Sparse-Distinctive Saliency Detection

Yongkang Luo, Peng Wang, *Member, IEEE*, Wenjun Zhu, and Hong Qiao, *Senior Member, IEEE*

***Abstract*—In this letter, we propose a novel saliency model for saliency detection, named sparse-distinctive (SD) saliency model. Different from the existing models that only consider sparsity or distinctness of image, the proposed model computes saliency based on sparsity and distinctness. The basic idea is that sparsity and distinctness contribute to saliency simultaneously and play different roles under different scenes. This sparse-distinctive saliency model is based on some key ideas introduced in this letter and supported by psychological evidence. Experimental results on public benchmark eye-tracking datasets show that considering the sparsity and distinctness for saliency can improve the accuracy of predicting human fixations, and the proposed model outperforms the mainstream models on predicting human fixations.**

***Index Terms*—Distinctness, saliency, sparsity, visual attention.**

I. Introduction

VISUAL attention is an important mechanism of human visual system, which enables human to efficiently process the information from complex environments. In recent years, visual attention has received extensive interests by cognitive psychologists, neurophysiologists and computer scientists. Many visual attention theories and computational models [1] have been proposed in the literatures. The computational models of visual attention provide wide applications in object detection [2], object recognition [3], image segmentation [4], image retargeting [5], [6], and video compression [7].

According to the mechanism of inferring image saliency, the computational models can be divided into two categories: bottom-up models and top-down models. Bottom-up models are data-driven and detect saliency from image features, such as contrast, location and texture. Top-down models are task-driven and require human interaction for saliency detection. In this letter, we focus on bottom-up saliency models. According to the method of obtaining saliency, the existing bottom-up saliency models can also be divided into two categories: sparsity-based models [8]–[15] and distinctness-based models [5], [16], [17]–[21].

The sparsity-based models assume that the saliency in image is sparse, and consider that salient regions are consistent with the locations which have sufficiently sparsity in image representation. These models include spectral analysis models [8]–[12] and spatial sparsity pursuit models [13]–[15]. The spectral analysis models use transforms such as Fourier transform or quaternion Fourier transform to get the frequency spectrum of the image, and then get the sparse description of the image in frequency domain, which is inversely transformed back into the spatial domain to get salient regions finally. These models are computationally efficient and get high prediction accuracy. However, these models highlight boundaries of salient regions rather than entire regions. Another sparsity-based models are spatial sparsity pursuit models, which treat saliency detection as a sparsity pursuit problem. They represent image patches with the learned overcomplete sparse bases and infer salient regions by low-rank and sparsity matrix decomposition. These spatial sparsity pursuit models produce accurate and reliable results, and can be easy to incorporate top-down priors. In saliency detection procedure, the sparsity-based models, either Fourier transform or sparse representation on a whole image, usually consider global properties more than local properties, which leads to coarse detection results. These models work well in the simple and sparse scene. However, when the scene is not sufficiently sparse, their performance may be affected, as illustrated in Fig. 1(b) and (c).

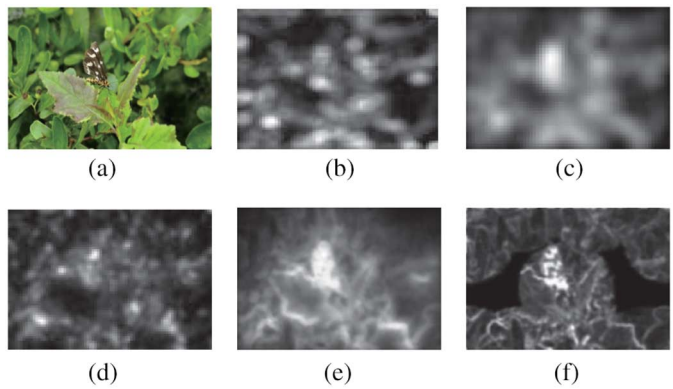


Fig. 1. Different models for saliency detection. (a) is an input image. (b) and (c) are the saliency maps produced by sparsity-based models in SR [8], QDCT [10]. (d) and (e) are produced by distinctness-based models in IT [16], CA [5]. (f) is produced by the proposed sparse-distinctive saliency model.

The distinctness-based models consider that salient regions consist of the pixels or patches which are distinctive from their surroundings or whole image. For example, models in [16], [17] calculate the local distinctness of pixels based on center-surround divergence. These models ignore global relations and structures, and they are more sensitive to high frequency contents like image edges and noises. Models in [5], [18] calculate the distinctness on patterns, colors and several additional high-level cues of the complete image by comparing

Manuscript received September 19, 2014; revised November 13, 2014; accepted December 11, 2014. Date of publication December 18, 2014; date of current version February 27, 2015. This work was supported in part by the National Natural Science Foundation of China under Grants 61100098, 61379097, 61033011, and 61210009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rui Liao.

The authors are with the Institute of Automation, Chinese Academy of Sciences, Beijing, China (e-mail: yongkang.luo@ia.ac.cn; peng_wang@ia.ac.cn; wenjun.zhu@ia.ac.cn; hong.qiao@ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2382755

each image patch with all other image patches. These models combine global and local distinctness to improve the detection performance. These distinctness-based models are usually biologically plausible since they adopt some relevant psychological conceptions and properties such as feature integration theory and center-surround saliency mechanism. However, when the image has lots of local distinctive regions without global distinctive regions, these distinctness-based models may not work well, as illustrated in Fig. 1(d) and (e).

Actually, if an image has salient regions, the salient regions in this image may have sufficiently sparsity or be distinctive with respect to both their local and global surroundings. Therefore, it would be better to consider both sparsity and distinctness for saliency to improve the detection performance, as illustrated in Fig. 1 (f).

In this letter, we propose a bottom-up saliency model based on both sparsity and distinctness. Our main contributions are:

1 Some key ideas to consider sparsity and distinctness for saliency detection simultaneously are introduced (Section II).
2 Based on the key ideas, we devise a sparse-distinctive (SD) saliency model (Section II).
3 We evaluate our model qualitatively and quantitatively on benchmark eye-tracking datasets, and demonstrate that our model outperforms the mainstream models on predicting human fixations (Section III).

II. Sparse-Distinctive Saliency Detection

A. Key Ideas of Sparse-Distinctive Saliency Detection

Palmer [22] suggests that the visual system does not passively process all the message available within the image, and human selectively attend to different aspects of image at different times with flexible strategies. Therefore, it would be better to adopt flexible strategies for computational saliency detection.

Our sparse-distinctive saliency model is based on four key ideas, which are supported by psychological evidence:

1) Saliency in image is sparse. The simple cell receptive field to be spatially localized, oriented and bandpass can be represented by sparse image code [23]. The sufficiently sparse foreground locations are consistent with the locations of human eye movement fixations [11].
2) Salient regions consist of pixels whose local neighborhoods are distinctive in image both locally and globally. This idea is supported by psychological evidence [24], [25].
3) According to Gestalt laws [26], the organization of visual forms may possess one or several gravity centers, so the salient regions may have several attention focuses.
4) Different attributes of image contribute to saliency simultaneously, and play different roles under different scenes [22], [27].

Related works typically just follow one or two of the first three ideas, such as saliency models in [8]–[15] are based on the first idea. Saliency models in [16], [17], [20] are based on the second idea. Saliency models in [5], [18] are based on the second and third ideas. However, the fourth idea is also important for saliency detection. When the image has lots of local distinctive regions, the sparsity of image contributes to saliency more significantly; when the image is cluttered or textured, the distinctive regions may attract more attention. Thus, it will be better to consider the sparsity and distinctness of image simultaneously for saliency detection.

B. Sparse-Distinctive Saliency Detection Method

Based on the four key ideas, we design a sparse-distinctive saliency model. Firstly, we use spectral analysis method to measure the sparse saliency and get candidate regions. Then we calculate the distinctness of the image in the candidate regions. Finally, we integrate the two saliency measurement with an adaptive method to get a final saliency map. The whole procedure is illustrated in Fig. 2.

Sparse Saliency: Since saliency in image is sparse (Idea 1), we first get the sparse saliency measurement S_s with Spectrum Residual method in SR [8] in CIE $L*a*b$ color space. This method transforms each color channel image f_c into frequency domain with Fourier transform: $f_c(x,y) \xrightarrow{\mathcal{F}} \mathcal{F}(f_c)(u,v)$, then gets the amplitude $\mathcal{A}_c(u,v) = |\mathcal{F}(f_c)|$ and the phase $\mathcal{P}_c(u,v) = angle(\mathcal{F}(f_c))$, and then calculates the log amplitude spectrum: $\mathcal{L}_c(u,v) = \log \mathcal{A}_c(u,v)$. The spectrum residual is calculated by

$$\mathcal{R}_c(u,v) = \mathcal{L}_c(u,v) - h_n * \mathcal{L}_c(u,v), \tag{1}$$

then the sparse saliency measurement can be obtained by

$$S_{s_c}(x,y) = \mathcal{F}^{-1}[\exp(\mathcal{R}_c(u,v) + i \cdot \mathcal{P}_c(u,v))], \tag{2}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the Fourier and inverse Fourier Transforms, and h_n denotes a local average filter. Different from SR [8], we just want to get the sparse saliency measurement, so we do not need the smooth procedure at the final step. Then we sum up the saliency measurements of all the color channels and normalize the summation to the range [0, 1] by

$$S_s(x,y) = \mathcal{N}[\sum_c S_{s_c}(x,y)], \tag{3}$$

where $\mathcal{N}(\cdot)$ denotes normalized operation.

Candidate Salient Regions: The organization of visual forms may possess one or several gravity centers, so the image may usually has multiple attention focuses (Idea 3). We consider the pixels whose sparse saliency values exceed a certain threshold T_1 as attention focuses:

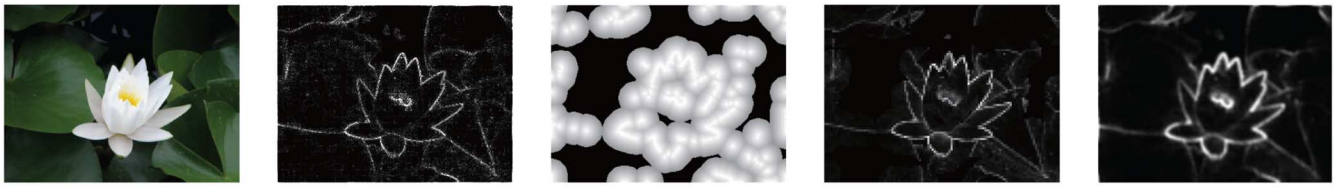
$$foci(x,y) = \begin{cases} 1 & \text{if} S_{\mathrm{s}}(x,y) \geq T_1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

In our experiments, we set $T_1 = 0.1$ empirically.

The regions surrounding the focuses are considered as candidate salient regions. We calculate the euclidean positional distance $d_{foci}(x,y)$ between pixel (x,y) and its closest focus, and normalize $d_{foci}(x,y)$ to the range $[0,1]$. Then the candidate salient region set U can be defined as follows:

$$CR(x,y) = \begin{cases} 1 - d_{foci}(x,y) & \text{if } d_{foci}(x,y) \leq T_2, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$U = \{(x,y) | CR(x,y) > 0\}, \tag{6}$$



(a) (b) (c) (d) (e)

Fig. 2. Illustration of the main phases of our model. (a) Input an image. (b) Calculate the sparse saliency at first. (c) Get the candidate regions. (d) Calculate the distinctive saliency in the candidate regions. (3) Integrate sparse and distinctness saliency for the final saliency map.

where T_2 is a neighborhood distance threshold. In our experiments, we set $T_2 = 0.7$ empirically.

Distinctive Saliency: Since salient regions consist of pixels whose local neighborhoods are distinctive in image both locally and globally (Idea 2), we calculate the distinctive saliency value in the candidate regions with a patch-based method inspired by [5]. Different from [5], we search salient regions in the narrow candidate regions instead of the whole image. In this way, we can take the advantage of global sparse saliency constraints to improve computational efficiency. Segment each scale image into 3×3 patches. Denote $P_{i\in U}$ as the patch centered at the pixel i in the candidate region set U and Q_k as the patch centered at pixel k. Let $d_{color}(P_{i\in U}, Q_k)$ be the euclidean distance between the vectorized patches $P_{i\in U}$ and Q_k in CIE $L*a*b$ color space. Denote $d_{position}(P_{i\in U}, Q_k)$ as the euclidean distance between the positions of patches $P_{i\in U}$ and Q_k. Normalize $d_{color}(P_{i\in U}, Q_k)$ and $d_{position}(P_{i\in U}, Q_k)$ to the range $[0, 1]$. Then, we define a dissimilarity measurement between a pair of patches as

$$d(P_{i\in U}, Q_k) = \frac{d_{color}(P_{i\in U}, Q_k)}{1 + d_{position}(P_{i\in U}, Q_k)}. \tag{7}$$

We search for the K most similar patches $\{Q_k\}_{k=1}^{K}$ in the image for every patch $P_{i\in U}$ according to $d_{color}(P_{i\in U}, Q_k)$. If patch $P_{i\in U}$ is distinctive in its K most similar patches, it will be distinctive in all other image patches [5]. We compute the distinctness value of the patch whose center pixel i is in the candidate region set U with

$$S_d(P_{i\in U}) = 1 - \exp\left\{-\frac{1}{K}\sum_{k=1}^{K} d(P_{i\in U}, Q_k)\right\}, \tag{8}$$

where $K = 4$ in our experiments. Then assign $S_d(P_{i\in U})$ to every pixel contained in patch $P_{i\in U}$, and set distinctness value of other pixels equal to 0:

$$S_d(x, y) = \begin{cases} S_d(P_{i\in U}) & \text{if}(x, y) \in P_{i\in U}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Sparse-Distinctive Saliency: In order to consider sparsity and distinctness for saliency detection simultaneously (Idea 4), we integrate these two kinds of saliency measurement with an adaptive method. The entropy of the saliency map can evaluate the distribution of saliency value, and the smaller entropy value means that the saliency value is more concentrated, and then the salient regions are easier to be extracted from the image. If the entropy value of sparse saliency map is smaller than μ times of entropy value of distinctive saliency, the sparse saliency will contribute to image saliency more significantly. Thus, we can use entropy values to weight the contributions of sparsity and distinctness for saliency. When the pixel is near the focuses, its sparse saliency will contribute to the saliency more significantly. We use $CR(x, y)$ in (5) to weight its contribution, and then define the saliency value as

$$S(x, y) = S_d(x, y) + \rho \cdot CR(x, y) \cdot S_s(x, y) \tag{10}$$

with

$$\rho = \mu \cdot \frac{entropy(S_d)}{entropy(S_s)}, \tag{11}$$

where $\mu = 3$ in our experiments, S_d is the distinctive saliency map, and S_s is the sparse saliency map.

Multiscale Saliency: If a region is salient, it is likely to be salient at multiple image scales. The good saliency map can highlight salient regions and suppress the common regions, and the entropy of the good saliency map will be small. Thus, we use the entropy values of the saliency maps at different scales to weight their contributions in the generation of final saliency. We calculate the saliency value $S^r(x, y)$ of each pixel in image at scale r by (10), and get a saliency map S^r. Since the size of patch is fixed with 3×3 pixels, when the image scale is smaller, the fineness of the saliency map will be lower. Thus, we emphasize $S^r(x, y)$ at big scale with higher significance. Finally, we define the multiscale saliency value as

$$\overline{S(x, y)} = \frac{\sum_{r\in R} \exp\{-entropy(S^r)\} \cdot r^2 \cdot S^r(x, y)}{\sum_{r\in R} \exp\{-entropy(S^r)\} \cdot r^2}, \tag{12}$$

where $R = \{1, 0.8, 0.6, 0.4\}$ denote the set of image scales.

III. Experimental Results

In this section, we evaluate the proposed saliency model qualitatively and quantitatively by using MIT saliency benchmark described in [28] and Hou's saliency benchmark described in [11]. The MIT saliency benchmark includes 300 natural indoor and outdoor scenes. The Hou's saliency benchmark includes the data set of human eye-tracking data introduced by Bruce and Tsotsos [30], and this data set consists of 120 images with the corresponding free-viewing eye-tracking data of 20 subjects for 4 seconds.

Firstly, we qualitatively compare our model with six mainstream models including sparsity-based models such as QDCT [10], SR [8], HFT [12] and distinctness-based models such as CA [5], GBVS [29], IT [16] on MIT saliency benchmark and Hou's saliency benchmark. Fig. 3 shows several images within the saliency benchmarks and the corresponding saliency maps.

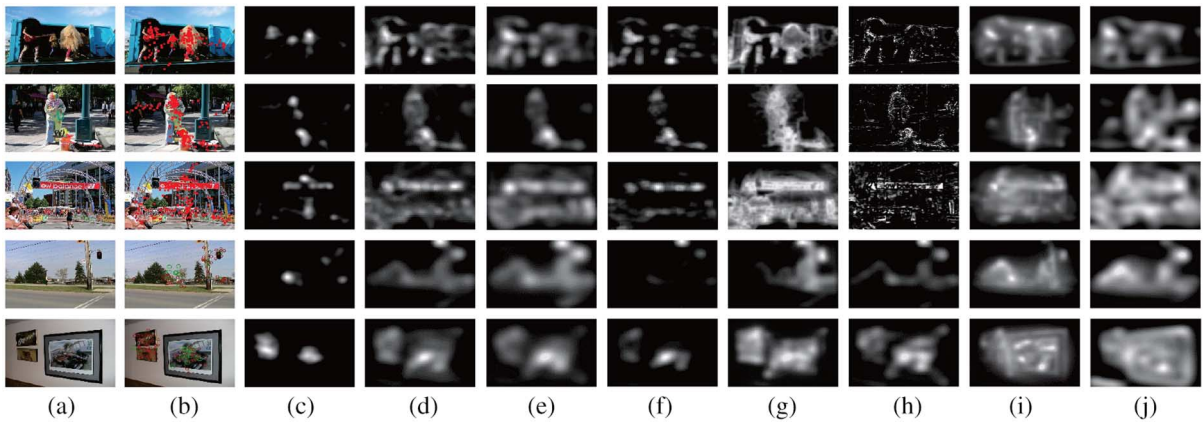


Fig. 3. Qualitative results for the datasets [28] (Row 1-3) and [11] (Row 4-5). (a) Original image. (b) Fixations in original image. (c) Human Fixations Map. From (d)-(i) are saliency maps generated by our model and mainstream models. (d) Our model SD. (e) QDCT [10]. (f) SR [8]. (g) CA [5]. (h) HFT [12]. (i) GBVS [29]. (j) IT [16]. Saliency maps in row 4 and 5 are blurred with optimal blurring factor σ which are shown in Fig. 4.

Fig. 3 illustrates that the proposed model SD has more separability and discriminability, and its predicting results are more close to the human fixations, compared with the mainstream models. In particular, saliency maps generated by sparsity-based models SR [8], HFT [12] (Fig. 3(f) and (h)) highlight boundaries of salient regions rather than entire regions. In addition, the distinctness-based models CA [5], GBVS [29] and IT [16] are sensitive to complex texture that lead to highlight most image regions (Fig. 3(g), (i) and (j)). In these outdoor scenes, the images are not sufficiently sparse or have lots of local distinctive regions. Thus, if only considering sparsity or distinctness, it is not easier to get the real salient regions. The proposed model SD considers both sparsity and distinctness, and obtains better performance in these scenes.

Secondly, on Hou's saliency benchmark [11], we quantitatively compare our model with the six models mentioned above and other related models such as DCT [11], SP and DIST. SP and DIST are the subprocedures of the proposed model SD, which only consider sparsity or distinctness for saliency. We follow the evaluation procedure proposed on the benchmark, which computes an ROC Area Under the Curve (AUC) score to evaluate the consistency between the saliency map and its fixations map. The ROC curve plots the false positive rate as a function of the true positive rate. It considers the thresholded saliency map as a binary classifier to separate the fixated points from not-fixated points [11]. In order to get a fair comparison, this evaluation method computes the AUC score under different blurring conditions parameterized by blurring factor σ. Fig. 4 and Table I show that our model outperforms the other models. The experiments demonstrate that considering both sparsity and distinctness for saliency detection can improve the precision of predicting human fixations. Furthermore, SD is more efficient than CA. In the experiments, CA scales the input images to the size of 250 pixels (height), and SD scales the input images to the size of 64 pixels (height). Eliminating the effect of scale factor, SD is about two times faster than CA.

For all the experiments described in this letter, the parameters are kept fixed. Our model is robust to the values of these parameters. For example, when K in (8) is range from 1 to 60, the AUC scores are all in [0.7202, 0.7243]; when μ in (11) is range from 1 to 20, the AUC scores are all in [0.7211, 0.7243];

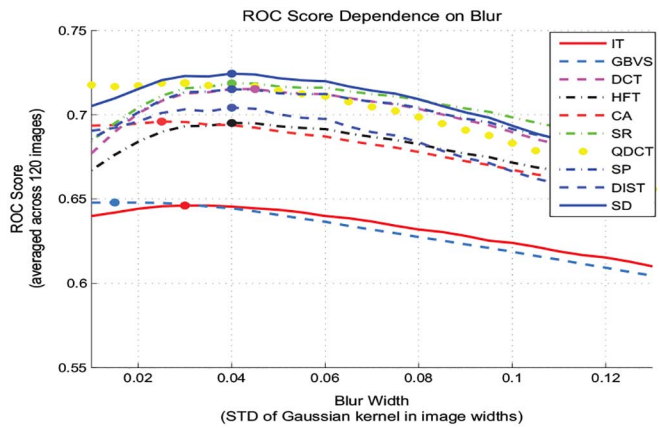


Fig. 4. ROC Area Under the Curve (AUC) Score. This metric follows the procedure proposed on the benchmark [11]. The curves show that our model SD outperforms the other models on predicting human fixations.

TABLE I
AUC SCORE AND TIME COST OF EACH MODEL (UNDER OPTIMAL σ)

Models	AUC	Time(s)	Models	AUC	Time(s)
IT [16]	0.6461	0.531	SR [8]	0.7186	0.124
GBVS [29]	0.6469	1.535	QDCT [10]	0.7187	0.187
DCT [11]	0.7152	0.026	**SP**	0.7151	0.078
HFT [12]	0.6950	0.279	**DIST**	0.7042	2.474
CA [5]	0.6959	46.304	**SD**	**0.7243**	2.329

when the number of scales R in (12) is two, three or four, the AUC scores are all with 0.7243, and if only one scale is used, the AUC score is 0.7235.

IV. CONCLUSION

In this letter, we introduce some key ideas based on psychological evidence, and propose a novel saliency model based on sparsity and distinctness, i.e., sparse-distinctive (SD) saliency model. The experimental results show that considering the sparsity and distinctness for saliency detection can improve the accuracy of predicting human fixations. Furthermore, our method achieves superior accuracy in comparison to mainstream approaches on predicting human fixations. Future works include the further improvement of the computational efficiency and the integration of multiple features in the proposed method.

REFERENCES

[1] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, 2013.

[2] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson, "Top-down control of visual attention in object detection," in *ICIP*, 2003, vol. 1, pp. I–253–6.

[3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," in *CVPR*, 2004, vol. 2, pp. II-37–II-44.

[4] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 141–145, 2006.

[5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, 2012.

[6] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, 2012.

[7] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, 2004.

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

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

[10] B. Schauerte and R. Stiefelhagen, "Quaternion-based spectral saliency detection for eye fixation prediction," in *ECCV*, 2012, pp. 116–129.

[11] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, 2012.

[12] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, 2013.

[13] Y. Li, Y. Zhou, L. Xu, X. Yang, and J. Yang, "Incremental sparse saliency detection," in *ICIP*, 2009, pp. 3093–3096.

[14] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *IEEE Signal Process. Lett.*, vol. 17, no. 8, pp. 739–742, 2010.

[15] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1327–1338, 2012.

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

[17] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *ICCV*, 2011, pp. 2214–2219.

[18] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?," in *CVPR*, 2013, pp. 1139–1146.

[19] S. Lu, C. Tan, and J. Lim, "Robust and efficient saliency modeling from image co-occurrence histograms," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 36, no. 1, pp. 195–201, 2014.

[20] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley, "Image saliency: From intrinsic to extrinsic context," in *CVPR*, 2011, pp. 417–424.

[21] C. Scharfenberger, A. Wong, K. Fergani, J. Zelek, and D. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *CVPR*, 2013, pp. 979–986.

[22] S. E. Palmer, *Vision Science: Photons to Phenomenology*. Cambridge, MA, USA: MIT press, 1999, vol. 1.

[23] 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.

[24] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[25] C. Koch and T. Poggio, "Predicting the visual world: Silence is golden," *Nature Neurosci.*, vol. 2, pp. 9–10, 1999.

[26] K. Koffka, *Principles of Gestalt Psychology*. Oxford, U.K.: Routledge, 1955.

[27] T. S. Horowitz and J. M. Wolfe, "What attributes guide the deployment of visual attention and how do they do it?," *Nature Rev. Neurosci.*, vol. 5, pp. 495–501, 2004.

[28] Z. Bylinskii, T. Judd, F. Durand, A. Oliva, and A. Torralba, *MIT saliency benchmark* [Online]. Available: http://saliency.mit.edu/

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

[30] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *NIPS*, 2005, pp. 155–162.