Salient Region Detection using Patch Level and Region Level Image Abstractions

Rajkumar Kannan, Gheorghita Ghinea, and Sridhar Swaminathan

***Abstract*—In this letter, a novel salient region detection approach is proposed. Firstly, color contrast cue and color distribution cue are computed by exploiting patch level and region level image abstractions in a unified way, where these two cues are fused to compute an initial saliency map. A simple and computationally efficient adaptive saliency refinement approach is applied to suppress saliency of background noises, and to emphasize saliency of objects uniformly. Finally, the saliency map is computed by integrating the refined saliency map with center prior map. In order to compensate different needs in speed/accuracy tradeoff, three variants of the proposed approach are also presented in this letter. The experimental results on a large image dataset show that the proposed approach achieve the best performance over several state-of-the-art approaches.**

***Index Terms*—saliency detection, color contrast, color distribution, center prior, adaptive saliency refinement.**

I. Introduction

Detecting salient regions in images is an interesting and difficult multidisciplinary problem. The field has considerable attention in the recent years, and has become an active area of research in Computer Vision due to its various applications in object detection, object recognition, adaptive image and video compression, and image retargeting. Many computational saliency detection models have been proposed over the years, which can be roughly categorized into *bottom-up* and *top-down* approaches [1].

Bottom-up saliency is data-driven and is often estimated using *color contrast* cue [1]-[5], since salient objects always pose high contrast from the background. Recent contrast based approaches estimate saliency of an image element by computing its contrast with respect to rest of the image elements in a global manner. Most recent approaches [1], [3], [5] compute saliency by estimating color contrast cue along with another important saliency cue called *color distribution*. Since color components of a salient object are always spatially compact rather than widely spread around the image, lower spatial distribution of a color component indicates its higher spatial saliency. Apart from these two cues, another widely used cue is *center prior* [1]. The center prior gives more weight to regions that are near to image center, since salient objects are placed near the image center most of the time.

Most existing salient region detection approaches operate on either *patch level* [1], [5] or *region level* [3], [6], [7] image abstractions. A major problem with patch level approaches is that, they often fail to suppress saliencies of textured background noises, and to highlight the saliency of objects uniformly. This issue can be solved by using region level image abstractions for saliency detection. Since these methods compute and assign saliency at region-level, imprecise region segmentation leads to degraded performance. Applications of saliency detection such as image thumbnail generation, object extraction and image retargeting do not need pixel accurate saliency maps, but require high speed saliency estimation.

In order to solve the aforementioned issues, this letter proposes a novel salient region detection approach. The major contributions of this paper are: 1) Salient region detection is achieved by exploiting both patch and region abstractions in a novel and unified way. 2) The proposed region abstraction approach makes the proposed saliency detection approach robust to different region segmentation methods and different numbers of regions. 3) The proposed computationally feasible saliency refinement approach effectively removes the background noises and highlights the salient objects uniformly. 4) The faster variants of the proposed approach also present fast and robust saliency detection performance.

II. Proposed Approach

Firstly, an image is abstracted at patch level, where the patch abstractions are used for region level image abstraction. Both patch and region abstractions are further used for estimation color contrast and color distribution cues. These two cues are fused to compute initial saliency of an image, which is further refined, and integrated with center prior map to generate the final saliency map. Fig. 1 depicts the main phases involved in the proposed approach.

A. Patch Level Image Abstraction

The given image I is segmented into homogenous patches using SLIC superpixel segmentation [8]. The number of superpixels N is set to 500. Each superpixel s_i is represented by a mean color sc_i (in CIELab) and a spatial position sp_i (x and y coordinates). Since SLIC suffers from slow computation speed, a faster patch level abstraction is achieved by uniformly segmenting the image into non-overlapping square patches of size $w \times w$, where w is set to 15.

B. Region Level Image Abstraction

The segmented superpixels are grouped into regions using spectral clustering [9]. Let $G = \{V, E\}$ be a weighted undirected graph, having nodes $V = \{s_1, s_2, s_3 \ldots s_n\}$ denote to the set of superpixels in an image, where the edges E represent the set of links that connect adjacent superpixels. An $N \times N$ affinity matrix A is constructed for G, where each element a_{ij}

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. This includes Matlab implementation of the proposed approach. This material is 228 KB in size.

Rajkumar Kannan is with the College of Computer Sciences and Information Technology, King Faisal University, Al Ahsa 31982, Kingdom of Saudi Arabia, Tel: +966-13-5899273 (e-mail: rkaruppan@kfu.edu.sa).

Gheorghita Ghinea is with the Department of Computer Science, Brunel University, Uxbridge, UB8 3PH, United Kingdom (e-mail: george.ghinea@brunel.ac.uk).

Sridhar Swaminathan is with the Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirappalli, India (email: sridarah@gmail.com).

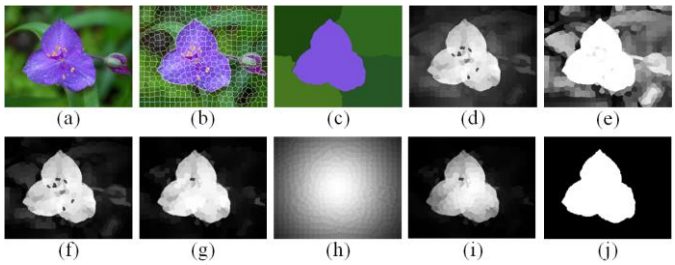

Fig. 1. Main phases of the proposed approach. (a) source image. (b) patch level abstraction. (c) region level abstraction. (d) color contrast cue. (e) color distribution cue. (f) fused saliency. (g) refined saliency. (h) center prior map. (i) final saliency map. (j) ground truth.

denotes the similarity between adjacent superpixels s_i and s_j, calculated as:

$$a_{ij} = \exp\left(-d(sc_i, sc_j) / 2\sigma_1^2\right) \tag{1}$$

where $d(sc_i,sc_j)$ is the Euclidean distance between colors of the superpixels which is normalized to [0,1] using min-max normalization. The scaling parameter σ_1 is set to 0.4. Then, a spectral clustering algorithm [9] is applied to cluster the graph G into M clusters. Here, Eigen-gap heuristic [9] is used to automatically determine M, which is still restricted to be within a specific range $[M_{min}, M_{max}]$, where they are set to 5 and 10 correspondingly. Each region r_j is represented using a prototype that comprises of a dominant color rc_j and a spatial position rp_j. Averaging superpixels' colors of a region is prone to region segmentation errors. Here, the region prototyping is formulated as a multivariate feature mediation problem. So, geometrical mediation is used to determine the dominant color rc_j of a region r_j which is defined as:

$$rc_j = \arg\min_{rc_j \in sc} \sum_{i=1}^{|r_j|} d(sc_i, rc_j) \tag{2}$$

where $|r_j|$ denotes the number of superpixels in region r_j and sc is the set of colors of superpixels in r_j. Equation (2) finds a superpixel that has the minimum color distance from the rest of the superpixels in r_j, and sets its color as dominant color rc_j. The spatial position rp_j (i.e. geographical midpoint) is also determined in the same manner as a center of minimum distance:

$$rp_j = \arg\min_{rp_j \in sp} \sum_{i=1}^{|r_j|} d(sp_i, rp_j) \tag{3}$$

where sp is a set of spatial positions of superpixels in r_j. Since the mediation based prototyping is robust to region segmentation errors, comparatively faster region segmentation can be achieved by uniformly segmenting image into $z \times z$ rectangle regions, where z is set to 2. The superpixels or the square patches that fall into a rectangle area are considered to belong to that region.

C. Color Contrast Estimation

The color contrast of a patch s_i is measured by computing the spatially weighted color contrast to all regions of the image except the region it belongs, which is formulated as:

$$con(s_i) = \sum_{j \neq i} \frac{|r_j|}{N} \cdot \exp\left(-d(sp_i, rp_j) \cdot \beta_1\right) \cdot d(sc_i, rc_j) \tag{4}$$

where $|r_j|$ is the number of superpixels in a region r_j, which is used to favour contrast to bigger regions to have more influence. The exponential function gives spatial weighting to the contrast measure, where contrast to the spatially near regions will be given more weight than the farther regions. The scaling parameter β_1 is empirically set to 2. The spatial distance $d(sp_i,rp_j)$ is normalized into [0,1] using the maximum dimension of the image. The function $d(sc_i,rc_j)$ return the color contrast of a patch to the region compared. Finally, the contrast cue $con(s_i)$ is normalized to a range [0,1] using min-max normalization.

D. Color Distribution Estimation

The color distribution of a patch is estimated by computing the spatial variance of its color [5]. Firstly, the weighted mean position of a superpixel's color sc_i is computed as:

$$msp_i = \frac{1}{M} \sum_{j \neq i} \exp\left(-d(sc_i, rc_j) \cdot \beta_2\right) \cdot rp_i \tag{5}$$

where the exponential function weights the position of each region based on its color similarity to s_i. The color distribution of the superpixel s_i is defined as:

$$cdis(s_i) = \sum_{j \neq i} \frac{|r_j|}{N} \cdot d(msp_i, rp_j) \cdot \exp\left(-d(sc_i, rc_j) \cdot \beta_2\right) \tag{6}$$

where $|r_j|$ is used to emphasize the color distribution of s_i comparing to bigger regions. Because, the higher similarity to bigger regions indicates wider distribution of a superpixel color sc_i. The spatial distance between a superpixel's mean color position and a region position $d(msp_i,rp_j)$ is normalized to a range [0,1] using the maximum dimension of the image. The exponential function returns the color similarity between superpixel s_i and a region r_j. The parameter β_2 is empirically set to 8. The color distribution of each superpixel is then normalized to range [0,1] using min-max normalization. The higher color distribution indicates that the color component is widely spread over the image, which is less likely to be the color of the salient object. So, the color distribution cue for saliency is defined as:

$$dis(s_i) = 1 - cdis(s_i) \tag{7}$$

The proposed color distribution cue estimation is similar to [5]. But the major difference here is [5] determines it only using patches.

E. Saliency Assignment and Adaptive Refinement

The saliency is computed by fusing two independent saliency cues using a simple multiplication defined as:

$$sal(s_i) = con(s_i) \cdot dis(s_i) \tag{8}$$

The spatial saliency $sal(s_i)$ is normalized to a range [0,1] using min-max normalization. There may be some noises in the fused saliency map due to small scale textured patterns in the background. Simply averaging the surrounding superpixels' saliencies [5] cannot preserve saliency near object boundaries. Since a salient object will be comprised of group of spatially connected salient superpixels, a superpixel surrounded by highly salient superpixels belongs to the salient object. Also, a superpixel surrounded by low salient superpixels belongs to the background. The refined saliency of a superpixel is defined as:

$$sal(s_i) = \begin{cases} \max\limits_{s_j \in ns} sal(s_j), & \text{if } \left(\frac{1}{|ns|} \sum\limits_{s_j \in ns} sal(s_j)\right) \geq 1 - \mu \\ \min\limits_{s_j \in ns} sal(s_j), & \text{if } \left(\frac{1}{|ns|} \sum\limits_{s_j \in ns} sal(s_j)\right) \leq \mu \\ sal(s_i), & \text{otherwise} \end{cases} \tag{9}$$

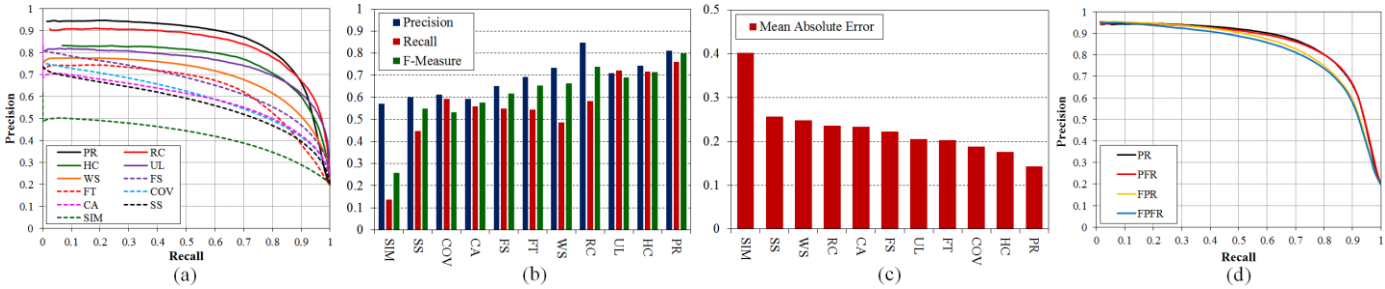

Fig. 2. Statistical comparison with 10 state-of-the-art methods on MSRA-1000 dataset. (a) Precision recall rates using fixed thresholding. (b) Precision, recall and f-measure values using adaptive thresholding. (c) Mean Absolute Errors of the different methods. (d) Precision-recall curves of variants of the proposed method.

Equation (9) first finds average saliency of the adjacent superpixels of s_i. If the average neighborhood saliency exceeds 1-μ, then it sets the maximum among neighborhood superpixels' saliencies as the saliency of s_i. If it is less than μ, then the minimum among the neighborhood saliencies is set as the saliency of s_i. The parameter *ns* denote the set of neighborhood superpixels of s_i, where $|ns|$ is the number of neighborhood superpixels. The parameters μ is set to 0.2 empirically. The average neighborhood saliency between 1-μ and μ denotes that superpixel s_i is adjacent to the boundary of salient object where the saliency of s_i remains the same after refinement. The adaptive saliency refinement method highlights the salient object uniformly, and detects the background efficiently by removing the textured noises from the background (Fig. 1(g)).

F. Incorporation of Center Prior

A widely used high-level prior called *center prior* [1] is incorporated into the saliency detection framework. The center prior gives more a weight to the regions that are nearer the image center than the regions near to the image boundaries. The center prior weight for a superpixel is defined as:

$$cen(s_i) = \exp\left(-d(sp_i, c)/2\sigma_2^2\right) \tag{10}$$

where $d(sp_i,c)$ is the Euclidean distance between a superpixel and the image center c. The parameter σ_2 is set to $\min(W,H)/2.5$, where W and H are the width and height of the image. The center prior cue integrated into the refined saliency cue using a simple multiplication defined as:

$$sal(s_i) = sal(s_i) \cdot cen(s_i) \tag{11}$$

The saliency $sal(s_i)$ is normalized into [0,1] using min-max normalization. The saliency values can be normalized to a range [0,255] to produce a grey scale saliency map.

III. Experimental Results

The experimental comparison is performed on the most widely used MSRA-1000 dataset [10] with pixel accurate ground truth annotations. The proposed **P**atch-**R**egion based saliency detection approach (**PR**) is compared with 10 state-of-the-art methods, RC[2], HC[2], CA[4], FT[10], COV[11], SS[12], WS[13], FS[14], SIM[15] and UL[16].

A. Quantitative Evaluation

Similar to [2], [10], performance of the proposed approach is evaluated using *precision recall* rate. Precision and recall rates are computed by comparing the binary saliency maps that are obtained using a number of fixed thresholds in [0,1,…,255] with the ground truth. These precision and recall values are averaged over all the images, which results in a precision-recall curve. Fig. 2(a) shows that the proposed method presents the best precision recall curve. The proposed method maintains more than 90% precision rate for higher thresholds. However precision decreases only after recall rate reaches 90%. This is due to the inclusion of some false positives in the binary maps for very lower thresholds.

Since precision recall analysis using fixed thresholding alone is not a sufficient measure for saliency evaluation, Similar to that in [2], [10], precision, recall and F-measure analysis using image dependent adaptive thresholding method is carried out. The adaptive threshold is defined as twice the mean saliency of the saliency map. The proposed method achieves the best performance in terms of recall and F-measure, while also maintaining fair precision (Fig. 2(b)). Even though the precision of the proposed method is slightly lower than that of RC[2], the proposed method outperforms RC[2] in terms of recall and F-measure. In many application of saliency detection, both high precision and high recall are always required, where the proposed method has a good balance between the three measures.

In order to evaluate the detection of the salient as well as non-salient pixels in an image, the *mean absolute error* (MAE) between the continuous saliency map and ground truth is measured as proposed in [5]. Fig. 2(c) depicts that the proposed method presents the smallest MAE. Fig. 2(d) shows the robust performance of the faster variants of the proposed approach. The variant method **PFR** uses superpixel based **P**atch segmentation and uniform sampling based **F**aster **R**egion segmentation, where **FPR** uses uniform sampling based **F**aster **P**atch segmentation and spectral clustering based **R**egion segmentation. The variant **FPFR** uses **F**aster **P**atch segmentation and **F**aster **R**egion segmentation.

Fig. 3(a) and 3(b) shows the performance of the proposed approach with different numbers of patches. Fig. 3(c) and 3(d) show the robustness of the proposed approach by varying the numbers of regions. Even though higher numbers of patches and regions result slightly improved performance, it also increases the overall computation time.

Fig. 4(a) and 4(b) depict the performance of the individual phases of the proposed approach, and the influence of the saliency refinement parameter μ respectively. Fig. 5 shows the visual comparison of the different methods. Despite the robust performance, the proposed approach sometimes fails to detect salient regions from highly cluttered background.

B. Computational Complexity and Running Time

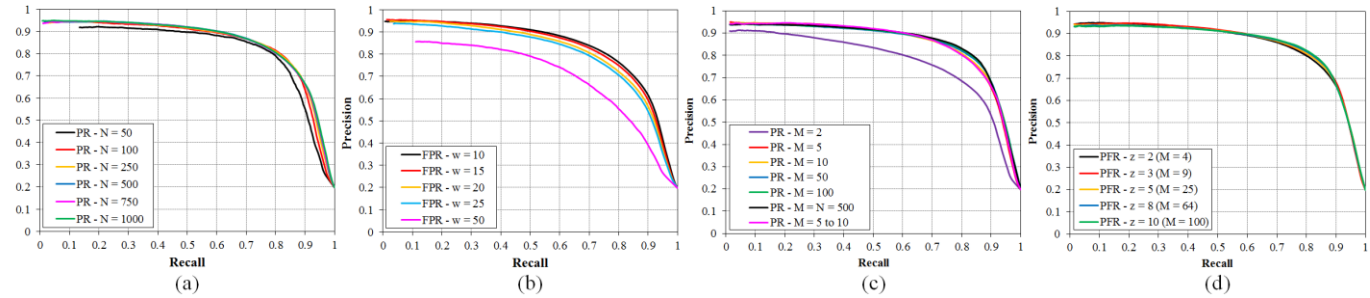

Fig. 3. Precision-recall curves for different parameter settings. (a) Proposed method with different numbers of superpixels N. (b) Proposed method with different values for w. (c) Proposed method with different numbers of regions M. (d) Proposed method with different values for z.

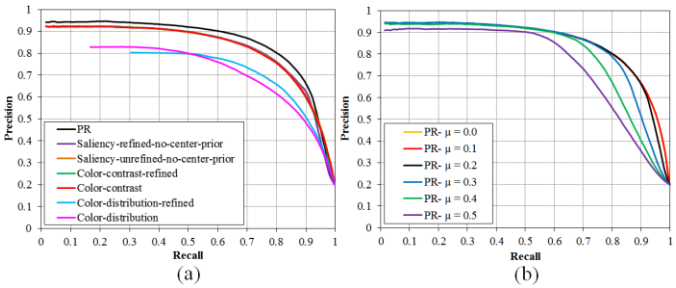

Fig. 4. (a) Precision-recall curves of individual phases of the proposed method. (b) Precision-recall curves of proposed method with different values for refinement parameter μ.

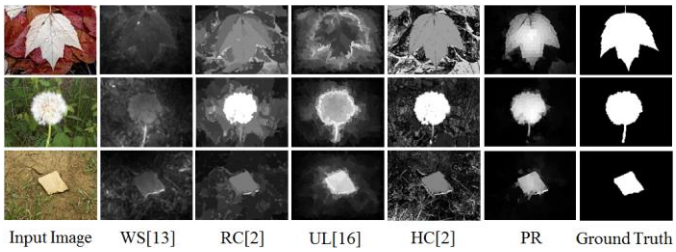

Fig. 5. Visual comparison of saliency maps of different methods.

In the proposed approach PR, approximate time complexity of superpixel segmentation, region abstraction and saliency estimation are $O(P)$, $O(N^3)$ and $O(NM)$ respectively, where P is the number of pixels in an image. So, the total time complexity of the proposed approach PR is $O(P+N^3+NM)$. The superpixel segmentation and region abstraction dominate the total time cost. Since M is a relatively smaller number, saliency estimation takes only less computation time.

Table 1 shows the average running times of methods on the MSRA-1000 dataset which are taken on a laptop with Intel i5 2.50 GHz CPU and 4 GB RAM. The proposed method is much faster than some of the previous methods. The proposed method PR takes 1.12s (44%), 0.89s (35%) and 0.54s (0.21%) for superpixel segmentation, region abstraction and saliency estimation respectively. Table 1 also shows that the FPFR presents the fastest performance among the variants of the proposed method.

TABLE I
AVERAGE RUNNING TIME MEASURED ON MSRA -1000 DATASET

Method	CA[4]	WS[13]	HC[2]	RC[2]	PR	PFR	FPR	FPFR
Time(s)	54.1	5.81	0.017	0.19	2.56	2.05	2.35	1.88
Code	Matlab	Matlab	C++	C++	Matlab	Matlab	Matlab	Matlab

IV. CONCLUSION

In this letter, a novel salient region detection approach based on patch level and region level image abstractions is presented. The proposed approach presents robust performance for different numbers of patches and regions. The adaptive refinement strategy greatly reduces noises in the saliency maps and emphasizes salient object uniformly. The experimental results have shown the potential performance of the proposed approach in comparison with 10 state-of-the-art methods. In addition, faster variants of the proposed approach were also presented for achieving high speed as well as robust saliency estimation. In future, other saliency cues such as *semantic prior*, *color prior*, and *background prior* will also be incorporated into the proposed approach.

REFERENCES

[1] K. Fu, C. Gong, J. Yang, Y. Zhou, and I. Yu-Hua Gu, "Superpixel based color contrast and color distribution driven salient object detection," *Signal Processing: Image Communication*, vol. 28, no. 10, pp. 1448-1463, 2013.

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

[3] 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. ICCV*, pp. 1529-1536, 2013.

[4] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE TPAMI*, vol. 34, no. 10, pp. 1915–1926, 2012.

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

[6] Z. Ren, S. Gao, L. T. Chia, and I. Tsang, "Region-based Saliency Detection and Its Application in Object Recognition," *IEEE TCSVT*, vol. 24, no. 5, pp. 769 – 779, 2014.

[7] J. G Yu, J. Zhao, J. Tian, and Y. Tan, "Maximal entropy random walk for region-based visual saliency." *IEEE Transaction on Cybernetics*, vol. 44, no. 9, pp.1661-1672, 2014.

[8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[9] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," in *Proc. NIPS*, pp. 849-856, 2002.

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

[11] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *Journal of Vision*, vol. 13, no. 4, article. 11, pp. 1-20, 2013.

[12] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE TPAMI*, vol. 34, no. 1, pp. 194-201, 2012.

[13] N. Imamoglu, W. Lin, and Y. Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Transactions on Multimedia*, vol. 15, no. 1, pp. 96-105, 2013.

[14] M. D. Levine, X. An, and H. He, "Saliency detection based on frequency and spatial domain analysis," in *Proc. BMVC*, pp. 86.1-86.11, 2011.

[15] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *Proc. CVPR*, pp. 433–440, 2011.

[16] P. Siva, C. Russell, T. Xiang, and L. Agapito, "Looking beyond the image: Unsupervised learning for object saliency and detection," in *Proc. CVPR*, pp. 3238-3245, 2013.