ON CONTRAST COMBINATIONS FOR VISUAL SALIENCY DETECTION

Quan Zhou, Ji Chen, Shiwei Ren, Yu Zhou, Jun Chen, Wenyu Liu

Department of Electronics and Information Engineering
Huazhong University of Science and Technology, Wuhan, China PR

ABSTRACT

Saliency detection is an important task in computer vision and image processing. The most influential factor in bottom-up visual saliency is contrast operation. In this paper, we propose a unified model to combine widely used contrast measurements, namely, center-surround, corner-surround and global contrast to detect visual saliency. The proposed model benefits from the advantages of each individual contrast operation, and thus produces more robust and accurate saliency maps. Extensive experimental results on natural images show the effectiveness of the proposed model for visual saliency detection task, and demonstrate the combination is superior than individual subcomponent.

***Index Terms*—** Saliency detection, human attention system, contrast combination, visual saliency

1. INTRODUCTION

The human visual system (HVS) is able to quickly detect the most interesting regions in a given scene. Computational modeling of this system suffices various applications in computer vision and image processing, e.g., object recognition [1, 2], image segmentation [3, 4], image/video compression [5, 6], image matching [7, 8], image editing [9], image retrieval [10] and video tracking [11, 12, 13, 14]. For this reason, considerable effort has been devoted to detecting salient regions over the last few years. Existing saliency models could be categorized into two classes: top-down and bottom-up. Top-down methods employ high-level cues (e.g., faces and pedestrians) [15, 16, 17, 18, 19, 20], it is hardly generalized, however, as the high-level cues are not available in every image. To cope with this problem, various bottom-up approaches have been introduced, which mainly estimate visual saliency using the contrast operation [9, 21, 22, 23, 24]. As a pioneer work, Itti *et al.* [21] introduced a biologically inspired saliency model using center-surround contrast (CESC) based on simple low-level features (e.g., luminance, color, and orientation). Zhou *et al.* [23] expend CESC via corner-surround contrast (CSC) to get more accurate saliency maps. Visual saliency is formulated based on Information Maximization scheme following the principles of information theory [25, 26]. Some models measure saliency in the frequency domain. Hou and Zhang [27] proposed a method based on relating extracted spectral residual features of an image in the spectral domain to the spatial domain. Guo *et al.* [5] employ the Phase spectrum of the Quaternion Fourier Transform (PQFT) to achieve better saliency prediction in the spatio-temporal domain.

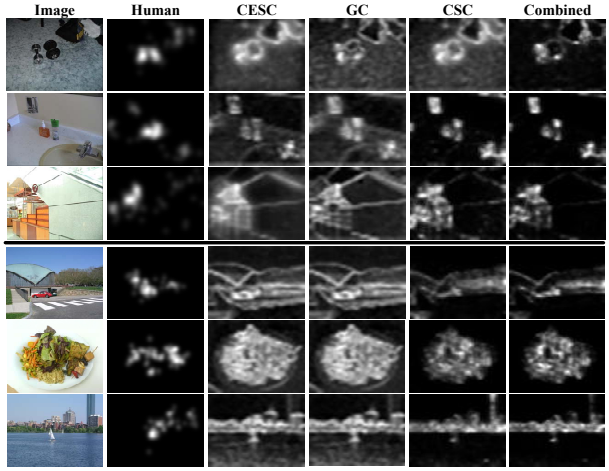

Fig. 1. **Illumination of our framework.** Saliency maps from TORONTO dataset (top) [26] and MIT dataset (bottom) [19]. (Best viewed in color)

The bottom-up models mentioned above further fall into three general categories: 1) models that calculate saliency by implementing local CESC (e.g., Itti *et al.* [21] and Gao *et al.* [22]), 2) models that predict saliency using local CSC (e.g., Zhou *et al.* [23]), and 3) models that find salient regions globally by measuring the rareness with respect to the entire image (e.g., Chen *et al.* [9] and Achanta *et al.* [24]). Our first contribution is to propose to incorporate all these contrast measurements into a unified model that benefits from the advantages of all the individual approaches, which thus far have been treated separately. Although the ideas of combining local and global contrast (GC) have been investigated in [9, 23, 28] for the task of salient object detection/segmentation approaches, but those have not yet been tested with human fixation prediction, which is the main goal of most models (including ours).

Most saliency models in literature utilize the color appear-

This work was supported by NSFC 61173120.

ance cues. Some have used RGB (e.g., [19, 21, 26, 29]) while others have employed CIELab (e.g., [30, 31, 32, 33]), or their combinations thereof [34]. We argue that employing just one RGB color system in our CSC always leads to successful outlier detection and significant performance improvement, as shown in Fig. 1. Hence, a yet unexplored strategy, which is our second contribution, is combining saliency maps from CSC using RGB color space.

We compare accuracy of our model with the mainstream models over two benchmark eye tracking datasets. These are top-ranked models that previous studies have shown to be significantly predictive of eye fixations in free viewing of natural scenes. In addition, we also analyze the contribution of the subcomponents of our models.

2. PROPOSED SALIENCY MODEL

Our framework is based on three saliency operations. The first one, CESC, considers the rarity of image patches with respect to their surrounding neighborhoods. The second one, CSC, extends CESC by considering the relative location between center patch and its surroundings. The third operation, GC, evaluates saliency of an image patch using its contrast over the entire image. Finally, these three contrast maps are consolidated. We first introduce image representation by projecting image patches to the space of a dictionary of image basis learned from a repository of natural scenes.

2.1. Image representation

It is well known that natural images can be sparsely represented by a set of localized and oriented filters. We thus employ sparse coding technique to represent images, which has been demonstrated as an effective tool for saliency detection task [25, 26, 29, 34].

The input image is first resized to $2^9 \times 2^9$ pixels. Let $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_N\}$ represent a series of N image patches from the top-left to bottom-right of image with no overlap. Then the reconstructive coefficients $\boldsymbol{\alpha}_i$ are calculated to represent patch $\mathbf{p}_i$ using the sparse coding algorithms [35]:

$$\boldsymbol{\alpha}_i^*(\mathbf{p}_i, \mathbf{D}) = \arg\min_{\boldsymbol{\alpha}_i \in \mathbb{R}^n} \frac{1}{2}||\mathbf{p}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda||\boldsymbol{\alpha}_i||_1 \quad (1)$$

where $||\cdot||_1$ denotes the ℓ_1-norm and λ is a regularization parameter. $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_n] \in \mathbb{R}^{m \times n}$ is a set of n m-dimensional basis function. Thus, $\mathbf{p}_i \sim \mathbf{p}_i' = \mathbf{D}\boldsymbol{\alpha}_i^*$, where $\mathbf{p}_i'$ is the estimation of $\mathbf{p}_i$. To learn the dictionary $\mathbf{D}$, considering a training set of q data samples $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_q] \in \mathbb{R}^{m \times q}$, an empirical cost function $g_q(\mathbf{D}) = \frac{1}{q}\sum_{i=1}^{q} l_u(\mathbf{y}_i, \mathbf{D})$ is minimized, where $l_u(\mathbf{y}_i, \mathbf{D})$ is

$$l_u(\mathbf{y}_i, \mathbf{D}) = \min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \frac{1}{2}||\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1 \quad (2)$$

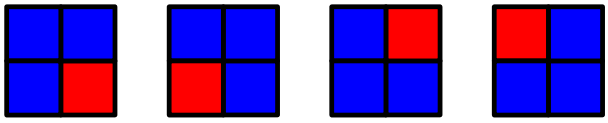

Fig. 2. **Illumination of four types of CSC.** From left to right are bottom-right, bottom-left, up-right, and up-left CSC, respectively. Each cell represents a 8×8 patch. Red cell denotes center patch and blue cells are surrounding patches.

In our implementation, we extracted 500000 8×8 image patches (for each sub channel of RGB color space) from 1500 randomly selected color images from natural scenes. Each basis function in the dictionary is a $8 \times 8 = 64D$ vector, and we learn $n = 200$ dictionary functions. The sparse coding coefficients $\boldsymbol{\alpha}_i$ are computed with the above learned basis using the LARS algorithm [36] implemented in the SPAMS toolbox [1].

2.2. Measuring visual saliency

In this section, we elaborate the details of our three contrast operations and their combinations thereof.

CESC saliency. CESC saliency $S_{ce}^c(\mathbf{p}_i)$ in our model is the average weighted dissimilarity between a center patch $\mathbf{p}_i$ and its surrounding L patches in a rectangular neighborhood:

$$S_{ce}^c(\mathbf{p}_i) = \frac{1}{L}\sum_{j=1}^{L} W_{ij}^{-1} D_{ij} \quad (3)$$

where W_{ij} is the the Euclidean distance between the location of center patch $\mathbf{p}_i$ and the surround patch $\mathbf{p}_j$. Thus, those patches further away from the center patch will have less influence on the saliency of the center patch. D_{ij} denotes the Euclidean distance between $\mathbf{p}_i$ and $\mathbf{p}_j$ in the feature space between $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_j$, vectors of coefficients for $\mathbf{p}_i$ and $\mathbf{p}_j$, respectively, derived from sparse coding algorithm. Superscript c denotes color sub channels in RGB color space.

CSC saliency. It often happens that CESC may assign high saliency value to background leading to incorrect detections. In order to overcome this shortcoming, we resort to CSC [23] to estimate visual saliency not only investigating the appearance difference between center patch and the surrounding neighborhoods but also considering their relative location. As a result, four types of CSC, namely, bottom-right, bottom-left, top-right, and top-left templates, are defined, as shown in Fig. 2. Let $S_{br}^c(\mathbf{p}_i)$, $S_{bl}^c(\mathbf{p}_i)$, $S_{tr}^c(\mathbf{p}_i)$ and $S_{tl}^c(\mathbf{p}_i)$ denote these four types of CSC, respectively, then the CSC saliency $S_c^c(\mathbf{p}_i)$ of patch $\mathbf{p}_i$ is calculated as:

$$S_c^c(\mathbf{p}_i) = S_{br}^c(\mathbf{p}_i) \times S_{bl}^c(\mathbf{p}_i) \times S_{tr}^c(\mathbf{p}_i) \times S_{tl}^c(\mathbf{p}_i) \quad (4)$$

For one specific CSC (e.g., bottom-right), we calculate the saliency in terms of the χ^2 distance from patch $\mathbf{p}_i$ (denote as

[1] http://www.di.ens.fr/willow/SPAMS/index.html

red cell in the first panel of Fig. 2) to its surrounding region $\mathbf{s}_i$ (denote as blue cells in the first panel of Fig. 2):

$$S^c_{br}(\mathbf{p}_i) \propto \chi^2(\mathcal{H}(\mathbf{p}_i), \mathcal{H}(\mathbf{s}_i))$$
$$\chi^2(\mathcal{H}(\mathbf{p}_i), \mathcal{H}(\mathbf{s}_i)) = \frac{1}{2}\sum_{b=1}^{B}\frac{(\mathcal{H}^b(\mathbf{p}_i) - \mathcal{H}^b(\mathbf{s}_i))^2}{\mathcal{H}^b(\mathbf{p}_i) + \mathcal{H}^b(\mathbf{s}_i)} \quad (5)$$

where $\mathcal{H}(\cdot)$ is the binned histogram ($B = 100$ bins here), and calculated from all of the patches based on the corresponding coefficients $\boldsymbol{\alpha}$, and $\mathcal{H}^b(\cdot)$ is the b^{th} element in $\mathcal{H}(\cdot)$. The same operation then applies to other three types of CSC.

GC saliency. Sometimes the appearance cues of local patch are similar to its neighbors but globally rareness with respect to the entire scene. Using only the local saliency may suppress areas within a homogeneous region resulting in blank holes. To remedy such drawback, we build our global saliency operator $S^c_g(\mathbf{p}_i)$ guided by the information-theoretic saliency measurement [26]. Instead of each pixel, here we calculate the probability of each patch $p(\mathbf{p}_i)$ over the entire scene and use its inverse as the global saliency:

$$S^c_g(\mathbf{p}_i) = p(\mathbf{p}_i)^{-1} = (\prod_{j=1}^{n} p(\alpha_{ij}))^{-1}$$
$$\log(S^c_g(\mathbf{p}_i)) = -\log p(\mathbf{p}_i) = -\sum_{j=1}^{n}\log(p(\alpha_{ij})) \quad (6)$$
$$S^c_g(\mathbf{p}_i) \propto -\sum_{j=1}^{n}\log(p(\alpha_{ij}))$$

The GC assumes that coefficients $\boldsymbol{\alpha}$ are conditionally independent from each other, which is to some extent guaranteed by the sparse coding algorithm [35]. For each coefficient of the patch representation vector (i.e., α_{ij}), first a binned histogram (also 100 bins) is calculated from all of the patches in the scene and is then converted to a probability density function ($p(\alpha_{ij})$) by dividing to its sum. If a patch is rare in one of the features, the above product will get a small value leading to high global saliency for that patch overall.

Saliency combination. For each contrast operation defined in Eqn. (3), Eqn. (4) and Eqn. (6), saliency values of patch $\mathbf{p}_i$ are assigned to the contained pixel $\mathbf{x}$, then the saliency map $S_*(\mathbf{x})$ is normalized and summed among all color channel in RGB color space:

$$S_*(\mathbf{x}) = \sum_{c\in R,G,B}\mathcal{N}(S^c_*(\mathbf{x})) \quad (7)$$

where $*$ denotes CESC, CSC and GC, respectively.

Since objects appear at different sizes, it is required to perform saliency detection at several spatial scales. To make our approach multi-scale, we calculate the saliency of images downsampled from the original image and then take the $\max$ operation after normalization:

$$S_*(\mathbf{x}) = \max_{m=1}^{M}\mathcal{N}(S^m_*(\mathbf{x})) \quad (8)$$

where $S^m_*(\mathbf{x})$ is the m^{th} scale saliency map ($M = 3$ here) resized from the result created by Eqn. (7). Then, three contrast saliency maps are normalized again and combined:

$$S(\mathbf{x}) = \mathcal{N}(S_{se}(\mathbf{x})) \circ \mathcal{N}(S_g(\mathbf{x})) \circ \mathcal{N}(S_c(\mathbf{x})) \quad (9)$$

where $\circ$ is an integration operation (i.e., $+$, $*$, $\max$, or $\min$). Through the experiments, we found that "$\max$" for first "$\circ$" and "$+$" for second "$\circ$" in this stage leads to the best performance. Finally, we smooth the resultant map by convolving it with a small Gaussian kernel for better visualization.

Normalization ($\mathcal{N}$). We first get the maximum and minimum value (denoted as $S_{\max}(\mathbf{x})$ and $S_{\min}(\mathbf{x})$, respectively) of saliency map $S(\mathbf{x})$, then $S(\mathbf{x})$ is normalized as:

$$S(\mathbf{x}) = \frac{S(\mathbf{x}) - S_{\min}(\mathbf{x})}{S_{\max}(\mathbf{x}) - S_{\min}(\mathbf{x})} \quad (10)$$

3. EXPERIMENTAL EVALUATION

To validate the effectiveness of our method, we conducted several experiments on two eye fixed benchmark datasets.

Evaluation metric. In our experiment, we adopt the widely used **shuffled AUC** [29] as evaluation metric. In shuffled AUC, human fixations for an image are considered as the positive set and other human fixations are used as the negative set. The saliency map is then treated as a binary classifier to separate the positive samples from the negatives. By thresholding over the saliency map and plotting true positive rate vs. false positive rate, an ROC curve is achieved and its underneath area is calculated as shuffled AUC value.

Datasets. We test our proposed model on two eye fixed datasets: (1) The **TORONTO** [26] dataset is the most widely used for model comparison. It contains 120 color images with resolution of 511×681 pixels from indoor and outdoor environments. Images are presented at random to 20 subjects for 3 seconds with 2 seconds of gray mask in between. (2) The **MIT** [19] dataset is the largest dataset containing 1003 images with resolution from 405×1024 to 1024×1024 pixels collected from Flicker and LabelMe datasets. There are 779 landscape and 224 portrait images. Fifteen subjects freely viewed images for 3 seconds with 1 seconds delay in between.

Overall results. Tab 1 reports the comparison results of our method with 9 state-of-the-art models in terms of shuffled AUC, and the contribution of each individual contrast component. It demonstrates our combined saliency model (CESC + CSC + GC) outperforms other models over two datasets. Our each individual saliency operator has less accuracy than the combined model but is still above several models (e.g., GB [37], IT [21], and SP [38]). Results show that GC saliency works better than CESC and CSC saliency over large datasets (MIT) while they are close to each other over TORONTO dataset. Among compared models, IC [25], AIM [26], and SR [27] performed higher than the rest. GB [37], IT [21], and SP [38] models are ranked at the bottom.

Table 1. **Quantitative comparison on two datasets.** Parameter settings: Histogram bin number $B = 100$; size of the surrounding neighborhood $L = 8$; scale number $M = 3$. Accuracies of the best model over each dataset are shown in bold font.

Dataset	AIM [26]	GB [37]	SR [27]	IC [25]	IT [21]	SD [39]	SUN [29]	SP [38]	LG [34]	CESC S_{ce}	CSC S_c	GC S_g	Combined S
TORONTO [26]	0.67	0.647	0.685	0.691	0.61	0.687	0.66	0.605	0.696	0.691	0.693	0.69	**0.738**
MIT [19]	0.664	0.637	0.65	0.666	0.61	0.646	0.649	0.642	0.678	0.653	0.668	0.676	**0.702**

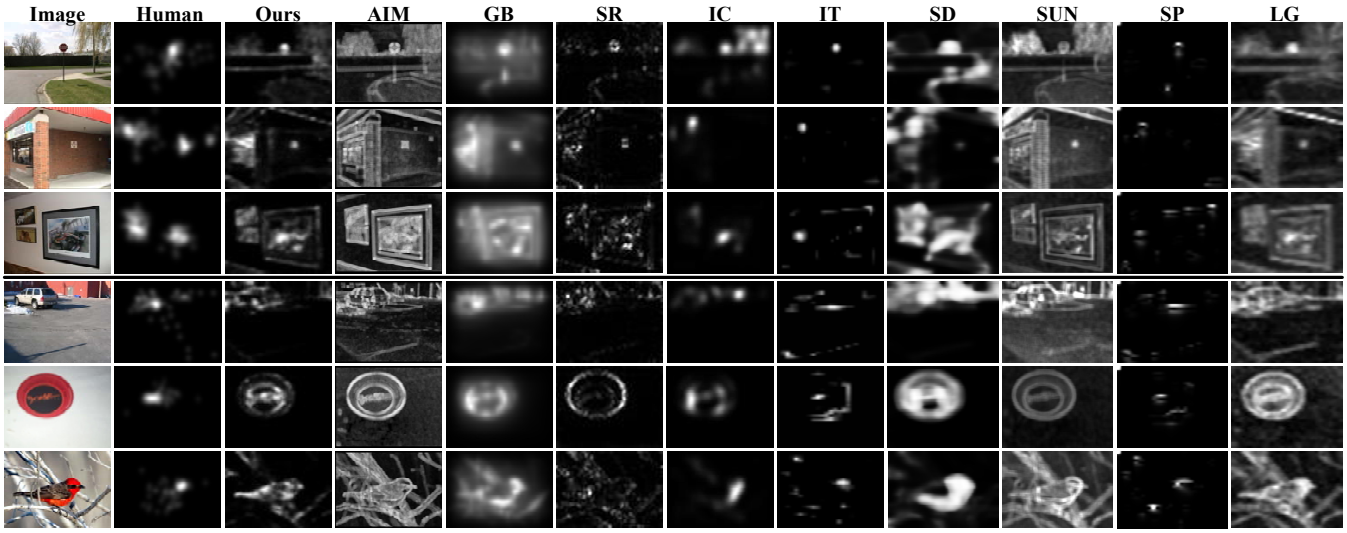

Fig. 3. **Some examples for visual comparison of previous models with our method from TORONTO (top) and MIT (bottom) datasets.** (Best viewed in color)

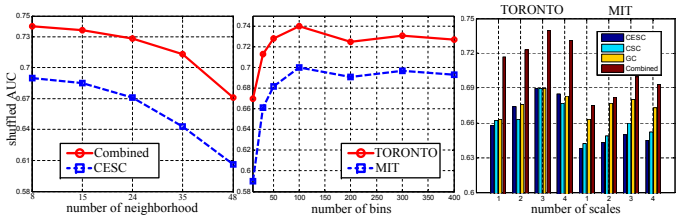

Fig. 4. **Parameter analysis.** Left: Effect of the number of surround neighborhoods over TORONTO dataset [26] (B =100 and $M = 3$). Middle: Effect of the bin number over two datasets ($M = 3$ and $L = 8$). Right: Influence of scale over two datasets ($B = 100$ and $L = 8$). (Best viewed in color)

In Fig.3, we exhibit the saliency maps of our combined saliency model and compared models for some sample images from TORONTO and MIT datasets. It demonstrates that our model is effective to exclude outlier backgrounds.

Parameter analysis. We also analyze how the size of surround neighborhoods, the bin number of histograms and number of spatial scales affect performance of our models. As the left diagram of Fig. 4 shows, increasing the number of neighbors reduces the accuracy of CESC saliency operator. Correspondingly, this reduces the accuracy of the combined model. The middle panel of Fig. 4 shows the accuracy of our method is insensitive to changes when bin number = 100, and any refinement to this parameter will result in slightly decrease of performance. As illustrated in the right panel of Fig. 4, increasing the number of scales enhances the performance and peaks at 3 scales (with resolution of [512×512, 256×256, and 128×128]), then the result drops.

4. CONCLUSION AND FUTURE WORK

In this paper, we enhance the state-of-the-art in saliency modeling by proposing a unified framework that incorporates different contrast measurements. We introduce three saliency operators, namely CESC, CSC and GC, that each represents a class of previous models to some extent. We conclude that integration of these saliency operators works better than just using either one individually, which encourages more research in this direction. Extensive experiments well validate the effectiveness of our framework on nature images.

In the future, we would like to explore more color space (e.g., perceptual CIELab color space), and adaptive weighting for combination operator using learning scheme [40].

5. REFERENCES

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

[2] Dashan Gao, Sunhyoung Han, and Nuno Vasconcelos, "Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition," *TPAMI*, vol. 31, no. 6, pp. 989–1005, 2009.

[3] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk, "Salient region detection and segmentation," *CVS*, vol. 5008, no. 1, pp. 66–75, 2008.

[4] Y. Fu, J. Cheng, Z. Li, and H. Lu, "Saliency cuts: An automatic approach to object segmentation," in *ICPR*, 2008, pp. 1–4.

[5] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *TIP*, vol. 19, no. 1, pp. 185–198, 2010.

[6] Andrew P Bradley and Fred WM Stentiford, "Visual attention for region of interest coding in jpeg 2000," *JVCIR*, vol. 14, no. 3, pp. 232–250, 2003.

[7] Timor Kadir and Michael Brady, "Saliency, scale and image description," *IJCV*, vol. 45, no. 2, pp. 83–105, 2001.

[8] Alexander Toshev, Jianbo Shi, and Kostas Daniilidis, "Image matching via saliency region correspondences," in *CVPR*, 2007, pp. 1–8.

[9] 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.

[10] E. Loupias, N. Sebe, S. Bres, and J.M. Jolion, "Wavelet-based salient points for image retrieval," in *ICIP*, 2000, pp. 518–521.

[11] Vijay Mahadevan and Nuno Vasconcelos, "Saliency-based discriminant tracking," in *CVPR*, 2009, pp. 1007–1013.

[12] Yu Zhou, Xiang Bai, Wenyu Liu, and Longin Jan Latecki, "Fusion with diffusion for robust visual tracking," in *NIPS*, 2012, pp. 2987–2995.

[13] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *TPAMI*, vol. 32, no. 1, pp. 171–177, 2010.

[14] Jia Li, Yonghong Tian, Tiejun Huang, and Wen Gao, "Probabilistic multi-task learning for visual saliency estimation in video," *IJCV*, vol. 90, no. 2, pp. 150–165, 2010.

[15] Robert J Peters and Laurent Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," in *CVPR*, 2007, pp. 1–8.

[16] A. Torralba, "Modeling global scene factors in attention," *JOSAA*, vol. 20, no. 7, pp. 1407–1418, 2003.

[17] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal, "Context-aware saliency detection," *TPAMI*, vol. 34, no. 10, pp. 1915–1926, 2012.

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

[19] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *ICCV*, 2009, pp. 2106–2113.

[20] 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.

[21] 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.

[22] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," in *ICCV*, 2007, pp. 1–6.

[23] Q. Zhou, N.Y. Li, Y. Yang, P. Chen, and W.Y. Liu, "Corner-surround contrast for saliency detection," in *ICPR*, 2012, pp. 1423–1426.

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

[25] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *NIPS*, 2008, pp. 681–688.

[26] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *NIPS*, 2006, pp. 155–162.

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

[28] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *CVPR*, 2011, pp. 473–480.

[29] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, and G.W. Cottrell, "Sun: A bayesian framework for saliency using natural statistics," *JOV*, vol. 8, no. 7, pp. 1–17, 2008.

[30] 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.

[31] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *ECCV*, 2010, pp. 366–379.

[32] A. Garcia-Diaz, X. Fdez-Vidal, X. Pardo, and R. Dosil, "Decorrelation and distinctiveness provide with human-like saliency," in *ACIVS*, 2009, pp. 343–354.

[33] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," in *CVPR*, 2010, pp. 2368–2375.

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

[35] B.A. Olshausen et al., "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[36] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *JMLR*, vol. 11, pp. 19–60, 2010.

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

[38] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," in *NIPS*, 2005, pp. 547–554.

[39] H.J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *JOV*, vol. 9, no. 12, pp. 1254–1259, 2009.

[40] Laurent Itti and Christof Koch, "Feature combination strategies for saliency-based visual attention systems," *JEI*, vol. 10, no. 1, pp. 161–169, 2001.