SALIENT OBJECT DETECTION VIA BACKGROUND CONTRAST

Quan Zhou[1,*], *Nianyi Li*[2,†], *Jianxin Chen*[1], *Shu Cai*[1], *and Longin Jan Latecki*[3]

[1]Key Lab of Ministry of Education for Broad Band Communication & Sensor Network Technology, Nanjing University of Posts & Telecommunications, Nanjing, P.R. China
[2]Department of Computer & Information Sciences, University of Delaware, Newark, USA
[3]Department of Computer & Information Sciences, Temple University, Philadelphia, USA

ABSTRACT

This paper addresses the problem of salient object detection. We introduce a novel framework which aims to automatically identify salient regions in natural images based on two key ideas. The first one is to consider the statistical spatial distribution of saliency and non-saliency regions as two complementary processes. The second one is based on the assumption that contrast saliency with respect to background regions outperforms those with respect to entire image. Experimental results demonstrate the effectiveness of our approach over 12 state-of-the-art models.

Index Terms— salient object detection, background contrast, spatial distribution

1. INTRODUCTION

Traditional image analysis processes often scan the image exhaustively, looking for familiar objects/regions of different location and size. This process can be implemented much more efficient if an attention mechanism assigns priorities to important image parts, leading to the challenging problem of *salient object detection* that is an important function for computer vision and image processing [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. According to whether the detection procedure requires human interaction or not, existing methods are divided into two categories: bottom-up (unsupervised) and top-down (supervised) approaches. The first category usually determines the saliency of a pixel based on low-level stimuli-driven features without any prior of the salient region or object [4, 5, 11]. On the contrary, the second one often describes the saliency by the visual knowledge constructed from the training process, and then use such knowledge for saliency detection on the test images [12, 13, 14].

Recent research has shown that the computational models based on bottom-up methods are quite successful and very suitable to extend to large scale datasets. Results from perceptual research [15, 16] and previous approaches [6, 7, 17, 18] indicate that the most influential factor in bottom-up visual saliency is *contrast*. The definition of contrast in previous works is mainly based on different types of image features, such as color variation, edges and gradients [4], spectral analysis [19], histograms [20], multi-scale descriptors [13], or combinations thereof [5, 21]. Both types of methods tend to rely solely on the local center-surround contrast [4, 13] or the global contrast [20, 22] with respect to the entire scene for estimating the saliency. However, we argue that the contrast based on background regions also plays an important role in this process. Furthermore, by exploring the statistical spatial distribution of salient and background regions, we found that the salient object regions tend to be more compact than the background regions.

In this paper, we introduce a bottom-up saliency detection framework using the contrast with respect to a series of background candidate regions. Fig. 1 outlines the proposed method, which differs significantly from previous methods in its motivation and methodology. The input image is first over-segmented into a series of homogeneous segments. Since the background regions often occupy large image areas, and al-

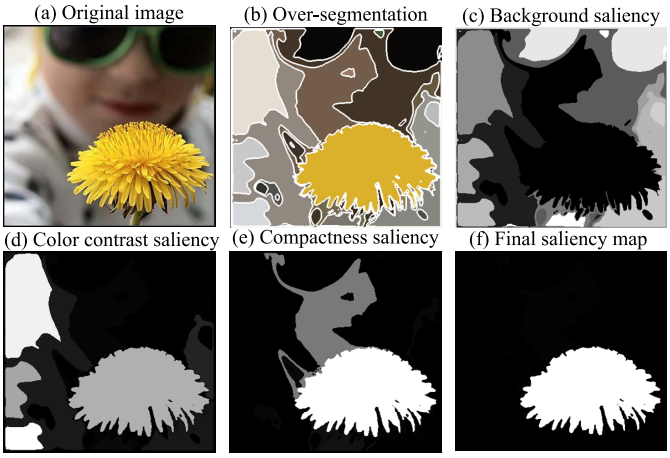

Fig. 1. Overview of the main steps of our method. White represents high saliency, while black indicates low saliency. (Best viewed in color)

*This work was supported by NSFC 61201165, 61271240, 61401228, 61401231, 61403350, PAPD, NY213067, and by NSF IIS-1302164.

†This author has equal contribution with first author.

ways stretch until the edge of the image, we generate the background saliency maps, as illustrated in Fig. 1(c), with regions far away from image center are assigned high background saliency value. In addition to the ability to encode background saliency using the spatial distribution of background regions, as will be seen, another advantage of our method is that it also adopts color contrast and compactness to encode foreground saliency. After a set of background candidate regions are abstracted based on background saliency map, the color contrast saliency, as shown in Fig. 1(d), is created by computing the contrast of one segments with respect to these background regions. The compactness saliency is produced according to gestalt laws, indicating the regions with regular shape tend to be more salient, otherwise not. We obtain the final saliency map by incorporating these three saliency maps, as shown in Fig. 1(f).

We evaluated our methods on publicly available benchmark dataset [13], and compared with 12 main-stream state-of-the-art saliency models [4, 12, 18, 19, 20, 22, 23, 24, 25, 26, 27] as well as with manually produced ground truth annotations. The experimental results demonstrate the effectiveness of our method for the task of salient object detection.

2. THE APPROACH

In this section, we first introduce image representation using over-segmentation technique, and then elaborate on the details of our saliency measurement.

2.1. Image Representation

As shown in Fig. 1(b), our first step is to form a set of segments from raw pixel intensities of original image. These segments correspond to small, homogeneous regions with accurate boundaries in the image, and have been found useful for salient object detection by other researchers [18, 20, 23].

The input images are first converted to the perceptually uniform CIELab color space, and then convolved with a 17-dimensional filter-bank, which consists of a series of Gaussian, derivatives of Gaussian and Laplacian of Gaussian filters. We employ the Euclidean-distance K-means clustering algorithm [28] to perform unsupervised clustering. Finally, each pixel is assigned to the nearest cluster center to generate an over-segmented image. Although those segmented regions tend to be highly irregular in size and shape, the advantage of this technique is that it can often group large homogeneous regions with similar appearance while dividing heterogeneous regions into many smaller ones.

2.2. Measuring Visual Saliency

Our method is based on three saliency operations, namely, background saliency, color contrast saliency, and compactness saliency. These three saliency maps are consolidated to predict final saliency. Note at each stage, maps are normalized before integration.

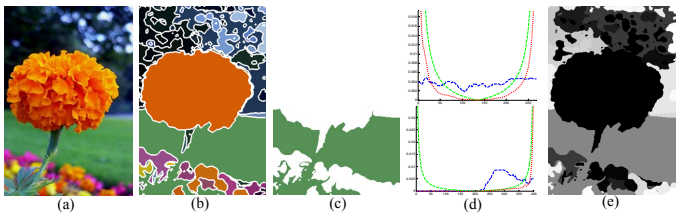

Fig. 2. Illustration of background saliency. (a) original image; (b) over-segmented image; (c) one background region; (d) spatial filtered distribution of the background region in horizontal (above) and vertical (below) direction, where the band suppression filter, spatial distribution, and filtered distribution are plotted as green, blue, and red dash line, respectively; (e) background saliency map. (Best viewed in color)

Background saliency (BS). Given an input image I with size $W \times H$ as illustrated in Fig. 2(a), where W and H are image width and height, respectively. We first establish the coordinate system with respect to up-left point of image as coordinate origin, and the horizontal and vertical directions denote x-axis and y-axis, respectively. Then a band suppression filter (BSF) $\mathbb{F}^h(x, W)$ in x-axis is defined as:

$$\mathbb{F}^h(x, W) = \tau \times \left(\frac{1}{\sqrt{1 + (\frac{x}{\eta})^2}} + \frac{1}{\sqrt{1 + (\frac{x-W}{\eta})^2}}\right) \quad (1)$$

where x is the x-coordinate. The parameter τ controls the upper bound and η controls the shape of the BSF, and $\mathbb{F}^v(y, H)$ in y-axis is similarly defined. For one specific region r_i, as shown in Fig. 2(c), which contains L pixels indexed by (x, y), we can compute the normalized spatial distribution $\mathbb{D}^B_{r_i}(x)$ and $\mathbb{D}^B_{r_i}(y)$ for r_i by counting the number of pixels with coordinate x and y in two axis, respectively. Then the background saliency is defined as the average weighted filtered responses among all pixels in r_i:

$$\mathbb{S}_{BS}(r_i) = \frac{1}{L}\Big(\sum_x \mathbb{D}^B_{r_i}(x) \times \mathbb{F}^h(x, W) + \sum_y \mathbb{D}^B_{r_i}(y) \times \mathbb{F}^v(y, H)\Big) \quad (2)$$

Thus, those large homogenous regions far away from image center will be assigned more saliency value than the center regions, as displayed in Fig. 2(e).

Color contrast saliency (CCS). It often happens that salient object may not perfectly locate in the center of image, while the color of object regions is still quite different with respect to the entire scene. Instead of computing saliency based on entire image [20], here we calculate the contrast based on background candidate regions. Thanks to the BS, we can choose the regions with relative higher saliency and lager size as robust background candidates.

Fig. 3. Illustration of color contrast saliency. Whether salient regions are homogenous or heterogeneous, CCS always assigns correct saliency that intuitively reflects human attention.

Let $\{B_1, B_2, \cdots, B_N\}$ be selected background candidate regions. Then we calculate the color contrast saliency (CCS) in RGB color space for segment r_i as:

$$\mathbb{S}_{CCS}(r_i) = \prod_{j=1}^{N} f_j, \quad f_j = \max\{d_j^R, d_j^G, d_j^B\} \tag{3}$$

where $d_j^R = (c^R(r_i) - c^R(B_j))^2$, $d_j^G = (c^G(r_i) - c^G(B_j))^2$, and $d_j^B = (c^B(r_i) - c^B(B_j))^2$. Here $c^R(\cdot)$, $c^G(\cdot)$, and $c^B(\cdot)$ denote the mean color in RGB channel, respectively. From Eqn. (3), if a region has great color difference in one channel with respect to background regions, the above product will get a large value leading to high contrast saliency for that region overall. Fig. 3 illustrates the results of calculating CCS.

Compactness saliency (CS). Intuitively, the background regions will be distributed over the entire image exhibiting a high spatial variance, whereas foreground objects are generally more compact with regular region shape [12, 13]. This local property allows us to define compactness saliency (CS), which renders image segments more saliency when they are grouped in a particular image region rather than evenly distributed over large image area. Instead of computing compactness in RGB feature space [18], we prefer to calculate it in location feature space:

$$\begin{aligned}\mathbb{S}_{CS}(r_i) = L(&\sum_x \mathbb{D}_{r_i}^F(x) \times \mathbb{N}^h(x, \mu_i, \sigma_i) + \\ &\sum_y \mathbb{D}_{r_i}^F(y) \times \mathbb{N}^v(y, \mu_i, \sigma_i))\end{aligned} \tag{4}$$

where L is the number of pixels belonging to r_i, $\mathbb{D}_{r_i}^F(x)$ and $\mathbb{D}_{r_i}^F(y)$ are, respectively, the pixel number distribution of r_i in x-axis and y-axis, as the blue dash line shown in Fig. 4(d). $\mathbb{N}(\cdot)$ denotes the Gaussian kernel with parameter μ_i and σ_i, where μ_i is the center location of minimum bounding box containing r_i, and σ_i is set as $\min\{W, H\}$.

Combined saliency. We assume that the three measurements are independent, and start by normalizing BS, CCS, and CS to the range $[0, 1]$. In practice, we found that BS to be of higher significance and discriminative power to represent backgrounds. Therefore, we use an exponential function in order to emphasize foreground saliency (FS) as:

$$\mathbb{S}_{FS}(r_i) = \exp\{-\alpha \cdot \mathbb{S}_{BS}(r_i)\} \tag{5}$$

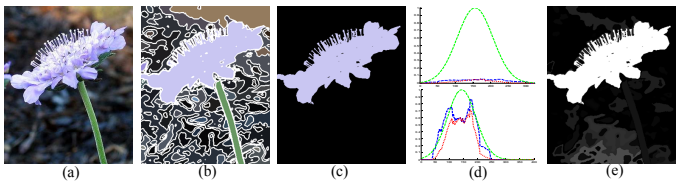

Fig. 4. Illustration of compactness saliency. (a) original image; (b) over-segmented image; (c) one specific foreground region; (d) spatial filtered distribution of the foreground region in horizontal (above) and vertical (below) direction, where the Gaussian filter, spatial distribution, and filtered distribution are plotted as green, blue, and red dash line, respectively; (e) Compactness saliency map. (Best viewed in color)

where α is the scaling factor.

In our experiments, using only the CCS and CS may highlight some background regions (as shown in Fig. 1) resulting in the incorrect saliency assignment. To remedy this shortcoming, we modify them by multiplying FS to eliminate the influence of backgrounds:

$$\begin{aligned}\mathbb{S}'_{CCS}(r_i) &= \mathbb{S}_{CCS}(r_i) \cdot \mathbb{S}_{FS}(r_i) \\ \mathbb{S}'_{CS}(r_i) &= \mathbb{S}_{CS}(r_i) \cdot \mathbb{S}_{FS}(r_i)\end{aligned} \tag{6}$$

The final saliency map is then defined as:

$$\mathbb{S}(r_i) = \mathbb{S}'_{CCS}(r_i) \cdot \mathbb{S}'_{CS}(r_i) \tag{7}$$

The saliency map $\mathbb{S}(r_i)$ is normalized to a fixed range $[0, 255]$, and each image pixel belonging to r_i is assigned a saliency value as $\mathbb{S}(r_i)$.

3. EXPERIMENTAL RESULTS

This section evaluate the effectiveness of our method.

Datasets. We test our proposed model on Microsoft Research Asia (MSRA) 1000 dataset [22], which contains 1000 images with resolution of approximate 400×300 or 300×400 pixels, and provides accurate object-contour-based ground truth.

Baselines. We selected 12 state-of-the-art models as baselines for comparison, including spectral residual saliency (SR [19]), spatiotemporal cues (LC [27]), attention measure (IT [4]), graph-based saliency (GB [25]), frequency-tuned saliency (FT [22]), saliency segmentation (AC [26]), context-aware saliency (CA [12]), global-contrast saliency (HC and RC [20]), saliency filter (SF [18]), low rank matrix recovery (LRMR [23]), and nature statistic saliency (SUN [24]).

Overall Results. The parameters are set as $K = 6$, $\alpha = 10$, $\tau = 0.6$ and $\eta = 0.19$, empirically, and we implemented all the 12 state-of-the-art models using a Dual Core 2.6 GHz machine with 4GB memory to generate saliency maps. The precision-recall curve (PRC) and F-measure[22] are illustrated in Fig. 5. It clearly shows that our method outperforms other approaches. It is interesting to note that the minimum

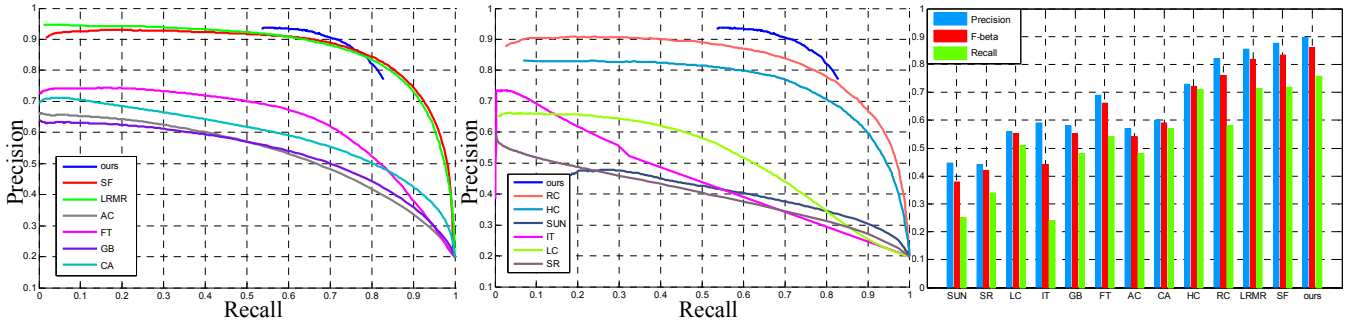

Fig. 5. Quantitative comparison for all methods with naive thresholding of saliency maps. Left and middle: PRC of our method compared with CA [12], AC [26], IT [4], LC [27], SR [19], GB [25], SF [18], LRMR [23], FT [22], SUN [24], HC and RC [20]. Right: Average precision, recall and F-measure with adaptive-thresholding segmentation. Our method shows high precision, recall, and F_β values over the MSRA 1000 dataset. (Best viewed in color)

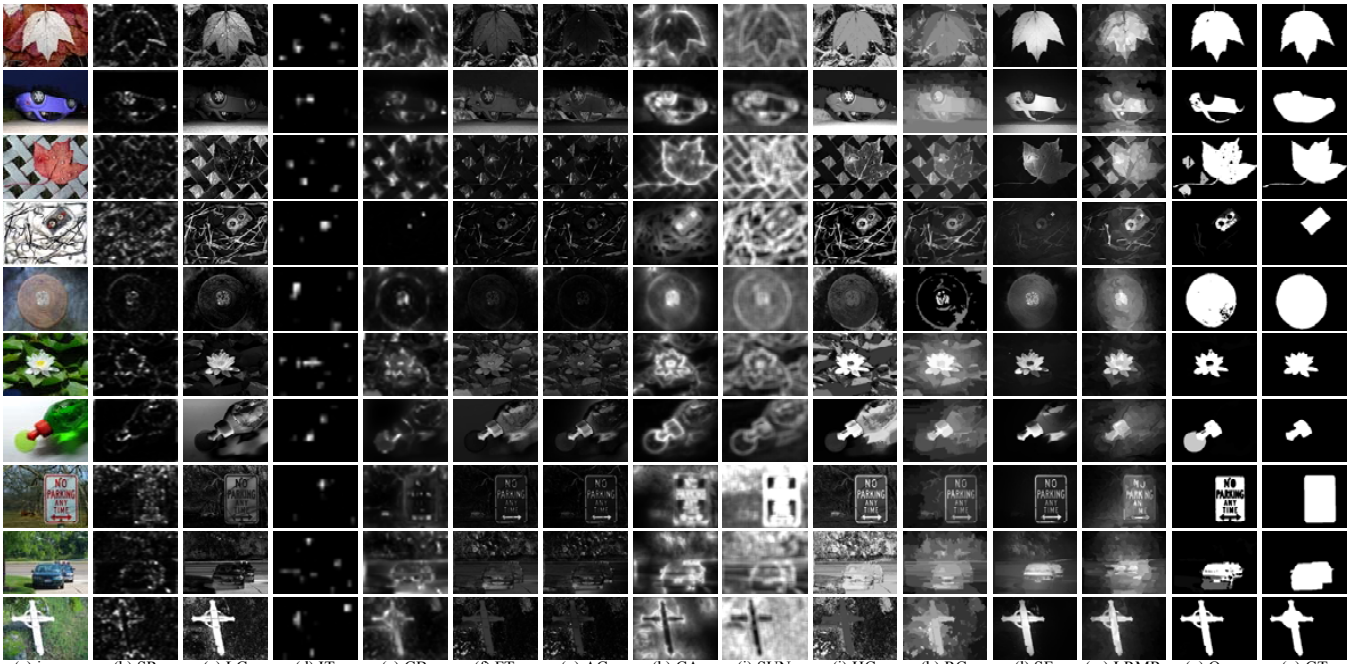

Fig. 6. Visual comparison of previous approaches with our method. See the legend of Fig. 5 for the references to all methods.

recall value of our methods starts from 0.53, and the corresponding precision is higher than those of the other methods. This probably because the saliency maps computed by our methods contain more pixels with the saliency value 255. Visual comparison with different methods is shown in Fig. 6. Our method generates uniformly highlighted salient regions, and produces saliency maps closest to the ground truth.

4. CONCLUSION AND FUTURE WORK

In this paper, we present a novel model to encode contrast with respect to background candidate regions for saliency detection. The key advantages of our method are: (1) candidate regions can be automatically produced according to the spatial distribution of backgrounds; (2) using the contrast with respect to backgrounds makes our method more robust than the methods computing contrast with respect to the entire image; (3) the compact measurement of region regularity plays an important role in improving the performance. Our method has been tested on MSRA 1000 dataset, and the results demonstrate the effectiveness of our approach on nature images.

There are two areas that we would like to improve upon. The first one is incorporating top-down priors to further improve the performance, as well as recent work [23] does. We are also interested in extending our model to predict region saliency in spatio-temporal domain (e.g., video sequence).

5. REFERENCES

[1] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu, "Saliency detection on light field," in *CVPR*, 2014, pp. 2806–2813.

[2] Oleg Muratov, Pamela Zontone, Giulia Boato, and Francesco GB De Natale, "A segment-based image saliency detection," in *ICASSP*, 2011, pp. 1217–1220.

[3] Ali Borji, Hamed R Tavakoli, Dicky N Sihite, and Laurent Itti, "Analysis of scores, datasets, and models in visual saliency prediction," in *CVPR*, 2013, pp. 921–928.

[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *TPAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.

[5] Z. Quan, C. Ji, R. Shiwei, Z. Yu, C. Jun, and Wenyu L., "On contrast combinations for visual saliency detection," in *ICIP*, 2013, pp. 2665–2669.

[6] Zhi Liu, Le Meur, and Shuhua Luo, "Superpixel-based saliency detection," in *WIAMIS*, 2013, pp. 1–4.

[7] Zhixiang Ren, Yiqun Hu, Liang-Tien Chia, and Deepu Rajan, "Improved saliency detection based on superpixel clustering and saliency propagation," in *ICMM*, 2010, pp. 1099–1102.

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

[9] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *CVPR*, 2013, pp. 1155–1162.

[10] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013, pp. 3166–3173.

[11] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *TPAMI*, vol. 28, no. 5, pp. 802–817, 2006.

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

[13] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.Y. Shum, "Learning to detect a salient object," *TPAMI*, vol. 33, no. 2, pp. 353–367, 2011.

[14] J. Van De Weijer, T. Gevers, and A.D. Bagdanov, "Boosting color saliency in image feature detection," *TPAMI*, vol. 28, no. 1, pp. 150–156, 2006.

[15] W. Einhauser and P. Koenig, "Does luminance-contrast contribute to a saliency map for overt visual attention?," *European Journal of Neuroscience*, vol. 17, no. 5, pp. 1089–1097, 2003.

[16] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision research*, vol. 42, no. 1, pp. 107–124, 2002.

[17] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

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

[19] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *CVPR*, 2007, pp. 1–8.

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

[21] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *CVPR*, 2012, pp. 478–485.

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

[23] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *CVPR*, 2012, pp. 853–860.

[24] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, and G.W. Cottrell, "Sun: A bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, pp. 1–20, 2008.

[25] Harel J., Koch C., and Pernoa P., "Graph-based visual saliency," in *NIPS*, 2006, pp. 545–552.

[26] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," *Computer Vision Systems*, pp. 66–75, 2008.

[27] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACMMM*, 2006, pp. 815–824.

[28] C. Elkan, "Using the triangle inequality to accelerate k-means," in *ICML*, 2003, pp. 147–153.