Region of interest detection based on visual saliency analysis and iteratively clustering for remote sensing images

Sun, Yang, Wang, Shuang, Zhang, Libao

Yang Sun, Shuang Wang, Libao Zhang, "Region of interest detection based on visual saliency analysis and iteratively clustering for remote sensing images," Proc. SPIE 11061, Automated Visual Inspection and Machine Vision III, 1106108 (21 June 2019); doi: 10.1117/12.2524701

SPIE.

Event: SPIE Optical Metrology, 2019, Munich, Germany

Region of Interest Detection Based on Visual Saliency Analysis and Iteratively Clustering for Remote Sensing Images

Yang Sun[a], Shuang Wang[a], Libao Zhang*[a]
[a] College of Information Science and Technology, Beijing Normal University, No. 19, Xinjiekouwai Road, Haidian District, Beijing 100875, China
*libaozhang@163.com

ABSTRACT

Saliency analysis is essential to detect common regions of interest (ROI) in remote sensing images. However, many methods imply saliency analysis in single images and cannot detect common ROI accurately. In this paper, we propose the joint saliency analysis based on iterative clustering (JSIC) method to detect common ROIs. Firstly, the size of superpixel patch is adaptively determined by texture feature. Secondly, color feature and intensity feature are utilized to get initial saliency maps and Otsu is utilized to obtain initial ROIs. Finally, iterative clustering is applied to obtain final ROI with less background inference. Quantitative and qualitative experiments results show that the iterative clustering joint saliency analysis method not only has better performance when compared to the other state-of-the-art methods, but also can eliminate image without ROI. Our contributions lie in three aspects as follows: 1) We propose a novel method to calculate the number of superpixel blocks adaptively. 2) A new joint saliency analysis method is proposed based on color feature and intensity feature. 3) We propose a novel saliency modification strategy based on the iterative cluster, which could reduce the background inference and eliminate images without ROIs.

Keywords: Machine Vision, image processing, region of interest, saliency analysis, clustering

1. INTRODUCTION

With the increase capacity to obtain high spatial resolution remote sensing images, the detection of ROIs in a group of remote sensing images has been changeable. Saliency is an effective tool to extract ROIs. Saliency analysis aims to identify the informative regions in an image for further processing. Since the saliency analysis method has been developed, many researchers proposed target detection methods for remote sensing images based on single image saliency [1, 2]. However, in the single saliency map the information between the images cannot be effectively utilized. Joint saliency detection aims at discovering the common and salient objects in multiple images. It explores not only intra-image but extra inter-image visual cues, and hence compensates the shortages in single-image saliency detection [3]. Through the common saliency analysis and calculation, we can make better use of the mutual information between images, so as to obtain more accurate detection results. In many cases, joint saliency detection is usually dependent on the single image saliency detection results [4]. Runmin Cong et al. propose an iterative RGBD joint saliency framework, which utilizes the existing intra-saliency maps as the initialization, and generates the final RGBD joint saliency map by using a refinement-cycle model [5].

Unlike single image saliency detection, joint saliency detection for multiple images leverages not only single image appearance evidence but also inter-image correspondences to locate common salient regions [6]. The joint saliency analysis method computes saliency in a set of images; its goal is to detect the common salient objects in multiple images. Such capability is helpful to object co-segmentation and similar image search. In Zhang's model [3], a novel multi-image saliency analysis (MSA) model based on multiple multispectral images clustering saliency analysis (MMCS) and panchromatic image co-occurrence histogram saliency analysis (PCHS) is proposed. Martin et al. propose a method to accurately detect and localize boundaries in natural scenes using local image measurements [7]. Le Callet and Niebur introduced the concepts of overt and covert visual attention and of bottom-up and top-down processing [8].Most existing bottom-up methods measure the foreground saliency of a pixel or region based on its contrast within a local context or the entire image [9]

In this paper, we present the joint saliency analysis based on iteratively clustering. Firstly, we propose a novel method to calculate the number of superpixel adaptively, which effectively remains the integrity of the target, and avoids the

Automated Visual Inspection and Machine Vision III, edited by Jürgen Beyerer,
Fernando Puente León, Proc. of SPIE Vol. 11061, 1106108 · © 2019 SPIE
CCC code: 0277-786X/19/$21 · doi: 10.1117/12.2524701

fragmentation in the extracted target regions. Secondly we implement the saliency analysis to obtain joint saliency maps, including single saliency maps based on color feature and intensity feature. Thirdly, we use iterative cluster computation to eliminate the background regions further, which neighbors the target region, thus obtain target extraction results with clear boundaries. Our model achieves good results on detecting target in remote sensing images. The framework of the proposed model is illustrated in Fig.1.

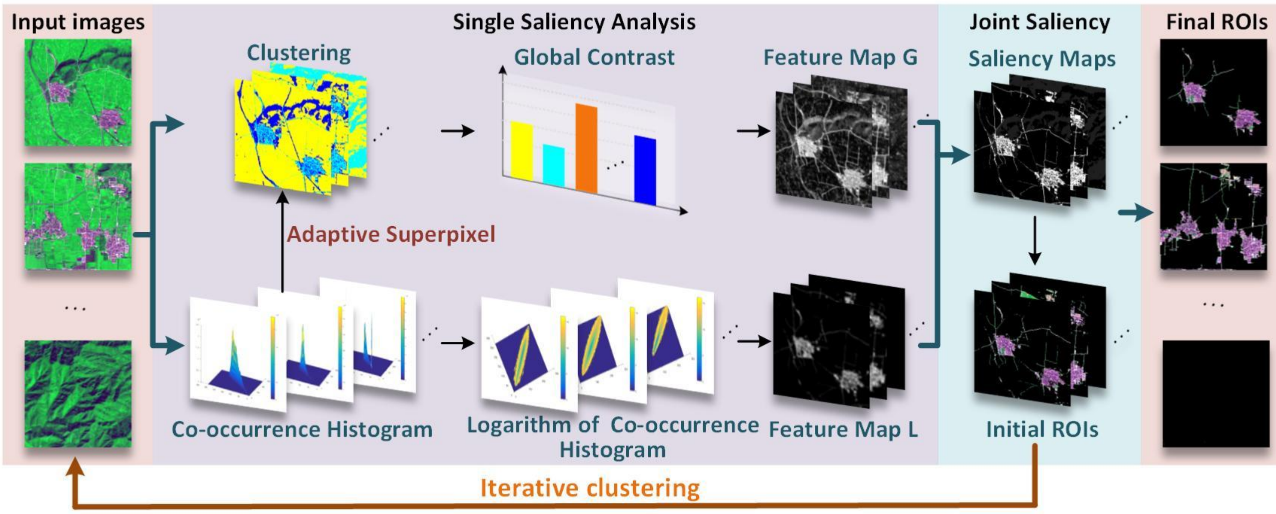


Figure 1. Framework of our proposed model.

The contributions of this paper lie in three aspects.

- We propose a novel method to calculate the number of superpixel blocks adaptively.
- A new joint saliency analysis method is proposed based on color feature and intensity feature.
- We propose a novel saliency modification strategy based on the iterative cluster, which could reduce the background inference and eliminate images without ROIs.

2. METHODOLOGY

2.1 Adaptively superpixel based on texture

Superpixel is a good way to put adjacent pixels that have similar colors, textures, or intensity to a pixel block [9]. When using the superpixel to split the image, setting the number of superpixel K is important. If K is relatively small, each size of super pixel will be larger, the boundaries of the target segmentation will be worse; if K is relatively large, superpixel size will be smaller, it may lead to overfitting. For different image, the number of superpixels required is diverse from image to image.

Co-occurrence histogram could capture the global and distribution of intensity values and local distribution of pixel with same intensity value. K is relevant to the contrast, energy, entropy and correlation of the gray level co-occurrence matrix. When the contrast and entropy parameters are relatively large and the energy and correlation are relatively small, it indicates that the texture of the image is relatively intensive, and the value of K should be large to keep the boundaries of superpixels well-defined. On the contrary, it indicates that the texture of the image is relatively sparse, and K should be small to avoid over-fitting. Thus, we calculate K by following equation:

$$\bar{K}_1 = \omega\sqrt{\left(\frac{M \times N}{\delta}\right)} \tag{1}$$

$$\omega = Asm \times Corr / Con \times Ent \tag{2}$$

where M, N denote the size of image, δ is the parameter to adjust the size of image, Asm, Corr, Con, Ent is the energy, correlation contrast, and entropy of the gray level co-occurrence matrix, respectively.

After calculating the number of pixel required adaptively, superpixel segmentation is applied to images to preserve the boundary and reduce computing complexity. For remote sensing images, the object are grouped into different clusters as shown in Fig.1.

2.2 Intra-saliency calculation

In general, people tend to pay more attention to these regions with sharp contrast to surroundings. That is, saliency of one region mainly depends on its contrast to the nearby regions, and contrasts to distant regions are less significant. Thus, for these images after superpixel segmentation, we calculate the color histogram for each cluster, and then calculate the saliency of each cluster based on color contrast and spatial information. The intra-saliency map $S_s(i, j)$ is calculated as shown below.

$$\mu(c_i) = \exp\left(\frac{-d_s(c_i, c_j)}{\sigma_s^{\ 2}}\right) \tag{3}$$

$$\mathrm{r}(c_i, c_j) = -\ln\left(1 - \frac{1}{2}\sum_{k=1}^{n}\frac{(f_{i,k} - f_{j,k})^2}{f_{i,k} + f_{j,k}}\right) \tag{4}$$

$$S_S(c_i) = \frac{\sum_{i \neq j} \mu(c_i)\omega(c_j) r(c_i, c_j)}{\omega(c_i)} \tag{5}$$

Where $\mu(c_i)$ is the spatial weighted information, $d_s(c_i, c_j)$ is spatial distance between regions. In the method, we use the European distance between centers of c_i and c_j. σ_s controls the impact of $\mu(c_i)$ weight in saliency calculation. $r(c_i, c_j)$ calculates difference between two color in Lab color space. $f_{i,k}$ denotes the frequency of kth range in the cluster c_i, $\mathrm{w}(c_j)$ is the ratio of pixels in each cluster to pixels in the image.

The intensity feature is also a significant cue for extracting the saliency regions. Firstly, we construct the probability density distribution of the two-dimension symbiotic histogram $coh(i, j)$ .For each pixel that the intensity value is equal to m in the image, computing the total number $sum(i, j)$ of pixels that the intensity value is n in the window $w \times w$, and the two-dimension symbiotic histogram is calculated by the normalized $sum(i, j)$.

$$coh(i, j) = \frac{sum(i, j)}{\sum_{i=1}^{L}\sum_{j=1}^{L} sum(i, j)} \tag{6}$$

For the region in image, its global and local intensity distribution indicate the saliency of region. In coh, each pairs of intensity represents one of the elements of the diagonal, by which we obtain the global distribution of image intensities. Meanwhile, each pixel pairs with other pixel that the value of intensity is not equal to its intensity to capture the local distribution of image intensities.

According to human visual mechanism, the lower the frequency of the pixel pair of the intensity is, the more salient this pixel pair is, that is to say, the saliency value is negative correlated with the frequency of pixel. Motivated by the

Boltzmann entropy theorem, we use logarithmic to calculate the saliency of each pairs of intensity $S_I(i,j)$ of the image as shown in equation (7)

$$S_I(i,j) = -\ln\left[coh(i,j)\right] \tag{7}$$

When a pair of pixels are inexistence, the pixel saliency is inexistence, that is, if $sum(i,j)=0$, $coh(i,j)=0$ and $S_I(i,j)=0$ for the reason that nonexistent intensity pairs contains no saliency.

Table 1. The proposed JSIC method.

Algorithm 1: The proposed JSIC

Input: The input remote sensing image set $\{I^j\}_{j=1}^M$
for each image **do**
Compute GLCM, Asm, Ent, and Corr, Then obtain initial number of super-pixel K_1.
Segment image into super-pixel with K_1, and regard the mean value of the region as the value of the region
end
for each cluster **do**
Clustering the super-pixel into k groups by k-means++
Compute single saliency S_s using **Eq**.(5) and S_I using **Eq**.(7)
end
Combine the two intra-saliency map using **Eq**.(8), obtain inter-saliency map S_{inter}
Segment inter-saliency map by otsu algorithm and obtain target extract results $I_{ROI}(i,j)$
Segment images by super-pixel cluster with $K_2 = \frac{2}{3}K_1$ iteratively
Output: The joint saliency map and target extraction results

After obtaining two intra-saliency maps, using multiplication will obtain the final inter-saliency map by equation (8).

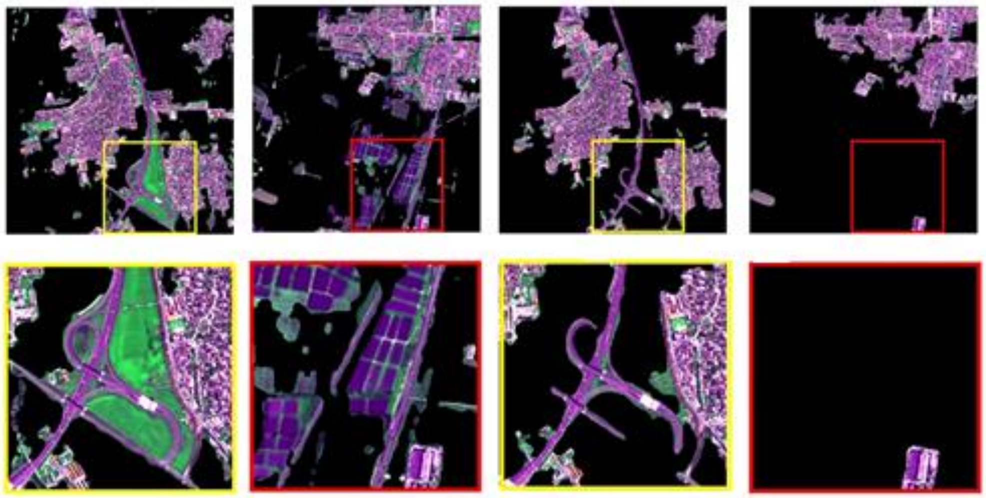
$$S_{inter} = S_S + S_I \tag{8}$$

Then, we apply the Otsu method for optimal global threshold processing as shown in equation (9).

$$Bw(i,j) = Otsu(S_{inter}(i,j)) \times g \tag{9}$$

The final extracted target area is obtained by multiplying the generated binary image with the original image. The detailed algorithm is described in below.

$$I_{ROI}(i,j) = I(i,j) \otimes Bw(i,j) \tag{10}$$



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

Figure 2. Compare between iterative clustering and only once clustering results. The second column images are the enlarged image of the first column images. (a), (b) are the result of only once clustering. (c), (d) are the results of twice clustering.

2.3 Iterative segmentation

For the images after the target segmentation, we regard them as the input images, and then carry on the superpixel clustering saliency analysis again, getting the final inter-saliency analysis result. In this section, because most backgrounds have been eliminated, the superpixel size is 2/3 of the size of the superpixel used in the first cluster.

$$sp(i,j) = slic_{K_2}\left(ROI_1(i,j)\right) \tag{11}$$

$$K_2 = \frac{2}{3}\bar{K}_1 \tag{12}$$

For the remote sensing image clusters, due to the color characteristics interference between the target and the background area, some background may be misclassified. Thus we cluster the image again after the target segmentation to obtain saliency maps accurately and eliminate image without ROIs effectively.

Fig.2 presents the target extraction of clustering by once and twice respectively. It can be seen that the region could not be eliminated by once clustering, and iterative clustering could eliminate background region effectively. Thus, iteratively clustering is necessary. The detail steps of our proposed method are presented in Algorithm 1.

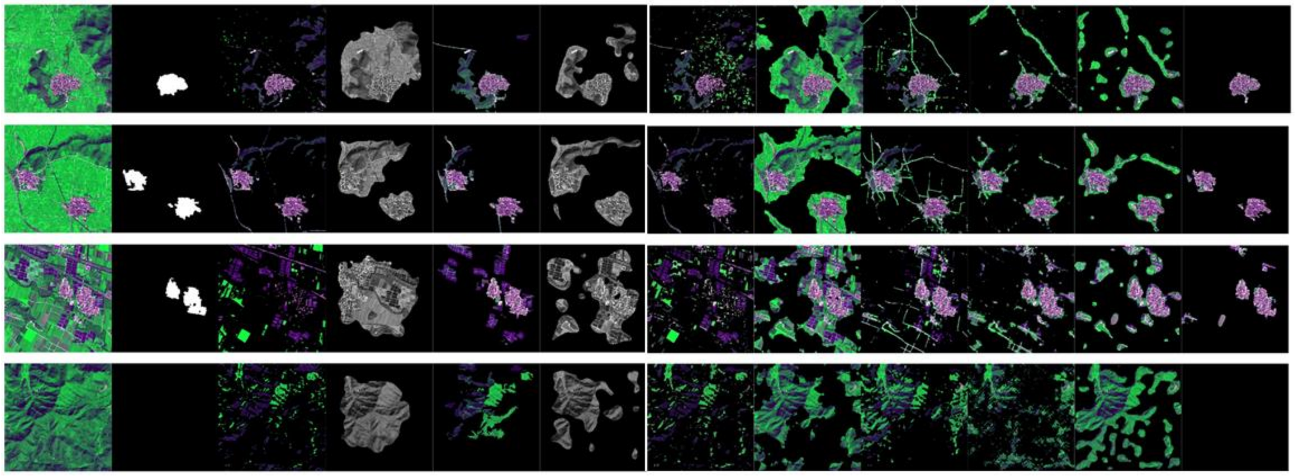
3. EXPERIMENTAL RESULTS

In this paper, all the experiments are proceeded for the remote sensing images. The image source contain the SPOT-5 and Google Earth. We use 200 images to the training stage and choose another 10 images as the testing set. The images are all size of 512×512. In experimental tests, our method is compared with nine classic and representative saliency methods (FT[10], GBVS[11], H[12], ITTI[13], MSSS[14], SIM[15], TMM[16], IWT-VF[17], and FDA[18]).

3.1 Qualitative Assessment

We evaluate the performance of the proposed method with nine state-of-the-art saliency analysis methods in residential area images. Due to the limitation of our paper, four testing images from SPOT5 are chosen for exhibition. Visual comparison of different target extraction obtained by various methods is shown in Fig.3.

Qualitative experiment results show that the proposed model can obtain highlighted region with well-defined boundaries, and suppress the background regions more effectively compared to other saliency maps. For the image without ROIs, our method gets accurate result while other methods misjudge inference as salient regions, such as the fourth row in Fig.3.



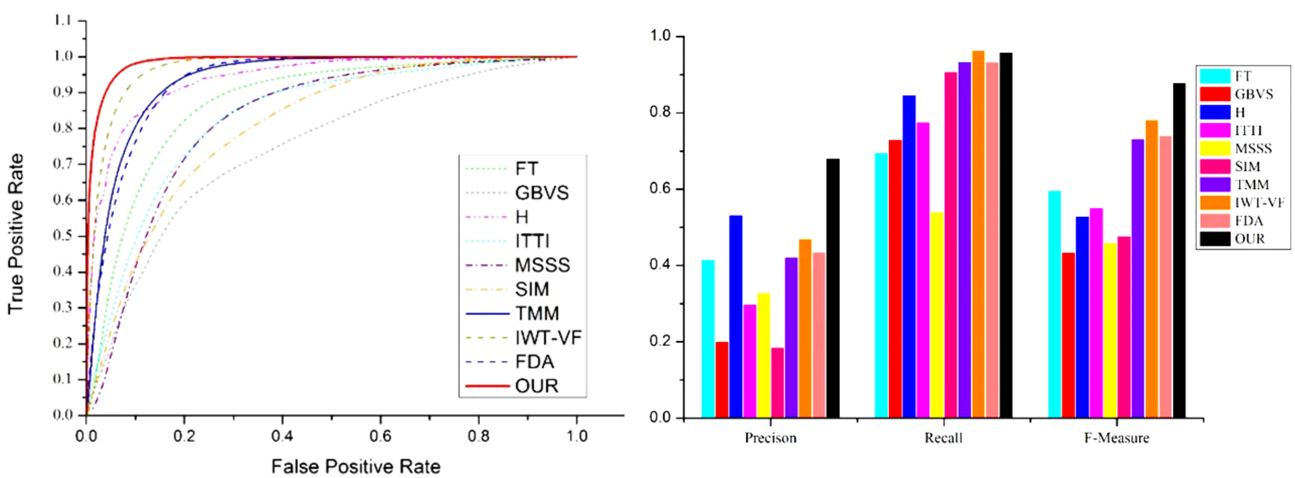
(a) Input (b) GT (c) FT (d) GBVS (e) H (f) ITTI (g) MSSS (h) SIM (i) TMM (j) IWT-VF (k) FDA (m) JSIC

Figure 3. Visual comparison of saliency target extraction results.

3.2 Quantitative Assessment

For the quantitative experiment results, we choose receiver operating characteristic curve (ROC curve) and PRF value for comparison. ROC graphs are useful for organizing classifiers and visualizing their performance. PRF value contains comprehensive evaluation F-measure (F), precision (P), and recall (R). F-Measure is the harmonic mean value of recall and precision.

Fig.4 (a) present the ROC curve, Fig.4 (b) show the F-measure of all models. From quantitative evaluation, it can be seen that the proposed method achieves better ROC and higher F-Measure than the other nine state-of-the-art saliency detection models.



(a) ROC curve (b) PRF values

Figure 4. Performance of proposed model compared with nine state-of-the-art models.

4. CONCLUSIONS

In this paper, we have presented the JSIC method. We first segment the image into object and the background region by superpixel cluster. The number of the super pixel block is determined by texture adaptively, aims to preserve the boundary and avoid the overfitting. After clustering, intra-saliency maps are obtained by color and intensity feature, and otsu algorithm is applied to produce initial ROIs. Then, we iteratively segment images by superpixel to suppress the background region further. Experimental results demonstrate that our model can detect target regions and eliminate background accurately. The proposed method is of great importance in further remote sensing image processing such as image pressing, image fusion and reduce their computational complexity.

5. ACKNOWLEDGEMENTS

This research was supported by the Beijing Natural Science Foundation under Grant L182029, and the National Natural Science Foundation of China under Grants 61571050.

REFERENCES

[1] M. Cheng, N. Mitra, X. Huang, P.H.S. Torr, and S. Hu, “Global contrast based salient region detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 569-582 (2015).

[2] L. Zhang, and J. Zhang, “A New Saliency-Driven Fusion Method Based On Complex Wavelet Transform for Remote Sensing Images,” IEEE Geoscience and Remote Sensing Letter, 14(12), 2433–2437 (2017).

[3] L. Zhang, and Q. Sun, “Saliency detection and region of interest extraction based on multi-image common saliency analysis in satellite images,” Neurocomputing, 283, 150-165 (2018).

[4] L. Zhang, Q. Sun and J. Zhang, “Region of interest extraction via common salient feature analysis and feedback reinforcement strategy for remote sensing images,” GIScience & Remote Sensing, 55(5), 745-762 (2018).

[5] R. Cong, J. Lei, H. Fu; W. Lin, Q. Huang; X. Cao, and C. Hou, “An Iterative co-saliency Framework for RGBD Images,” IEEE Transactions on Cybernetics, 49(1), 233-246 (2019).

[6] J. Chen, L. Zhang, “Joint Multi-Image Saliency Analysis for Region of Interest Detection in Optical Multispectral Remote Sensing Images,” Remote Sensing, 8(6), 461 (2016).

[7] D.R. Martin, C.C. Fowlkes, J. Malik “Learning to detect natural image boundaries using local brightness color, and texture cues,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(5), 530–549 (2004).

[8] P.L. Callet, and E. Niebur, “Visual Attention and Applications in Multimedia Technology,” Proceedings of the IEEE, 101(9), 2058-2067 (2013).

[9] M. Cheng, N. Mitra, X. Huang, P. Torr, and S. Hu, “Global contrast based salient region detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 569-582 (2015).

[10] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in Computer Vision and Pattern Recognition, 2009. CVPR’09. IEEE Conference on, 1597-1604 (2009).

[11] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," Advances in Neural Information Processing System, 19, 545–552 (2007).

[12] J. Shi, Q. Yan, L. Xu and J. Jia, “Hierarchical image saliency detection on extended CSSD,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(4), 717-729 (2016).

[13] L. Itti, C. Koch and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), 1254-1259 (1998).

[14] R. Achanta, and S. Susstrunk, “Saliency detection using maximum symmetric surround,” in 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, China, 2653–2656 (2010).

[15] N. Murray, M. Vanrell, X. Otazu, C.A. Parraga, “Saliency estimation using a non-parametric low-level vision model,” in 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, USA, 433–440 (2011).

[16] N. Imamoglu, W. Lin and Y. Fang, “A saliency detection model using low-level features based on wavelet transform,” IEEE Transactions on Multimedia, 15(1), 96–105 (2013).

[17] Y. Zhang, L. Zhang and X. Yu. "Region of interest extraction based on multiscale visual saliency analysis for remote sensing images," Journal of Applied Remote Sensing, 9, 095050 (2015).
[18] L. Zhang, and K. Yang, "Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image," IEEE Geoscience and Remote Sensing Letters, 11(5), 916-920 (2014).