Co-Saliency Detection via Co-Salient Object Discovery and Recovery

Linwei Ye, Zhi Liu, *Senior Member, IEEE*, Junhao Li, Wan-Lei Zhao, and Liquan Shen

***Abstract*—This letter proposes a novel co-saliency model to effectively discover and highlight co-salient objects in a set of images. Based on the gross similarity which combines color features and SIFT descriptors, some co-salient object regions are first discovered in each image as exemplars, which are exploited to generate the exemplar saliency maps with the use of single-image saliency model. Then both local recovery and global recovery of co-salient object regions are performed by propagating the exemplar saliency to the matched regions, and border connectivity is further exploited to generate the region-level co-saliency maps. Finally, the foci of attention area based pixel-level saliency derivation is used to generate the pixel-level co-saliency maps with even better quality. Experimental results on two benchmark datasets demonstrate that the proposed co-saliency model outperforms the state-of-the-art co-saliency models.**

***Index Terms*—Co-saliency detection, exemplar, object discovery, object recovery, saliency model.**

I. Introduction

VISUAL attention mechanism enables human observers to capture the most attractive objects in a scene automatically, and motivated the research on computational models of saliency detection, which has been a booming research topic in the last decades. Saliency models benefit a number of saliency-based applications including salient object segmentation [1], content-based image retrieval [2], content-aware image compression [3], etc. However, most saliency models [4]–[8] are applicable to detect salient objects in a single image.

In recent years, co-saliency detection becomes an interesting issue which aims to detect co-salient objects in a set of images. Co-salient objects not only show distinctiveness in each individual image but also repetitively occur across multiple images. Co-saliency is proposed to detect the common objects with local structure changes in a pair of images with the constraint of similar background in [9]. In [10], a linear combination model integrates saliency maps generated by three single-image saliency models and a multi-image saliency map based on a co-multi-layer graph to detect co-salient objects in image pairs. For an image set with more than two images, individual saliency map is first generated for each image, and then the salient regions frequently occurring in most images are detected as co-salient in [11]. In [12], contrast cue, spatial cue and corresponding cue are measured at the cluster level to generate co-saliency maps for the image set. Multiple saliency maps generated using different single-image saliency models are combined based on the rank-one constraint to obtain co-saliency maps in [13], and saliency maps generated by one single-image saliency model are fully exploited under a two-stage query scheme to guide co-saliency detection in [14]. In [15], an unsupervised random forest is used to extract the rough contours of common objects, and then single saliency map and inter-saliency map are combined into co-saliency map.

Segmentation has been proven as a useful preprocessing to provide meaningful regions for effective saliency detection. In [16], multiscale segmentation based intra-image saliency is integrated with region match based inter-image saliency for co-saliency detection. In [17], region similarity and region contrast are measured on the fine segmentation and object prior is measured on the coarse segmentation, and these measures are integrated with global similarity among regions to obtain co-saliency maps. In [18], both hierarchical segmentation based region-level fusion and pixel-level refinement are combined to improve the co-saliency detection performance.

As a generic object detection method, objectness [19] has shown its effectiveness to complement saliency detection in a single image. For example, a graphical model is constructed to optimize an energy function by combining saliency, objectness and interaction terms in [20], which improves the quality of original saliency maps. In order to elevate saliency detection performance, objectness measures of windows are integrated with uniqueness and focusness for saliency estimation in [21], and are exploited to enforce the consistency among salient regions in [22].

Different from the existing co-saliency models, this letter proposes a novel co-saliency model illustrated in Fig. 1, which first discovers co-salient exemplars for generating exemplar saliency maps, then exploits co-salient object recovery and border connectivity to generate region-level co-saliency maps, and finally utilizes the foci of attention area to generate pixel-level co-saliency maps. Our main contribution lies in the following three aspects. First, inspired by the effectiveness

Manuscript received June 11, 2015; accepted July 16, 2015. Date of publication July 20, 2015; date of current version July 23, 2015. This work was supported by National Natural Science Foundation of China under Grants 61171144 and 61471230, and by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antoni Chan.

L. Ye, Z. Liu, J. Li, and L. Shen are with the School of Communication and Information Engineering, Shanghai University, Shanghai, China (e-mail: yelinweimail@163.com; liuzhisjtu@163.com; jxlijunhao@163.com; jsslq@163.com).

W.-L. Zhao is with the School of Information Science and Technology, Xiamen University, Xiamen, China (e-mail: wlzhao@xmu.edu.cn).

(*Corresponding author: Z. Liu.*)

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.2015.2458434

Fig. 1. Illustration of the proposed co-saliency model. (a) Input images; (b) co-salient exemplar discovery; (c) exemplar saliency maps; (d) region-level co-saliency maps; (e) pixel-level co-saliency maps.

of integrating objectness with single-image saliency models, we introduce the objectness into co-saliency detection, and propose a novel framework of co-salient object discovery and recovery. Second, different from pervious methods which directly use objectness measures of all generated bounding boxes, our co-saliency model measures the gross similarity of bounding boxes in the image set and selects some exemplar bounding boxes to generate the exemplar saliency map. Last, our co-saliency model is also robust to different object proposal methods, which can be used to replace the objectness method, and consistently outperforms the state-of-the-art co-saliency models. The rest of this letter is organized as follows. Section II details the four stages of the proposed co-saliency model shown in Fig. 1. Experimental results and analysis are presented in Section III, and conclusion is given in Section IV.

II. Proposed Co-Saliency Model

A. Co-Salient Exemplar Discovery

Given a set of M images $\{I_m\}_{m=1}^{M}$, color quantization is performed on each image in the *Lab* color space, which is close to human perception with one luminance channel and two chrominance channels. Each color channel is uniformly quantized into 16 bins, and thus the total number of quantized colors is reduced to $K_C = 16^3$ for calculating color histograms. Besides, the SIFT descriptor, which has been widely used in image retrieval and image recognition for its robust invariance to scale, rotation and illumination, is densely extracted on the basis of 16×16 patches over the image set in the *RGB* color space according to [23]. All the extracted SIFT descriptors are clustered to K_S codewords where K_S is adaptively set to $3M$. Then the objectness method [19] is used to generate bounding boxes of object candidates with non-maximum suppression for each image I_m. Let $B_{m,i}$ denote the ith bounding box in I_m, and the two histograms, $H_{m,i}^{C}$ with K_C bins and $H_{m,i}^{S}$ with K_S bins, are calculated for $B_{m,i}$ based on the quantized colors and the clustered SIFT descriptors, respectively. Then the similarity between each pair of bounding boxes, $B_{m,i}$ and $B_{n,j}$, is defined as follows:

$$Sim(B_{m,i}, B_{n,j}) = 1 - \frac{1}{2}\left\{\chi^2\left[H_{m,i}^{C}, H_{n,j}^{C}\right] + \chi^2\left[H_{m,i}^{S}, H_{n,j}^{S}\right]\right\}, \quad (1)$$

where $\chi^2[\cdot]$ is the chi-square distance between histograms. The gross similarity of each bounding box $B_{m,i}$ over the image set is then defined as follows:

$$\lambda(B_{m,i}) = \sum\nolimits_{n=1, n\neq m}^{M} \max_{j} \left[Sim(B_{m,i}, B_{n,j})\right]. \quad (2)$$

According to the repetitiveness property of co-salient objects, the representative co-salient objects, which are the most similar over the image set, are contained in some exemplar bounding boxes with a higher gross similarity. Specifically, the top T_B bounding boxes with the highest gross similarity in each image I_m are selected as the exemplar bounding boxes $B_{m,i}^{E}(i = 1, \ldots, T_B)$, and constitute the bounding box set B_m^E. For the input images in Fig. 1(a), some red football players, which are the co-salient objects in the image set, are successfully discovered as exemplar co-salient objects in each image and marked with cyan bounding boxes in Fig. 1(b).

B. Exemplar Saliency Generation

Due to the better saliency detection performance of single-image saliency models which are built on the basis of segmented regions, the *gPb-owt-ucm* segmentation method [24] is used to generate the ultrametric contour map, which is then thresholded to generate a sparse segmentation result G_m^S and a dense segmentation result G_m^D, with less than 20 regions and 100 regions, respectively. Let $R_{m,k}$ denote the kth region in G_m^S, $R_{m,k}$ is defined as an exemplar region of co-salient objects if it has an adequate overlap with any exemplar bounding box $B_{m,i}^{E}$. Specifically, $R_{m,k}$ is added into the exemplar region set Φ_m^E if the area ratio of $R_{m,k} \cap B_{m,i}^{E}$ to $B_{m,i}^{E}$ is greater than the area ratio of $B_{m,i}^{E}$ to I_m.

For saliency measurement in each image I_m, the saliency tree model [4] is adopted to generate the initial pixel-level saliency map S_m^I. Then both Φ_m^E and B_m^E are exploited to calculate the exemplar saliency map S_m^E on the basis of S_m^I. Specifically, for each pixel p that falls outside all regions in Φ_m^E, its exemplar saliency $S_m^E(p)$ is set to zero; otherwise, for each pixel p in a region $R_{m,k} \in \Phi_m^E$, its exemplar saliency based on each exemplar bounding box $B_{m,i}^{E}$ is defined as follows:

$$S_{m,i}^{B}(p) = \begin{cases} \sum_{p\in\Omega_{m,k,i}} S_m^I(p)/|\Omega_{m,k,i}|, \forall p \in \Omega_{m,k,i} \\ \sum_{p\in\bar{\Omega}_{m,k,i}} S_m^I(p)/|\bar{\Omega}_{m,k,i}|, \forall p \in \bar{\Omega}_{m,k,i} \end{cases}, \quad (3)$$

where $\Omega_{m,k,i}$ is the set of pixels covered by both $R_{m,k}$ and $B_{m,i}^{E}$, and $\bar{\Omega}_{m,k,i}$ is the set of pixels covered by $R_{m,k}$ but outside $B_{m,i}^{E}$. Then the exemplar saliency for pixel p is defined as follows:

$$S_m^E(p) = \sum\nolimits_{i=1}^{T_B} S_{m,i}^{B}(p). \quad (4)$$

Based on the exemplar discovery results in Fig. 1(b), the exemplar saliency maps are shown in Fig. 1(c), in which some exemplar co-salient object regions are highlighted with well-defined boundaries and most background regions are suppressed. However, some object regions which are not co-salient are also detected in some exemplar bounding boxes (such as the yellow player in the rightmost column of Fig. 1), and thus they are unavoidably highlighted in the exemplar saliency map. The following two subsections are exploited to suppress such falsely highlighted regions as well as to better highlight co-salient regions over the image set.

C. Region-level Co-Saliency Generation

On the basis of dense segmentation results $\{G_m^D\}_{m=1}^M$, we try to recover co-salient object regions by exploiting the similarity between regions using Eq. (1). Let $r_{m,i}$ denote the ith region in G_m^D, its most matched region in the image I_n is denoted as $z_{m,i}^n$, which is found using the following criterion:

$$z_{m,i}^n = \arg\max_{r_{n,j}} \left[Sim(r_{m,i}, r_{n,j}) \right]. \tag{5}$$

The process of recovering co-salient object regions consists of local recovery and global recovery. In the local recovery stage, for each region $r_{m,i}$, the most matched region $z_{m,i}^m$ in the same image is exploited to highlight those co-salient object regions that are suppressed in the exemplar saliency map. The locally recovered saliency for $r_{m,i}$ is defined as follows:

$$S_m^L(r_{m,i}) = \max \left[S_m^E(r_{m,i}), S_m^E(z_{m,i}^m) \cdot Sim(r_{m,i}, z_{m,i}^m) \right], \tag{6}$$

where $S_m^E(r_{m,i})$ is the average exemplar saliency value of all pixels in $r_{m,i}$. In the global recovery stage, for each region $r_{m,i}$, the most matched regions $z_{m,i}^n (n \neq m)$ in all the other images transfer their locally recovered saliency measures to $r_{m,i}$ weighted by their similarities with $r_{m,i}$, and the globally recovered saliency for $r_{m,i}$ is defined as follows:

$$S_m^G(r_{m,i}) = \frac{\sum_{n=1,n\neq m}^M Sim(r_{m,i}, z_{m,i}^n) \cdot S_n^L(z_{m,i}^n)}{\sum_{n=1,n\neq m}^M Sim(r_{m,i}, z_{m,i}^n)}. \tag{7}$$

In natural images, salient object regions usually have a less connectivity with image borders than background regions have. The connectivity between a region and image borders is helpful for region-based saliency detection on single images such as in [4], [7], [17], and thus a border connectivity based object prior is similarly evaluated on each region $R_{m,k}$ in the sparse segmentation results $\{G_m^S\}_{m=1}^M$ as follows:

$$BP(R_{m,k}) = \exp\left(-|R_{m,k} \cap IB_m| / |\partial R_{m,k}|\right), \tag{8}$$

where $\partial R_{m,k}$ is the boundary of $R_{m,k}$, and IB_m is a band region within 15 pixels to the borders of image I_m. Then for each region $r_{m,i}$ in G_m^D, its object prior is assigned as $BP(r_{m,i}) = BP(R_{m,k}), \forall r_{m,i} \subseteq R_{m,k}$. With such an object prior for further background suppression, the region-level co-saliency measure is defined as follows:

$$S_m^{CR}(r_{m,i}) = S_m^G(r_{m,i}) \cdot BP(r_{m,i}). \tag{9}$$

Based on the exemplar saliency maps in Fig. 1(c), the region-level co-saliency maps calculated using Eqs. (5)–(9) are normalized into the range of [0, 1] and shown in Fig. 1(d). Some co-salient object regions which are not effectively highlighted in Fig. 1(c) (see the two red players in the rightmost column) recover to a higher saliency in Fig. 1(d) due to a higher similarity with the highlighted exemplar regions. Besides, region-level co-saliency maps in Fig. 1(d) effectively suppress irrelevant object regions (see the yellow player in the rightmost column) which lack sufficient exemplar regions with high similarities in the whole image set, and also better suppress background regions such as stands and playfields.

D. Pixel-level Co-Saliency Generation

According to Gestalt laws, the regions that are close to the foci of attention should be more salient than far away regions. Based on the suggestion in [8], the foci of attention area Θ_m for the image I_m is obtained by thresholding the normalized region-level co-saliency map S_m^{CR} with $\max(T_m, 0.5)$ where T_m is the mean saliency value of S_m^{CR}. The co-saliency measure for each pixel p in I_m is then derived as follows:

$$S_m^{CP}(p) = S_m^{CR}(p) \cdot \exp\left[-\gamma \cdot \min_{q \in \Theta_m} (\|\mu_p - \mu_q\|) \right], \tag{10}$$

where μ_p is the spatial position of pixel p, and the attenuation coefficient γ is set to 6 as suggested by [6]. Using Eq. (10), those regions with higher saliency values but far away from the foci of attention area are suppressed effectively. Compared to Fig. 1(d), the pixel-level co-saliency maps in Fig. 1(e) still highlight the co-salient object regions well and suppress the irrelevant background regions more effectively.

III. Experimental Results

Our co-saliency model has been evaluated on two public benchmark datasets, i.e., iCoseg [25] and MSRC [26]. The iCoseg dataset contains 643 images from 38 object classes with varying image number per class. The MSRC dataset contains 14 object classes with about 30 images per class, and the pixel-wise binary ground truths are provided by [27]. We compared with the state-of-the-art co-saliency models including CB [12], HS [17], RFPR [18] and SACS [13]. For a fair comparison, all co-saliency maps are normalized into the same display range of [0, 255].

In our model, the number of selected exemplar bounding boxes, i.e., T_B, is important to the co-saliency detection performance. Therefore, we first analyzed the influence of T_B on performance and a series of precision-recall (PR) curves of our pixel-level co-saliency maps with different values of T_B are plotted in Fig. 2(a). Specifically, the precision and recall are measured on a series of binary object masks, which are obtained by varying threshold on co-saliency maps from 0

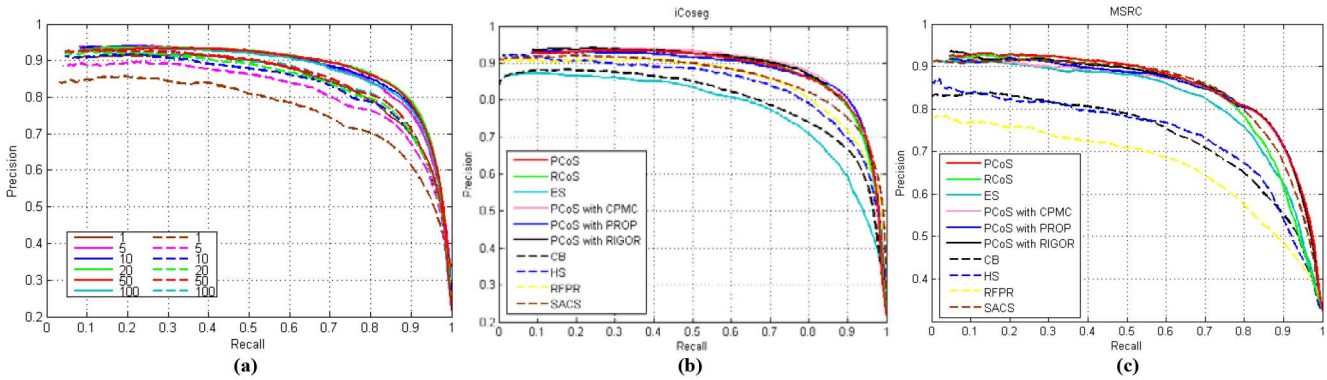

Fig. 2. (better viewed in color). (a) Influence of the number of selected bounding boxes on iCoseg (solid) and MSRC (dashed). Precision-recall curves of different models on (b) iCoseg dataset and (c) MSRC dataset.

to 255. At each integer threshold, the average precision and recall measures of all co-saliency maps generated by each model on each dataset are plotted to obtain the PR curves. As shown in Fig. 2(a), on both datasets, the co-saliency detection performance tends to be stable when T_B is not less than 10. For a comparison with other models, T_B is selected as 50, which shows the overall better performance on the two datasets.

As shown in Fig. 2(b) and (c), the PR curves of our exemplar saliency maps (ES), region-level co-saliency maps (RCoS) and pixel-level co-saliency maps (PCoS) exhibit the progressive improvement on co-saliency detection performance from ES, RCoS to PCoS. This clearly demonstrates the contribution of each stage of our model. Compared with CB, HS, RFPR and SACS, our model (PCoS) consistently outperforms all these co-saliency models on both datasets.

Fig. 3 intuitively shows the visual comparison of our model with all the other co-saliency models on some image sets. Compared with all the other models, we can see that our model highlights co-salient objects more consistently and suppresses background more effectively. It can be seen from Fig. 3 that CB fails to highlight nonhomogeneous objects which have distinctly contrasted regions, such as the two image sets *Panda* and *Cat*, while HS, RFPR and SACS cannot perform well when some background regions frequently occur over the image set, such as the darker background regions in the image set *Car* and the sky or grass regions in the image set *Tree*.

We also evaluated the robustness of our model to different object proposal methods [28]–[30]. We replace the objectness method with CPMC [28], PROP [29] and RIGOR [30] for exemplar saliency map generation. As shown in Fig. 2(b) and (c), our pixel-level co-saliency maps with the three object proposal methods, i.e., PCoS with CPMC/PROP/RIGOR, show similar co-saliency detection performance as PCoS, and also consistently outperform all the other co-saliency models.

Our model is implemented using Maltab on Ubuntu 14.04 with 3.5 GHz CPU. We performed the statistic of computation time on MSRC dataset, in which each image has a resolution of about 320×213. The average processing time per image is 31.93s, in which the preprocessing step including objectness, color/SIFT histograms and *gPb-owt-ucm* takes 26.08s, and the generation of ES, RCoS and PCoS takes 2.43s, 3.13s and 0.29s, respectively. The time-consuming part is the preprocessing step, which occupies 81.7% processing time. Fortunately, we believe that the preprocessing step can be substantially accelerated by using the parallel programming with CUDA for *gPb-owt-ucm*[31], SIFT [32] and fast algorithm for objectness [33].

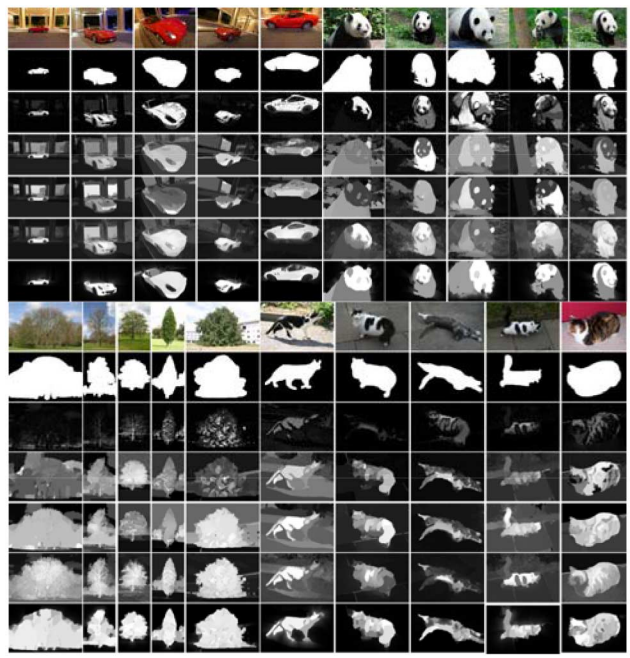

Fig. 3. Comparison with state-of-the-art co-saliency models on iCoseg dataset (top) and MSRC dataset (bottom). From top to bottom: original images, ground truths, co-saliency maps generated using CB [12], HS [17], RFPR [18], SACS [13] and our model, respectively.

IV. Conclusion

This letter presents a novel framework for co-saliency detection by co-salient object discovery and recovery. The reliable exemplars for co-salient objects are discovered for generating exemplar saliency maps. Then both local and global recovery of co-salient object regions, border connectivity and foci of attention area are effectively exploited to generate co-saliency maps for the image set. Experimental results on two benchmark datasets demonstrate that our co-saliency model outperforms the state-of-the-art co-saliency models.

REFERENCES

[1] Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1275–1289, Aug. 2012.

[2] 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.

[3] J. Xue, C. Li, and N. Zheng, "Proto-object based rate control for JPEG2000: An approach to content-based scalability," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1177–1184, Apr. 2011.

[4] 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.

[5] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE CVPR*, Jun. 2014, pp. 2814–2821.

[6] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE CVPR*, Jun. 2012, pp. 733–740.

[7] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, "Segmentation driven low-rank matrix recovery for saliency detection," in *Proc. BMVC*, Sep. 2013, article 79.

[8] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE CVPR*, Jun. 2010, pp. 2376–2383.

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

[10] 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.

[11] K. Chang, T. Liu, and S. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Proc. IEEE CVPR*, Jun. 2011, pp. 2129–2136.

[12] 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.

[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] 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.

[15] S. Du and S. Chen, "Detecting co-salient objects in large image sets," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 145–148, Feb. 2015.

[16] H. Li, F. Meng, and K. N. Ngan, "Co-salient object detection from multiple images," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1896–1909, Dec. 2013.

[17] 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.

[18] 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.

[19] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?," in *Proc. IEEE CVPR*, Jun. 2010, pp. 73–80.

[20] K. Chang, T. Liu, H. Chen, and S. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. IEEE ICCV*, Nov. 2011, pp. 914–921.

[21] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: Uniqueness, focusness and objectness," in *Proc. IEEE ICCV*, Dec. 2013, pp. 1976–1983.

[22] Y. Jia and M. Han, "Category-independent object-level saliency detection," in *Proc. IEEE ICCV*, Dec. 2013, pp. 1761–1768.

[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. ICCV*, Sep. 1999, pp. 1150–1157.

[24] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[25] 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.

[26] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *Proc. ECCV*, May 2006, pp. 1–15.

[27] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, "Unsupervised joint object discovery and segmentation in internet images," in *Proc. IEEE CVPR*, Jun. 2013, pp. 1939–1946.

[28] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1312–1328, Jul. 2012.

[29] I. Endres and D. Hoiem, "Category-independent object proposals with diverse ranking," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 36, no. 2, pp. 222–234, Feb. 2014.

[30] A. Humayun, F. Li, and J. M. Rehg, "RIGOR: Reusing inference in graph cuts for generating object regions," in *Proc. IEEE CVPR*, Jun. 2014, pp. 336–343.

[31] B. Catanzaro, B.-Y. Su, N. Sundaram, Y. Lee, M. Murphy, and K. Keutzer, "Efficient, high-quality image contour detection," in *Proc. IEEE ICCV*, Sep. 2009, pp. 2381–2388.

[32] M. Bjorkman, N. Bergstrom, and D. Kragic, "Detecting, segmenting and tracking unknown objects using multi-label MRF inference," *Comput. Vis. Image Understand.*, vol. 118, pp. 111–127, Jan. 2014.

[33] M. M. Cheng, Z. Zhang, W. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300 fps," in *Proc. IEEE CVPR*, Jun. 2014, pp. 3286–3293.