SALIENT OBJECT DETECTION VIA A LINEAR FEEDBACK CONTROL SYSTEM

Shuwei Huo[1], *Yuan Zhou*[1,2] * *and Sun-Yuan Kung*[2]

[1] School of Electrical and Information Engineering, Tianjin University, Tianjin, China.
[2] Electrical Engineering Department, Princeton University, Princeton, USA.
Email: zhouyuan@tju.edu.cn

ABSTRACT

Linear feedback control systems (LFCS) have been widely applied in signal analysis, filtering, and error correction. Many functional properties of LFCS are amenable to numerous object recognition and detection tasks. In fact, there exists an intimate relationship between control states and salient values. This prompts us to adopt the linear feedback control system to detect salient object in static images. Via an innovative iteration method, the system gradually converges an optimized stable state, which is associating with an accurate saliency map. In addition, to initialize the system, we propose the so called boundary homogeneity based on a priori knowledge on the boundary to estimate the background likelihood and indirectly depict a foreground (saliency) map. By our experimental results, we demonstrates that such feedback control model can bring about noticable improvement in salient object detection.

***Index Terms*—** saliency detection, linear feedback control system, iterative optimization.

1. INTRODUCTION

Salient object detection aims at generating a saliency map that highlight the objects attracting visual attention in a scene while suppressing the background. Recently, visual saliency has gained significant popularity for its great potential usefulness in many computer vision applications, such as visual tracking, object recognition, image question answering, and image parsing. Generally, methods for salient object detection can be categorized as top-down task-driven [1, 2, 3, 4] or bottom-up stimuli-driven [5, 6, 7, 8, 9, 10, 11, 12, 13, 14] approaches. The top-down approaches are task-driven and require supervised learning with the human-labelled ground truth. Thanks to abundant external databases, many recent algorithms take advantage of many high-level features and supervised methods [4] to establish the model and achieve great performance. On the contrary, the bottom-up methods generally do not rely on external databases. They utilize some empirical assumptions (also called *prior*) based on experiments and data statistics. We focus on bottom-up visual saliency models in this work.

Following the basic bottom-up principles (such as contrast prior [14] and boundary prior [15]), researchers have carried out further explorations aiming at one of the following problems in object size variety, object invisibility, unhomogeneous object, cluttered background, multi-objects scene, etc. In order to further improve robustness for complex foreground and unknown-scale object detection, many processing methods are proposed as complementary to contrast prior and boundary prior. Several works integrate the feature contrast measure calculated under multi-scale region definition [14, 16, 12], or fuse the results generated by various principles. Some other methods take the image regions along the image boundary as background samples, and explore semi-supervised learning approaches to generate the saliency map [10, 11]. However, those methods are still insufficient to generate a satisfactory saliency map for complex scenarios, such as low global contrast, unhomogeneous objects and tanglesome background. Thus, we are motivated to explore high-performance saliency analysis methods to improve the overall saliency detection performance for wide ranging scenes.

Linear feedback control system is the basic automatic control theory model which has been widely applied in the fields of system analysis and signal processing. Inspired by its advantage in signal filtering and error correction, in this paper, we apply a feedback control system model to generate smooth and accurate saliency maps. The framework of the proposed method is illustrated in Figure 1. This figure also serves to illustrate our contributions.

1) As shown in Figure 1, our main idea hinges upon the development of an iterative method, where we make good use of a linear feedback control system mathematical model to establish a relationship between control states and salient object detection. Thus we design a feedback control system for the saliency detection procedure. The system state is updated and gradually converged to an optimized stable value, which is associated with an accurate saliency map.

2) In addition, another contribution lies in developing a boundary homogeneity model for saliency probability estimation. This model introduces the image geometric theory as a complementary of boundary prior in order to accurately depict the general position and contour of the salient objects. This model is applied to properly initialize system input.

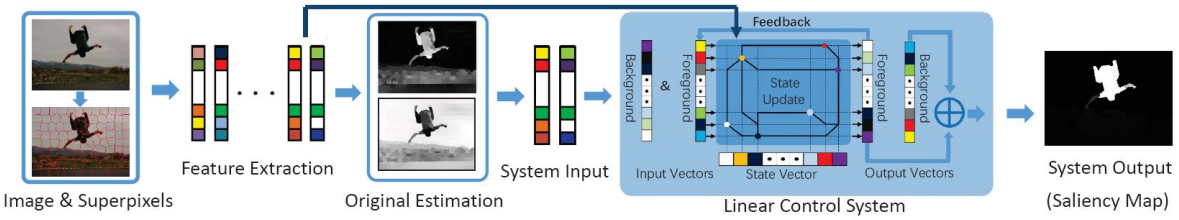

Fig. 1. Illustration of the proposed method. We present boundary homogeneity measurement to estimate the foreground and background likelihood, then utilize the linear control system model to generate refined saliency map.

2. DISCRETE LINEAR CONTROL SYSTEM

Discrete linear control system model describes the relationship of system state, input and output. We define a discrete linear control system which has M_* input channels and P_* output channels, and N_* significant system nodes are selected for state observation. At system time t $(t = 0, 1, 2, \cdots)$, the input vector, output vector and state vector are indicated as $\mathbf{u}(t) = [u_i(t)]_{M_* \times 1}$, $\mathbf{y}(t) = [y_i(t)]_{P_* \times 1}$, $\mathbf{x}(t) = [x_i(t)]_{N_* \times 1}$, respectively. The property of discrete linear control system model can be formulated as a state equation and an output equation, shown as follows

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{cases} \tag{1}$$

where $\mathbf{A}(t), \mathbf{B}(t), \mathbf{C}(t)$, and $\mathbf{D}(t)$ are all time-variant system parameters.

To promote the self-adjustment ability of the system, the linear feedback is often brought into the system framework to establish a closed-loop control. The feedback signal effects on the system input through the following equation,

$$\mathbf{u}(t+1) = \mathbf{u}(t) + \mathbf{K}(t)\mathbf{y}(t). \tag{2}$$

The system model is illustrated by Fig. 2. In particular, by properly configurating the system parameters and input signal, the system may be designed as a stable system.

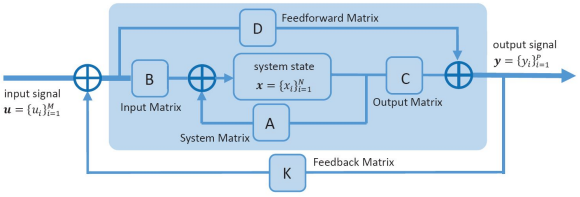

Fig. 2. Illustration of linear control system model.

3. LFCS-BASED SALIENCY DETECTION

3.1. Image feature and saliency cues extraction

To better capture intrinsic structural information and reduce the computing cost, an input image is firstly segmented into N superpixels using SLIC algorithm [17]. The mean color features in CIE-Lab space (denoted by $\{\mathbf{c}_i\}_{i=1}^N = [L_i, a_i, b_i]_{i=1}^N$) is used to describe each superpixel. For system construction, each superpixel of image is abstracted into a node of LFCS.

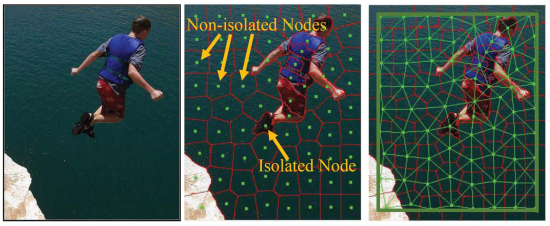

Fig. 3. Image feature and saliency cues extraction. (a) input image; (b) superpixels (for image) and nodes (for system), this figure also exemplifies the isolated and non-isolated nodes (superpixels); (c) connectivity relationship between adjacent node (superpixel) pairs.

To explore saliency cues, we define the color similarity between each superpixel pair (r_i, r_j) in both local and global scale. Local color similarity is defined only between the superpixel pair neighboring each other (shown in Fig. 3(c)), it can be computed using following formula,

$$w_{ij} = \frac{e^{-\gamma\|\mathbf{c}_i - \mathbf{c}_j\|} \cdot Bin\{r_j \in \text{Adj}(r_i)\}}{\sum_j e^{-\gamma\|\mathbf{c}_i - \mathbf{c}_j\|} \cdot Bin\{r_j \in \text{Adj}(r_i)\}} \tag{3}$$

where $\|\mathbf{c}_i - \mathbf{c}_j\|$ is the Euclidean distance between the i-th and j-th superpixel in CIE-Lab color space; $\text{Adj}(r_i)$ denotes the superpixels neighboring r_i, $Bin\{r_j \in \text{Adj}(r_i)\}$ is the binary index, it equals to 1 when the superpixel $r_j \in \text{Adj}(r_i)$ or 0 otherwise. Global color similarity is defined for all superpixel pairs, shown as follow,

$$s_{ij} = \frac{e^{-\gamma D_{ij}^{geo}}}{\sum_j e^{-\gamma D_{ij}^{geo}}} \tag{4}$$

where D_{ij}^{geo} denotes the geodesic distance [15] between a superpixel pair (r_i, r_j), which is defined as $D_{ij}^{geo} = \min_{p_1=i, p_2, \cdots, p_n=j} \sum_{k=1}^{n-1} w_{p_k p_{k+1}}, s.t.\ w_{p_k p_{k+1}} \neq 0;\ w_{ij} = 0;$ and γ is a constant.

Besides, we also explore the non-isolatism cue to reflect whether a superpixel is isolated like a spot (shown in Fig. 3(b)). It is defined using the local color similarity, and we use *m-norm* to emphasize the strong local connectivity. The definition is as follow,

$$g_i^* = \sqrt[m]{\sum_{j=1}^{N}(w_{ij})^m} \tag{5}$$

We limit its range to $[\alpha_1, \alpha_2] \subseteq [0, 1]$ as (6),

$$g_i = \alpha_1 \times \text{Norm}(g_i^*) + (\alpha_2 - \alpha_1) \tag{6}$$

where $\text{Norm}(\cdot) = \frac{(\cdot)-\min(\cdot)}{\max(\cdot)-\min(\cdot)}$ is the normalizing function.

3.2. Input Signal and State Initialization

We initialize the system input signal using the original saliency estimation. The input signal matrix $\mathbf{u}(t)$ consists of the foreground likelihood vector (the first column) and background likelihood vector (the second column). For i-th superpixel, its original background likelihood is defined as,

$$\mathbf{u}(t) = [\mathbf{u}^f(t)\ \mathbf{u}^b(t)] = \begin{bmatrix} u_1^f(t) & \cdots & u_i^f(t) & \cdots & u_N^f(t) \\ u_1^b(t) & \cdots & u_i^b(t) & \cdots & u_N^b(t) \end{bmatrix}^T \tag{7}$$

We estimate the original background likelihood of each superpixels using the boundary homogeneity matching — a novel formulation of the boundary prior, which is defiend as follow,

$$u_i^{b*}(0) = \frac{\sum_{j=1}^{N} s_{ij} \cdot Bin\{r_j \in Bnd\}}{\sum_{j=1}^{N} \max\limits_{r_k \in Adj(r_j)} \{s_{ij} - s_{ik}\} \cup \{0\}} \tag{8}$$

where $Bin\{r_j \in Bnd\}$ is the binary index, it equals to 1 when r_j is on the image boundary or 0 otherwise.

The original foreground likelihood is estimated based on the background estimation. We search the regions with high contrast to the background template. That is,

$$u_i^f(0) = \sum_{j=1}^{N} u_j^{b*}(0) \times \|\mathbf{c}_i - \mathbf{c}_j\|. \tag{9}$$

We utilize the normalized foreground and background estimation to initialize the system input signal, that is, $\mathbf{u}^b(0) = \text{Norm}(\mathbf{u}^{b*}(0))$, and $\mathbf{u}^f(0) = \text{Norm}(\mathbf{u}^{f*}(0))$.

3.3. System Parameters

The property of the linear control system hinges on the time-varying system parameters $\mathbf{A}(t)$, $\mathbf{B}(t)$, $\mathbf{C}(t)$, $\mathbf{K}(t)$. Thus we focus on the parameters definition to construct a specialized linear control system for salient region detection.

As shown in (1), $\mathbf{A}(t)$ and $\mathbf{B}(t)$ adjust the effect of the previous state value $\mathbf{x}(t)$ and the system input $\mathbf{u}(t)$. Usually, a non-isolated image region shares similar saliency value with its homochromatic neighbors. Similarly, a non-isolated node of LFCS tends to share common state value with its strongly connected neighbors. In contrast, for an isolated node, the situation is opposite. Therefore, for a non-isolated region, we need to assign larger value to $\mathbf{A}(t)$ and smaller value to $\mathbf{B}(t)$; for an isolated region, $\mathbf{A}(t)$ should be smaller but $\mathbf{B}(t)$ should be larger. In addition, $\mathbf{A}(t)$ should also have the function to homogenize the state of similar node pairs. Considering all the mentioned factors, we define $\mathbf{A}(t)$ and $\mathbf{B}(t)$ as constants, given as follows,

$$\mathbf{A} = \mathbf{A}(t) \equiv \mathbf{GW},\ \mathbf{B} = \mathbf{B}(t) \equiv \mathbf{I} - \mathbf{G}, \tag{10}$$

where $\mathbf{W} = [w_{ij}]_{N\times N}$ is the adjacent connectivity matrix which is able to homogenize the state of nodes, $\mathbf{G} = \text{diag}([g_1, g_2, \cdots, g_N])$ is the non-isolation matrix.

Parameter $\mathbf{C}(t)$ reveals the relationship between system state and output vector. State equation in (1) tends to homogenize the saliency values of similar superpixels by controlling the state update, but meanwhile, the distinction between foreground and background will be shorten. To distinctly distinguish the foreground and background, we emphasize the contrast of the global likelihood in the definition of $\mathbf{C}(t)$, which can accuralize the system output $\mathbf{y}(t)$ by comparing the foreground and background likelihood.

In our system, we parallelly measure the foreground and background likelihood of each superpixels. We define $\mathbf{C}(t)$ differently for foreground and background conditions ($\mathbf{C}^f(t)$ and $\mathbf{C}^b(t)$). In addition, $\mathbf{C}(t)$ is time-variant parameter for better adaption. Let $\mathbf{x}^f(t)$ and $\mathbf{x}^b(t)$ denote the system state vector corresponding to the foreground and background, respectively. The two $N \times N$ diagonal matrixes $\mathbf{C}^f(t)$ and $\mathbf{C}^b(t)$ are as follows,

$$\mathbf{C}^f(t)\text{=diag}([C_i^f(t)]_{i=1}^N), \mathbf{C}^b(t)\text{=diag}([C_i^b(t)]_{i=1}^N) \tag{11}$$

where $C_i^f(t)$ and $C_i^b(t)$ can be computed using

$$C_i^\chi(t)\text{=}\left(x_i^\chi(t\text{-}1) + \frac{\sum_j s_{ij} x_j^{\overline{\chi}}(t\text{-}1)}{\sum_j s_{ij} x_j^{\chi}(t\text{-}1)}(1\text{-}x_i^\chi(t\text{-}1))\right)^{-1} \tag{12}$$

where superscript $\chi, \overline{\chi} \in \{f, b\}, \chi \neq \overline{\chi}$.

Parameters $\mathbf{A}$ and $\mathbf{B}$ are able to promote local consistency, thus the feedback coefficient matrix $\mathbf{K}(t)$ is designed to adjust input matrix and accelerate convergence, thereby enhance global consistency through feedback control.

$$\mathbf{K}(t) = \tau\text{diag}\left([\sigma_1^t, \cdots, \sigma_i^t, \cdots, \sigma_N^t]\right), \tag{13}$$

where $\sigma_i^t = \text{Sign}(i,t) \cdot \sum_{j=1}^{N} s_{ij}|y_i(t) - y_j(t)|$, $\text{Sign}(i,t)$ equals $+1$ when $y_i(t) \geqslant y_i(t-1)$ or -1 otherwise, and τ is a constant.

3.4. Algorithm Framework

The details of the proposed algorithm framework are described in **Algorithm 1**. We firstly estimate the foreground and background likelihood to initialize system input using

the method proposed in Sec 3.2. Then the system will be activated and converge to steady state. The final saliency map will be generated from steady output.

Algorithm 1 LFCS-based Saliency Detection

Input: Input Image $\boldsymbol{I}$, Superpixel Number N

1: Segment $\boldsymbol{I}$ into N superpixels
2: Compute time-invariant parameters $\mathbf{W}$, $\mathbf{S}$, $\mathbf{G}$, $\mathbf{A}$, $\mathbf{B}$
3: Initialize: $\mathbf{u}^{\chi}(0)$, $\chi \in \{f, b\}$ using (8) and (9),
Set $x_i^{\chi}(0) = 0$, $t = 0$.
4: **Parallel** $\chi = f$ and $\chi = b$
do
$\mathbf{x}^{\chi}(t+1) = \mathbf{A}\mathbf{x}^{\chi}(t) + \mathbf{B}\mathbf{u}^{\chi}(t)$
$t \leftarrow t + 1$
Update $\mathbf{C}^{\chi}(t)$, and $\mathbf{y}^{\chi}(t) = \mathbf{C}^{\chi}(t)\mathbf{x}^{\chi}(t)$
Update $\mathbf{K}(t)$, and $\mathbf{u}^{\chi*}(t) = \mathbf{u}^{\chi}(t-1) + \mathbf{K}(t)\mathbf{y}^{\chi}(t)$
Normalization: $\mathbf{u}^{\chi}(t) = \text{Norm}(\mathbf{u}^{\chi*}(t))$
While $\sum_i |y_i^{\chi}(t) - y_i^{\chi}(t-1)| > \varepsilon \sum_i y_i(t-1)$
end Parallel
5: Final System Output: $\mathbf{y} = [\text{diag}(\mathbf{y}^f + \mathbf{y}^b)]^{-1}\mathbf{y}^f$
6: Contrast Enhancement: $Sal(i) = \frac{0.8}{1+e^{-15(y_i-0.4)}} + 0.2y_i$
(Similar model is also proposed and explained in [8])

Output: Saliency Map (saliency values $\{Sal(i)\}_{i=1}^{N}$)

4. EXPERIMENTAL RESULTS

We choose a complex dataset ECSSD [16] and a simple ASD [18] (2000 images in total) to test the performance when our proposed method (abbreviated to **LCS**) is used in different scene. We compare our method with 10 classic or state-of-the-art saliency detection algorithms, they are MST [7], MBD [8], RBD [9], DSR [19], GMR [20], LPS [10], MAP [11], MS [12], GS [15], BL [13], CB [14].

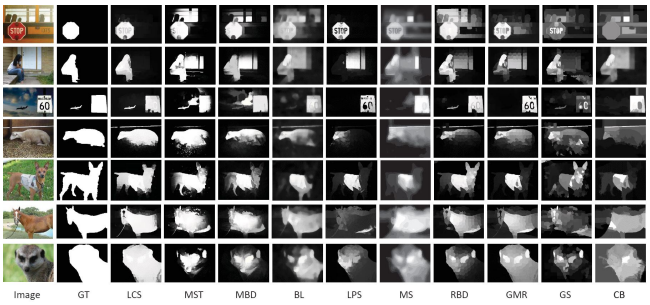

Fig. 4. Visual comparison among different methods. GT is short for ground truth.

We quantitatively evaluate using two credible saliency metrics: $F_\beta$measure-Threshold curve [7] and weighted F-Measure [21]. The quantitative performance evaluation results on ECSSD and ASD dataset are shown in Fig. 5. The figures show that our method achieves the highest F_β-Measure over most thresholds, and gives highest weighted F-Measure score among all compared state-of-the-arts.

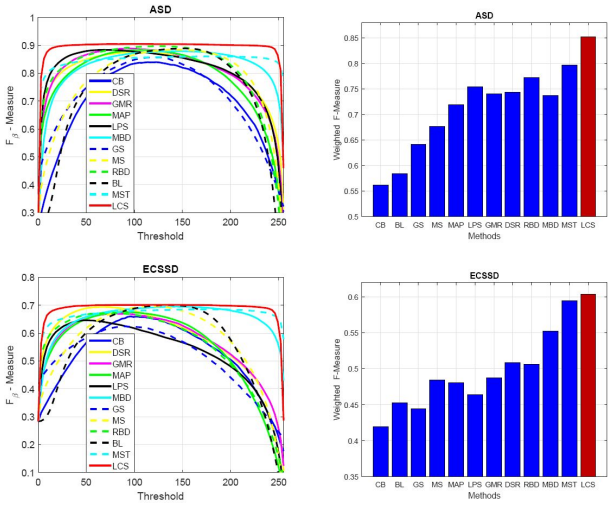

Fig. 5. Quantitative comparison among different saliency detection methods on ASD and ECSSD datasets. The left column shows comparison of "$F_\beta$measure-Threshold" curves, the right column shows comparison of weighted F-measures.

For intuitive comparison, we also choose some sample saliency maps of our method and compared method, as shown in Fig. 4. These maps prove that our method can effectively distinguish the salient objects from the background, even in complex scene.

5. CONCLUSION

Linear feedback control system (LFCS) model, the basic model in automatic control theory, has been widely used in the fields of system analysis and signal processing. In this paper, we explore the internal relation between LFCS and visual saliency. The proposed LFCS-based approach is devoted to overcoming the challenging problems in saliency detection and improving overall performance. To achieve this goal, we constructed a specialized LCFS model by analyzing multiple saliency cues from image information and structural features. This design ensures that the system reaches the steady state on an accurate salient objects segmentation. Experiments on different standard datasets demonstrate that our proposed method can achieve above goals and consistently outperform the state-of-the-art methods.

6. ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (No. 61571326, 61471262, 61520106002) and National Natural Science Foundation of Tianjin (No. 16JCQNJC00900).

7. REFERENCES

[1] Shengfeng He, Rynson WH Lau, Wenxi Liu, Zhe Huang, and Qingxiong Yang, "Supercnn: A superpixelwise convolutional neural network for salient object detection," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 330–344, 2015.

[2] Lijun Wang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Deep networks for saliency detection via local estimation and global search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3183–3192.

[3] Nian Liu and Junwei Han, "Dhsnet: Deep hierarchical saliency network for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 678–686.

[4] Shuang Li, Huchuan Lu, Zhe Lin, Xiaohui Shen, and Brian Price, "Adaptive metric learning for saliency detection," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3321–3331, 2015.

[5] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[6] Zhi Liu, Wenbin Zou, and Olivier Le Meur, "Saliency tree: A novel saliency detection framework," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1937–1952, 2014.

[7] Wei-Chih Tu, Shengfeng He, Qingxiong Yang, and Shao-Yi Chien, "Real-time salient object detection with a minimum spanning tree," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, pp. 2334–2342, 2016.

[8] Jianming Zhang, Stan Sclaroff, Zhe Lin, Xiaohui Shen, Brian Price, and Radomir Mech, "Minimum barrier salient object detection at 80 fps," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1404–1412.

[9] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun, "Saliency optimization from robust background detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2814–2821.

[10] Hongyang Li, Huchuan Lu, Zhe Lin, Xiaohui Shen, and Brian Price, "Inner and inter label propagation: salient object detection in the wild," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3176–3186, 2015.

[11] Jingang Sun, Huchuan Lu, and Xiuping Liu, "Saliency region detection based on markov absorption probabilities," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1639–1649, 2015.

[12] Na Tong, Huchuan Lu, Lihe Zhang, and Xiang Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1035–1039, 2014.

[13] Na Tong, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Salient object detection via bootstrap learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1884–1892.

[14] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, and Shipeng Li, "Automatic salient object segmentation based on context and shape prior.," in *BMVC*, 2011, vol. 6, p. 9.

[15] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *European Conference on Computer Vision*. Springer, 2012, pp. 29–42.

[16] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1155–1162.

[17] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, "Slic superpixels," Tech. Rep., 2010.

[18] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*. IEEE, 2009, pp. 1597–1604.

[19] Xiaohui Li, Huchuan Lu, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via dense and sparse reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2976–2983.

[20] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3166–3173.

[21] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal, "How to evaluate foreground maps?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 248–255.