Graph-based Salient Object Detection Using Background and Foreground Connectivity Cues

M. Rezaei Abkenar, *Student Member, IEEE,* H. Sadreazami, *Member, IEEE,* and M. Omair Ahmad, *Fellow, IEEE*
Department of Electrical and Computer Engineering
Concordia University, Montreal, Quebec, Canada H3G 1M8

***Abstract*—Salient object detection is an active research topic due to several potential applications in image compression, scene understanding, image retrieval, and so forth. In this paper, a salient object detection method is proposed by leveraging the recent advances in graph signal processing. Since, the image boundary regions generally belong to the image background, a distribution-based boundary contrast map is generated. Also, the graph representation of the image is used to compute the connectivity of the image regions to the image boundary as well as those to their local neighbors and the image foreground. The connectivity maps obtained are fused with the boundary contrast map in order to obtain the image saliency map. Several experiments are conducted to evaluate the performance of the proposed salient object detection method and to compare it with the state-of-the-arts. Results on datasets of images demonstrate that the proposed method achieves superior performance to the state-of-the-art methods in terms of precision, recall, and mean absolute error values.**

***Index Terms*—Salient object detection, graph signal processing, boundary connectivity.**

I. Introduction

Salient object is part of an image that captures the human visual system's (HVS) attention due to its high contrast with respect to other regions in the image. The main goal of a saliency detection method is obtaining a saliency map, in which the salient object is highlighted, while the background region is suppressed. With several applications, such as object of interest image segmentation [1], adaptive image compression [2], object-based image retrieval [3], content-aware image resizing [4]-[5], and medical imaging [6], salient object detection has attracted a great deal of interest in image processing and computer vision research communities, and several salient object detection methods have been proposed [7]–[11].

The key point in salient object detection for an image is to compare its different regions using appropriate measures in order to discriminate the salient region from the non-salient ones.

Recent advances in graph signal processing has provided an opportunity to revisit traditional image processing solutions by providing a new framework for representing the relations among different pixels or regions of an image [12]–[14]. In a graph-based framework, the image samples are represented on a weighted graph. Such a weighted graph can be utilized to capture the similarities between image pixels or image regions. In view of this, some graph-based salient object detection methods have been proposed [15]–[19]. In these methods, the image boundary connectivity cue has been shown to be effective in detecting the salient object. Considering the image boundary connectivity, these methods assume that the image boundary regions mostly belong to the image background. Therefore, the regions that are closer and more similar to the image boundary are most likely non-salient. In other words, the non-salient regions provide a larger boundary connectivity value in comparison to the salient region.

In the methods which are based on the boundary connectivity cue, each region is only compared to the image boundary. Therefore, the saliency value of each region is evaluated in a global manner. However, evaluating the similarity of each region to its local neighborhood is also required. The local saliency value has been overlooked in most of the existing graph-based schemes. In addition, the boundary connectivity cue fails in images in which the salient object touches the image boundary. In view of this, in this paper, by leveraging the new framework of graph signal processing, we address the salient object detection problem by considering the distance of each region to the image boundary, its local neighborhood and the image foreground. Moreover, in order to further improve the performance specially in images with salient objects located around the image boundary, a distribution-based boundary contrast value is calculated for all the image regions including the ones located at the image boundary. The background and foreground cues obtained are then integrated in a graph-based optimization framework. Several experiments are conducted to evaluate the performance of the proposed graph-based salient object detection method.

II. Proposed Salient Object Detection Method Using Foreground and Background Cues

In this section, the proposed graph-based salient object detection method is presented. In the proposed method, the saliency value of each region, is determined based on its connectivity to the image background as well as that to the image foreground, where the regions around the image boundary are assumed to belong to the image background. Boundary contrast of each region is first measured using a distribution-based algorithm. Then, a graph-based strategy is

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the Regroupement Strategique en Microelectronique du Quebec (ReSMiQ).

978-1-7281-0397-6/19/$31.00 ©2019 IEEE

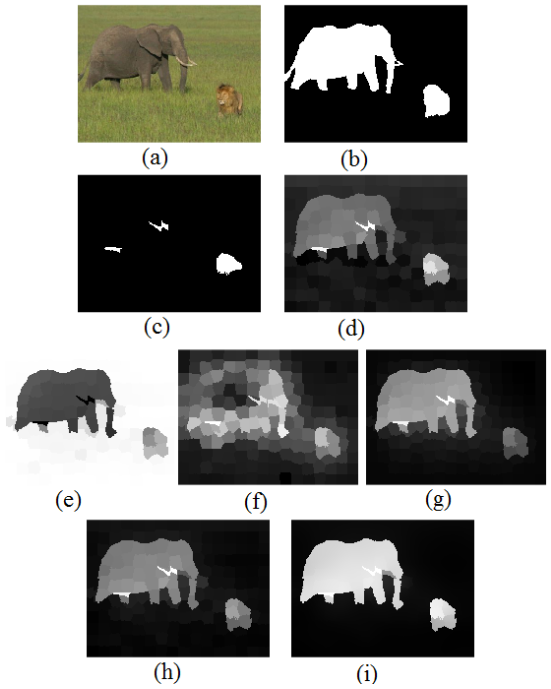


Fig. 1: (a) A sample image. (b) Ground truth. (c) The rough foreground region. (d) The distribution-based boundary contrast map. (e) The graph-based background connectivity map. (f) The graph-based foreground connectivity map. (g) Optimized fused map. (h) The map obtained after combining the optimized fused map and the boundary contrast map. (i) The final saliency map.

employed to evaluate the connectivity of each region to the image foreground and background. Fig. 1 shows a sample image, its corresponding ground truth and output of different steps of the proposed method.

In the proposed method, the input image is first converted to the commonly-used CIELAB color space, having luminance, red/green, and blue/yellow channels, denoted by L, a, and b, respectively. In the saliency detection problem, the main goal is detecting the most distinctive region of the image rather than the individual pixels. In view of this, the input image is segmented into n nearly regular regions called super-pixels using the SLIC algorithm [20].

A. Graph Construction

In the proposed method, an undirected weighted graph $G = (V, E, K)$ is constructed consisting of a finite set V of vertices (image super-pixels) and a finite set E of edges with the corresponding weights $k_{pq} \in \boldsymbol{K}$, which denote similarity between vertices (super-pixels) p and q in the graph. The similarity weights are represented as $\boldsymbol{K} = [k_{pq}]$, where k_{pq} is defined by the standard Gaussian kernel as

$$k_{pq} = 1 - exp\left(-\left[\frac{d_{pq}^2}{2\sigma^2}\right]\right), \tag{1}$$

where $d_{pq}^2 = (x_p - x_q)^2$ and σ controls the level of similarity achieved by (1), x_p and x_q denote the mean Lab color vectors of super-pixels p and q, respectively. Since in the proposed method length of the shortest path between the nodes are measured, the weights in (1) are defined such that a larger weight is equivalent to being more different. Moreover, in order to increase the connectivity value of the boundary super-pixels, they are all linked together by adding edges in the graph.

Having the graph similarity matrix $\boldsymbol{K}$, the corresponding graph Laplacian matrix is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{K}$, where $\boldsymbol{D} = \text{diag}\left\{\sum_q k(1,q), ..., \sum_q k(n,q)\right\}$. The graph Laplacian matrix plays an important role in describing the underlying structure of the graph signal.

B. Distribution-based Boundary Contrast Map

Distribution-based boundary contrast map is constructed to detect the regions that are different from the image boundary. To this end, the pixels located in four strips, each s pixels wide, around the four image boundaries are assumed to be the boundary region. Considering the distribution of pixels in such a non-salient boundary region, the Mahalanobis distance between each super-pixel and the boundary region distribution is computed. The mean Lab color vector, $\bar{\mathbf{x}}_{bnd}$, and the color covariance matrix, $\mathbf{C}_{bnd}$, of the boundary region, and also the mean color vector of the pixels inside each super-pixel, $\bar{\mathbf{x}}_p$, are computed. The distribution-based boundary contrast value BC_{dist}, of each super-pixel is then obtained as

$$BC_{dist}(p) = \sqrt{(\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_{bnd})\,\mathbf{C}_{bnd}^{-1}\,(\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_{bnd})^T}. \tag{2}$$

These values are then normalized by scaling between 0 and 1.

In the contrast map obtained, super-pixels with a value larger than a threshold $th_{BC_{dist}}$ are roughly selected as the image foreground. This rough foreground region, FG_{region}, is further used in the following graph-based saliency computation. Fig. 1(c) and (d) show the rough foreground region and the boundary contrast map for the sample image in Fig. 1(a).

It should be noted that the boundary contrast value is computed for all the super-pixels including the ones located at the image boundary. This makes the proposed algorithm robust especially in images in which the salient region touches the image boundary.

C. Graph-based Foreground and Background Connectivity Maps

Using a graph representation of the image discussed in Section II. A, two maps, namely, foreground and background connectivity maps, are generated in order to evaluate the similarity of each super-pixel to the image foreground and background regions.

The shortest path between the two super-pixels p and q in the graph, is calculated as the accumulated edge weights along the shortest path between them, as given by

$$d(p,q) = \min_{p_1=p, p_2, \ldots, p_n=q} \sum_{i=1}^{n-1} k_{p_i p_{i+1}}. \tag{3}$$

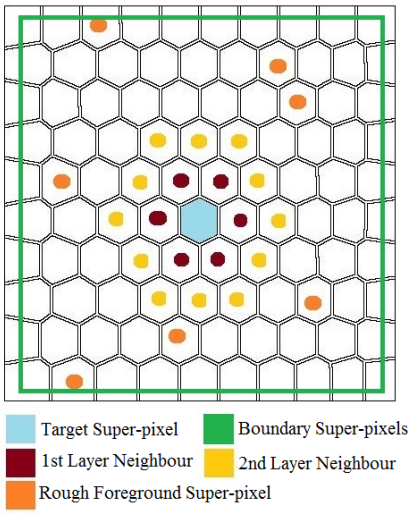


Fig. 2: A synthetic image representing a sample super-pixel, the boundary super-pixels, the 2 layers adjacent super-pixels, and the rough foreground super-pixels.

In order to obtain the background map, super-pixels located around the image boundary are regarded as the background region as shown in Fig. 2. The value of a super-pixel p in the graph-based background connectivity map B_{map}, is computed based on the shortest path length from p to all the boundary super-pixels as

$$B_{map}(p) = -\left\{\sum_{i=1}^{N} d(p,p_i)\,|p_i \in Boundary\right\}, \quad (4)$$

where N is number of the image boundary super-pixels. The background map obtained is scaled between 0 and 1. Fig. 1(e) shows the graph-based background connectivity map for the sample image in Fig. 1(a).

The value of a super-pixel p in the graph-based foreground map F_{map}, is obtained using the shortest path length from p to the super-pixels of the rough foreground region FG_{region}, obtained in Section II.B, and also to its local super-pixels $Local_{region}$ (shown in Fig. 2), as given by

$$F_{map}(p) = -\left\{\sum_{i} d(p,p_i)\,|p_i \in FG_{region}\right\} + \left\{\sum_{i} d(p,p_i)\,|p_i \in Local_{region}\right\}, \quad (5)$$

where the local region $Local_{region}$ of each super-pixel consists of its adjacent super-pixels in two layers and also its κ-nearest neighbors set of super-pixels. The foreground map obtained is then normalized to $[0,1]$ range. The foreground connectivity map for the sample image in Fig. 1(a), is shown in Fig. 1(f).

D. Fusion of Foreground and Background Cues

The foreground and background connectivity maps are combined by considering three following constraints in an optimization framework [16] The super-pixels having large values in the foreground map are most probably salient. 2) The super-pixels having large values in the background map are most probably non-salient. 3) The super-pixels that are adjacent and also similar in appearance should have the same saliency values. Accordingly, the cost function is defined as

$$\sum_{i=1}^{n} F_{map}^{i}\left(\hat{X}_i - 1\right)^2 + \sum_{i=1}^{n} B_{map}^{i}\left(\hat{X}_i\right)^2 + \hat{\boldsymbol{X}}^T \boldsymbol{L}\hat{\boldsymbol{X}} \quad (6)$$

where $\hat{X}_i$ is the optimized value of the super-pixel i in the fused map and $\hat{\boldsymbol{X}}^T \boldsymbol{L}\hat{\boldsymbol{x}}$ denotes the graph Laplacian regularizer or smoothness term between adjacent super-pixels.

The optimized fused map is computed by solving Eq. (6) in a least-square error sense, as given by

$$\hat{\mathbf{X}} = \mathbf{F}_{map}\left(\mathbf{F}_{map} + \mathbf{B}_{map} + \mathbf{L}\right)^{-1} \quad (7)$$

Fig. 1(g) shows the optimized fused map obtained by applying Eq. (7).

In order to leverage the boundary contrast cue, the distribution-based boundary contrast map, $\mathbf{BC}_{dist}$, is then added to the optimized fused map thus obtained, $\hat{\mathbf{X}}$, i.e., $\mathbf{S} = (1/entropy(\hat{\mathbf{X}})) * \hat{\mathbf{X}} + (1/entropy(\mathbf{BC}_{dist})) * \mathbf{BC}_{dist}$, where the entropy-based weights are used to assign a larger weight to the map with smaller number of gray levels [10]. Fig. 1(h) shows the map obtained by the weighted sum of the boundary contrast map, Fig. 1(d), and the optimized fused map, Fig. 1(g).

E. Post-processing

In order to smooth the saliency map $\mathbf{S}$, while retaining the details of the boundaries between salient and non-salient regions, a bilateral filtering is applied to the fused map [21]. Also, accounting for the center-bias criterion, the map obtained is pixel-wise multiplied with a parameter free centeredness map [18]. Finally, to increase the contrast between salient and non-salient regions, the contrast is further enhanced by applying a sigmoid function, defined as $f(x) = \frac{1}{1+exp(-\alpha(x-0.5))}$, to the saliency map, where α is a parameter to control the contrast level. The final saliency map after applying the post-processing step for the sample image in Fig 1(a), is shown in Fig. 1(i).

III. Experimental Results

Several experiments are conducted on images of two challenging and widely used datasets, DUT-OMRON [17] (containing 5168 images) and HKU-IS [22] (containing 4447 images). Performance of the proposed salient object detection method is evaluated and compared to those of the other existing methods. In the experiments, number of super-pixels in each image n is 600. Also, σ in Eq. (1) is set to 0.5. The width of the boundary strips s is set to $0.1 \times (min(H,W))$ where H and W are the two dimensions of the image. The threshold to find the rough foreground region $th_{BC_{dist}}$ is 0.7. To find the $Local_{Region}$ of each node, κ is found adaptively by considering all the nodes connected to the target node with

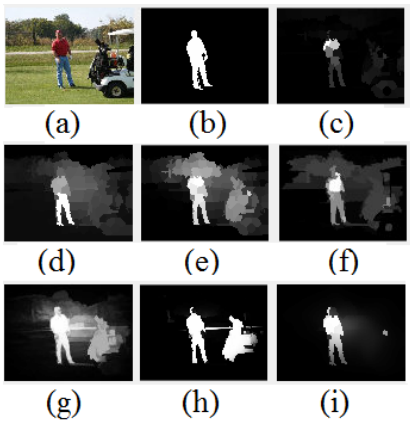


Fig. 3: Saliency maps obtained by applying different methods on a test image from DUT-OMRON dataset. (a) Original image. (b) Ground truth. (c) SF [24]. (d) MR [17]. (e) SO [16]. (f) RC [23]. (g) MBD+ [18]. (h) MST [19]. (i) Proposed method.

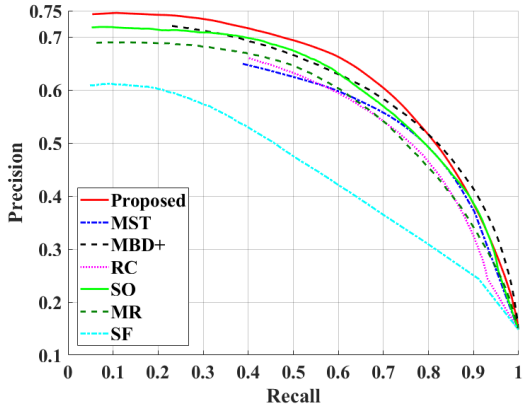
an edge value smaller than $0.1 \times (mean\,(all\,the\,edges))$. In the post processing step we set $\alpha = 10$.

Performance of the proposed method is compared with that of a number of related and state-of-the-art methods, namely, MST [19], MBD+ [18], RC [23], SO [16], MR [17], and SF [24]. The saliency maps obtained using the proposed method as well as that of the other methods for a test images is shown in Fig. 3. It is seen from this figure that the salient object is detected more accurately by applying the proposed method as compared to the other methods.

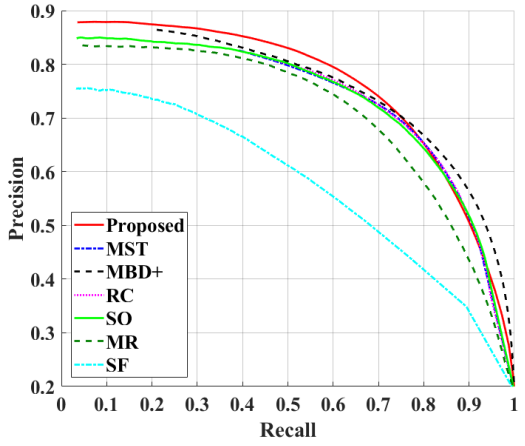
In order to objectively compare the gray level saliency maps with the binary ground truth, the saliency maps are converted to binary maps using fixed and adaptive threshold values [10], [25]. Figs. 4 (a)-(b) illustrate averaged precision-recall curves obtained by using various methods over the images in the two datasets. It can be seen from these figures that for the images in the two datasets, the proposed method provides larger precision values for a wide range of recall values as compared to the other methods. The F_β [25] and mean absolute error (MAE) values obtained using the proposed method as well as that of the other existing methods are given in Table 1. It is seen from this table that the proposed salient object detection method is superior to the other existing methods, as indicated by the larger F_β when applied to the images in the two datasets. Also, the proposed method provides smaller values of MAE as compared to the other methods for images in the both datasets.

IV. Conclusions

In this paper, a salient object detection method has been proposed by using the foreground and background connectivity cues in a graph-based framework. To this end, each superpixel's distance the boundary superpixels and its local region have been computed using the image graph representation. The two connectivity maps have been fused in an optimization framework and the result has been combined with a distribution-based boundary contrast map. Experimental results have shown that the proposed method outperforms the other methods in terms of precision, recall, F-measure and mean absolute error values. The improved performance of the proposed method is attributed to utilizing the effective graph-based representation, the boundary connectivity cues, and local connectivity computation.



(a)



(b)

Fig. 4: Precision-recall curves obtained by applying the proposed saliency detection method and the other methods for images (a) DUT-OMRON and (b) HKU-IS datasets.

TABLE I: F_β and MAE values obtained by applying the proposed method and the other methods on images in the two datasets.

Dataset	**DUT-OMRON**		**HKU-IS**	
Method	F_β	**MAE**	F_β	**MAE**
Proposed	**0.6064**	**0.1202**	**0.7306**	**0.1151**
MST	0.5754	0.1434	0.7117	0.1216
MBD+	0.5575	0.1448	0.6957	0.1235
RC	0.5400	0.1462	0.6933	0.1233
SO	0.5619	0.1345	0.7046	0.1168
MR	0.5670	0.1314	0.4733	0.1537
SF	0.4733	0.1537	0.5697	0.1618

REFERENCES

[1] B. Ko and J.-Y. Nam, "Object-of-interest image segmentation based on human attention and semantic region clustering," *Journal of the Optical Society of America A,* 2006, vol. 23, no. 10, pp. 2462-2470.

[2] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing,* 2004, vol. 13, no. 10, pp. 1304-1318.

[3] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," *Proc. ACM International Conference on Multimedia,* 2003, pp. 374-381.

[4] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on graphics (TOG),* 2007, vol. 26, no. 3.

[5] G. Zhang, M. Cheng, Sh. Hu and R. Martin, "A shape-preserving approach to image resizing," *Computer Graphics Forum,* 2009, vol. 28, no. 7.

[6] Y. Yuan, J. Wang, B. Li and M.-H Meng, "Saliency based ulcer detection for wireless capsule endoscopy diagnosis," *IEEE Transactions on Medical Imaging,* 2015, vol. 34, no. 10, pp. 2046-2057.

[7] M. Rezaie Abkenar and M. O. Ahmad, "Quaternion-based salient region detection using scale space analysis," in *Proc. Signal Processing and Intelligent Systems Conference (SPIS),* pp. 78-82, 2015.

[8] M. Rezaie Abkenar and M. O. Ahmad, "Superpixel-based salient region detection using the wavelet transform," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 2719-2722, 2016.

[9] M. Rezaie Abkenar, H. Sadreazami, and M. O. Ahmad, "Patch-based salient region detection using statistical modeling in the non-subsampled contourlet domain," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 1-4, 2017.

[10] M. Rezaie Abkenar and M. O. Ahmad, "Salient region detection using efficient wavelet-based textural feature maps," *Multimedia Tools and Applications,* vol. 77, no. 13, pp. 16291-16317, 2018.

[11] M. Rezaie Abkenar, H. Sadreazami, and M. O. Ahmad, "Salient region detection using feature extraction in the non-subsampled contourlet domain," *IET Image Processing,*, 2018.

[12] H. Sadreazami, A. Mohammadi, A. Asif and K. N. Plataniotis, "Distributed graph-based statistical approach for intrusion detection in cyber-physical systems," *IEEE Transactions on Signals and Information Processing over Networks,* vol. 4, no. 1, pp. 137-147, 2018.

[13] H. Sadreazami, A. Asif and A. Mohammadi, "Iterative graph-based filtering for image abstraction and stylization," *IEEE Transactions on Circuits & Systems II: Express Briefs,* vol. 65, no. 2, pp. 251-255, 2018.

[14] H. Sadreazami, A. Asif and A. Mohamamdi, "Data-adaptive color image denoising and enhancement using graph-based filtering," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 1-4, 2017.

[15] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. European Conference on Computer Vision (ECCV),* pp. 29-42, 2012.

[16] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 2814-2821, 2014.

[17] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 3166-3173, 2013.

[18] J. Zhang, et. al. "Minimum barrier salient object detection at 80 fps," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1404-1412.

[19] W. C. Tu, S. He, Q. Yang and S. Y. Chien, "Real-time salient object detection with a minimum spanning tree," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2334-2342.

[20] R. Achanta, et al. "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, no. 11, pp. 2274-2282, 2012.

[21] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp 839-846, 1998.

[22] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5455-5463.

[23] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr and S. M. Hu "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2015, vol. 37, no. 3, pp. 569-582.

[24] F. Perazzi, P. Krahenbuhl, Y. Pritch and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 733-740, 2012.

[25] R. Achanta, S. S. Hemami, F. J. Estrada and S. Susstrunk, "Frequency tuned salient region detection," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2009, pp. 1597-1604.