Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Sparsity-guided saliency detection for remote sensing images

Danpei Zhao
Jiajia Wang
Jun Shi
Zhiguo Jiang

SPIE.

Sparsity-guided saliency detection for remote sensing images

Danpei Zhao,[a,b,*] Jiajia Wang,[a,b] Jun Shi,[a,b] and Zhiguo Jiang[a,b]
[a]Beihang University, School of Astronautics, Image Processing Center, 37 Xueyuan Road, Haidian District, Beijing 100191, China
[b]Beijing Key Laboratory of Digital Media, 37 Xueyuan Road, Haidian District, Beijing 100191, China

Abstract. Traditional saliency detection can effectively detect possible objects using an attentional mechanism instead of automatic object detection, and thus is widely used in natural scene detection. However, it may fail to extract salient objects accurately from remote sensing images, which have their own characteristics such as large data volumes, multiple resolutions, illumination variation, and complex texture structure. We propose a sparsity-guided saliency detection model for remote sensing images that uses a sparse representation to obtain the high-level global and background cues for saliency map integration. Specifically, it first uses pixel-level global cues and background prior information to construct two dictionaries that are used to characterize the global and background properties of remote sensing images. It then employs a sparse representation for the high-level cues. Finally, a Bayesian formula is applied to integrate the saliency maps generated by both types of high-level cues. Experimental results on remote sensing image datasets that include various objects under complex conditions demonstrate the effectiveness and feasibility of the proposed method. [DOI: 10.1117/1.JRS.9.095055]

Keywords: sparsity-guided saliency model; global cues; background prior; sparse representation; Bayesian integration; remote sensing images.

Paper 15202 received Mar. 15, 2015; accepted for publication Aug. 11, 2015; published online Sep. 11, 2015.

1 Introduction

Object detection in remote sensing images is of vital importance and has great potential in many fields such as navigation reconnaissance, autonomous navigation, scene understanding, geological survey, and precision-guided systems. Remote sensing images are captured by sensors on an airplane or other aircraft as an aerial view under various luminance and viewing angle conditions. In contrast to natural scene images taken from the ground, remote sensing images have more complex backgrounds (e.g., forests, lakes, sand, roads, and lawns) that sometimes share similar characteristics with the interesting objects. In addition, remote sensing images with down-looking or front-downward views are more likely to be disturbed by noise, luminance fluctuation, fog, cloud cover, and blur caused by flight vibration. Therefore, it is difficult and time-consuming to precisely and quickly extract objects from complex backgrounds in practical applications. In order to achieve automatic, rapid, and accurate remote sensing target detection, saliency detection was introduced to the remote sensing field in the last decade.[1–6] This method imitates human visual attention to identify the attention-grabbing regions that may contain candidate objects.[7–9]

There are two main types of models for saliency detection: data-driven bottom-up models[10–22,23,24] and task-driven top-down models.[25] The bottom-up model has shown that low-level cues (e.g., frequency[26,27] and contrast[10,11,13–20,22,28,29,30,31]) are quite useful for saliency detection. Itti et al.[10] exploited the contrast of the center and its surroundings at multiple scales with multiple features to detect salient regions in an image. Bruce and Tsotsos[11]

*Address all correspondence to: Danpei Zhao, E-mail: zhaodanpei@buaa.edu.cn

extracted the local Shannon's self-information to generate the saliency map. Color contrast (e.g., RGB or LAB)[10,13–20,22,28,29,30,31,32] has been utilized to form low-level cues, and many studies[7,15–18,20,22,28,31] have shown that the LAB color space is more suitable for human visual perception. Compared with local contrast,[12,13] which highlights the object boundaries, global contrast[8,17] usually highlights the entire prominent region, but it easily mistakes noisy regions as salient parts. Most recently, methods[16–18,22] exploiting foreground and background priors have proven to be efficient. In particular, the extraction of background information[16–18] provides a background template and achieves unsupervised saliency detection. Despite all this, models employing only low-level cues fail to generate object-level saliency maps. To discover more effective cues for detecting salient regions, high-level saliency cues have been investigated. Shen and Wu[19] designed a unified model based on low-rank matrix recovery to obtain the saliency map. Margolin et al.[20] computed saliency by exploiting the reconstruction error of the principle component analysis to analyze the distinctness of a region. Xie et al.[21] proposed a Bayesian model via low and midlevel cues to produce a saliency map. Borji and Itti[22] detected the salient regions by calculating local and global patch rarities after reconstructing the image using a sparse representation. Li et al.[18] achieved efficient saliency maps with dense and sparse reconstruction errors. In contrast to low-level cues, these high-level cues can generate a better saliency detection performance. Some researchers tend to combine existing saliency models to detect saliency. Sun et al.[1] employed a combination of edge- and graph-based visual saliency models by fusing two saliency maps to detect salient regions in remote sensing images. Zhang and Yang[6] proposed a method based on frequency domain analysis and salient region detection to extract salient regions. However, the methods that fuse two saliency maps generated by different saliency models can easily lead to a less effective performance of saliency detection in remote sensing images because of the complex and abundant image content. Consequently, it is important to seek new cues that effectively predict salient regions where candidate objects are likely to exist in remote sensing images.

Because the objects in remote sensing images are different from complex backgrounds in the visible spectrum, we attempt to discover persuasive cues to extract salient regions from complex backgrounds. In this paper, we propose a sparsity-guided saliency model (SGSM) that combines global cues with background priors for saliency detection in remote sensing images. Our proposed model takes a sparse representation approach by measuring the relationship between image patches and a dictionary to generate an objective saliency map. This method exploits a sparse representation to produce high-level cues via global-based and background-based dictionaries. These two dictionaries are, respectively, obtained by low-level cues based on global cues and the background prior, and they contain the category information (i.e., object or background). Hence, high-level cues can reveal the intrinsic similarity of images and determine the categories of patches. Using the patch category information, the saliency map is obtained by a clustering algorithm. As there are no benchmark datasets for saliency detection in remote sensing images, we constructed two datasets to validate the efficiency of our proposed model. The images in the datasets contain various objects (e.g., house or vehicle) captured by Google Earth under varying conditions. The single-object dataset (SOD) contains 500 images of a single object, while the multiple-object dataset (MOD) contains 1000 images of multiple objects.

The remainder of this paper is organized as follows: Sec. 2 demonstrates the theory and motivation of our proposed model first and then illustrates the specific implementation of the proposed model. In Sec. 3, the experimental results and analysis are shown. Finally, Sec. 4 provides the conclusion.

2 Sparsity-Guided Saliency Model

This section presents the theoretical basis of SGSM in detail.

First, we provide the general theory that is necessary to understand our proposed model. SGSM exploits a combination of global cues and background prior information to provide global and background information, respectively. With the global cues, the false positive detection of regions that contain candidate objects can be avoided, especially when these regions are similar to the background. In addition, by using the background prior information, regions that are

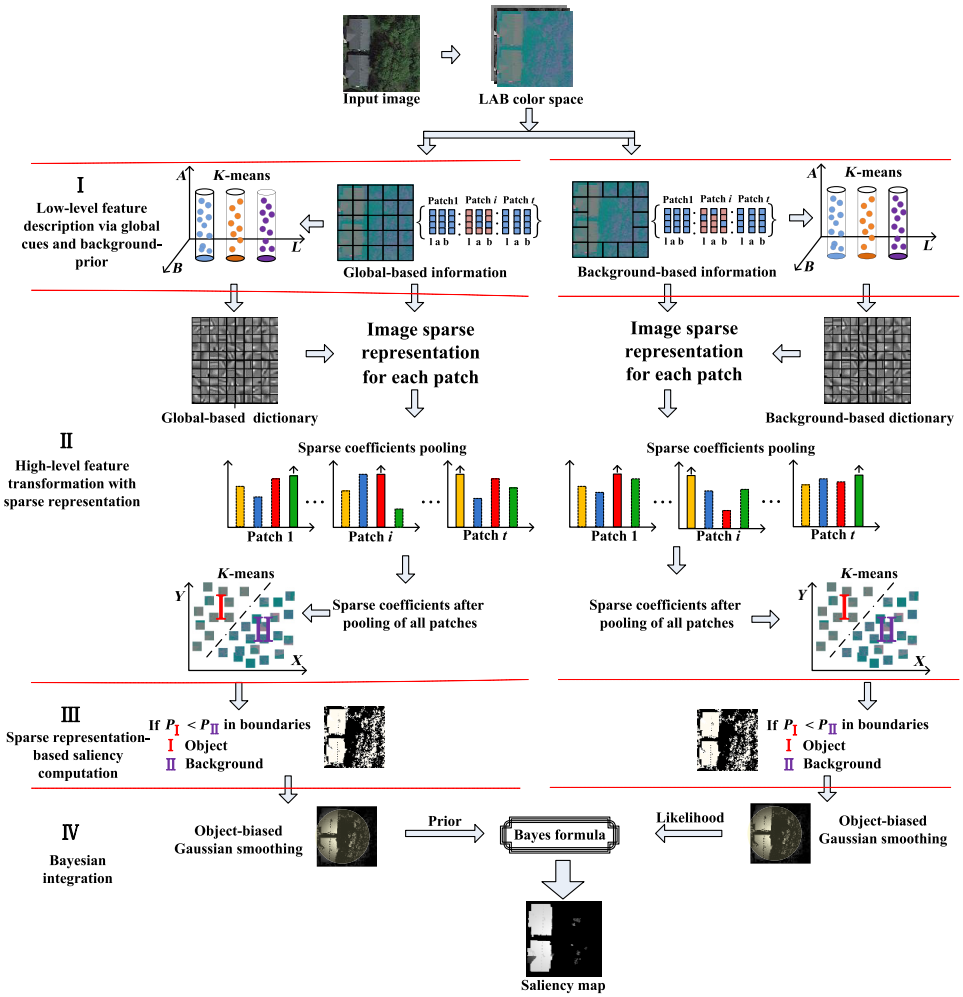


Fig. 1 Sparsity-guided saliency model saliency detection.

different from the background stand out. The low-level cues based on global cues and background priors are, respectively, clustered into global-based and background-based dictionaries. These two dictionaries separately contain the category information (i.e., object or background) of global and background cues. Based on these two low-level dictionaries, high-level cues are generated using a sparse representation. Finally, these high-level cues are clustered to obtain a saliency map. The overall procedure is presented in Fig. 1.

2.1 *Low-level Feature Description via Global Cues and Background Prior*

In order to determine the visual uniqueness of image regions, we decompose the image into nonoverlapping patches of uniform size. Because the LAB color space[7,15,16–18,20,22,28,31] corresponds more closely to human vision, we chose it for the low-level representation. Generally, global information comes from global cues, and background information stems from the background prior. According to the background prior assumptions[17,18] that salient objects usually appear in the center of the image and the boundaries are mostly background, we use boundary-based cues to extract background information.

Given a color image I of size $T = W \times H$ (W and H are, respectively, the image width and height), we first divide it into nonoverlapping patches of size $T_p = P \times Q$ such that the whole image contains $t(t = T/T_P)$ patches. There are then $n = 2(W/P + H/Q) - 4$ patches at the four boundaries to form the background set. For the i'th $(1 \le i \le t)$ patch containing T_p pixels, the values of all pixels in the three LAB channels form the rows of matrix $G^{\text{lab}}(i)$, and the pixel values of the j'th $(1 \le j \le n)$ patch in the background set form the rows of matrix $B^{\text{lab}}(j)$.

$$G^{\text{lab}}(i) = \begin{bmatrix} G^l_{1i} & G^l_{2i} & \cdots & G^l_{T_p i} \\ G^a_{1i} & G^a_{2i} & \cdots & G^a_{T_p i} \\ G^b_{1i} & G^b_{2i} & \cdots & G^b_{T_p i} \end{bmatrix}, \qquad i = (1,2,\cdots,t), \tag{1}$$

$$B^{\text{lab}}(j) = \begin{bmatrix} B^l_{1j} & B^l_{2j} & \cdots & B^l_{T_p j} \\ B^a_{1j} & B^a_{2j} & \cdots & B^a_{T_p j} \\ B^b_{1j} & B^b_{2j} & \cdots & B^b_{T_p j} \end{bmatrix}, \qquad j = (1,2,\cdots,n). \tag{2}$$

Furthermore, all t patches form the global information set G^{lab}, and all n patches at the four boundaries of the image form the background information set B^{lab}.

$$G^{\text{lab}} = \begin{bmatrix} \bar{G}^l_{k1} & \bar{G}^l_{k2} & \cdots & \bar{G}^l_{kt} \\ \bar{G}^a_{k1} & \bar{G}^a_{k2} & \cdots & \bar{G}^a_{kt} \\ \bar{G}^b_{k1} & \bar{G}^b_{k2} & \cdots & \bar{G}^b_{kt} \end{bmatrix}, \qquad k = (1,2,\cdots,T_p), \tag{3}$$

$$B^{\text{lab}} = \begin{bmatrix} \bar{B}^l_{k1} & \bar{B}^l_{k2} & \cdots & \bar{B}^l_{kn} \\ \bar{B}^a_{k1} & \bar{B}^a_{k2} & \cdots & \bar{B}^a_{kn} \\ \bar{B}^b_{k1} & \bar{B}^b_{k2} & \cdots & \bar{B}^b_{kn} \end{bmatrix}, \qquad k = (1,2,\cdots,T_p). \tag{4}$$

The two matrices G^{lab} and B^{lab} are clustered into global-based dictionary D_{Global} and background-based dictionary $D_{\text{Background}}$, respectively, using K-means with clustering number K_D. These two dictionaries, respectively, contain the global and background information. The details of this procedure are illustrated in part I of Fig. 1.

2.2 *High-Level Feature Transformation Using a Sparse Representation*

Sparse representation[22,33,34,35,30] has been a focus of research in the area of computer vision and pattern recognition. Based on a dictionary consisting of a set of bases, sparse representation can represent an image by a sparse coefficient vector. A nonzero element in the vector reflects the correlation between the image and the bases in the dictionary. As we divide the image into patches, the sparse coefficients of each patch can be learned by sparse representation. We choose one group of sparse coefficients to express the patch-dictionary relationship by max pooling the T_p groups of sparse coefficients in every patch. These sparse coefficient vectors are used to compute the patch categories.

Concretely, we represent the image using a sparse representation by minimizing the l_1-norm using a given dictionary. Every patch in the global-based set G^{lab} can be represented by the corresponding global coefficients α_{Global} from the global-based dictionary D_{Global}. Similarly, each patch in the background-based set B^{lab} can be represented by the corresponding background coefficients $\alpha_{\text{Background}}$ from the background-based dictionary $D_{\text{Background}}$. This representation is shown as follows:

$$G^{\text{lab}}(i) = D_{\text{Global}}\alpha_{\text{Global}}(i) \quad B^{\text{lab}}(i) = D_{\text{Background}}\alpha_{\text{Background}}(i). \tag{5}$$

We then encode all the patches in image I by

$$\begin{aligned} &\min_{\alpha} \|D_{\text{Global}}\alpha_{\text{Global}}(i) - G^{\text{lab}}(i)\|_2 \quad \text{s.t. } \|\alpha_{\text{Global}}(i)\|_1 \le \beta \\ &\min_{\alpha} \|D_{\text{Background}}\alpha_{\text{Background}}(i) - B^{\text{lab}}(i)\|_2 \quad \text{s.t. } \|\alpha_{\text{Background}}(i)\|_1 \le \beta, \end{aligned} \qquad i = (1,2,\cdots,t), \tag{6}$$

where $\beta \ge 0$ is a tuning parameter. The sparse coefficients of all patches α_{Global} and $\alpha_{\text{Background}}$ are optimized using the least absolute shrinkage and selection operator (Lasso).[36] After max pooling in every patch, we obtain the global-based coefficient set $\alpha^{\max}_{\text{Global}}$ and background based coefficients' set $\alpha^{\max}_{\text{Background}}$. Coefficients' sets $\alpha^{\max}_{\text{Global}}$ and $\alpha^{\max}_{\text{Background}}$ are separately clustered into two

categories (i.e., object and background) by K-means to determine the patch category labels. We obtain global-based estimate maps $\mathrm{EM}_{\mathrm{Global}}$ and background-based estimate maps $\mathrm{EM}_{\mathrm{Background}}$ by returning the category labels to the corresponding patches. The saliency map integration procedure is shown in part II of Fig. 1.

2.3 *Sparse Representation-Based Saliency Computation*

According to the background prior principle[16,18,22] mentioned in Sec. 2.1, we assume that the edges of the image are generally background. We then obtain the patch object probability P_{Object} by calculating the ratio of the patches confirmed as objects to all edge patches. Similarly, we obtain the patch background probability $P_{\mathrm{Background}}$ by calculating the ratio of the patches confirmed as background to all edge patches. These probabilities, respectively, form the estimated maps $\mathrm{EM}_{\mathrm{Global}}$ and $\mathrm{EM}_{\mathrm{Background}}$. According to the background prior, P_{Object} should be less than $P_{\mathrm{Background}}$. Therefore, we define the parts with lower probability to be objects and the parts with higher probability to be background. We then form a binary object map $\mathrm{BM}(i)$ of the clustered pixel patches defined as follows:

$$\mathrm{BM}(i) = \begin{cases} 1 & P_{\mathrm{Object}} < P_{\mathrm{Background}} \\ 0 & \text{Otherwise} \end{cases}, \qquad i = 1,2,\ldots,t. \tag{7}$$

The mean values of the sparse coefficient vectors after pooling show the degree of the patch-dictionary relationship. If the patches are similar, their pooling coefficients are analogous, and the mean values of the sparse coefficients indicate only slight differences. We then define the mean values of sparse coefficients after pooling to be the saliency scores of the patches. A labeled map $S(z)$ is obtained by returning saliency scores to the corresponding patches if they are confirmed as objects

$$S(z) = \begin{cases} \mathrm{mean}(\alpha_i^{\max}) & \mathrm{BM}(i) = 1 \\ 0 & \mathrm{BM}(i) = 0 \end{cases}, \quad i = 1,2,\ldots,t, \quad z = 1,2,\ldots,T, \tag{8}$$

where $\alpha_i^{\max}$ denotes the sparse coefficients of the i'th patch after max pooling.

The primary saliency maps $S_{\mathrm{Global}}(z)$ and $S_{\mathrm{Background}}(z)$ are, respectively, obtained from α_{Global} and $\alpha_{\mathrm{Background}}$ according to Eq. (8). This high-level feature transformation is illustrated in part III of Fig. 1.

2.4 *Saliency Map Integration*

Because remote sensing images are captured by sensors in aircraft, there is no certainty regarding the location of the objects in the images. Therefore, an object-biased Gaussian model[18] is more suitable than a center-biased Gaussian model[22] for erasing interference. Finally, we employ a Bayesian formula to integrate primary saliency maps $S_{\mathrm{Global}}(z)$ and $S_{\mathrm{Background}}(z)$ using posterior probability.

2.4.1 *Object-biased Gaussian smoothing*

We employ object-biased Gaussian smoothing to erase the interference judged to be noise. Borji and Itti[22] noted that a center-bias exists in some saliency detection datasets and hence removes noise by the Gaussian model

$$G(z) = \exp\left\{-\left[\frac{(x_z - x)^2}{2\sigma_x^2} + \frac{(y_z - y)^2}{2\sigma_y^2}\right]\right\}, \tag{9}$$

where σ_x and σ_y denote the covariances, (x, y) denotes the coordinates of the object center, and (x_z, y_z) are the coordinates of any pixel in the map, where $x = 0$ and $y = 0$ indicate the image center. Li et al.[18] refined the model to be object-biased with dense and sparse reconstruction errors. In this paper, we adopt patch labels from Eq. (7) instead of dense and sparse

reconstruction errors to determine a more accurate object center. We set the coordinates (x, y) of the object center to be the position determined using the labels of the image region as

$$\begin{cases} x = \sum_i x_i * S(i) / \sum_j S(j) \\ y = \sum_i y_i * S(i) / \sum_j S(j) \end{cases}. \tag{10}$$

An object-biased Gaussian model is generated using Eq. (9) with coordinates (x, y) in Eq. (10). The final result S is a convolution of the primary saliency map $S(z)$ and refined object-biased Gaussian model $G(z)$. We refine global-based saliency map (G-map) S_{Global} and background-based saliency map (B-map) $S_{\text{Background}}$ via this object-biased Gaussian model with its more accurate object centers.

$$S_{\text{Global}} = G(z) * S_{\text{Global}}(z), \qquad S_{\text{Background}} = G(z) * S_{\text{Background}}(z). \tag{11}$$

2.4.2 *Bayesian integration*

As illustrated in Ref. 18, an effective saliency map is obtained by the Bayesian integration of two given saliency maps. Bayes' formula states that

$$p(F|S_{\text{map}}) = \frac{p(F)p(S_{\text{map}}|F)}{p(F)p(S_{\text{map}}|F) + [1 - p(F)]p(S_{\text{map}}|B)}, \tag{12}$$

where $p(F)$ is the prior probability, namely, the saliency map $p(S_{\text{map}}|F)$ is the probability of foreground for the whole saliency map, and $p(S_{\text{map}}|B)$ is the respective probability of background.

We utilize a global-based saliency map S_{Global} or background-based saliency map $S_{\text{Background}}$ as the prior, and, respectively, either $S_{\text{Background}}$ or S_{Global} is then used to compute the likelihood. Together, these maps determine the final saliency map S

$$S = p(F_{\text{Global}}|S_{\text{Background}}) + p(F_{\text{Background}}|S_{\text{Global}}), \tag{13}$$

where F_{Global} and $F_{\text{Background}}$, respectively, denote the foreground segmented by the mean saliency value from S_{Global} and $S_{\text{Background}}$. The saliency map integration procedure is shown in part IV of Fig. 1.

2.5 *Algorithm*

The full SGSM algorithm consists of the following steps:

Step 1: Divide input color image I into patches of size $P \times Q$.

Step 2: Extract global information G^{lab} and background information B^{lab} from the three LAB channels and then, respectively, cluster them into dictionaries D_{Global} and $D_{\text{Background}}$ using K-means with clustering number K_D.

$$G^{\text{lab}} \xrightarrow{K\text{-means}} D_{\text{Global}}, \qquad B^{\text{lab}} \xrightarrow{k\text{-means}} D_{\text{Background}}.$$

Step 3: Learn coefficients α_{Global} and $\alpha_{\text{Background}}$ using Eq. (2) via a sparse representation based on D_{Global} and $D_{\text{Background}}$.

Step 4: Cluster sparse coefficients after separately max pooling $\alpha^{\max}_{\text{Global}}$ and $\alpha^{\max}_{\text{Background}}$ into two categories by K-means to get estimated maps $\text{EM}_{\text{Global}}$ and $\text{EM}_{\text{Background}}$.

Step 5: Compute the patch saliency values to get primary saliency maps $S_{\text{Global}}(z)$ and $S_{\text{Background}}(z)$ by Eqs. (7) and (8).

Step 6: Smooth $S_{\text{Global}}(z)$ and $S_{\text{Background}}(z)$ using an object-biased Gaussian model by Eq. (11) to get S_{Global} and $S_{\text{Background}}$, respectively.

Step 7: Obtain saliency map S by a Bayesian integration of S_{Global} and $S_{\text{Background}}$ in Eq. (13).

2.6 *Multiple Scales Integration*

We obtained different results at different spatial scales for objects at different depths and of different sizes, hence we divided the input image into patches of size $(k * P) \times (k * Q)$ at the k'th scale to generate the SGSM saliency map at that scale. Large patches contribute to the definition of properties for the image region, but they generate jagged edges because of the few pixels that do not have the same property as the majority of the pixels within that patch. The final saliency map was obtained by fusing the maps at the k scales as follows:

$$S_{\text{final}} = \varepsilon S_{\text{scale}\,1} + \psi S_{\text{scale}\,2} + \cdots + \vartheta S_{\text{scale}\,k}, \qquad \varepsilon + \psi + \cdots + \vartheta = 1, \tag{14}$$

where $\varepsilon, \psi, \cdots, \vartheta$ are the weights for different scales. We then normalized the saliency map S_{final} to the range of [0,1] to obtain the final saliency map S_{final}.

3 Experiments

This section presents the database used to validate the efficiency of our proposed method and evaluates it with respect to 10 other state-of-the-art methods.

3.1 *Databases*

SGSM aims to detect salient objects in remote sensing images that mainly contain houses and oil tanks. All images were collected from Google Earth and were captured under conditions of diverse illumination and various viewpoints. We collected images taken at heights of 300 to 2000 m, the resolution is about 0.4 to 1.9 m. It is important to ensure that detailed images can be captured. There are 500 images containing a single object and 1000 images containing multiple objects. Each group of images forms a database, respectively, called the SOD and MOD, and their corresponding binary ground truth GT is manually obtained. In remote sensing images, all kinds of interesting objects have different appearances and shapes, but the objects share a great deal in common with surrounding backgrounds in color, texture, and shape. Complicated backgrounds (such as forest, lakes, sand, roads, and lawns) and various conditions (including fog, shadow, and luminance fluctuation) easily lead to false detection. Sample images from the two datasets are shown in Fig. 2.

3.2 *Experimental Setup*

The database test images were resized to 400×400 pixels. For these experiments, we set the patch size $P = 2$, $Q = 2$, the first clustering number $K_D = 10$, the parameters $\sigma_x = 100$ and $\sigma_y = 100$ in Eq. (9).

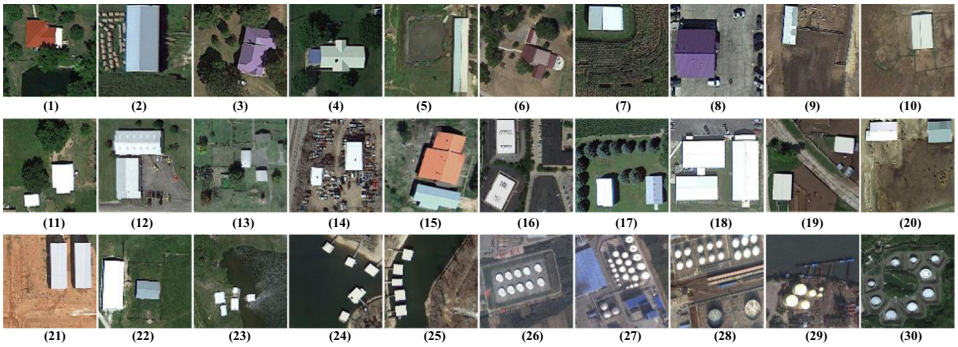


Fig. 2 Samples from the databases: (1)–(10) are from the single-object dataset (SOD) that contains 500 single-object images and (11)–(30) are from the multiple-object dataset (MOD) that contains 1000 multiobject images. These objects have different shapes, colors, and illumination.

We carried out the experiments to certify the efficiency of the combination of global cues and background prior, and the experimental results are detailed in Sec. 3.2.1. We note that the selection of patch size affects the performance of SGSM, and there are different outputs at different scales. Hence, we employed multiple scales to produce a better saliency map. The selection of these multiple scales is based on the experimental results of Sec. 3.2.2.

3.2.1 *Combining global cues and background prior information*

The global-based saliency map (G-map), background-based saliency map (B-map), and final saliency map (C-map) obtained by combining both maps were obtained for all 1500 images from the SOD and MOD. The performance of these three saliency maps is shown in Fig. 3(a), where it can be seen that the information selected to generate the dictionaries affects the results of saliency detection. The sparse coefficients computed by the global- and background-based dictionaries show different relationships among the same patches. The objects are easily confused with the background if they have sparse coefficients that are similar to it. In addition, sparse coefficients computed by the global-based dictionary interpret the relationship between image patches and all the categories that the image contains, while sparse coefficients computed by background-based dictionary interpret the relationship between image patches and the categories that the background contains. The C-map clearly generates the best results. From Fig. 3(b), we can see that the integration of global cues and background prior information results in better precision and recall (PR) values and detects salient regions more accurately and efficiently.

3.2.2 *Selection of multiple scales*

We can obtain k saliency maps with the procedure in Sec. 2.5 in k scales, and chose the scale on the basis of experimental analysis. According to the results of different scales shown in Fig. 4, we chose $k = 2$ to generate SGSM saliency maps at two scales in order to obtain an efficient and accurate saliency map. Furthermore, we set $\varepsilon = 0.2$ and $\psi = 0.8$ in Eq. (14).

3.3 *Experimental Evaluation Measures*

3.3.1 *Precision and recall curves and F-measure*

We evaluated the results of our algorithm to a manually generated ground truth using the PR curve[28,37] and F-measure.[28,37] Precision measures the ratio of correctly assigned salient pixels to

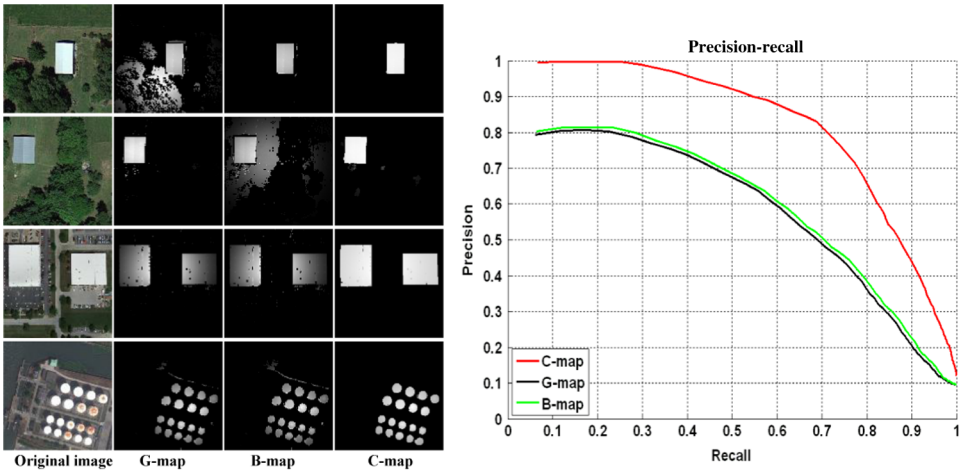


Fig. 3 Comparison of G-map, B-map, and C-map: (a) saliency maps computed from different clustering dictionaries and (b) average precision and recall (PR) curves of 1500 images from the SOD and MOD.

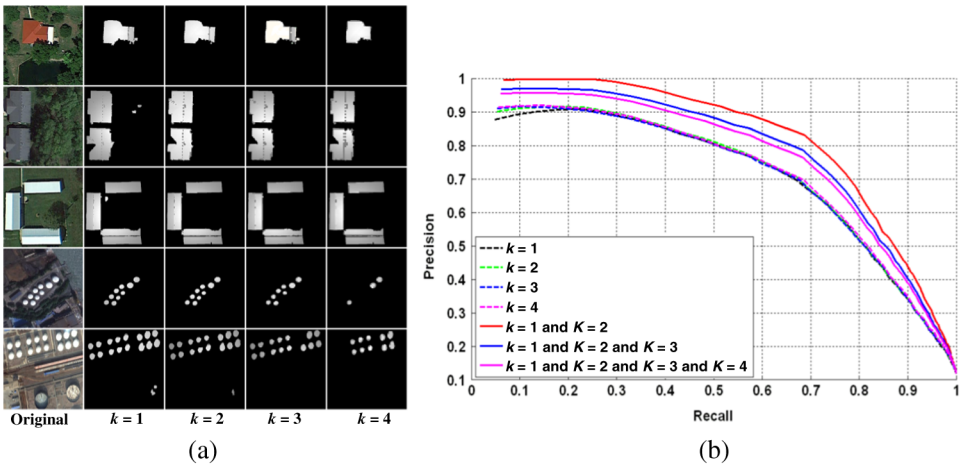


Fig. 4 Comparison of saliency maps at different scales: (a) visual results of four scales from SOD and MOD and (b) average PR curves in four scales and the combination of multiple scales.

all pixels of the extracted regions. Recall measures the percentage of detected salient pixels to the salient ground truth in the same image. A binary map is generated with the threshold $T \in [0,255]$ and then compared to the ground truth image to obtain the average PR values of all the images in the datasets to measure the overall performance. The F-measure is computed as the weighted harmonic of precision and recall and is defined as:

$$F_\beta = \frac{(1+\beta^2)\text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}}. \tag{15}$$

We set β^2 to 0.3 for these experiments.[9,15,16,28]

3.3.2 *Mean absolute error*

Similar to Ref. 28, we also evaluated the mean absolute error (MAE) between the binary ground truth GT and final saliency map S_{final} to obtain a more balanced comparison. MAE is defined as:

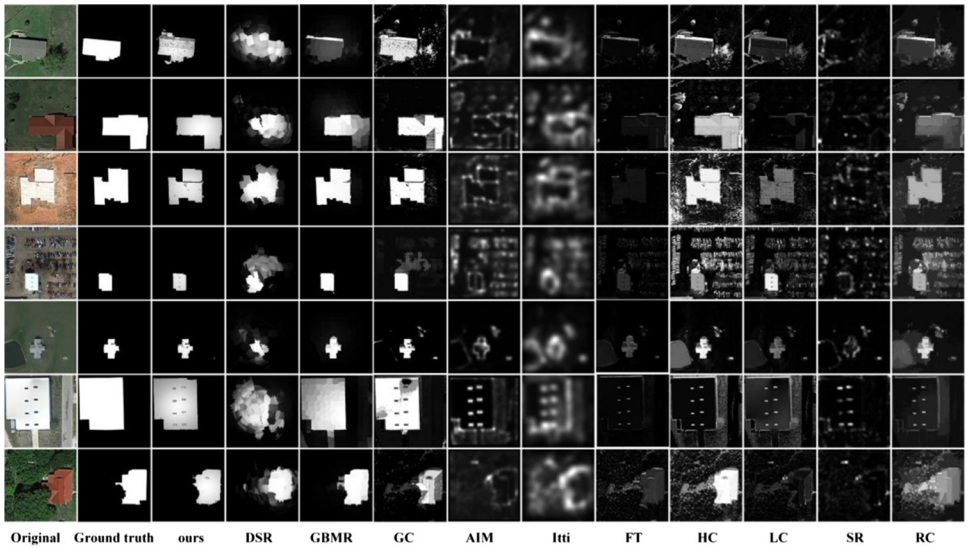


Fig. 5 Saliency maps of the proposed method and 10 state-of-the-art methods for SOD images.

$$\mathrm{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S_{\mathrm{final}}(x, y) - \mathrm{GT}(x, y)|, \tag{16}$$

where W and H, respectively, denote the width and height of the saliency map and ground truth image.

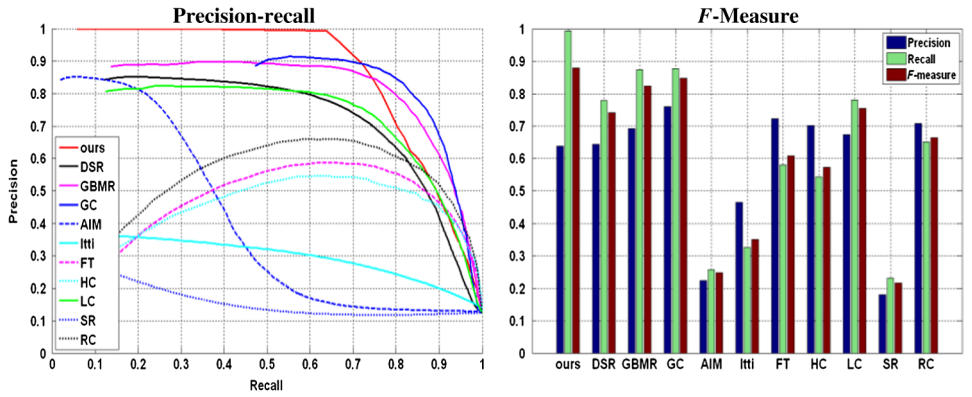
3.4 *Comparison with 10 State-of-the-Art Methods*

We compared our proposed method (SGSM) with 10 state-of-art methods: dense and sparse reconstruction (DSR),[18] graph-based manifold ranking (GBMR),[16] global cues (GC),[14] a model of information maximization (AIM),[11] saliency-based visual attention (Itti),[10] frequency-tuned (FT),[27] histogram-based contrast (HC),[15] spatial attention model (LC),[8] spectral residual (SR),[26] and region-based contrast (RC).[15] We find that it is very difficult for all these methods to exactly detect the saliency region in remote sensing images. Two experiments were performed to validate the efficiency of the proposed method. The first experiment detected a single salient object from the SOD, while the second group detected multiple salient objects from the MOD. The results of single object detection are illustrated in Figs. 5 and 6, and those of multiple object detection are illustrated in Figs. 7 and 8. Figures 5 and 7 show 11 saliency models, and Figs. 6 and 8 show the PR curves and F-measure values. Table 1 lists the MAE results for both the SOD and MOD.

3.4.1 *Single salient object detection*

Methods exploiting low-level cues such as AIM, Itti, FT, and SR tend to find the boundaries of the salient object. Methods employing global cues such as GC, RC, and LC are likely to mistake background noise as salient points. Methods based on background priors such as DSR and GBMR fail to accurately detect salient regions, specifically when the salient regions have a similar appearance to the background. Our method, exploiting both global cues and background prior information, produces a more precise saliency map. It can distinguish features when the object and background regions share the same appearance. The high-level cues of the patches, which are learned from the global and background dictionaries, can precisely reveal the category of the patches. Therefore, the categories of all patches can be obtained by the machine learning method.

Figure 6 shows that our proposed SGSM can highlight the entire salient region of an object. Furthermore, it has a higher F-measure. It is superior to 10 state-of-art methods both in terms of integrity and accuracy of object segmentation. When the object has similar color and a different structure compared with the background, SGSM can detect the differences and highlight the



(a) (b)

Fig. 6 Performance of the proposed method and 10 state-of-the-art methods: (a) average PR curves and (b) F-measures.

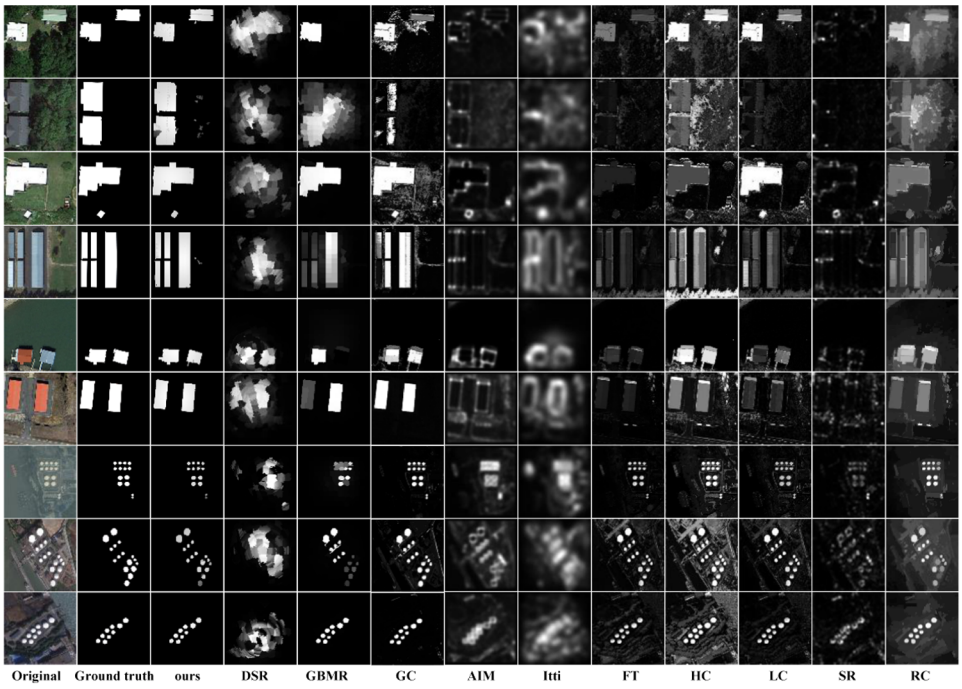


Fig. 7 Saliency maps of proposed method and 10 state-of-the-art methods for MOD images. The images show man-made objects including houses and oil depots.

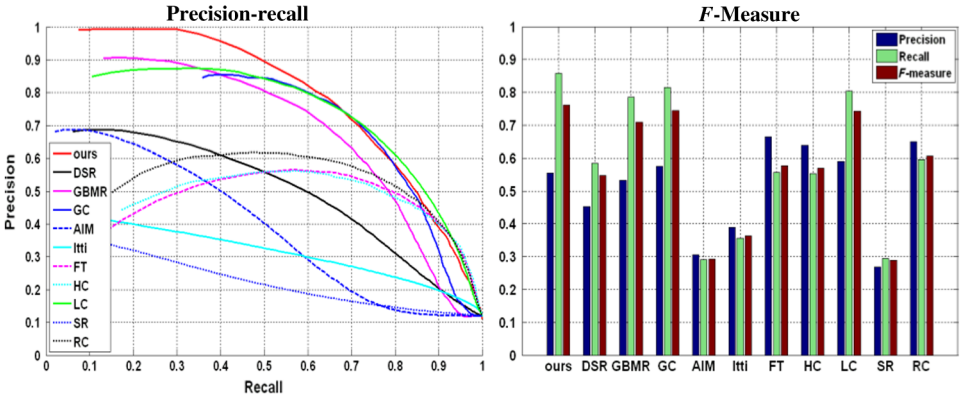


Fig. 8 Performance of the proposed method compared to 10 other methods: (a) average PR curves and (b) *F*-measures.

corresponding regions. All kinds of interesting objects in the testing database have different colors, sizes, type attributes, and forms, which makes every detection task unique and difficult. Facing these complicated situations, our method can still acquire the better saliency detection results. From Table 1, we can see that our method is closer to the ground truth and reduces MAE by 24.44% with respect to the previous best method, GBMR.

Table 1 Mean absolute errors of the proposed method and 10 state-of-the-art methods.

	Ours	DSR	GBMR	GC	AIM	Itti	FT	HC	LC	SR	RC
SOD	0.0306	0.0842	0.0405	0.0637	0.1833	0.2203	0.1348	0.1842	0.1147	0.1547	0.2024
MOD	0.0320	0.1169	0.0639	0.0971	0.1682	0.2187	0.1267	0.1697	0.1079	0.1418	0.1973

3.4.2 *Multiple salient object detection*

In contrast to the detection of a single object, it is difficult to identify two or more objects with different colors and shapes in one image. Figure 7 shows that our model achieves the best results visually of all the saliency models. Methods exploiting low-level cues such as AIM, Itti, FT, and SR hardly detect the objects at all. Methods employing global cues such as GC, RC, and LC cannot generate accurate saliency maps because of noise interference.

GBMR is unable to detect the objects if they have different appearances because of its dependence on ranking with queries, as it is likely to mistake objects with a lower ranking score as background. GC fails to detect objects that have an analogous appearance to the background because it relies on the color histogram. But our method can avoid these situations, because it stems from the machine learning theory. It can precisely categorize the patches though there are multiple salient objects in one image.

Figure 8 shows that our model also has better PR values and F-measures than the other 10 saliency methods. Because it ignores background patches judged to be salient parts by the others, the final saliency maps of SGSM are much closer to the ground truth. In comparison to single salient object detection, it is more difficult to detect multiple salient objects in one image because not only do the objects have different types and appearances, but a portion of object areas are similar to the background. In addition, the objects may be covered by the fog, sheltered by the trees, or interfered with by their own shadows. The test results demonstrate that our proposed method can weaken these interferences and precisely detect the edge of the multiple salient objects. Table 1 proves that our method has less error when detecting multiple objects, reducing the error by 52.11% with respect to the second-best method, GBMR.

4 Conclusion

In this paper, we proposed a sparsity-guided saliency detection method based on global cues and background prior information for remote sensing images. This method uses a sparse representation to obtain high-level global and background cues, and then integrates the saliency maps generated by both of these cues using a Bayesian formula. Consequently, SGSM not only considers the global and background properties of the image content, but also introduces a sparse representation for high-level cues. The proposed method was evaluated on a database of remote sensing images that contained diverse textures, structures, and complex conditions. Experimental results showed that our method outperforms 10 state-of-the-art saliency detection methods, yielding higher precision and better recall rates, in particular when multiple salient objects have analogous appearances. But our propose method is not very effective for low-resolution remote sensing images with fewer detail features. Furthermore, the problem of the time consumed problem also urgently needs to be resolved. In the next work, we intend to use enforcement learning or a deep learning algorithm to obtain more high-level cues and obtain fast and precise saliency detection results.

In addition, rather than performing a traversal search, quickly and accurately extracting some salient object regions can be useful for large data volumes of remote sensing images, which in turn will improve the object detection and recognition rate in cluttered scenes. Hence, our future work will also focus on how to automatically detect and recognize objects (e.g., houses and oil depots) based on SGSM.

Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities and the National Natural Science Foundation of China (Nos. 60802043, 61071137, and 61271409), National Basic Research Program (also called the 973 Program, No. 2010CB327900), Aviation Science Foundation Project, and Space Support Foundation Project.

References

1. J. Sun et al., "Salient region detection in high resolution remote sensing images," in *Proc. Wireless and Optical Communications Conf.*, pp. 1–4 (2010).

2. Y. J. Shi et al., "Multiview saliency detection based on improved multi manifold ranking," *J. Electron. Imaging* **23**(6), 061113 (2014).
3. M. H. Tian, S. H. Wan, and L. H. Yue, "A novel approach for change detection in remote sensing image based on saliency map," in *Proc. Computer Graphics, Imaging and Visualisation*, pp. 397–402 (2010).
4. C. B. Chen et al., "Saliency modeling via outlier detection," *J. Electron. Imaging* **23**(5), 053023 (2014).
5. J. B. Zhao et al., "Unsupervised saliency detection and a-contrario based segmentation for satellite images," in *Proc. Seventh Int. Conf. on Sensing Technology*, pp. 678–681 (2013).
6. L. B. Zhang and K. N. Yang, "Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image," *IEEE Geosci. Remote Sens. Lett.* **11**(5), 916–920 (2014).
7. X. Li et al., "Contextual hypergraph modeling for salient object detection," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 3328–3335 (2013).
8. Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatio temporal cues,"in *Proc. 14th Annual ACM Int. Conf. on Multimedia*, p. 815 (2006).
9. M. M. Cheng et al., "Salient object detection and segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 409–416 (2011).
10. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998).
11. N. Bruce and J. Tsotsos, "Saliency based on information maximization," *Adv. Neural Inf. Process. Syst.* 155–162 (2005).
12. S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2376–2383 (2010).
13. V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color orientation," *IEEE Trans. Multimedia* **11**(5), 892–905 (2009).
14. M. M. Cheng et al., "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1529–1536 (2013).
15. M. M. Cheng et al., "Global contrast based salient region detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 409–416 (2011).
16. C. Yang et al., "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3166–3173 (2013).
17. Y. S. Chen and A. B. Chan, "Adaptive figure-ground classification," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 654–661 (2012).
18. X. H. Li et al., "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 2976–2983 (2013).
19. X. H. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 853–860 (2012).
20. R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1139–1146 (2013).
21. Y. L. Xie, H. C. Lu, and M. H. Yang, "Bayesian saliency via low and mid levels cues," *IEEE Trans. Image Process.* **22**(5), 1689–1698 (2013).
22. A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 478–485 (2012).
23. C. Scharfenberger et al., "Statistical textural distinctiveness for salient region detection in natural images," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 979–986 (2013).
24. J. Zhu et al., "Unsupervised object class discovery via saliency guided multiple class learning," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3218–3225 (2012).
25. T. Liu et al., "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 353–367 (2011).
26. X. D. Hou and L. Q. Zhang, "Saliency detection: a spectral residual approach," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–8 (2007).
27. R. Achanta et al., "Frequency-tuned salient region detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1597–1604 (2009).

28. F. Perazzi et al., "Saliency filters: contrast based filtering for salient region detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 733–740 (2012).
29. Q. Yan et al., "Hierarchical saliency detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1155–1162 (2013).
30. P. Jiang et al., "Salient region detection by UFO: uniqueness, focusness and objectness," *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1976–1983 (2013).
31. Y. Q. Jia and M. Han, "Category-independent object-level saliency detection," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1761–1768 (2013).
32. R. S. Hunter, "Photoelectric color difference meter," *Proc. J. Opt. Soc. Am.* **48**(12), 985–993 (1958).
33. K. Kreutz-Delgado et al., "Dictionary learning algorithms for sparse representation," *Neural Comput.* **15**(2), 349–396 (2003).
34. M. Zheng et al., "Graph regularized sparsecoding for image representation," *IEEE Trans. Image Process.* **20**(5), 1327–1336 (2011).
35. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code of natural images," *Nature* **381**, 607–609 (1996).
36. S. L. Kukreja, J. Lofberg, and M. J. Brenner, "A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification," in *Proc. 14th IFAC Symp. on System Identification*, pp. 814–819 (2006).
37. R. S. Hunter, "Accuracy, precision, and stability of new photoelectric color difference meter," *J. Opt. Soc. Am.* **38**(12), 1094 (1948).

Danpei Zhao received her PhD in optical engineering from Changchun Institute of Optics, Fine Mechanics and Physics of Chinese Academy of Sciences in 2006. From 2006 to 2008, she was in Beihang University for postdoctoral research. Until now, she has been engaged in teaching and research work in Beihang University. Her research interests include automatic remote sensing image understanding technology, moving target detection, and tracking and recognition for complicated scenes.

Jiajia Wang received her BS degree in electrical and information engineering from North China Institute of Aerospace Engineering in 2010 and her MS degree from Beihang University in 2012. Her research interests include saliency detection and object detection for remote sensing image.

Jun Shi received his BS degree from Huainan Normal University,China, in 2007 and his MS degree from Yangzhou University,China, in 2011. Currently, he is pursuing his PhD degree in the Image Processing Center, School of Astronautics, Beijing University of Aeronautics and Astronautics, China. His research interests include computer vision, pattern recognition and machine learning.

Zhiguo Jiang is a professor at Beihang University, and has been the vice dean of the School of Astronautics at Beihang University since 2006. Currently, he serves as a standing member of the Executive Council of China Society of Image and Graphics and also serves as a member of the Executive Council of Chinese Society of Astronautics. He is an editor for the Chinese Journal of Stereology and Image Analysis. His current research interests include remote sensing image analysis, target detection, tracking and recognition, and medical image processing.