# Co-Saliency Detection Based on Hierarchical Segmentation

Zhi Liu, *Member, IEEE*, Wenbin Zou, Lina Li, Liquan Shen, and Olivier Le Meur

*Abstract*—Co-saliency detection, an emerging and interesting issue in saliency detection, aims to discover the common salient objects in a set of images. This letter proposes a hierarchical segmentation based co-saliency model. On the basis of fine segmentation, regional histograms are used to measure regional similarities between region pairs in the image set, and regional contrasts within each image are exploited to evaluate the intra-saliency of each region. On the basis of coarse segmentation, an object prior for each region is measured based on the connectivity with image borders. Finally, the global similarity of each region is derived based on regional similarity measures, and then effectively integrated with intra-saliency map and object prior map to generate the co-saliency map for each image. Experimental results on two benchmark datasets demonstrate the better co-saliency detection performance of the proposed model compared to the state-of-the-art co-saliency models.

*Index Terms*—Co-saliency detection, global similarity, hierarchical segmentation, regional similarity, saliency model.

## I. INTRODUCTION

THE research on saliency detection was originally motivated by simulating human visual attention mechanism to predict human fixations [1]. In the past decade, a number of saliency models for saliency detection in a single image have been proposed and widely used for salient object detection and segmentation, content-aware image retargeting, object-based image retrieval, etc [2]. Recently, an emerging and interesting issue in saliency detection is to detect the common salient objects in a set of images, and is formally defined as co-saliency detection [3]–[9]. Compared with saliency models for a single image, co-saliency models also exploit the relevance between different images to highlight the common salient objects, which simultaneously occur with similar appearances in different images of the image set. The output of co-saliency model is a set of co-saliency maps, and can be used for object co-segmentation and co-recognition, which are key to the internet visual media retrieval [10].

In a pair of images, which are captured for the same scene, the local structure changes caused by salient objects between the two images are exploited for co-saliency detection [3]. However, its obvious limitation is that the two images need to have highly similar background. Without such a constraint, the joint information provided by the image pair is utilized under a preattentive scheme to find co-salient object regions [4]. In [5], the co-saliency is formulated as a linear combination of single-image saliency map, which is calculated using the three available saliency models, and multi-image saliency based on a co-multilayer graph. However, the above two models [4][5] are still applicable only to image pairs. Regarding with the image set containing more than two images, individual saliency map is first generated for each image, and then those salient parts that frequently occur in most images are considered as co-salient regions in [6]. For a collection of images, group saliency [7] is estimated by extracting salient objects that maximize between-image similarities and within-image distinctness, and inter-similarity with reference images is exploited in the narrative saliency detection [8] for image collage. In [9], a cluster-based co-saliency model first integrates contrast cue, spatial cue and corresponding cue to measure the cluster-level co-saliency, and then incorporates the likelihoods of pixels belonging to clusters, to generate pixel-level co-saliency maps.

For saliency detection in a single image, some recent models benefit from measuring saliency on over-segmented regions/superpixels and object/background priors. For example, kernel density estimation based region model [11], region based contrast [12], distinctiveness and compactness of regional histograms [13], region segmentation based object prior [14] and geodesic saliency of superpixels using background priors [15] have shown their effectiveness to improve the saliency detection performance. Inspired by the above works, we propose an effective hierarchical segmentation based co-saliency model applicable to saliency detection in a set of images. Our main contribution lies in the following two aspects. First, as the basis of building co-saliency model, we introduce the hierarchical segmentation, as a further improvement, to evaluate regional contrast based intra-saliency on the fine segmentation and object prior on the coarse segmentation, respectively. Second, for co-saliency measurement, we propose to derive the global

Z. Liu is with the School of Communication and Information Engineering, Shanghai University, Shanghai, China, and also with IRISA, Campus Universitaire de Beaulieu, Rennes, France (email: liuzhisjtu@163.com).

W. Zou is with IETR, UMR CNRS 6164, INSA de Rennes, Université Européenne de Bretagne, 35708 Rennes, France.

L. Li and L. Shen are with the School of Communication and Information Engineering, Shanghai University, Shanghai, China.

O. Le Meur is with the University of Rennes 1 and IRISA, Campus Universitaire de Beaulieu, Rennes, France.
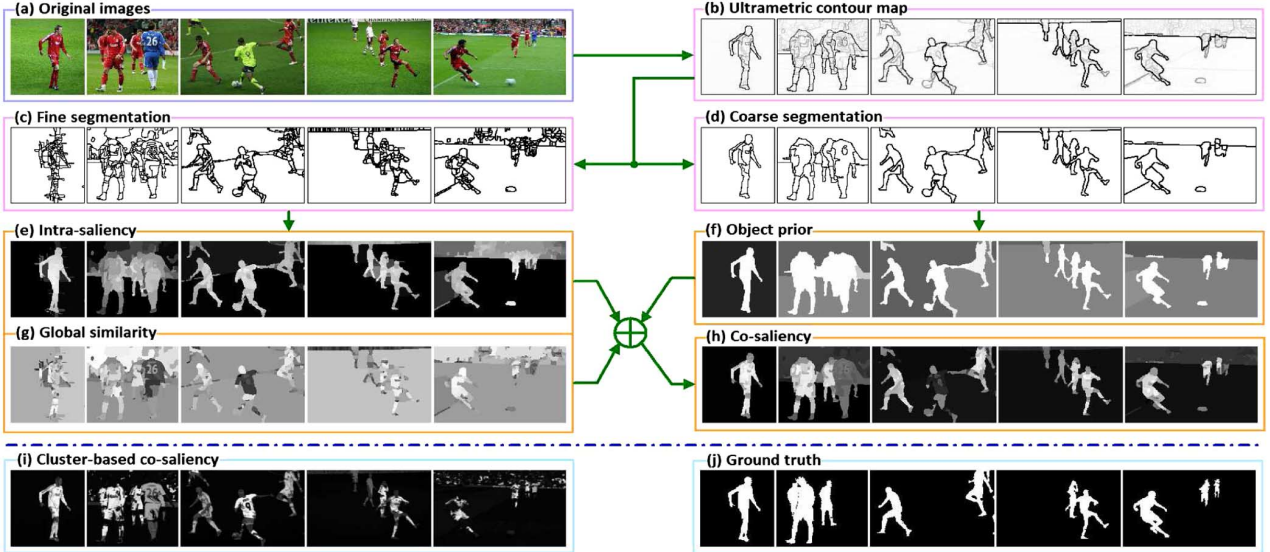
Fig. 1. Illustration of the proposed hierarchical segmentation based co-saliency model.

similarity measures of regions over the image set based on regional similarity measures, and effectively integrate with intra-saliency and object prior for co-saliency map generation, which consistently outperforms the state-of-the-art co-saliency models on two benchmark datasets.

The rest of this letter is organized as follows. Section II describes the proposed co-saliency model in detail. Experimental results and analysis are presented in Section III, and conclusion is given in Section IV.

## II. PROPOSED CO-SALIENCY MODEL

The proposed co-saliency model is illustrated in Fig. 1. The following five subsections will describe the hierarchical segmentation, regional similarity measure, the generation process of intra-saliency map, object prior map and co-saliency map, respectively.

### A. Hierarchical Segmentation

Given a set of images $\{I_m\}_{m=1}^M$, we choose the hierarchical image segmentation approach in [16] to calculate for each image $I_m$ the real-valued ultrametric contour map $UCM_m$, which weighs each boundary by its scale of disappearance. For the example image set (5 out of 33 images in this set are shown) in Fig. 1(a), the corresponding UCMs are shown in Fig. 1(b), in which darker boundaries disappear at a coarser scale of region segmentation than lighter boundaries. The thresholding operation is performed on each UCM, which is normalized into the range of [0, 1], with the threshold increasing from 0 to 1 with an interval of 0.01. The fine segmentation result and the coarse segmentation result are obtained when the generated region number is just below $\tau_F$ and $\tau_C$, which are set to 200 and 20, respectively, for a sufficient over-segmentation and a more meaningful moderate segmentation. Based on UCMs in Fig. 1(b), the fine and coarse segmentation results are respectively shown in Fig. 1(c) and (d), in which different regions are delineated using black boundaries.

### B. Regional Similarity

Using all images in the image set $\{I_m\}_{m=1}^M$, each color channel in the *Lab* color space is uniformly quantized into $q$ bins, to obtain the normalized global histogram $\mathbf{h}_G$ with $q^3$ bins for the image set. The quantized color of each bin in $\mathbf{h}_G$ is calculated as the mean color of those pixels that fall into the $k$th bin. Then the first $t$ high-probability bins that accumulate just above the threshold $\alpha$, are selected as the representative colors. Finally, each of the remaining bins is merged into one of the selected $t$ bins with the smallest difference between their quantized colors, to update $\mathbf{h}_G$ with $t$ bins. $q$ and $\alpha$ are set to 16 and 0.95, respectively, for a moderate color quantization of images.

For each region $r_F^{m,i}(i = 1, \ldots, N_F^m)$ in the fine segmentation of each image $I_m$, its regional histogram $\mathbf{h}_F^{m,i}$ is calculated using the quantized colors of all pixels in $r_F^{m,i}$, similarly as the calculation of $\mathbf{h}_G$. For each pair of regions, $r_F^{m,i}$ and $r_F^{n,j}$, the regional similarity is defined as the Bhattacharyya coefficient between their regional histograms, i.e.,

$$\lambda(r_F^{m,i}, r_F^{n,j}) = \sum_{k=1}^t \sqrt{\mathbf{h}_F^{m,i}(k) \cdot \mathbf{h}_F^{n,j}(k)} \qquad (1)$$

### C. Intra-Saliency

We observed from a variety of natural images that salient object regions usually show noticeable contrast with background regions. Therefore for each region $r_F^{m,i}$, its intra-saliency is evaluated based on the regional contrasts with all the other regions in $I_m$, taking into account the weights of region areas and spatial distances between regions as follows:

$$S_{\text{intra}}(r_F^{m,i}) = \sum_{j=1,j\neq i}^{N_F^m} \left| r_F^{m,j} \right|$$
$$\cdot \left[ D^m - d_F^{m,i,j} \right] \cdot \left[ 1 - \lambda(r_F^{m,i}, r_F^{m,j}) \right] \quad (2)$$

where $\left| r_F^{m,j} \right|$ is the area of region $r_F^{m,j}$, $D^m$ is the diagonal length of image $I_m$, and $d_F^{m,i,j}$ is the distance between the spatial center of $r_F^{m,i}$ and $r_F^{m,j}$. Equation (2) indicates that those

regions, which are larger and spatially closer to $r_F^{m,i}$, have a relatively larger contribution to the evaluation of intra-saliency of $r_F^{m,i}$. As shown in Fig. 1(e), salient objects such as players and football in each image are reasonably highlighted while most background regions such as fields are suppressed.

### D. Object Prior

Besides, background regions generally have a higher ratio of connectivity with image borders than salient object regions in a variety of images. Such a prior of image layout was exploited to modulate the feature matrix used in low-rank matrix recovery for saliency detection in [14], and to formulate the saliency of patch/superpixel as the length of its shortest path to image borders in [15]. Following our previous work [14], the object prior is more suitably evaluated on the coarse segmentation, which partitions the image into relatively fewer regions, to obtain a more meaningful object prior map. Specifically, for each region $r_C^{m,i}(i = 1, \ldots, N_C^m)$ in the coarse segmentation of image $I_m$, its object prior is defined as follows:

$$p_C^{m,i} = \exp\left(-\gamma \frac{\left|r_C^{m,i} \cap B^m\right|}{b_C^{m,i}}\right) \quad (3)$$

where $B^m$ denotes the borders of $I_m$, and $b_C^{m,i}$ denotes the perimeter of $r_C^{m,i}$. We experimentally observed that the performance of our co-saliency model is not sensitive to the decay factor $\gamma$ when it is between 1.5 and 2.5, and thus $\gamma$ is set to 2 in our experiments. The object prior maps for coarse segmentations in Fig. 1(d) are shown in Fig. 1(f), which moderately suppresses background regions such as fields and stands. Then for each region $r_F^{m,j}(j = 1, \ldots, N_F^m)$ in the fine segmentation, its object prior is assigned with that of the corresponding region in the coarse segmentation, i.e., $p_F^{m,j} = p_C^{m,i}, \forall r_F^{m,j} \subseteq r_C^{m,i}$.

### E. Co-Saliency

For co-saliency detection, the regional similarity measures between regions in different images are exploited to highlight co-salient regions, for which highly similar regions can be found in other images of the image set, and to effectively suppress those regions, for which similar regions rarely occur in other images. Specifically, for each region $r_F^{m,i}$, the most similar region in each of the other images $\{I_n\}_{n=1,n\neq m}^M$ is first searched by maximizing the regional similarity measure. Then those maximum regional similarity measures found on the $M-1$ images are summed to define the global similarity measure of $r_F^{m,i}$ over the image set as follows:

$$\varphi(r_F^{m,i}) = \sum_{n=1,n\neq m}^M \max_{j=1,\ldots,N_F^n}\left\{\lambda(r_F^{m,i}, r_F^{n,j})\right\} \quad (4)$$

As shown in Fig. 1(g), the players with red jersey, which are co-salient objects in the image set, and fields, which also occur in almost all images in the image set, are highlighted in the global similarity maps, while other regions such as players without red jersey are suppressed.

On the basis of intra-saliency measures over the image set, the co-saliency for each region $r_F^{m,i}$ is measured by
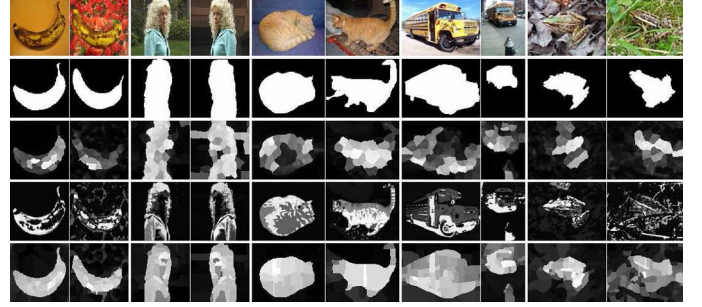


Fig. 2. Examples of co-saliency detection on CP dataset. From top to bottom: original images in five image pairs, binary ground truths, co-saliency maps generated using Li's model [5], Fu's model [9] and our model, respectively.

integrating regional similarity, global similarity and object prior as follows:

$$S_{co}(r_F^{m,i}) = p_F^{m,i} \cdot \frac{\sum_{n=1}^M \sum_{j=1}^{N_F^n}\left[S_{intra}(r_F^{n,j}) \cdot \varphi(r_F^{n,j}) \cdot \lambda(r_F^{m,i}, r_F^{n,j})\right]}{\sum_{n=1}^M \sum_{j=1}^{N_F^n} \lambda(r_F^{m,i}, r_F^{n,j})} \quad (5)$$

Eq. (5) indicates that the intra-saliency of each region $r_F^{n,j}$ is globally adjusted using its global similarity measure to obtain $S_{intra}(r_F^{n,j}) \cdot \varphi(r_F^{n,j})$. Then for each region $r_F^{m,i}$, its co-saliency is a sum of the globally adjusted intra-saliency measures of all regions weighted by their regional similarities with $r_F^{m,i}$, and finally regulated using its object prior. Co-saliency maps generated by integrating Fig. 1(e), (f) and (g) are shown in Fig. 1(h), which effectively highlights co-salient objects and suppresses irrelevant regions including fields, stands and other players compared to intra-saliency maps in Fig. 1(e). For a visual comparison, co-saliency maps generated using Fu's co-saliency model [9] and binary ground truths for co-salient objects are also shown in Fig. 1(i) and (j), respectively.

### III. Experimental Results

The proposed co-saliency model has been tested on two datasets with binary ground truths for co-salient objects. Specifically, the Co-saliency Pairs (CP) dataset [5] contains 210 images, which consists of 105 image pairs, and the CMU Cornell iCoseg dataset [17] contains 643 images from 38 object classes, each of which has 5 to 41 images. We have compared our model with two state-of-the-art co-saliency models, i.e., Li's model [5] (only suitable for image pairs) and Fu's model [9].

For a subjective comparison, co-saliency maps generated for several examples in CP and iCoseg are shown in Fig. 2 and Fig. 3, respectively. For a fair comparison, all co-saliency maps are normalized into the same range of [0, 255]. As shown in Figs. 2–3, our co-saliency maps highlight the complete co-salient objects with well-defined boundaries and suppress background regions more effectively compared to Li's and Fu's co-saliency maps. However, our model as well as other two models cannot effectively suppress some background regions, which either show similar color with the object such as the nearby regions surrounding the frog in the rightmost column of Fig. 2, or appear frequently in the image set and visually salient such as the bush regions compared to the major background
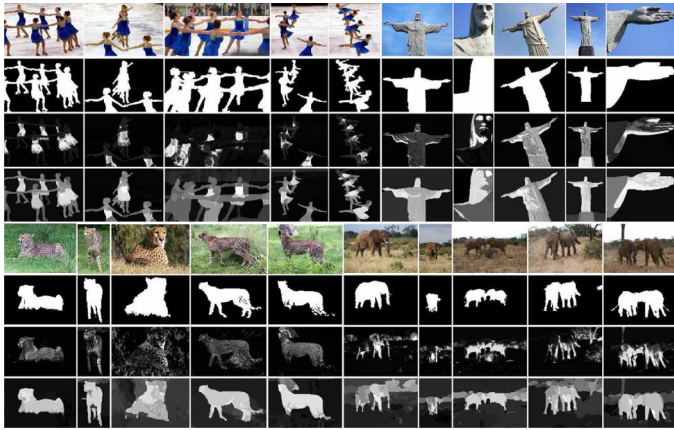
Fig. 3. Examples of co-saliency detection on iCoseg dataset. From top to bottom: some original images in four image sets, binary ground truths, co-saliency maps generated using Fu's model [9] and our model, respectively.

such as grassland and sky regions in the bottom-right example of Fig. 3. Since our model exploits the color based regional similarity for co-saliency measurement, such background regions are not effectively suppressed.

For an objective comparison, we performed thresholding with a series of fixed integers from 0 to 255 on co-saliency maps to obtain a set of binary co-salient object masks, and calculated the precision and recall measures using the binary ground truths in CP and iCoseg as reference masks. Then for each model, at each threshold, the precision/recall values for all co-saliency maps in the dataset are averaged to plot the precision-recall (PR) curves. Fig. 4 shows that on both datasets, the PR curves of our model are consistently higher than that of Li's and Fu's model, and thus demonstrates that our model outperforms other models on the co-saliency detection performance.

To clearly evaluate the contribution of hierarchical segmentation in Section II-A, object prior in Section II-D and co-saliency measurement in Section II-E, some additional PR curves are shown in Fig. 4 and analyzed as follows:

1) We used a conventional segmentation approach, i.e., mean shift [18], to replace the hierarchical segmentation. For mean shift, we set its parameter of allowable minimum region area to 0.1% and 5% of image area, to obtain the fine segmentation and coarse segmentation, respectively. As shown in Fig. 4, our model using mean shift still outperforms Li's and Fu's model, while the use of hierarchical segmentation can further improve the co-saliency detection performance of our model.

2) The PR curves of our co-saliency maps generated without using object prior show that the co-saliency detection performance is lowered, but still higher than Li's and Fu's model on the two datasets. In other words, the use of object prior further improves the co-saliency detection performance.

3) The PR curves of our co-saliency maps (with or without object prior) are obviously higher than the PR curves of our intra-saliency maps and saliency maps generated using the state-of-the-art saliency models for a single image (RC [12], KDE [11] and SLR [14]). This demonstrates the advantage of co-saliency detection, which analyzes a set of images together, over saliency detection on the basis of individual image.
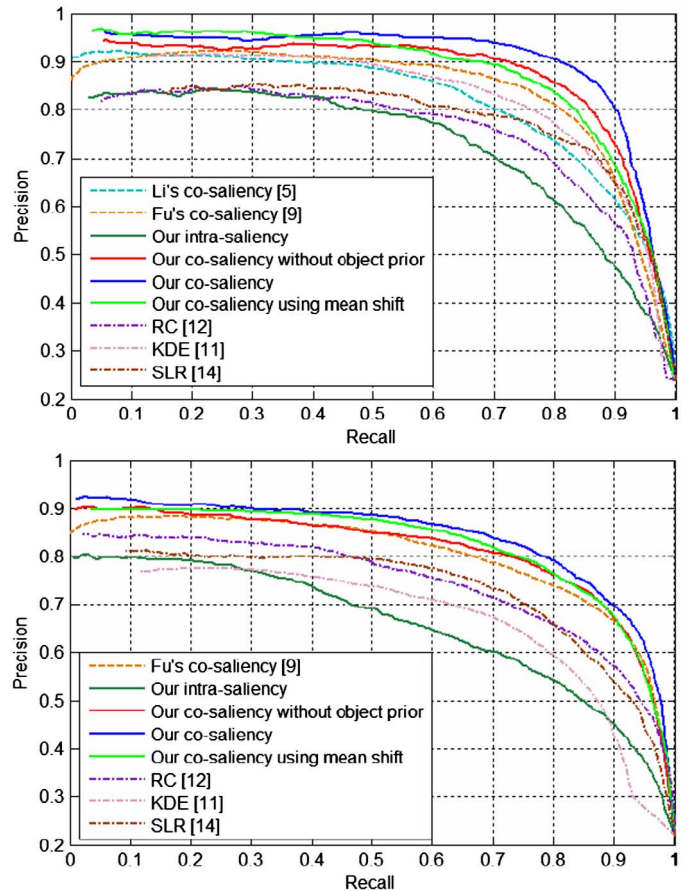


Fig. 4. (Better viewed in color) Precision-recall curves of different saliency models on CP dataset (top) and iCoseg dataset (bottom).

TABLE I
AVERAGE PROCESSING TIME (SECOND) PER IMAGE

| Model | Li [5] | Fu [9] | Our (HS) | Our (MS) |
|---|---|---|---|---|
| CP | 440.40 | 1.65 | 15.73 (14.38) | 4.45 (0.19) |
| iCoseg | N/A | 4.08 | 103.36 (91.57) | 25.99 (2.90) |

The average processing time of different models using Matlab implementations on a PC with Intel i7-3770 3.4 GHz CPU and 4 GB RAM are shown in Table I, in which the values in the brackets are the processing time taken by hierarchical segmentation (HS) and mean shift segmentation (MS). Table I shows that Fu's model has the highest computational efficiency, while our model is degraded due to the use of HS, which occupies 91.4% and 88.6% processing time on CP and iCoseg, respectively. Nonetheless, our model using MS efficiently elevates the computational efficiency, with a compromise of co-saliency detection performance.

## IV. CONCLUSION

In this letter, we have presented a co-saliency model based on hierarchical segmentation. Regional similarity, global similarity and intra-saliency, which are measured on the fine segmentation, and object prior, which is evaluated on the coarse segmentation, are effectively integrated to generate co-saliency maps for a set of images. Experimental results demonstrate the better co-saliency detection performance of our model.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[2] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 33, no. 11, pp. 2131–2146, Nov. 2011.

[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. 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] 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.

[7] M. M. Cheng, N. J. Mitra, X. Huang, and S. M. Hu, "SalientShape: Group saliency in image collections," *Vis. Comput.*, Aug. 2013, 10.1007/s00371-013-0867-4.

[8] L. Zhang and H. Huang, "Hierarchical narrative collage for digital photo album," *Comput. Graph. Forum*, vol. 31, no. 7, pp. 2173–2181, Sep. 2012.

[9] 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.

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

[11] 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.

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

[13] Z. Liu, O. Le Meur, S. Luo, and L. Shen, "Saliency detection using regional histograms," *Opt. Lett.*, vol. 38, no. 5, pp. 700–702, Mar. 2013.

[14] 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.

[15] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. ECCV*, Oct. 2012, vol. 3, pp. 29–42.

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

[17] 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.

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