SALIENT-REGION DETECTION IN A MULTI-LEVEL FRAMEWORK OF IMAGE SMOOTHING WITH OVER-SEGMENTATION

Hong-Yun Gao and Kin-Man Lam

Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

Saliency detection is one of the extraordinary abilities of the human visual system; it also provides a powerful tool for predicting where people tend to focus in the free-viewing process. In this paper, we propose a novel salient-object detection method which applies an over-segmentation-based saliency detection algorithm to multi-level smoothed images. The original image is initially subjected to smoothing based on multi-level L_0 gradient minimization; this can characterize its fundamental constituents while diminishing the insignificant details. Then, segment-based saliency computation is applied to the multi-level smoothed images to produce a series of intermediate saliency maps. The final saliency map is generated by combining the intermediate saliency maps. The proposed method is compared with six existing saliency models, and achieves the best performance in terms of Precision, Recall and F-measure, as well as in terms of the area under the ROC curve (AUC).

Index Terms— Salient-region detection, image smoothing, over-segmentation, multi-level framework

1. INTRODUCTION

Visual saliency detection has been a fundamental problem in neuroscience, image processing and computer vision for a long time, and it remains a challenging problem. Originally, saliency detection was used as a tool to predict eye-fixations in neuroscience. In recent years, it has been extended to many applications in image processing and computer vision, e.g. object detection and recognition [1], image cropping [2], photo collage [3], and image compression [4].

One of the earliest computational models was proposed by Itti *et al*. [5]. Their algorithm is based on the center-surround contrast, which is implemented as the difference between the fine and coarse scales in the intensity, color and orientation maps. The saliency map is generated by the combination of the three conspicuity maps in the three feature channels. Ma and Zhang [6] simulated human perception and proposed a saliency model based on local contrast analysis. Harel *et al*. [7] proposed a visual saliency model in a graph theory framework, namely Graph-Based Visual Saliency. Hou and Zhang [8] proposed a frequency-domain saliency model based on spectral residual. Achanta *et al*. [9] proposed using Difference of Gaussian (DoG) filters to eliminate redundant information, and to output full-resolution saliency maps with well-defined boundaries of salient objects. Liu *et al*. [10] proposed a new learning algorithm based on the conditional random field to combine features – e.g. multi-scale contrast, color spatial distribution and center-surround histogram for salient-object detection. Goferman *et al*. [11] proposed a context-aware saliency model by comparing each image patch with surrounding image patches. Zhang *et al*. [12] proposed a compactness measure based on over-segmented the image to detect salient objects. Hou *et al*. [13] proposed a image signature definition, which is defined as the sign function of the discrete Cosine transform of an image, to predict human-fixation points. Imamoglu *et al*. [14] proposed using low-level features obtained in the wavelet transform domain to construct the saliency map.

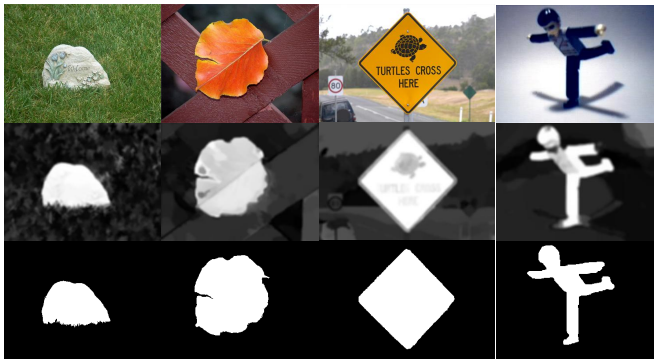


Fig. 1 Saliency map. From top to bottom: the original sample images, our saliency map, ground-truth images.

In this paper, we propose a multi-level saliency-detection framework that combines the saliency maps of multi-level smoothed images to form the final saliency map. Our approach consists of three steps. In the first step, the original image is subjected to multi-level smoothing from fine to coarse, based on L_0 gradient minimization. This can preserve the important information and eliminate noise in images. Secondly, over-segmentation-based saliency detection is applied to each smoothed image to generate multi-level intermediate saliency maps. Finally, these intermediate saliency maps are combined to form the final saliency map.

The main contributions of this paper are twofold: firstly, we combine multi-level image smoothing with over-segmentation to form a holistic framework for salient-object detection; secondly, we propose a new over-segmentation-based saliency computation method based on the color and the spatial similarity, which can effectively locate salient objects. Compared with some existing methods, our proposed method provides can detect the whole salient object accurately, as can be seen in Fig. 1.

The rest of this paper is organized as follows. In Section 2, we describe the proposed method in detail. The experimental results for our method and the comparison with several existing methods are presented in Section 3. Finally, Section 4 concludes this paper.

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

Fig. 2. From top to bottom: images from the MSRA dataset [10]; smoothed images by L_0 smoothing, with parameter λ at 0.01; and smoothed images by L_0 smoothing, with parameter λ at 0.05.

2. THE PROPOSED METHOD

2.1. Image Smoothing via L_0 Gradient Minimization

Most of the existing salient-object detection algorithms do not pay much attention to the smoothing process. However, noise is common in all images, resulting in high-contrast regions that are unimportant from either a local or a global perspective, e.g. in Fig. 2, the sea in (a), the grass in (b) and (c), and the woods in (d). Xu *et al*. [15] proposed an image-smoothing method for sharpening major edges by increasing the steepness of the transition while eliminating a manageable number of low-amplitude structures, as can be seen in the second and third rows of Fig. 2. The basic idea of the L_0 smoothing filtering is to confine the number of intensity changes to the neighboring pixels; this mathematically links to the L_0 norm in terms of the information-sparsity pursuit. For an input image I, and the computed result S, the gradient of each pixel p is calculated as the color difference between neighboring pixels in the x and y directions. The number of pixels whose gradient magnitudes are not zero is defined as:

$$C(S) = \#\{p \mid\mid \partial_x S_p \mid + \mid \partial_y S_p \mid \neq 0\}, \quad (1)$$

where $\partial_x S_p$ and $\partial_y S_p$ are the gradients for pixel p along the x and the y direction, respectively. Then, with this definition, S is obtained by solving:

$$\min_S \left\{ \sum_p (S_p - I_p) + \lambda \cdot C(S) \right\}, \quad (2)$$

where λ is a smoothing parameter to control the significance of $C(S)$. For color images, the gradient magnitude is defined as the sum of gradient magnitudes in the r, g, and b channels.

2.2. Saliency Detection through Over-Segmentation

Many saliency detection methods involve segmentation as a pre-processing step to preserve the boundary of objects, e.g. [16], [12] and [17]. Xie and Lu [16] used superpixels as the minimum processing unit in the prior map calculation. Zhang *et al.* [12] combined mean-shift segmentation and quad mesh in producing small size segments, and proposed an over-segmentation-based salient-object detection method. Sun *et al.* [17] proposed a boundary-based prior map and a soft-segmentation-based convex hull to improve the saliency detection result.

With the L_0 smoothing filtering, the performance of the mean-shift segmentation [18] is greatly improved since high-contrast background is partly smoothed by L_0 filtering to avoid false-alarms. However, since the segments may vary widely in the size and shape, it is desirable to set an upper limit on the size of the segments. The maximum number of pixels in each segment can be confined with one parameter in the mean-shift segmentation algorithm, but the shape cannot be controlled by the algorithm. Therefore, we extend the idea of [12] by over-segmenting the original image into 256 small rectangular blocks, with the height and width of the blocks being $H/16$ and $W/16$, respectively. With the initial results of mean-shift segmentation, we can further segment the fixed-size blocks into smaller-sized segments if a block contains more than one label. In that case, each label is considered as one segment.

After over-segmenting the original image into multiple segments, we define saliency on these segments as the basic processing units, based on the color and spatial similarity.

2.2.1. Color Similarity

We represent the color feature at each pixel (x, y) as $F(x, y) = (KW(x, y), RG(x, y), BY(x, y))$, where $KW(x, y)$, $RG(x, y)$, and $BY(x, y)$ are the black-white, red-green, and blue-yellow color opponents, respectively. We propose to use the luminance map from the *YUV* color space as the black-white color opponent in our method, which is defined as follows:

$$KW(x, y) = 0.299r(x, y) + 0.587g(x, y) + 0.114b(x, y), \quad (3)$$

where $r(x, y)$, $g(x, y)$ and $b(x, y)$ are the three channels of the *RGB* color space. As for $RG(x, y)$ and $BY(x, y)$, we follow the analysis of [5], and these two color-opponents can be defined as follows:

$$RG(x, y) = R(x, y) - G(x, y), \quad (4)$$

$$BY(x, y) = B(x, y) - Y(x, y), \quad (5)$$

where $R(x, y)$, $G(x, y)$, $B(x, y)$ and $Y(x, y)$ are four broadly-tuned color channels which can be defined as follows:

$$R(x, y) = r(x, y) - (g(x, y) + b(x, y)), \quad (6)$$

$$G(x, y) = g(x, y) - (r(x, y) + b(x, y)), \quad (7)$$

$$B(x, y) = b(x, y) - (r(x, y) + g(x, y)), \quad (8)$$

$$Y(x, y) = (r(x, y) + g(x, y))/2 + |r(x, y) - g(x, y)|/2 - b(x, y). \quad (9)$$

We define color similarity (CS) of region S_i as:

$$CS(i) = \sum_{j=1, j\neq i}^{n} e^{-\|F(i)-F(j)\|}, \quad (10)$$

where n is the number of segments in the image, and $F(i)$ and $F(j)$ are the mean color feature vectors of segments S_i and S_j, respectively. Equation (10) indicates that a segment with its color feature vector similar to the remaining segments will be assigned a larger color similarity value, while a segment with discrepant color feature vector will be assigned a smaller color similarity value.

2.2.2. Spatial Similarity

Besides the color similarity, the spatial similarity of segments also plays an important role in the determination of saliency value since the contrast of two discrepant segments attenuates as the distance between the two segments increases. Also, visual attention is often biased towards the center of an image, which is known as center-bias effect [19]. After considering the above two effects, we define spatial similarity as:

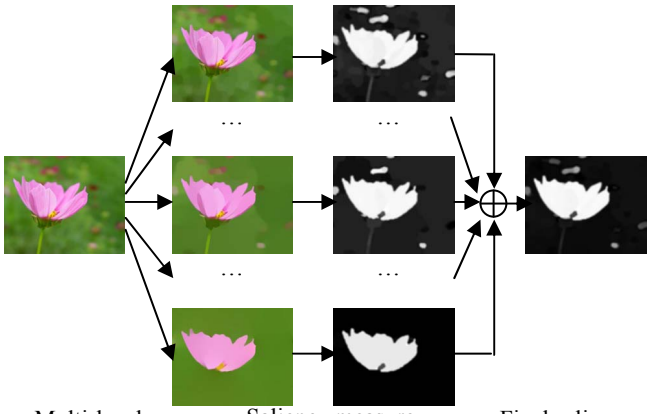


Fig. 3. Framework of the proposed method.

$$SS(i) = (n-1) - \sum_{j=1, j\neq i}^{n} \frac{\| S(i) - S(j) \|}{d}, \quad (11)$$

where n is the number of segments in the image, d is the diagonal length of the image, and $S(i)$ and $S(j)$ are the centroids of the segments S_i and S_j, respectively.

2.2.3. Saliency Map

As mentioned previously, the color and the spatial similarity are correlated with the saliency value of a particular segment. The saliency map for each segment is thus defined as:

$$SM(i) = SS(i)(1 - CS(i)), \quad (12)$$

where $SS(i)$ and $CS(i)$ are the spatial and the color similarity, respectively. Equation (12) indicates that the segment with a large spatial similarity will generate large saliency, while a segment with a small color similarity will produce small saliency since those segments tend to be background.

2.3. Saliency Map by Multi-Level Image Smoothing

Having generated the intermediate saliency map based on the color and the spatial similarity, we propose to produce the final saliency map using a multi-level L_0 smoothing framework, inspired by the multi-level segmentation framework in [20]. Fig. 3 demonstrates the framework of our proposed method. Firstly, M smoothing levels of the original image are created using the L_0 smoothing filter, which progressively smoothes the image from fine to coarse. As suggested in [15], the parameter λ is tuned in the range [0.001 0.1]. Hence, we propose the number of levels M to be 7, and the corresponding values of λ are (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1). Then, the smoothed images of different levels are individually subjected to saliency detection, as described in the previous sub-section. Finally, those intermediate saliency maps of different smoothing levels are combined together to form the final saliency map. The reason for using multi-level smoothing to enhance saliency performance is to reduce false segmentation results.

3. EXPERIMENTAL RESULTS

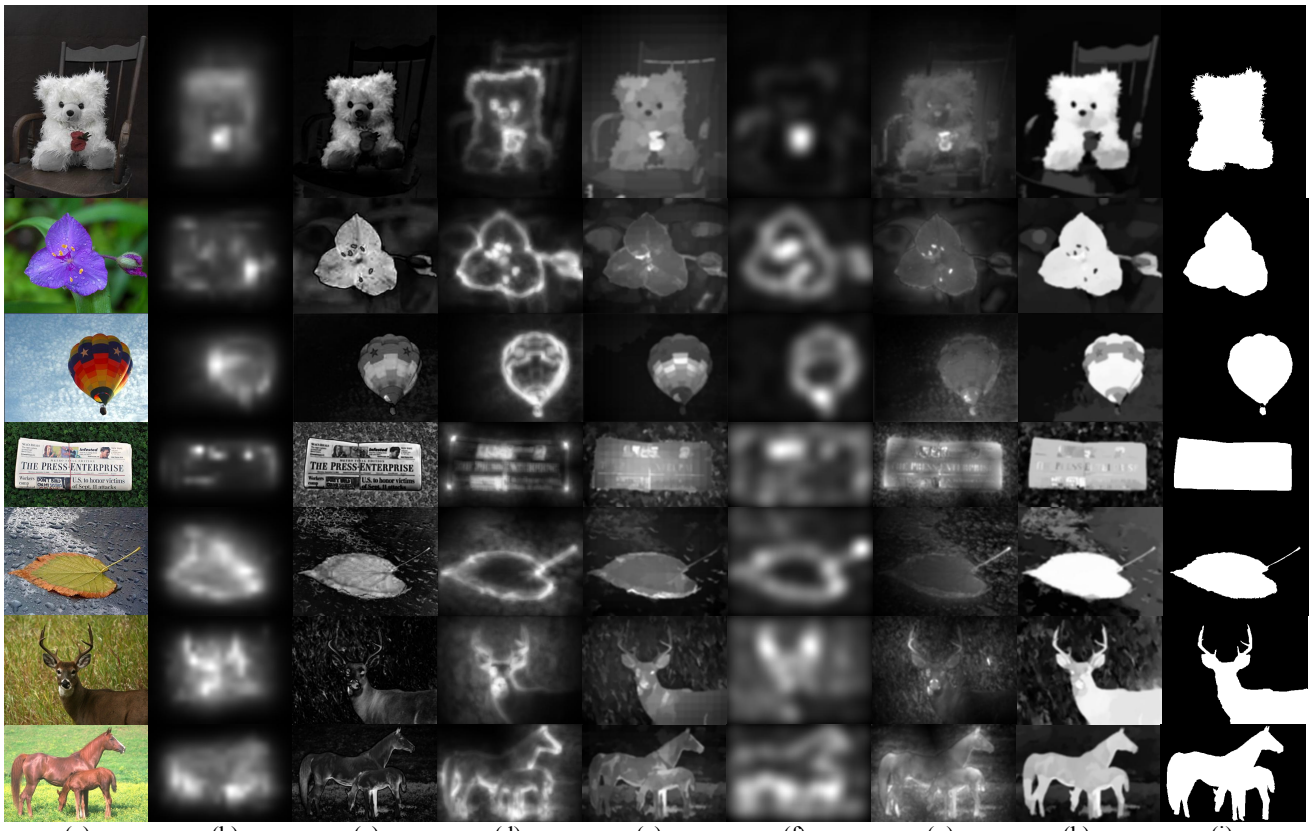
We evaluate the performance of our proposed method on two datasets, namely the MSRA dataset [10] and the ECSSD dataset [21]. The MSRA dataset [10] contains 5,000 color images, as well as the ground-truths. Besides, Achanta *et al.* [9] refined 1,000 of the ground-truths more accurately by segmenting salient objects entirely, instead of depicting them with rectangular boxes only. The ECSSD dataset [21] contains 1,000 color images and the corresponding ground-truths. In this section, we perform both qualitative and quantitative comparisons of our method with several state-of-the-art saliency-detection algorithms, namely the graph-based center surround contrast model GB [7], difference-of-Gaussian-based model FT [9], local contrast model CA [11], over-segmentation model OS [12], the spectral domain model based on discrete cosine transform DCT [13], and wavelet model WS [14].

The qualitative comparison is shown in Fig. 4, which displays some images from the MSRA and the ECSSD dataset, respectively. The first five images are from the MSRA dataset, and the last two are from the ECSSD dataset. Since our method is based on segments, the pixels in the same segment can always share the same saliency value, which means that our approach can outline an entire object, while the patch-based saliency models always have large saliency values at an object's boundary, e.g. column (d) in Fig. 4. As can be seen in the fourth and the fifth row, our method can detect the objects well with complex texture and cluttered background.

The quantitative comparison is based on three benchmarks. The first is the Precision, Recall, and F-measure; the second is the Receiver Operating Characteristic (ROC) curve; and the third is the Area Under ROC Curve (AUC). The comparisons are made on those images with accurate mean-shift segmentation results. The results are shown in Fig. 5, with independent evaluations on the MSRA dataset and the ECSSD dataset, respectively. First of all, we evaluate the performance of our method quantitatively in terms of Precision, Recall and F-measure. Precision corresponds to the percentage of the salient points that coincide with the ground-truth. Recall is the percentage of the ground-truth points that are correctly detected. F-measure is the weighted harmonic mean of Precision and Recall, and is used as an overall performance indicator. The Precision, Recall and F-measure are defined as follows:

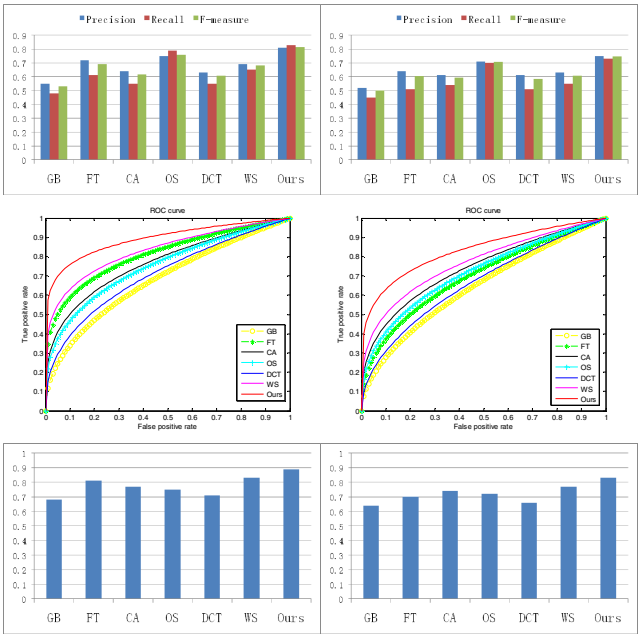
$$\text{Precision} = \frac{\sum_{x,y} s(x,y)g(x,y)}{\sum_{x,y} s(x,y)}, \quad \text{Recall} = \frac{\sum_{x,y} s(x,y)g(x,y)}{\sum_{x,y} g(x,y)},$$
$$F = \frac{(1+\alpha)\times \text{Precision} \times \text{Recall}}{\alpha \times \text{Precision} + \text{Recall}}, \quad (13)$$

where $s(x, y)$ and $g(x, y)$ are the values of the saliency mask and the ground-truth image at location (x, y), respectively. α is a positive parameter to determine the relative importance of Precision and Recall. α is chosen empirically as 0.3 in our experiments. The saliency mask is obtained by using an adaptive thresholding scheme, where the threshold is set empirically as the mean value of the final saliency map. As shown in Fig. 5, our proposed method outperforms other methods in terms of the Precision, Recall, and F-measure values on both datasets. We improve the Precision and Recall by 8% and 5.06%, respectively, on the MSRA dataset. For the ECSSD dataset, the performance is lower than the MSRA dataset since the images are more complex, with the performance on Precision and Recall increased by 5.63% and 4.29%, respectively. Next, we evaluate the performance using the ROC curve, and the overall performance can be reflected by the AUC score. The ROC curves and AUC on the two datasets are shown in Fig. 5. From Fig. 5, it can be seen that our method has the largest ROC area, with the performance increased by 7.23% and 7.79%



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 4. Comparisons of our algorithm with six state-of-the-art saliency detection methods. From left to right: (a) original images from the MSRA and the ECSSD datasets, (b) GB [7], (c) FT [9], (c) CA [11], (e) OS [12], (f) DCT [13], (g) WS [14], (h) our proposed method, and (i) ground-truth images.



MSRA dataset [10] ECSSD dataset [19]

Fig. 5. Quantitative comparisons of the proposed method with other methods. From top to bottom: the Precision, Recall, and F-measure; ROC curves; and AUC scores.

on the two respective datasets. Therefore, our proposed approach achieves the best performance compared to the other methods.

4. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel salient-object detection method in a multi-level framework based on image smoothing and over-segmentation. Our method works well due to two reasons: firstly, we combine multi-level image smoothing with over-segmentation to improve the performance; secondly, we propose a new saliency measure for segments based on color and spatial similarity.

Our future work will focus on the following aspects. Firstly, the shape of the over-segments can be changed from rectangle to other shapes, e.g. convex hull and polygon. Secondly, the saliency measure based on color and spatial similarity can be analyzed to construct a more advanced mathematical model for the saliency measure of the segments. Thirdly, the smoothing parameter for the multi-level framework can be selected adaptively to accommodate the method to images with different characteristics. Finally, some weighting and normalization mechanisms can be used to combine those intermediate saliency maps into the final saliency map.

5. ACKNOWLEDGEMENT

This work is supported by an internal grant from the Hong Kong Polytechnic University (Grant No. G-YN21).

6. REFERENCES

[1] D. Gao and N. Vasconcelos, "Integrated learning of saliency, complex features, and object detectors from cluttered scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[2] L. Marchesotti, C. Cifarelli, and G. Csurka. "A framework for visual saliency detection with applications to image thumbnailing," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[3] J. Wang, L. Quan, J. Sun, X. Tang, and H.-Y. Shum, "Picture collage," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

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

[5] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1254-1259, 1998.

[6] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *ACM International conference on Multimedia*, 2003.

[7] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, 2006.

[8] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

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

[10] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang and H-Y. Shum, "Learning to detect a salient object," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353-367, 2011.

[11] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-aware saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[12] X. Zhang, Z. Ren, D. Rajan and Y. Hu, "Salient object detection through over-segmentation," in *IEEE International Conference on Multimedia and Expo*, 2012.

[13] X. Hou, J. Harel and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194-201, 2012.

[14] Nevrez Imamoglu, Weisi Lin, and Yuming Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Trans. on Multimedia*, vol. 15, no. 1, pp. 96-105, 2013.

[15] L. Xu, C. Lu, Y. Xu and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Trans. on Graphics*, vol. 30, no. 6, 2011.

[16] Y. Xie and H. Lu, "Visual saliency detection based on Bayesian model," in *IEEE International Conference on Image Processing*, 2011.

[17] J. Sun, H. Lu and S. Li, "Saliency detection based on integration of boundary and soft-segmentation," in *IEEE International Conference on Image Processing*, 2012.

[18] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, 2002.

[19] T. Foulsham and G. Underwood, "What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition," *Journal of Vision*, vol. 8, no. 2, pp. 1-17, 2008.

[20] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng and S. Li, "Salient object detection: a discriminative regional feature integration approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[21] Q. Yan, L. Xu, J. Shi and J. Jia, "Hierarchical saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.