(ii) The higher precision the saliency model achieves, the better performance of the co-saliency model is. That's why we select [10] in our first step.

B. Running Time

In Table I we compare the average running time of our co-saliency model with others. Experiments are taken on a laptop with an Intel 1.8 GHz CPU and 8 GB RAM. PC [4] and CG [5] are only designed for image pairs. From Table I, it is clearly observed that the computational cost is significantly reduced due to the simple matrix multiplication of EMR [11]. One more critical factor that determines the efficiency is the anchor generation step. To find clustering centers from large amount of pixel data, the computational cost is reduced by setting the maximal iteration time of K-means to only 5 in our implementation.

IV. CONCLUSION

In this letter, we propose a co-saliency model that makes existing saliency models work well in co-saliency scenarios. Given single image saliency maps as initial queries, a two-stage scheme is proposed to obtain the guided saliency maps of the image set through a ranking framework. Then the guided saliency maps are fused to highlight common salient objects and to suppress non-common parts. Experimental results demonstrate the better performance of our proposed model in terms of both accuracy and efficiency.

REFERENCES

[1] 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 *Proc. IEEE CVPR*, Jun. 2011, pp. 2129–2136.

[2] S. M. Hu, T. Chen, K. Xu, M. M. Cheng, and R. Martin, "Internet visual media processing: A survey with graphics and vision applications," *Vis. Comput.*, vol. 29, no. 5, pp. 393–405, May 2013.

[3] D. Jacobs, D. Goldman, and E. Shechtman, "Cosaliency: Where people look when comparing images," in *Proc. ACM UIST*, Oct. 2010, pp. 219–228.

[4] H. T. Chen, "Preattentive co-saliency detection," in *Proc. IEEE ICIP*, Sep. 2010, pp. 1117–1120.

[5] 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.

[6] 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.

[7] 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.

[8] L. Li, Z. Liu, W. Zou, X. Zhang, and O. Le Meur, "Co-saliency detection based on region-level fusion and pixel-level refinement," in *Proc. IEEE ICME*, Jul. 2014, pp. 1–6.

[9] X. Cao, Z. Tao, B. Zhang, H. Fu, and X. Li, "Saliency map fusion based on rank-one constraint," in *Proc. IEEE ICME*, Jul. 2013, pp. 1–6.

[10] Y. Li, K. Fu, L. Zhou, Y. Qiao, J. Yang, and B. Li, "Saliency detection based on extended boundary prior with foci of attention," in *Proc. IEEE ICASSP*, May 2014, pp. 2798–2802.

[11] B. Xu, J. Bu, C. Chen, D. Cai, X. He, W. Liu, and J. Luo, "Efficient manifold ranking for image retrieval," in *Proc. ACM SIGIR*, Jul. 2011, pp. 525–534.

[12] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. ECCV*, Oct. 2012, pp. 414–429.

[13] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE CVPR*, Jun. 2013, pp. 3166–3173.

[14] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[15] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[16] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "Icoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proc. IEEE CVPR*, Jun. 2010, pp. 3169–3176.

[17] M. M. Cheng, J. Warrell, W. Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE ICCV*, Dec. 2013, pp. 1529–1536.

[18] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE CVPR*, Jun. 2009, pp. 1597–1604.

[19] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *Proc. IEEE CVPR*, Jun. 2011, pp. 409–416.

[20] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?," in *Proc. IEEE CVPR*, Jun. 2013, pp. 1139–1146.

[21] Z. Liu, W. Zou, and O. Le Meur, "Saliency tree: A novel saliency detection framework," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1937–1952, May 2014.