PROPAGATION BASED SALIENCY DETECTION FOR INFRARED PEDESTRIAN IMAGES

Yu Zheng, Fugen Zhou, Lu Li, Xiangzhi Bai**

Image Processing Center, Beihang University, 100191, Beijing, China

ABSTRACT

Saliency detection is popular in image processing, but it is still a challenging problem for infrared pedestrian images. In this paper, an effective saliency detection method for infrared pedestrian images is proposed. Taking into consideration the characteristics of pedestrians including luminance and shape, the MSER-based local stableness (MLS) is firstly introduced. Then vertical edge-weighted contrast (VEC) is calculated. Finally, an intra-scale and inter-scale neighborhood based saliency propagation method is constructed to optimize and integrate the two features. Extensive experiments demonstrate the effectiveness of the proposed saliency method for infrared pedestrian images.

Index Terms— infrared images, pedestrian, saliency, propagation

1. INTRODUCTION

Saliency detection focuses on the important and distinctive regions in images, which has been applied in many visual tasks, like object segmentation [1], image compression [2], and object detection [3]. However, it is still challenging to detect saliency in infrared pedestrian images, which is significant in intelligent transportation system.

Previous saliency methods mainly tried to explore the low-level cues of saliency object, like color [4], orientation [5], and texture [6]. With the further development of image understanding, some prior principles are introduced, like center prior [7] boundary prior [8] and so on.

Recently, probability propagation algorithms attract more and more attention in saliency detection and achieve state-of-the-art performance. Markov chains [9], manifold ranking [10], and random walks [11] are the most frequently used propagation method. Moreover, Gong [12] employed learning-to-teach and teaching-to-learn strategies as the saliency propagation method which improved the saliency. Li [13] proposed a co-transduction algorithm to fuse both the boundary and objectness labels through an inter propagation scheme. Qin [14] proposed a cellular automata based saliency propagation to exploit the intrinsic relevance between neighboring cells to improve the saliency.

Even though various saliency models are proposed, most of them are designed for visible images. And when they are directly applied to infrared images, the performance is bad due to the low contrast, low SNR, and lack of color information. KO [15] calculated the luminance saliency map which estimated the luminance contrast using a center-surrounded scheme. Zhang [16] proposed an associative saliency, generated from both region and edge contrast. However, most of the existing saliency models for infrared images cannot extract effective and robust features to represent the salient objects. Therefore, they are not effective for complicate scenes. Taking the above problems into consideration, this paper proposed a novel saliency detection model for pedestrians in infrared images. The main contributions of this paper can be concluded as:

- Two features, MSER-based local stableness (MLS) and vertical edge-weighted contrast (VEC), are proposed to combine the characteristics of infrared pedestrians.
- An intra-scale and inter-scale neighborhood based saliency propagation method is constructed to optimize and integrate MLS and VEC.

2. METHOLOGY

Fig. 1 shows the diagram of the proposed method. Firstly, the infrared image is over-segmented into superpixels. Then, MSER-based local stableness (MLS) and Vertical edge-weighted contrast (VEC) are proposed to describe the saliency of pedestrians. Finally, a mutual guidance-based propagation method is introduced to obtain the final saliency.

2.1. MSER-based local stableness (MLS)

By observing the luminance distribution of pedestrians in infrared images, we can find that pedestrians have distinctive difference with their neighboring regions. Besides, the pedestrian regions are usually completely surrounded by regions with lower luminance. These features can be taken as the local stableness of pedestrians which can make the pedestrians more prominent. In our experiments, we find that pedestrians can be well presented by MSER [17], which is defined by an extremal property of its intensity function in the region and on its outer boundary.

To calculate MSER in an input image I, the extremal regions are considered and defined as R_i:

$$\forall p \in R_i, \forall q \in boundary(R_i) \rightarrow I(p) \geq I(q), \qquad (1)$$

where $I(p)$ is the luminance of pixel p in the region

Corresponding authors: lilu@buaa.edu.cn, jackybxz@buaa.edu.cn

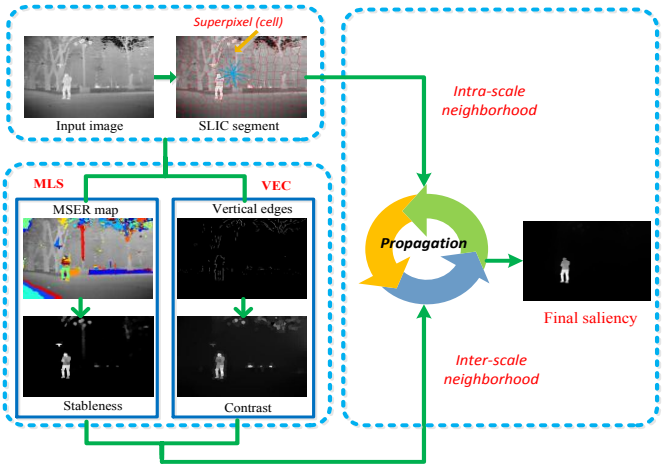


Fig. 1. The diagram of the proposed method

and $I(q)$ is the luminance of pixel q on its outer boundary. The extremal regions are identified as connected regions within binary threshold images I_{bin}^{g}:

$$I_{bin}^{g} = \begin{cases} 1 & I \geq g \\ 0 & otherwise \end{cases} \qquad g \in [\min(I) \ \ \max(I)]. \tag{2}$$

To generate MSER from R_i, a stableness value ψ is calculated for each connected region:

$$\psi(R_i^g) = (|R_i^{g+\delta} - R_i^{g-\delta}|)/|R_i^g|, \tag{3}$$

where R_i^g is a region obtained by thresholding with value g, and δ is a stability range parameter. If ψ is below an empirical threshold, R_i^g can be taken as MSER. Thus, the final MSER contain K stable regions $SR = \{SR_1, SR_2, \cdots SR_K\}$.

(a) (b) (c)

Fig. 2. (a) Pixel based stableness F (b) Superpixel based stableness without luminance filter (c) Superpixel based stableness with luminance filter.

And then, how many stable regions each pixel belongs to are applied to measure the stableness F. Pixels with higher F value are more likely to belong to pedestrians:

$$F(p) = \sum_{i=1}^{K} E_i(p) \qquad E_i(p) = \begin{cases} 1 & p \in SR_i \\ 0 & otherwise \end{cases}, \tag{4}$$

where $E_i(p)$ represents whether pixel p belongs to the i^{th} stable region. As the internal distribution of luminance in infrared pedestrian is imhomogeneous, the SLIC is applied to segment the image into N homogeneous superpixels $SP = \{SP_1, SP_2, \cdots SP_N\}$. And then, the saliency value of each superpixel can be calculated by mapping the stableness of pixels to the superpixels which they belong to. Fig. 2(b) proves that superpixel based stableness can reduce the imhomogeneous saliency distribution inside human body and partly exclude small noises of background.

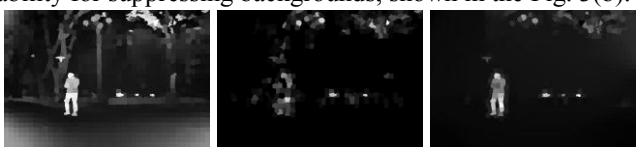
However, there are some stable regions, like street lamps, tree trunks, that may be mistakenly distributed with high saliency values. To suppress these interferences in background, the luminance filter is combined to enhance the stableness. The final MLS can be calculated as:

$$S_{MLS}(i) = \frac{\sum_{p \in SP_i} F(p)}{|SP_i|} \left| \frac{\sum_{p \in SP_i} I(p)}{|SP_i|} - I_{\mu} \right|^2, \tag{5}$$

where I_{μ} is the average luminance of the input image, and $|SP_i|$ is the area of the i^{th} superpixel. Since luminance inside the pedestrian is much higher than background, the luminance filter performs well in improving the stableness, as can be seen in Fig. 2(c).

2.2. Vertical edge-weighted contrast (VEC)

Although infrared images have low contrast and low resolution, the luminance distributions of pedestrians are different from background. Thus, the luminance contrast is an important feature of infrared pedestrians. Meanwhile, vertical shape feature is another remarkable characteristic of pedestrians, and regions with strong vertical edges are more likely to belong to pedestrians. Therefore, a vertical edge - weighted luminance contrast is proposed, which has a good ability for suppressing backgrounds, shown in the Fig. 3(b).



(a) (b) (c)

Fig. 3. (a) Contrast without weight (b) Contrast with vertical edge weight (c) Contrast with both vertical edge weight and luminance weight.

However, background regions along the edges may be wrongly highlighted by vertical edge weight at the same time. Also, regions inside the human body may not contain edges, thus superpixels in pedestrian regions may be mistakenly suppressed. Considering that infrared pedestrians have higher luminance than background, the luminance of each superpixel is applied as a complementary factor to solve those problems. Therefore, the VEC can be calculated as:

$$S_{VEC}(i) = w_i \cdot \sum_{j=1}^{N} |I_i - I_j| \cdot \exp(|c_i - c_j|), \tag{6}$$

where c_i and I_i respectively represent the coordinates and average luminance of the i^{th} superpixel, and w_i is the weight composed of the vertical edge weight and luminance weight. Obviously, Fig. 3(c) shows that the luminance weight can not only fill the holes caused by the edge weight, but also suppress superpixels around the pedestrians.

To calculate w_i, the probability of boundary [18] is firstly used to detect the boundary map M^{pb} of the input

image. And then, the vertical edge (VE) is obtained by detecting the vertical gradient of M^{pb} . Therefore, the vertical edge weight of each superpixel can be defined as:

$$w_i^{ve} = \frac{1}{|E_i|}\sum_{p\in E_i} p^{ve}, \tag{7}$$

where p^{ve} is the vertical gradient value of pixel p, and E_i represents the edge pixel set of the i^{th} superpixel. Then, we take the average gray value of each superpixel as its luminance weight, $w_i^l = I_i$. At last, by integrating the vertical edge weight and luminance weight, w_i can be calculated as:

$$w_i = w_i^{ve} + w_i^l. \tag{8}$$

This equation formulates the rule that superpixels with more vertical edges and higher luminance value have higher probability belonging to the pedestrian.

2.3. Mutual guidance based saliency propagation

To integrate MLS and VEC, the mutual guidance based saliency propagation is introduced, which is inspired by cellular automata [14]. Cellular automata is composed by three factors: cell, neighborhood and updating rules.

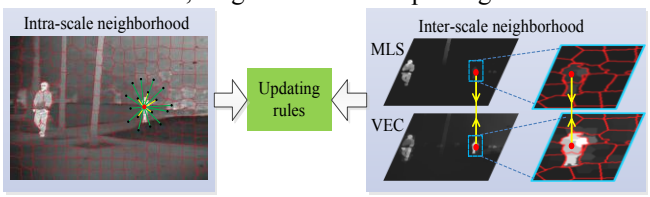


Fig. 4. Example of intra-scale neighborhood and inter-scale neighborhood.

In this work, each superpixel is taken as a cell, which is generated by SLIC [19]. Different from previous works just taking the neighboring superpixels as neighborhood, the neighborhood in our method consists of two parts: intra-scale neighborhood and inter-scale neighborhood.

Intra-scale neighborhood. Based on the intuition that neighboring cells are likely to share similar saliency values, the saliency of each cell should be influenced by its neighbors. Therefore, the intra-scale neighborhood of a cell (red dot in Fig. 4) is defined as its direct neighboring superpixels (yellow connections) and the direct neighbors of these superpixels (green connections). Also, a cell should depend more on the neighbors which have similar luminance with it. Thus, the similarity of gray level is used to measure the influence strength of each superpixel to its neighboring cell. This can be treated as a similarity matrix $M = [m_{ij}]_{N\times N}$:

$$m_{ij} = \begin{cases} \exp(\frac{|I_i - I_j|}{\sigma^2}) & j \in NB(i) \\ 0 & i = j \ or \ otherwise \end{cases}, \tag{9}$$

where $NB(i)$ is the intra-scale neighborhood of the i^{th} cell. Therefore, if a cell is mostly surrounded by cells with high saliency, its saliency will gradually approach to a higher value. Thus cells inside pedestrians will be gradually highlighted. Fig. 5(a) and (b) show that pedestrians can be continuously highlighted by the intra-scale neighborhood. But some small targets may be wrongly suppressed because of the strong influence of their surrounding background. To solve this problem, the inter-scale neighborhood is proposed.

Inter-scale neighborhood. The final saliency value of each cell should be approximately consistent with their corresponding values of MLS and VEC. Therefore, cells with the same coordinate in MLS map and VEC map are also taken as the inter-scale neighborhood of each other (red dots in Fig.4 is a pair of neighborhoods). Only using intra-scale neighborhood, cells like the examples in Fig. 4 are easy to be mistakenly suppressed, because they are mostly surrounded by the backgrounds. However, by utilizing the inter-scale neighborhood, MLS and VEC can complement each other to produce a more accurate result. Seen from Fig. 5, the inter-scale neighborhood can solve the problem of wrongly suppressing small targets.

Updating rules. To balance the influence strength of intra-scale and inter-scale neighborhoods, coherence vector $C = \{c_1, c_2, \cdots, c_N\}$ is defined as: $c_i = 1/\max(m_{ij})$, $j \in [1, N]$. Like the definition of C, if a cell has large difference with its intra-scale neighborhood, its saliency value in the next state should be more dependent on its inter-scale neighborhood. Therefore, to propagate the saliency of each cell in MLS and VEC, the updating rules are defined as:

$$\begin{cases} S_{MLS}^t = S_{MLS}^{t-1} + \underbrace{(1-C)\cdot M \cdot S_{MLS}^{t-1}}_{Intra-scale} + \underbrace{C \cdot S_{VEC}^{t-1}}_{Inter-scale} \\ S_{VEC}^t = S_{VEC}^{t-1} + \underbrace{(1-C)\cdot M \cdot S_{VEC}^{t-1}}_{Intra-scale} + \underbrace{C \cdot S_{MLS}^{t-1}}_{Inter-scale} \\ S^t = S_{MLS}^t \cdot S_{VEC}^t \end{cases}, \tag{10}$$

MLS and VEC can influence each other at every iteration. It is also important to decide when to stop the iteration. Previous cellular automata usually set a fixed value as the maximum iteration, but this way is not always suitable. In the proposed method, the termination of the iteration is decided by checking the average variance of $\{S^1, S^2, \cdots S^t\}$:

$$check = \mathrm{var}(S^1, S^2, \cdots S^t), \tag{11}$$

And when the value of $check$ reaches the empirical threshold θ at the T^{th} calculation, the iteration stops At last, the final saliency map can be obtained as: $S_{final} = S^T$.

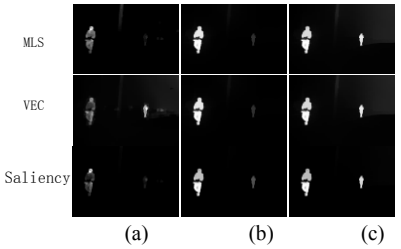


Fig. 5. (a) The original result of Eq. 5 and Eq. 6 (b) Propagation with only intra-scale neighborhood (c) Propagation with both intra-scale and inter-scale neighborhood.

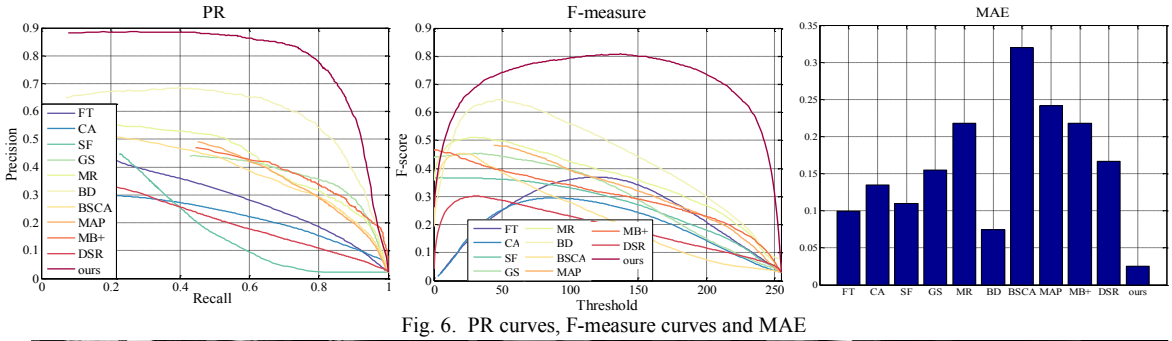


Fig. 6. PR curves, F-measure curves and MAE

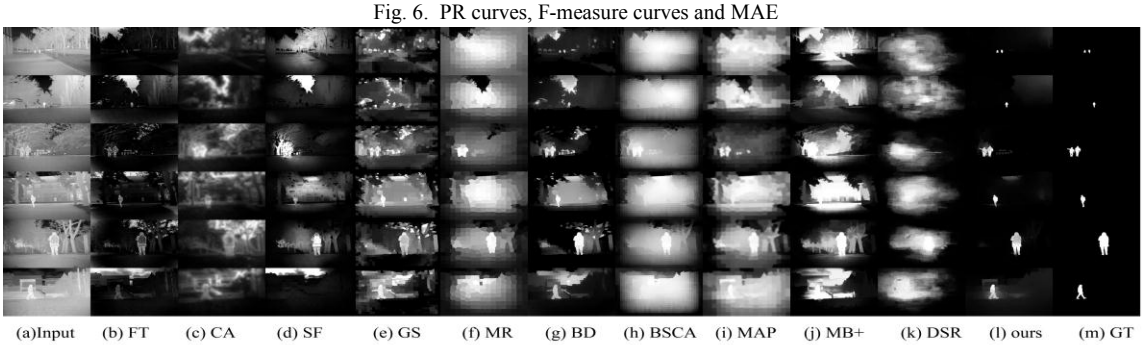


Fig. 7. Visual comparisons

3. EXPERIMENTS

3.1. Dataset

To evaluate the effectiveness of the proposed method, a dataset containing 400 infrared images is used in this work. The images were scanned by a Tau 2 LWIR camera. There are totally 634 pedestrians in the dataset, which contains 210 single-person images and 180 multiple-person images in 31 different scenes. Also, different postures, and various sizes are all taken into consideration. Moreover, the human-segmented ground truths (GT) are also provided.

Ten state-of-art saliency models are used as comparison method, including FT [20], CA [21], SF [22], GS [23], MR [10], BD [24], BSCA [14], MAP [25], MB+ [26], DSR [27].

3.2. Quantitative comparisons

To quantitatively compare different saliency methods, the precision-recall curve (PR curve) [20] is firstly introduced. It is obvious in Fig. 6 that the proposed method achieves the highest precision in almost all the recall range [0, 1]. This means most of the detected regions belong to the pedestrians. From the point of recall, our method also performs the best in the entire precision range, which implied that our method can detect more pedestrian regions than other methods.

Considering that high precision and high recall are always both required, F-measure[28] is applied as a complement, which is the combination of precision and recall,. From Fig. 6, it is easy to find that the proposed method obtains a much higher F score than other saliency methods.

Furthermore, Mean Absolute Error (MAE) [22] is also used for performance evaluation, which measures the discrepancy between the saliency map and ground truth. The lower the MAE value is, the more accurate the final saliency will be. As can be seen in Fig. 6, the proposed method achieves the best performance among all the methods.

3.3. Subjective comparisons

The visual comparison is presented in the Fig. 7, which obviously demonstrates that the proposed method outperforms all the other methods. We can find that most of the methods assign high saliency value to the background and the pedestrians are suppressed. Though BD [24] can separate the salient region from the background to some extent, highlighted background still exists substantially. However, our proposed method can both highlight the foreground and suppress the background well. Also, our method not only performs well in simple images with single large pedestrian, but also gets robust in complicated cases with multiple small pedestrians.

4. CONCLUSION

In this paper, by analyzing the luminance and shape features of pedestrians in infrared images, a novel saliency detection method is proposed. Two features: MSER-based local stableness and vertical edge weighted contrast, are proposed to describe the saliency of pedestrians. Also, an intra-scale and inter-scale based saliency propagation method is introduced to optimize and integrate the two features. All the experiments on the infrared pedestrian dataset verified the effectiveness of the proposed method.

5. REFERENCES

[1] X. Liao, H. Xu, Y. Zhou, K. Li, W. Tao, Q. Guo, and L. Liu, “Automatic image segmentation using salient key point extraction and star shape prior,” *Signal Processing,* vol. 105, no. 12, pp. 122-136, 2014.

[2] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," *IEEE International Conference on Computer Vision.* pp. 2232-2239, 2009.

[3] B. Alexe, T. Deselaers, and V. Ferrari, “V.: What is an object,” *IEEE Conference on Computer Vision and Pattern Recognition. ,* vol. 73-80, pp. 73-80, 2010.

[4] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, "Global contrast based salient region detection," *IEEE Conference on Computer Vision and Pattern Recognition. .* pp. 409-416, 2015.

[5] V. Gopalakrishnan, Y. Hu, and D. Rajan, “Salient Region Detection by Modeling Distributions of Color and Orientation,” *IEEE Transactions on Multimedia,* vol. 11, no. 5, pp. 892-905, 2009.

[6] C. Scharfenberger, A. Wong, K. Fergani, and J. S. Zelek, "Statistical Textural Distinctiveness for Salient Region Detection in Natural Images," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 979-986, 2013.

[7] C. Yang, L. Zhang, and H. Lu, “Graph-Regularized Saliency Detection With Convex-Hull-Based Center Prior,” *IEEE Signal Processing Letters,* vol. 20, no. 7, pp. 637-640, 2013.

[8] J. Wang, H. Lu, X. Li, N. Tong, and W. Liu, “Saliency detection via background and foreground seed selection,” *Neurocomputing,* vol. 152, pp. 359-368, 2015.

[9] B. Jiang, L. Zhang, H. Lu, C. Yang, and M. H. Yang, "Saliency Detection via Absorbing Markov Chain," *IEEE International Conference on Computer Vision.* pp. 1665-1672, 2013.

[10] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, "Saliency Detection via Graph-Based Manifold Ranking," *2013 IEEE Conference on Computer Vision and Pattern Recognition.* pp. 3166-3173, 2013.

[11] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs to model saliency in images," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 1698-1705, 2009.

[12] C. Gong, D. Tao, W. Liu, and S. J. Maybank, "Saliency propagation from simple to difficult," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 2531-2539, 2015.

[13] H. Li, H. Lu, Z. Lin, and X. Shen, “Inner and Inter Label Propagation: Salient Object Detection in the Wild,” *IEEE Transactions on Image Processing,* vol. 24, no. 10, pp. 3176-3186, 2015.

[14] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via Cellular Automata," *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[15] B. C. Ko, D. Y. Kim, and J. Y. Nam, “Detecting humans using luminance saliency in thermal images,” *Optics letters,* vol. 37, no. 20, pp. 4350-4352, 2012.

[16] L. Zhang, Y. Zhang, W. Wei, and Q. Meng, "An associative saliency segmentation method for infrared targets," *IEEE International Conference on Image Processing.* pp. 4264-4268, 2013.

[17] M. Donoser, and H. Bischof, "Efficient Maximally Stable Extremal Region (MSER) Tracking," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 553-560, 2006.

[18] D. R. Martin, C. C. Fowlkes, and J. Malik, “Learning to detect natural image boundaries using local brightness, color, and texture cues,” *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 26, no. 5, pp. 530-49, 2004.

[19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 34, no. 11, pp. 2274-2282, 2012.

[20] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," *IEEE International Conference on Computer Vision and Pattern Recognition.* pp. 1597-1604, 2009.

[21] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 34, no. 10, pp. 1915-26, 2012.

[22] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 733-740, 2012.

[23] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic Saliency Using Background Priors," *European Conference on Computer Vision.* pp. 29-42, 2012.

[24] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency Optimization from Robust Background Detection," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 2814-2821, 2014.

[25] J. Sun, H. Lu, and X. Liu, “Saliency Region Detection Based on Markov Absorption Probabilities,” *IEEE Transactions on Image Processing,* vol. 24, no. 5, pp. 1639-1649, 2015.

[26] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum Barrier Salient Object Detection at 80 FPS," *IEEE International Conference on Computer Vision.* pp. 1404-1412, 2015.

[27] H. Lu, X. Li, L. Zhang, and R. Xiang, “Dense and Sparse Reconstruction Error Based Saliency Descriptor,” *IEEE Transactions on Image Processing,* vol. 25, no. 4, pp. 1592, 2016.

[28] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical Saliency Detection," *IEEE Conference on Computer Vision and Pattern Recognition.* pp. 1155-1162, 2013.