Improving Saliency Detection Via Multiple Kernel Boosting and Adaptive Fusion

Xiaofei Zhou, Zhi Liu, *Senior Member, IEEE*, Guangling Sun, Linwei Ye, and Xiangyang Wang

***Abstract*—This letter proposes a novel framework to improve the saliency detection performance of an existing saliency model, which is used to generate the initial saliency map. First, a novel regional descriptor consisting of regional self-information, regional variance, and regional contrast on a number of features with local, global, and border context is proposed to describe the segmented regions at multiple scales. Then, regarding saliency computation as a regression problem, a multiple kernel boosting method based on support vector regression (MKB-SVR) is proposed to generate the complementary saliency map. Finally, an adaptive fusion method via learning a quality prediction model for saliency maps is proposed to effectively fuse the initial saliency map with the complementary saliency map and obtain the final saliency map with improvement on saliency detection performance. Experimental results on two public datasets with the state-of-the-art saliency models validate that the proposed method consistently improves the saliency detection performance of various saliency models.**

***Index Terms*—Adaptive fusion, multiple kernel boosting, regional descriptor, saliency detection, support vector regression.**

I. Introduction

SALIENCY detection has received more and more attention in the recent years due to its wide use in a number of applications such as salient object detection and segmentation [1], [2], content-aware image/video retargeting [3]–[6], content-based image/video compression [7], [8], to name a few. Most saliency models are bottom-up stimuli-driven, i.e., only exploit the features extracted from the image for saliency computation. For example, the center-surround scheme [9] on various features with different formulations has been widely exploited to measure saliency in a number of saliency models such as the regional contrast-based model [10] that simultaneously evaluates global contrast and spatial coherence of regions for estimating saliency. Recently, hierarchical saliency models via hierarchical inference [11] and saliency tree [12] enable hierarchical representation of saliency and improve the saliency detection performance. In [13], the graph-based manifold ranking is exploited to generate saliency map by incorporating local grouping cues and boundary priors.

Some efforts have been made on the incorporation of top-down information into saliency models via machine learning. The fusion weights of saliency features are learned by using the conditional random field (CRF) in [1] and a mixture of linear Support Vector Machines (SVMs) is learned to detect salient objects in [14]. In [15], joint CRF and dictionary learning are exploited for measuring saliency. In [16] and [17], the random forest regression is used to map the regional discriminative feature vector to the saliency score of each region and saliency scores across multiple levels are fused to obtain the final regional saliency. Regarding saliency computation as a classification problem, a bootstrap learning (BL) method [18] with the use of multiple kernel boosting based on SVM is exploited to generate the strong saliency map on the basis of an existing saliency model.

However, the saliency detection performance of the existing saliency models is obviously degraded on the complicated images. Therefore, we propose a novel framework to effectively improve the saliency detection performance of the existing saliency models via multiple kernel boosting and adaptive fusion. An example of the proposed method is shown in Fig. 1. Our main contribution lies in the following three aspects:

1) We propose a novel regional descriptor with the introduction of more effective features, the incorporation of local, global, and border context for computing regional contrast and the concatenation of regional self-information, regional variance, and regional contrast on the basis of each feature. The proposed regional descriptor is considerably different from those used in the previous works [16], [17] and can better describe the region for saliency computation.
2) Different from the BL method [18] which also uses the multiple kernel boosting, we regard saliency computation as a regression problem and propose a novel multiple kernel boosting method based on support vector regression (MKB-SVR), which is consistent with multiple kernel learning for taking full advantage of multiple features via the definition of regression error and the update of sample weights.
3) We propose an adaptive fusion method via learning a quality prediction model to obtain the fusion weights for initial saliency map and complementary saliency map and finally enable the proposed method to consistently improve the saliency detection performance of various saliency models.

Manuscript received October 01, 2015; revised December 08, 2015; accepted February 26, 2016. Date of publication March 01, 2016; date of current version March 21, 2016. This work was supported by National Natural Science Foundation of China under Grants 61171144 and 61471230, and by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Guy Gilboa. *(Corresponding author: Zhi Liu.)*

The authors are with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China (e-mail: zxforchid@outlook.com; liuzhisjtu@163.com; sunguangling@shu.edu.cn; yelinweimail@163.com; wangxiangyang@shu.edu.cn).

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

Digital Object Identifier 10.1109/LSP.2016.2536743

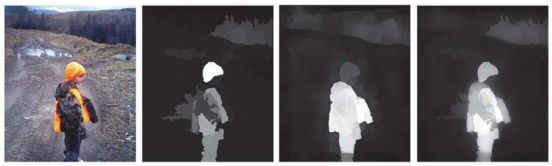

Fig. 1. Example of our method. (a) Input image. (b) Initial saliency map generated using [11]. (c) Complementary saliency map. (d) Final saliency map.

II. Proposed Method

To deal with the issue of scale, multiscale segmentation results $S_m(m = 1, \ldots, 4)$ are generated by using the simple linear iterative clustering (SLIC) method [19] at four scales with region number of 100, 150, 200, and 250, respectively.

A. Regional Descriptor

To comprehensively represent a region, we introduce a novel regional descriptor that measures both regional property (regional self-information and regional variance) and regional contrast on eight features ($n_F = 8$) as shown in Table I. On the basis of each region, we extract color features in the three color spaces including RGB, HSV, and Lab and two texture features including the local binary pattern (LBP) feature [20] and the responses of Leung and Malik (LM) filters [21]. Besides, scale-invariant feature transform (SIFT) descriptor [22] and histogram of oriented gradient (HOG) feature [23] are extracted to represent each region the local properties including scale and orientation and the geodesic distance (GD) from each region to image borders is calculated as a background prior [24]. Specifically, we extract the dense SIFT feature at pixel level using [25], compute the patch-level HOG feature on a dense grid of uniformly spaced cells and the patch-level GD using the GS Grid algorithm. The HOG feature and GD of each patch are assigned to all pixels falling into the patch.

For each region $R_i \in S_m$, we compute on the eight features a set of regional self-information $x^s_{i,k}(k = 1, \ldots, n_F)$ and regional variance $x^v_{i,k}(k = 1, \ldots, n_F)$ by using all pixels in R_i, i.e.,

$$x^v_{i,k} = \sum\nolimits_{p_j \in R_i} \|z_{j,k} - \theta_{i,k}\|^2 / |R_i|, \quad \theta_{i,k} = \sum\nolimits_{p_j \in R_i} z_{j,k} / |R_i| \tag{1}$$

where $z_{j,k}$ denotes the *k*th feature of each pixel p_j falling into R_i and $|R_i|$ is the number of pixels in R_i.

The regional contrast $x^c_{i,k}(k = 1, \ldots, n_F)$ for each region R_i on the *k*th feature is evaluated by comparing R_i with regions in the three contexts and is defined as follows:

$$x^c_{i,k} = \left[x^{c,L}_{i,k}, x^{c,G}_{i,k}, x^{c,B}_{i,k}\right], \quad x^{c,N}_{i,k} = \sum\nolimits_{R_j \in C^N_i} \lambda_j w_{ij} D_k(f_i, f_j) \tag{2}$$

where the superscript *N* may denote *L*, *G*, and *B* for local context, global context, and border context, respectively. For each region R_i, C^N_i may denote its local context (regions spatially adjacent with R_i) with $N = L$, its global context (all the other regions in S_m) with $N = G$, or its border context (regions along the image borders) with *N*=*B*. The regional contrast $x^c_{i,k}$ is the concatenation of contrasts computed with the above three contexts, and thus in Table I, the dimension of regional contrast on each feature is with a multiplication factor of three "×3". For calculating $x^{c,N}_{i,k}$ in any context *N*, λ_j is the normalized area of region R_j; $w_{ij} = \exp(-||\rho_i - \rho_j||^2/2)$ is the factor of spatial distance between ρ_i and ρ_j, i.e., the normalized centroid of R_i and R_j, respectively; and f_i denotes any regional feature of R_i. The feature difference $D_k(f_i, f_j)$ for a pair of regional features with n_D dimensions can take the three distance metrics: 1) $d_1(f_i, f_j) = \left[\left|f^1_i - f^1_j\right|, \ldots, \left|f^{n_D}_i - f^{n_D}_j\right|\right]$; 2) the Euclidean distance $d_2(f_i, f_j) = ||f_i - f_j||$; and 3) the chi-square distance $\chi^2(f_i, f_j)$. Please refer to Table I for details. Finally, on the basis of each feature, regional self-information, variance, and contrast are concatenated to form the regional descriptor $x_{i,k}$ for each region R_i on the *k*th feature as follows:

$$x_{i,k} := \left[x^s_{i,k}, x^v_{i,k}, x^c_{i,k}\right]. \tag{3}$$

B. Multiple Kernel Boosting Based on Support Vector Regression

Regarding saliency computation as a regression problem and taking full advantage of multiple features, we propose a novel MKB-SVR to estimate the complementary saliency map. Four types of kernels $\{K_m\}^{n_k}_{m=1}$ (linear, radial basis function (RBF), sigmoid, and polynomial, $n_k = 4$) with eight features are included in our MKB-SVR method.

For training the regressor, we select the training set $\{x_i, y_i\}^{n_s}_{i=1}$, where n_s is the number of samples based on the initial saliency map M_I, which can be generated using any saliency model for images. For the *i*th sample, $x_i := \{x_{i,k}\}^{n_F}_{k=1}$ is the regional descriptor of each region R_i and y_i is the saliency value of R_i as a positive or negative sample. Specifically, μ denotes the mean saliency value of M_I and μ_i denotes the mean saliency value of R_i in M_I. R_i is labeled as a positive sample and y_i is set to $+1$ if μ_i is higher than a larger threshold 1.5μ, or labeled as a negative sample and y_i is set to -1 if μ_i is lower than a smaller threshold 0.05.

Following the framework of MKB [26], the regression error of a single kernel SVR with a feature, specifically, the *m*th kernel with the *k*th feature at the *j*th iteration is defined as

$$\varepsilon^{(j)}_{m,k} = \sum\nolimits_{i=1}^{n_s} w^{(j)}_{i,k} \Psi[|h_{m,k}(x_{i,k}) - y_i| - \Delta] \tag{4}$$

where the sign function $\Psi[r]$ equals to 1 when $r > 0$ and 0 otherwise. $w^{(j)}_{i,k}$ is the weight for the *k*th feature of the *i*th training sample at the *j*th iteration and $h_{m,k}(x_{i,k})$ is the output of SVR on $x_{i,k}$. The threshold Δ is defined as

$$\Delta = (1.5/n_s) \times \sum\nolimits_{i=1}^{n_s} \|y_i - h_{m,k}(x_{i,k})\|^2. \tag{5}$$

The whole process of our MKB-SVR method is summarized in Algorithm 1, which obtains the regressor $F(x)$ via integrating the most discriminative features and the corresponding kernels adaptively. The number of boosting iterations *J* is adaptively set to $n_k \times n_F$. Then for each region R^*, using its regional descriptor as the input, $x = x^*$, the regressor outputs the estimated saliency value y^*, which is assigned to all pixels in R^*. The regressor is applied on all test samples (all the

TABLE I
REGIONAL DESCRIPTOR

Feature descriptor	Regional property				Regional contrast		Total dim
	Regional self-information	Dim	Regional variance	Dim	Definition	Dim	
Color $x_{i,k}$ $k=1,2,3$	Average color $x^s_{i,k}$ ($k=1,2,3$ corresponds to RGB, HSV and Lab color space, respectively.)	3	Color variance $x^v_{i,k}$	3	$d_1(x^s_{i,k}, x^s_{j,k})$, $d_2(x^s_{i,k}, x^s_{j,k})$	(3+1)×3	21
					$\chi^2(\phi_{i,k}, \phi_{j,k})$	1×3	
LM $x_{i,4}$	Absolute response $x^{s,\text{abs}}_{i,4}$	15	Variance of the response of LM filters $x^v_{i,4}$	15	$d_1(x^{s,\text{abs}}_{i,4}, x^{s,\text{abs}}_{j,4})$	15×3	93
	Max response histogram $x^{s,\max}_{i,4}$	15			$\chi^2(x^{s,\max}_{i,4}, x^{s,\max}_{j,4})$	1×3	
LBP $x_{i,5}$	LBP histogram $x^s_{i,5}$	59	Variance of LBP features $x^v_{i,5}$	1	$\chi^2(x^s_{i,5}, x^s_{j,5})$	1×3	63
SIFT $x_{i,6}$	Average SIFT descriptor $x^s_{i,6}$	128	Variance of SIFT descriptors $x^v_{i,6}$	128	$d_2(x^s_{i,6}, x^s_{j,6})$	1×3	259
HOG $x_{i,7}$	Average HOG feature $x^s_{i,7}$	32	Variance of HOG features $x^v_{i,7}$	32	$d_2(x^s_{i,7}, x^s_{j,7})$	1×3	67
GD $x_{i,8}$	Average GD $x^s_{i,8}$	1	Variance of GD values $x^v_{i,8}$	1	$d_1(x^s_{i,8}, x^s_{j,8})$	1×3	5

Note that $\phi_{i,k}$ $(k=1,2,3)$ denotes the regional histogram of each region R_i in the RGB, HSV, and Lab color space, respectively, and $x^s_{i,4} := \left[x^{s,\text{abs}}_{i,4}, x^{s,\max}_{i,4}\right]$.

Algorithm 1. Pseudo Code of MKB-SVR

Input: The training set $\{x_i, y_i\}_{i=1}^{n_S}$ and kernel functions $\{K_m\}_{m=1}^{n_K}$.

Output: The regressor $F(x)$.

1: For each kernel function with each feature, a single-kernel SVR $h_{m,k}$ is trained with the kernel K_m on the k^{th} feature by using the entire training set $\{x_{i,k}, y_i\}_{i=1}^{n_S}$, and a total of $n_K \times n_F$ SVRs are generated to form a candidate set Ω.

2: Initialize the weights for samples as $w^{(1)}_{i,k} = 1/n_S$ $(i = 1,\ldots,n_S,\ k = 1,\ldots,n_F)$ and initialize the regressor as $F(x) = 0$.

3: **for** $j = 1 : J$

For each SVR $h_{m,k}$, calculate the regression error $\varepsilon^{(j)}_{m,k}$ using Eq. (4). Select the optimal SVR as $H^{(j)} = \arg\min_{h_{m,k}\in\Omega}\left\{\varepsilon^{(j)}_{m,k}\right\}$, and let $\delta^{(j)} = \min_{h_{m,k}\in\Omega}\{\varepsilon^{(j)}_{m,k}\}$.

if $\delta^{(j)} < 0.5$

Compute the weight $\beta_{(j)} = \frac{1}{2}\log\frac{1-\delta^{(j)}}{\delta^{(j)}}$ for $H^{(j)}$;

Update the regressor as $F \leftarrow F + \beta_{(j)}H^{(j)}$.

else

break;

end if

Update each weight as $w^{(j+1)}_{i,k} := w^{(j)}_{i,k}\beta_{(j)}^{\left\{1-\Psi\left[\left|H^{(j)}(x_{i,k})-y_i\right|-\Delta\right]\right\}}$, and then perform normalization on all weights.

end for

4: The regressor is obtained as $F(x) = \sum_j \beta_{(j)}H^{(j)}(x)$

segmented regions at four scales of the input image) and the complementary saliency maps at four scales are generated as $\{M_{Cm}\}_{m=1}^{4}$. Finally, the complementary saliency map of the input image is calculated as $M_C = \sum_{m=1}^{4} M_{Cm}/4$.

C. Adaptive Fusion

The proposed adaptive fusion method evaluates the quality of saliency maps to calculate the fusion weights for adaptively fusing initial saliency map with complementary saliency map. We adopt the six quality measures [27] for saliency map and concatenate them together to form a quality feature. A "one-vs.-all" SVM [28] with linear kernel is trained on a set of samples to predict the quality of saliency map, and we randomly select $n_T = 3000$ images from Microsoft Research Asia (MSRA) dataset [1] to train this quality prediction model. Specifically, for each training image $T_i (i = 1,\ldots,n_T)$, the initial saliency map M^i_{TI} and the complementary saliency map M^i_{TC} are generated using the procedure in Sections II-A and II-B. The two training samples $\{q_{2i-1}, l_{2i-1}\}$ and $\{q_{2i}, l_{2i}\}$ are generated for M^i_{TI} and M^i_{TC}, respectively. Here, q_{2i-1} denotes the quality feature for M^i_{TI} and l_{2i-1} is the binary label indicating the relative quality of M^i_{TI} over its counterpart M^i_{TC}. A similar definition is applicable to $\{q_{2i}, l_{2i}\}$. For both M^i_{TI} and M^i_{TC}, we calculate the area under the receiver operating characteristic (ROC) curve (AUC), A^i_{TI} and A^i_{TC}, respectively, to quantitatively measure their quality and determine the binary labels as follows:

$$\begin{aligned} l_{2i-1} = 1 \text{ and } l_{2i} = -1, \quad &\text{if } A^i_{TI} > A^i_{TC} \\ l_{2i-1} = -1 \text{ and } l_{2i} = -1, \quad &\text{otherwise.} \end{aligned} \tag{6}$$

The learned quality prediction model is used to predict the quality score of a given saliency map. For the initial saliency map M_I and the complementary saliency map M_C of the test image, we first calculate the two quality features, q_I and q_C, respectively, and then use the learned model to obtain the two decision values as their quality scores, s_I and s_C, respectively. The fusion weights for M_I and M_C are then defined as

$$\alpha_I = s_I/(s_I + s_C), \quad \alpha_C = s_C/(s_I + s_C) \tag{7}$$

and the final saliency map M_F is generated as follows:

$$M_F = \alpha_I \times M_I + \alpha_C \times M_C. \tag{8}$$

III. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed method over two benchmark datasets, i.e., MSRA [1] and PASCAL-1500 [29]. Since 3000 images from MSRA dataset are used for

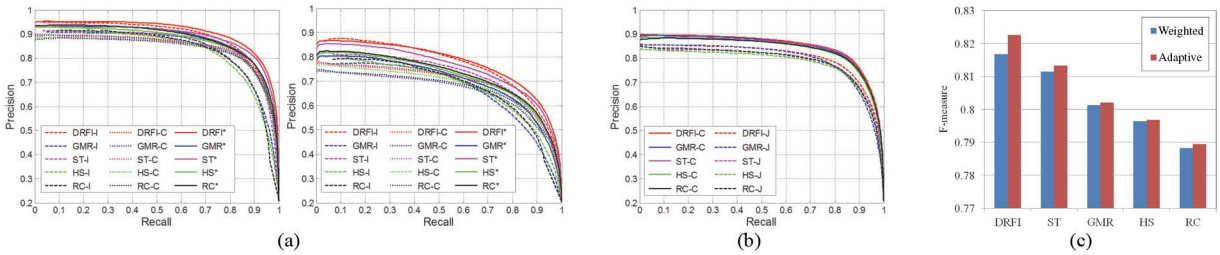

Fig. 2. (a) PR curves of initial, complementary, and final saliency maps on MSRA (left) and PASCAL-1500 (right). (b) PR curves generated using Jiang's and our regional descriptor. (c) F-measures achieved by the weighted combination and our adaptive fusion method.

training the quality prediction model in Section II-C, the remaining 2000 images are used for test. Following the evaluation measures used in [30], we evaluated the saliency detection performance using the precision-recall (PR) curve and the F-measure by setting its coefficient β^2 to 0.3. We use five saliency models including discriminative regional feature integration (DRFI) [17], graph-based manifold ranking (GMR) [13], saliency tree (ST) [12], hierarchical saliency (HS) [11], and regional contrast (RC) [10] to generate initial saliency maps, and use the proposed method to generate the complementary and final saliency maps. For example, using DRFI model, the initial, complementary, and final saliency maps are denoted as DRFI-I, DRFI-C, and DRFI*, respectively.

A comparison among initial, complementary, and final saliency maps is shown in Fig. 2(a). We can see from Fig. 2(a) that the complementary saliency maps with the four models (excluding DRFI model) achieve the higher performance than the corresponding initial saliency maps in the zone of high precision and high recall, which is critical for salient object detection. DRFI model is with the highest performance among the state-of-the-art saliency models [30], but the performance of DRFI-C is somewhat lower than DRFI-I. Nonetheless, the most important characteristic of our method is that over all the five saliency models on both datasets, our final saliency maps consistently outperform the corresponding initial and complementary saliency maps on the saliency detection performance.

Fig. 2(b) shows the performance comparison between the complementary saliency maps generated using our and Jiang's [17] regional descriptors (denoted as DRFI-J). Using our method, we only replaced our descriptor with Jiang's descriptor for the comparison. We can see from Fig. 2(b) that our descriptor consistently outperforms Jiang's over all the five saliency models. The F-measures shown in Fig. 2(c) indicate that our adaptive fusion method consistently outperforms the weighted combination method, in which the same weights are assigned to initial and complimentary saliency maps for their equal importance, over all the five saliency models. Note that Fig. 2(b) and (c) shows the results on MSRA dataset, while the same conclusion can be also drawn from the results on PASCAL-1500 dataset. Besides, to identify which feature makes the most contribution to the final results, we grouped features within all eight features using the leave-one-out scheme. Therefore, each feature set contains seven features and there are in total eight feature sets that are exploited to generate eight classes of final saliency maps using our method. Then we evaluated their saliency detection performances by using the PR curve and F-measure and found that the most important feature is GD, by removing which the performance is most degraded comparing to using all features.

Fig. 3. Visual comparison of saliency maps. In each example, the original image and the ground truth are shown in the first row, the saliency maps of DRFI-I, GMR-I, ST-I, HS-I, and RC-I are shown from the second to the sixth row in the odd columns, and the saliency maps of DRFI*, GMR*, ST*, HS*, and RC* are shown from the second to the sixth row in the even columns.

Fig. 3 shows the saliency maps of some example images. It can be seen that the saliency maps generated using our method (even columns) can better highlight salient object regions and suppress background regions than the initial saliency maps (odd columns). Our method can better handle some complicated images with low-contrast objects, heterogeneous objects, and cluttered background. Although our final saliency maps hinge on the quality of initial saliency maps, our method effectively improves the quality of saliency maps.

Our method is implemented using MATLAB on a PC with an Intel Core i7 4.0 GHz CPU and 16 GB RAM. The training time is around 47 h and the average testing time for an image with a resolution of 400×300 is 56.95 s including the generation of complementary saliency map (Sections II-A and B take 47.05 s and 0.73 s, respectively) and the adaptive fusion procedure (Section II-C takes 9.17 s).

IV. Conclusion

This letter proposes a novel framework to improve the saliency detection performance of the existing saliency models. The novel regional descriptor better describes a region for saliency estimation, the MKB-SVR algorithm effectively integrates the most discriminative features and the corresponding kernels to generate the complementary saliency map, and the adaptive fusion via learning a quality prediction model generates the final saliency map with the higher quality. Experimental results on two public datasets with five state-of-the-art saliency models demonstrate the effectiveness and the robustness of our method on improving the saliency detection performance.

REFERENCES

[1] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[2] Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1275–1289, Aug. 2012.

[3] A. Shamir and S. Avidan, "Seam carving for media retargeting," *Comm. ACM*, vol. 52, no. 1, pp. 77–85, Jan. 2009.

[4] H. Du, Z. Liu, J. Jiang, and L. Shen, "Stretchability-aware block scaling for image retargeting," *J. Vis. Commun. Image Represent.*, vol. 24, no. 4, pp. 499–508, May 2013.

[5] 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, Sep. 2012.

[6] Y. Fang, W. Lin, B.-S. Lee, C.-T. Lau, Z. Chen, and C.-W. Lin, "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum," *IEEE Trans. Multimedia*, vol. 12, no. 1, pp. 187–198, Feb. 2012.

[7] 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, Jan. 2010.

[8] L. Shen, Z. Liu, and Z. Zhang, "A novel H.264 rate control algorithm with consideration of visual attention," *Multimedia Tools Appl.*, vol. 63, no. 3, pp. 709–727, Apr. 2013.

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

[10] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'11)*, Jun. 2011, pp. 409–416.

[11] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'13)*, Jun. 2013, pp. 1155–1162.

[12] Z. Liu, W. Zou, and O. Le Meur, "Saliency tree: A novel saliency detection framework," *IEEE Trans. Image Process.*, vol. 23, No. 5, pp. 1937–1952, May 2014.

[13] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'13)*, Jun. 2013, pp. 3166–3173.

[14] P. Khuwuthyakorn, A. Robles-Kelly, and J. Zhou, "Object of interest detection by saliency learning," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV'10)*, Sep. 2010, pp. 636–649.

[15] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'12)*, Jun. 2012, pp. 2296–2303.

[16] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'13)*, Jun. 2013, pp. 2083–2090.

[17] H. Jiang, Z. Yuan, M.-M. Cheng, Y. Gong, N. Zheng, and J. Wang, "Salient object detection: A discriminative regional feature integration approach," arXiv: 1410.5926 [cs. CV], Oct. 2014, http://arxiv.org/abs/1410.5926.

[18] N. Tong, H. Lu, X. Ruan, and M.-H. Yang, "Salient object detection via bootstrap learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'15)*, Jun. 2015, pp. 1884–1892.

[19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, May 2012.

[20] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, Mar. 2009.

[21] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, Jun. 2001.

[22] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV'99)*, Sep. 1999, pp. 1150–1157.

[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'05)*, Jun. 2005, pp. 886–893.

[24] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV'12)*, Sep. 2012, pp. 29–42.

[25] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proc. 18th ACM Multimedia*, Oct. 2010, pp. 1469–1472.

[26] F. Yang, H. Lu, and Y.-W. Chen, "Human tracking by multiple kernel boosting with locality affinity constraints," in *Proc. 10th Asian Conf. Comput. Vis.*, Nov. 2011, pp. 39–50.

[27] L. Mai and F. Liu, "Comparing salient object detection results without ground truth," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV'14)*, Sep. 2014, pp. 76–91.

[28] L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, and L. Jackel, "Comparison of classifier methods: A case study in handwritten digit recognition," in *Proc. IEEE Int. Conf. Pattern Recognit. (ICPR'94)*, Oct. 1994, pp. 77–82.

[29] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, "Segmentation driven low rank matrix recovery for saliency detection," in *Proc. Br. Mach. Vis. Conf. (BMVC'13)*, Sep. 2013, pp. 1–13.

[30] A. Borji, M.-M. Cheng, H Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.