IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2019, accepted February 22, 2019, date of publication February 26, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901776

Subjective Saliency Model Driven by Multi-Cues Stimulus for Airport Detection

DANPEI ZHAO, (Member, IEEE), JIAYI LI, ZHENWEI SHI, (Member, IEEE), ZHIGUO JIANG, AND CAI MENG

Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China
Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technologies, Ministry of Education, Beihang University, Beijing 100191, China

Corresponding author: Cai Meng (tsai@buaa.edu.cn)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFC 1405600, in part by the National Natural Science Foundation of China under Grant 61671037, and in part by the Aeronautical Science Foundation of China under Grant 2017ZC51046.

ABSTRACT Traditional saliency models designed for natural scene images usually use human visual characteristics to detect the target, but those salient areas in the remote sensing images (RSIs) may not be the targets we are really interested in. Taking both remote sensing image attributes and airport characteristics into account, we put forward a subjective saliency model driven by multi-cues stimulus for the airport's detection (MCS-SSM). Different from traditional saliency models, this model mainly relies on the subjective target detection task to find specific target area eliminating disturbance from other salient targets. Based on the low-level features, we train an LDA classifier by only small target samples and then build an object feature map. In the meantime, the shape information based on line density is extracted to get a shape map. Depending on the fusion result of two saliency maps, we optimize the subjective saliency map with SVM classifier. Moreover, MST density map is generated to suppress background and highlight interesting airport regions in the subjective saliency map. Consequently, MCS-SSM can respectively take the target, the background, and the detection task as multiple cues to quickly locate interest airport targets in RSI with a large cover area. MCS-SSM breaks through the limitations on color, texture, and other low-level characteristics compared with traditional saliency models, which are more targeted to detect the specific targets. The extensive experimental results demonstrate that the proposed MCS-SSM outperforms nine state-of-the-art saliency models. Besides, it has a higher detection rate and better effective performance than other three airport detection approaches.

INDEX TERMS Airport detection, LDA topic model, multi-cues, subjective saliency model.

I. INTRODUCTION

In recent years, the airport detection has greater value and directive significance in both military and civil fields as an important branch of remote sensing application. The airport areas are usually accompanied by a very complex background including rivers, mountains, roads and buildings in the low-resolution remote sensing images (RSIs). These background regions may share the same features with the airports in different resolution of images, which brought difficulties to the detection and recognition. So far, many methods for airport detection problems have been proposed. These methods can be mainly classified into two categories: either geometrical features or machine-learning strategy. The first one is usually based on the long straight airport runways. Reference [1] use line feature as a priori knowledge. Reference [2] proposed a new way to extract lines for airport detection. Considering the line feature as the most important, these methods can achieve good results for high-resolution RSIs. However, this feature will cause some confusion if the background region contains rivers, houses or roads which can be extracted a large number of line features. Moreover, it is difficult to find the long straight runways for low-resolution RSIs. The second one is based on the machine learning. In recent years, machine learning is widely used in the target detection, especially when deep convolutional neural network (CNN) shows excellent performance in extracting representative features. Such as [3]. This is the first time to apply deep CNN for airport detection and it has achieved better results than other traditional methods in the condition

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

of complex background. However, deep learning methods usually rely on enough samples too much and require high computation expense, which is hardly satisfied in actually RSIs application. Then, saliency models are introduced to airport detection, which can quickly locate airport areas in large scale RSIs.

Saliency detection has become increasingly popular due to its importance in the field of computer vision. Visual saliency attempts to discover some important parts where people firstly pay attention to when they see a image. Consequently, traditional saliency models are usually used as a pre-processing procedure during detecting a target in an image. They can minimize time and cost by only concerning visual salient regions instead of the whole image, especially when it comes into RSIs.

Saliency models can be briefly divided into two types: traditional saliency models or deep learning-based saliency models. Traditional saliency models started from the article [4] proposed by Itti *et al.* This article laid the basic idea of saliency problem: first, extract feature; second, normalization; third, calculate the saliency value and last, divide regions of interest. Model LC [5], SR [6], FT [7] were proposed relatively early. These models' algorithms are easy to understand. The calculation amount is small, but the detection results are not good, which can only roughly locate the target. In the last few years, many novel saliency models provide new ideas and show great performance. Reference [8] proposed a salient object detection via global and local cues. The learning classifier studies some characteristics of pixels in a single image. Reference [9] introduce the minimum spanning tree (MST) to image representation. It has been proved its effectiveness in image matching based on the graph and the filter. And it can expedite processing time. But when using these saliency models for RSIs or low-resolution images, the results are not as good as natural images. The second type of saliency models, the deep learning-based saliency models, have emerged after the CNN was widely used. Moreover, many saliency models use AlexNet [10], GoogleNet [11], and VGGNet [12] as a pre-trained model. Such as [13], use VGGNet for capturing high-level features and another net for low-level features, then put these features into a two-layer perceptron for saliency detection. Reference [14] propose a method that uses both fixation prediction and salient object detection to lead the network. These deep learning-based saliency models can show great performance in precision and accuracy. But these saliency models do not have the ability to distinguish different salient targets, that is to say, they cannot choose to pay attention to those targets which they are most interested in. As for airport detection in RSIs, we can encounter other difficult problems, for example, insufficient training data set or feature variations at different resolutions.

To solve above problems, we propose a novel airport detection framework based on subjective saliency model driven by multiple cues stimulus (MCS-SSM). MCS-SSM is constructed with the background cue and the target cues which include shape map and object feature map. Shape map is built by the line density which can lead our model to locate specific target quickly. And object feature map is generated with LDA classifier based on small sample learning, which improves the robustness by introducing the target cues into the saliency model. As for the background cue, a Minnimum Spinning Tree based density map is used to suppress background and reduce areas of interference.

The main contributions of our approach are summarized in two aspects below:

1) In this paper, we propose a new frame of subjective saliency model which models the intrinsic properties of salient objects in shape and feature learning manner. Different from traditional visual saliency models, we introduce human's subjective intentions into the saliency model on the basis of the low-level visual characteristics. This subjective saliency model realized selectively highlight specific target in low-resolution RSIs which mainly relies on the detection task and object attribute. The proposed framework can also be extended to other target detection only using circle or other shape instead of straight line.
2) We present MCS-SSM driven by multi-cues stimulus and multi-level learning. Multi-cues perform collaborative saliency map by combing shape map, object feature map and the background map, which realize to selectively locate airport target. Multi-level learning includes object feature learning based on LDA and saliency optimization based on Support Vector Machine (SVM) learning, resulting in the saliency performance improvement in object detection. Even if the airport target is not salient in terms of color or texture for the whole RSIs, MCS-SSM can also accurately detect it.

II. MCS-SSM

In this section, we will discuss the proposed MCS-SSM for airport target detection in detail. We can see the whole framework clearly from the flowchart. In the first step, a shape map can be extracted based on the line features and line density map, thus, the non-linear regions are excluded. Second, an LDA topic model is trained to get the object feature map that based on visual and subjective dual perception. These two maps are merged as a subjective perception-based initial result. Furthermore, the subjective saliency map is optimized using SVM learning. Third, the MST density map is constructed to remove the environmental disturbance by background cues.

A. SUBJECTIVE SALIENCY MODEL

1) SUPERPIXEL SEGMENTATION

Superpixel segmentation is used to reduce the computational complexity of subsequent processing. In this paper, we adopt the Simple Linear Iterative Clustering (SLIC) superpixels segmentation. Four different numbers (Nsp = 100, 150, 200 and 250) of the superpixels are chosen for each image so that the result will be more accurate.

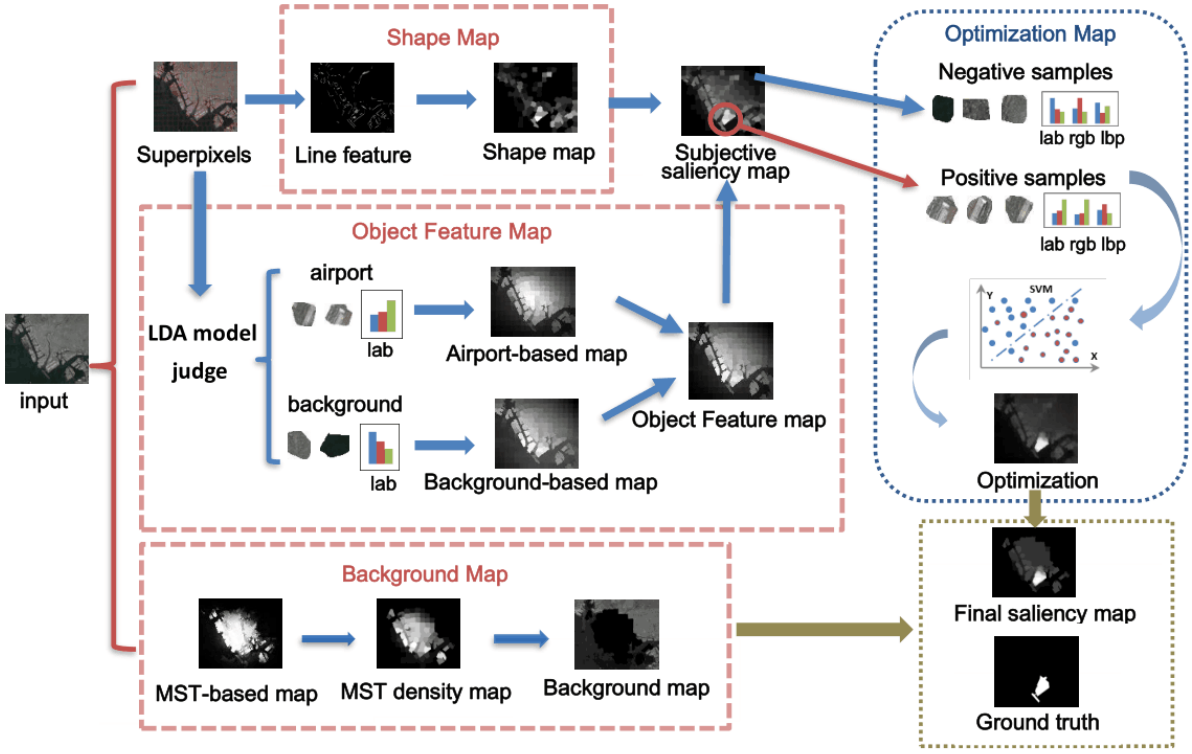

FIGURE 1. Flowchart of MCS-SSM.

2) OBJECT FEATURE MAP BASED ON LDA LEARNING

Different from the traditional visual saliency model, subjective saliency model use the target's own characteristics as the driving cues and use only about a hundred images instead of tens of thousands for training the LDA model, so we can achieve the subjectivity map according to the distribution characteristics of the actual RSIs. Therefore, this step will save more time in training process and avoid the problem that the number of training images is too small.

First, the Lab features of both the background and the airport target are extracted to train the LDA model. Then, the trained LDA model will classify all the superpixels into two kinds: LDA-background and LDA-target. The superpixels of the boundary area in the LDA-background are selected as background region (BR), and the rest of superpixels as the target region (TR). We calculate the Euclidean color distances ($d(c_i, c_n)$) and spatial distances ($d(s_i, s_n)$) between the superpixels belonging to the target region(TR) and the background region(BR), as the weight of each superpixel in the Background-based map, as follows:

$$S_R(i) = \frac{1}{N} \sum_{n,n\in BR,i\in TR} d(c_i, c_n)(1 - d(s_i, s_n)), \qquad (1)$$

N represents the number of superpixels in BR, and both distances are normalized to [0,1]. The result $S_R(i)$ means the weight of every pixel in the superpixels.

An adaptive threshold of the weight is set by [15]. And through the threshold, the new target region can be achieved. Then delete the superpixels that belong to the LDA-background in the target region. Calculate the Euclidean color and spatial distances as well between the airport region (AR) and the remaining region (RR) as the weight of each superpixel in the Airport-based map, as follows:

$$S_T(i) = \sum_{n,n\in AR,i\in RR} \frac{1}{d(c_i, c_n) + d(s_i, s_n)}, \qquad (2)$$

Then these two maps are added to get the Object feature map: $S_s(i)$.

3) SHAPE MAP BASED ON LINE DENSITY

In our subjective saliency model, special shape features of the target are used to characterize the uniqueness of the target such as the line, the round, the square or any shape feature that belongs to the target as a complement of the learning process. The line features are extracted as airport shape feature. However, the shape feature sometimes can be similar with some background regions, so we modify the shape feature map into shape density map. The value in shape density map can be calculated as:

$$S(shape) = \frac{N_1}{N}, \qquad (3)$$

Which N_1 represents the number of pixels with lines, and N represents the total number of pixels in this superpixel.

In addition, we also use four different superpixel sizes in this step so that we include the most accurate result.

4) SUBJECTIVE SALIENCY MAP

The subjective saliency map includes two different parts, one of which is shape map reflecting the geometrical shape characteristics of the target, the other of which is object feature map representing the learning characteristics of the target after training. By fusion, the subjective saliency map will be able to highlight specific salient targets. The fusion formula is as follows:

$$S(drive) = \frac{1}{2}S(shape) + \frac{1}{2}S_s(i), \tag{4}$$

B. LEARNING-BASED TARGET REGION OPTIMIZATION

In order to get subjective clues as for the target characteristics, we use a small number of airport samples to train the LDA model, which makes our saliency model fully represent the characteristics of the target itself. Because there are obvious differences in illumination, resolution and texture details between the actual test images and those training samples. And, only a hundred of training samples are used to train the classifier. It results in inaccurate target area in subjective saliency map. To solve this problem, a learning-based method is put to use to optimize Subjective saliency map, which makes the target region to be more accurate and more complete in new map.

Because the size of the airport in RSIs is not sure, four sizes of superpixels are used in each image in the beginning, so four maps are got from the previous steps. According to these maps, we can increase the chance that we include the most appropriate size for the airport detection. Meanwhile, we add more weight to the superpixels that has more possibility to be the target in the SVM training process.

According to the drive map we generated from the last step, we calculate each superpixel of the average value, and set two thresholds: v_p, v_n ($v_p > v_n$). In the experiments, v_p is 4 times the average value of the whole Sample map, v_n is set to a fixed 0.05 (all values have been conducted in the normalized operation). In the drive map, if the value of superpixels is higher than v_p, we consider it as a positive example; if the value of superpixels is lower than v_n, we consider it as a negative example; others are classified to the test set. All of the positive examples in four different superpixels size are added as the final training sample set and the negative examples as well. So all the examples come from the same pictures and it ensures all the feature is adapt to the test images. In this paper, we use LAB, RGB and LBP feature in the training process. Because of the airport in one remote sensing image can have large differences in another remote sensing image, it does not have a fix features; there is a difficulty to choose the kernel in SVM. So, we decided to use four different SVM kernels (linear kernel, polynomial kernel, radial basis function, and sigmoid functions) for each feature and then use AdaBoost to integration those results as MKL method [16]. After this process, an optimization map (M_i) is obtained.

C. MST-BASED BACKGROUND SUPPRESSION

In our saliency model, the object feature map and the shape map are both based on the target characteristics. The similarities between the targets are explored from the view of the detection task. In the meantime, we use MST to describe the image and then extract the background regions in pixel-level as a complement. Consequently, we can better suppress the background and get clearer boundaries of saliency target.

In the optimization map M, we noticed that the airport region can be extracted precisely, but some pixels of background regions still have higher values. So it is important to restrain remaining background regions. As we use superpixel-level in optimization map, we specially employ the pixel-level feature to represent the original image so as to make up for some deficiency caused by superpixels.

We take advantage of MST to represent the original image, and then use a barrier distance to get the maximum and the minimum values of every node. As is mentioned above, the border of the image is considered as the background, thus boundary pixels are clustered into 3 groups(because the boundary of RSIs may include rivers, forest, and airport) by their color values in Lab color space. So the distance can be calculated as follows:

$$S_k(i) = \sqrt{(I_i - \mu_k)C_k^{-1}(I_i - \mu_k)^T}, \tag{5}$$

And normalize it as the value in the MST-based map. Where μ_k represents the mean color and C_k represents the covariance matrix.

Like line density map, an MST density map can be get. For the superpixel i, the value DM_i in the MST density can be calculated as follows:

$$DM_i = \frac{\sum value}{N}, \tag{6}$$

where N is the total number of pixels in the region. In this paper, we compute DM_i in four different sizes which are the same as the sizes in the first step.

The MST density map is used to suppress the background in order that the MST map can find the background exactly right, but it cannot give a clear target area especially when the images have a low contrast or extreme illumination conditions.

D. FINAL MAP

We transform the MST density map into a binary image by setting a small enough threshold. Only when the value in the MST density map is zero, we consider it zero, and if the value bigger than zero, we consider it one as follows:

$$\theta = \begin{cases} 0 & value_i = 0 \\ 1 & value_i > 0, \end{cases} \tag{7}$$

We use θ to modify the target area map as follows:

$$S_i = \theta \times M_i, \quad i = 1, 2, 3, 4, \tag{8}$$

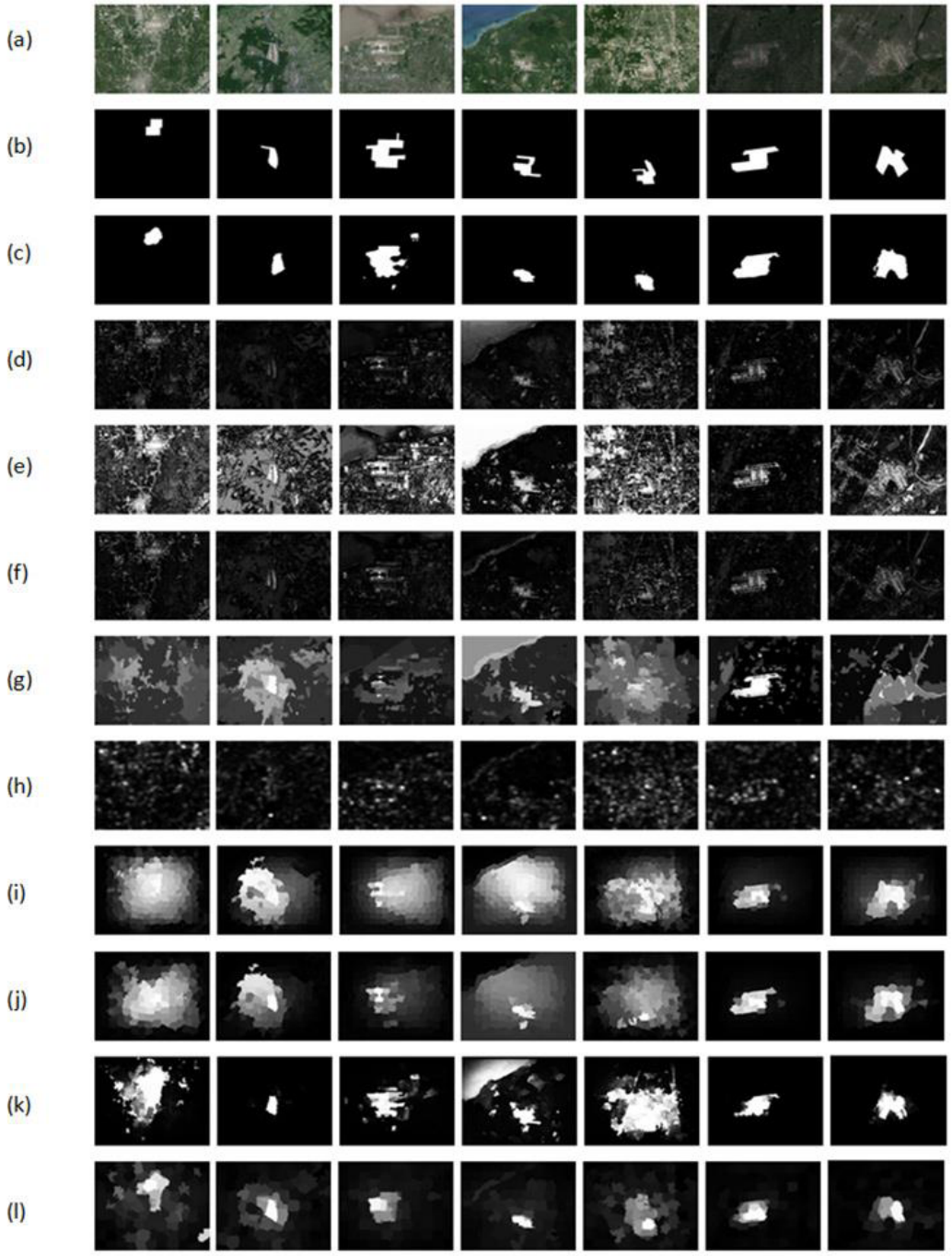

FIGURE 2. Comparison results of several saliency models. From top to down: (a)Input (b)GT (c)ours (d)FT (e)HC (f)LC (g)RC (h)SR (i)BSCA (j)GBMR (k)MST (l)GL.

The final saliency map is an image which fuses the four maps:

$$S = \frac{1}{4}\sum S_i, \tag{9}$$

By the three cues of this model, we can detect the airport target well and suppress the interference from the surrounding background.

III. EXPERIMENTS

Our proposed method is tested on an airport data set which includes 360 images containing 28 airports with multi-resolution (from 30 meters to 50 meters) and different lighting conditions. This data set includes forest, building, river, sea, and lands, which almost cover every possibility in RSIs. In addition, most of these images have complex

backgrounds. The ground-truth(GT) dataset was manually formed. Each image is 500*600 in size.

A. COMPARISON WITH OTHER SALIENCY MODELS

Our saliency model is compared with other nine saliency models (FT [7], HC [17], LC [5], RC [4], SR [6], BSCA [18], GBMR [19], MST [9], GL [8]) with precision and recall (P-R) curve, MAE and F-measure.

Figure 2 shows that MCS-SSM methods can find the target more accurately in different kinds of background than the other nine models. Those nine models can achieve good results only for some images of the data set but not all, usually less than half, especially when there are both land and sea in one image. Because the contrast between airports and the rest of the land is not as obvious as the contrast between land and sea, the saliency models tend to take the small one between the two areas (land and sea) as the saliency area, but not to find the real airport target. Instead of considering the higher contrast as the most important factor of the target, MCS-SSM emphasize the shape features and learning features of the target itself as well. Therefore, MCS-SSM is more robust to suppress interference background. MCS-SSM are based on both superpixel level (four sizes of superpixel) and pixel level so that the final saliency map can always hit the target region and be more similar with the ground truth.

PR-curve and F-measure are used for evaluating performance of saliency models. Precision represents the proportion of actual salient targets in the saliency regions. Recall represents the detected proportion in actual salient targets. The F-measure is computed as follows:

$$F-measure = \frac{1.3 \times Precision \times Recall}{0.3 \times Precision + Recall} S_i, \qquad (10)$$

We can learn from figure 3, the five curves at the bottom are the classical saliency models which were proposed earlier, these models consider only one or two underlying characteristics which cannot deal with complex images, so their F-measure values in the figure 4 are lower than others. The models corresponding to the middle four curves have been proposed in recent years, which can obtain better performance for natural images than classic saliency models in RSI. But if there are some distinct regions in the RSI, such as

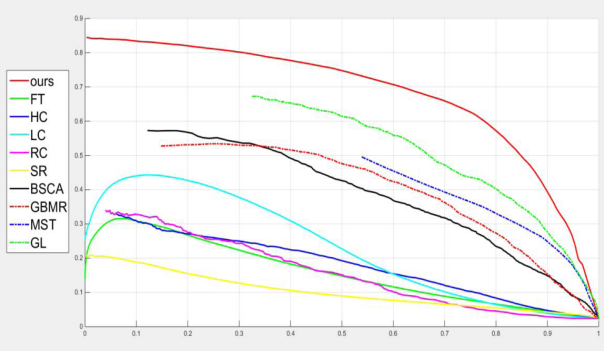

FIGURE 3. PR-curve.

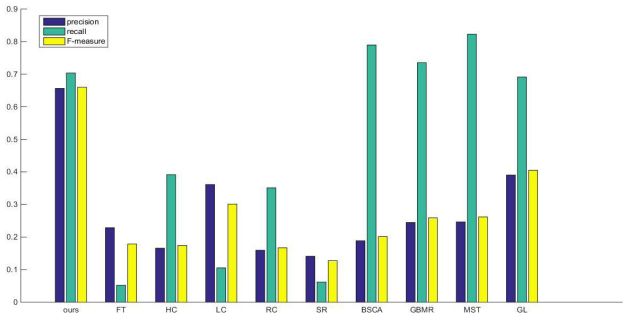

FIGURE 4. F-measure.

rivers, roads or buildings, these models will tend to select those salient areas with the most obvious color or shape rather than distinguishing real target from all the saliency regions. Consequently, in figure 4, these models may have a higher recall, but the precision and F-measure is lower. The MCS-SSM is specially proposed for RSI, from and figure 3 and figure 4, we can find it outperforms other nine saliency models in terms of comprehensive performance.

The MAE is used to get average pixel difference between saliency map and ground truth, it can be computed as follows:

$$MAE = \frac{1}{500 \times 600} \sum_{i=1}^{500} \sum_{j=1}^{600} |S(i,j) - GT(i,j)|, \qquad (11)$$

TABLE 1. MAE results of several saliency models.

Method	MCS-SSM	FT	HC	LC	RC
MAE	0.0599	0.0857	0.2042	0.0759	0.2169
Method	SR	BSCA	GBMR	MST	GL
MAE	0.0990	0.2188	0.2016	0.1306	0.1268

From Table 1, the MAE of our methods is much smaller than others, it can prove that our method is much better than others when used into RSIs. RSIs have no such obvious color difference or special texture which is usually considered as dominant factors for most saliency models. Our MCS-SST combines the subjective features (shape map and object feature map) to give a more accurate and effective cues to drive the model. And also, while facing difference resolutions, the test image is learned in the optimization step to segment the most appropriate salient areas.

B. COMPARISON WITH OTHER AIRPORT DETECTION METHODS

We compared our method with other three airport detection (Reference[20], Reference[1], Reference[21]) and three detection methods based on saliency model in this section.

We compute the IOU of every connected domain in the final map. Then we consider if the IOU>0.4, this region is a correct detection; if the IOU<0.2, this region is a false

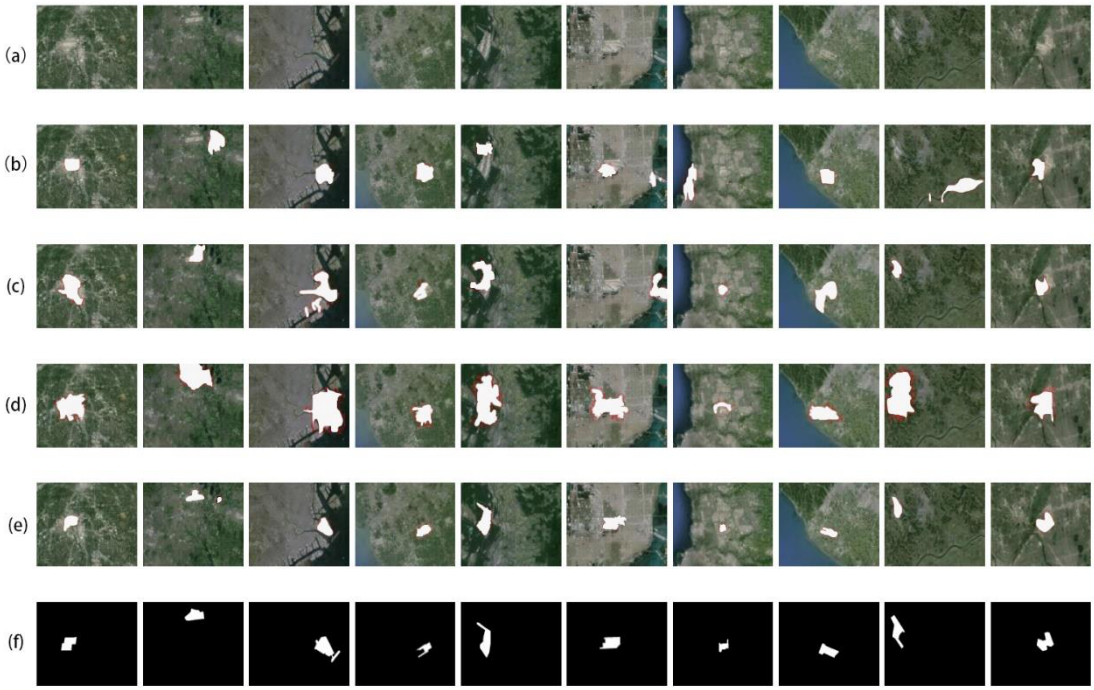

FIGURE 5. Comparison with other airport detection methods. (a)original (b)Reference[20] (c)Reference[1] (d)Reference[21] (e)ours (f)GT.

detection. And if there is no region's IOU>0.2 in one image, the airport target is missing.

We can see from Figure 4 and Table2, our method outperforms the other six methods (three airport detection methods and three saliency model). The precision and recall show that our method can give the target accurately with the least error result. MCS-SSM has several advantages when compared with other detection methods. It takes advantage of the subjective target detection task: LDA topic model and shape map, to generate the target feature map, neither just using some special texture belonging to the airports as most of the airport detection methods nor just using low-level features as other saliency models. Most of the airport detection methods are only applicable to high-resolution RSIs with the small field of view, but our method is suitable for low-resolution images with very large cover areas. Furthermore, it can quickly locate airport targets and eliminate buildings, roads and other interferences in large size RSIs.

TABLE 2. Comparison of airport detection results.

Method	Recall	Precision
MCS-SSM	0.8914	0.8839
Reference [20]	0.2274	0.2314
Reference [1]	0.3978	0.1437
Reference [21]	0.1753	0.0226
MST	0.3733	0.0007
GL	0.4251	0.0008
GBMR	0.1935	0.0003

C. COMPARISON WITH TWO DEEP LEARNING METHODS

In this section, we compare our MCS-SSM with two methods based on deep learning, i.e. FCN [22] (Fully Convolutional Networks) and SSD [23] (Single Shot MultiBox Detector). FCN and SSD have good performance respectively in image segmentation and target detection. Because our database only contains image data of 28 airports which are different in terms of features and appearance, it is insufficient for network training of the deep learning methods. Therefore, we add four airports and expand the original images by rotating and scaling. We utilize 18 airports (430 images) for training and 14 airports (201 images) for testing. Moreover, we use transfer learning to train neural network model of FCN and SSD. The database covers multiple resolutions, multiple illumination conditions, different background interference and the target sizes. To ensure the fairness in testing, we also tested our method with the same test dataset. The precision and recall value are shown in Table 3. The experimental results prove that the detection performance of FCN is not as good as our method although using more than 430 images for training, while SSD method is close to our method in precision, but at the cost of low recall.

TABLE 3. Comparison with two methods based on deep learning.

Method	Recall	Precision
MCS-SSM	0.8955	0.9091
FCN	0.8227	0.8564
SSD	0.6368	0.9014

FIGURE 6. Comparison results with two deep-learning methods. (a) our results (b) FCN results (c) SSD results.

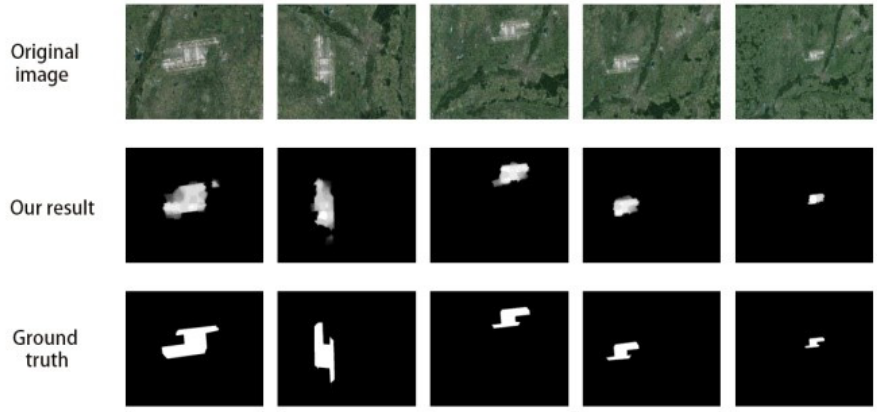

FIGURE 7. Comparison results of different resolution.

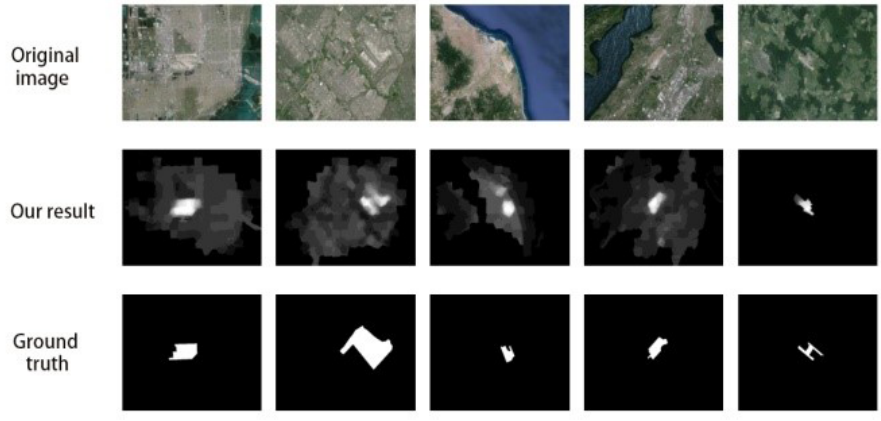

FIGURE 8. Comparison results of low contrast situation.

Through the experimental results, we find that SSD method obtain more accurate detection result for high-resolution, appropriate light and relatively large targets with clear texture. But for low-resolution, dark light condition or small target, the SSD method is difficult to accurately detect targets, and even fails. For example, none of an airportąś images among the 14 testing airports can be identified by SSD method because of the background interference around the

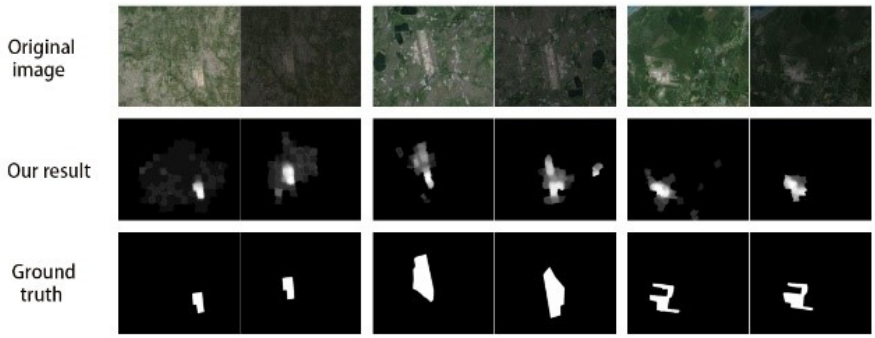

FIGURE 9. Comparison results of different illumination situation.

airport. Comparatively, the performance of the FCN is better which can locate most of the airport targets. But the detection effect becomes worse than our method when the resolution and light intensity get worse. Consequently, experimental results prove that our method has advantages for small targets with low-resolution and poor light condition in the case of less training sample number. When compared to the methods based on deep-learning, our method can obtain an almost same or better result but take less time. Moreover, MCS-SSM can find the target when we do not have sufficient target samples to train deep-learning network.

We also choose five images to show the detection results in Figure 6. The first and second one, all three methods worked well. The airport target in the third image and fourth image cannot be detected by SSD, but can be detected by our method and FCN method. Although the airport in the last picture can be detected by the three methods, the result of FCN method is quite different from the ground truth, so it is considered as the wrong result after calculating the IOU and it marked in red.

D. PERFORMANCE UNDER DIFFERENT RESOLUTIONS, ILLUMINATION AND LOW CONTRAST

From the Figure 7, Figure 8 and Figure 9, we find that MCS-SSM can detect the target precisely in different resolutions, low contrast and different lighting conditions, but the traditional saliency models cannot do it. The accurate results are due to our multi-cue processes. Moreover, we use the results of subjective saliency map to give the SVM classifier samples to optimize the results. And this can greatly improve the performance of our model.

IV. CONCLUSION

In this paper, we proposed an MCS-SSM approach for airports detection in RSIs with low-resolution. Compared with the traditional saliency models, MCS-SSM can only highlight interesting targets and eliminate disturbance from other salient areas according to different detection task. Driven by subjective motivation, this model combines shape cue, object feature cue and background cue to suppress other salient areas except for airport target. Compared with traditional saliency models only relying on the visual features, MCS-SSM can selectively detect specific airport target even if the target is not distinctive. Experiment result shows that MCS-SSM has a higher detection accuracy and stronger robustness for low-resolution RSI. Moreover, our model has good scalability. In the future work, we will extend this model to detect multi-class targets at the same time.

REFERENCES

[1] D. Zhu, B. Wang, and L. Zhang, "Airport target detection in remote sensing images: A new method based on two-way saliency," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1096–1100, May 2015.
[2] Z. Li, Z. Liu, and W. Shi, "Semiautomatic airport runway extraction using a line-finder-aided level set evolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4738–4749, Dec. 2014.
[3] P. Zhang, X. Niu, Y. Dou, and F. Xia, "Airport detection on optical satellite images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1183–1187, Aug. 2017.
[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[5] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th ACM Int. Conf. Multimedia*, 2006, pp. 815–824.
[6] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE CVPR*, Jun. 2007, pp. 1–8.
[7] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE CVPR*, Jun. 2009, pp. 1597–1604.
[8] N. Tong, H. Lu, Y. Zhang, and X. Ruan, "Salient object detection via global and local cues," *Pattern Recognit.*, vol. 48, pp. 3258–3267, Oct. 2015.
[9] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien, "Real-time salient object detection with a minimum spanning tree," in *Proc. IEEE CVPR*, Jun. 2016, pp. 2334–2342.
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
[11] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. CVPR*, Jun. 2015, pp. 1–9.
[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.
[13] L. Gayoung, T. Yu-Wing, and K. Junmo, "Deep saliency with encoded low level distance map and high level features," in *Proc. CVPR*, Jun. 2016, pp. 660–668.
[14] S. S. Kruthiventi, V. Gudisa, J. H. Dholakiya, and R. V. Babu, "Saliency unified: A deep architecture for simultaneous eye fixation prediction and salient object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5781–5790.
[15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[16] F. R. Bach, G. R. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *Proc. ICML*, 2004, pp. 1–8.
[17] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.
[18] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in *Proc. IEEE CVPR*, Jun. 2015, pp. 110–199.
[19] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE CVPR*, Jun. 2013, pp. 3166–3173.
[20] X. Yao, J. Han, L. Guo, S. Bu, and Z. Liu, "A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF," *Neurocomputing*, vol. 164, pp. 162–172, Sep. 2015.
[21] C. Tao, Y. Tan, H. Cai, and J. Tian, "Airport detection from large IKONOS images using clustered SIFT keypoints and region information," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 128–132, Jan. 2011.
[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[23] W. L. B. Liu, J. Matas, N. Sebe, and M. Welling, "SSD: Single shot multibox detector," in *Computer Vision* (Lecture Notes in Computer Science), vol. 9905. Cham, Switzerland: Springer, 2016.

DANPEI ZHAO received the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2006. From 2006 to 2008, she was a Postdoctoral Researcher with Beihang University. She was with the Department of Computer Science, Rutgers, The State University of New Jersey, USA, as a Visiting Scholar from 2014 to 2015. She currently serves as a Standing Member of the Executive Council of the Beijing Society of Image and Graphics. She is also an Associate Professor with Beihang University, where he has been the Vice Director of the Center of Image Processing. Her research interests include saliency detection and its application in remote sensing target detection, remote sensing image understanding, target detection, tracking, and recognition.

JIAYI LI received the B.S. degree from the School of Astronautics, Beihang University, Beijing, in 2017, where she is currently a graduate student. Her research interests include remote sensing image analysis, target classification, detection, and recognition.

ZHENWEI SHI received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, USA, from 2013 to 2014. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing, China. He has authored or co-authored more than 100 scientific papers in refereed journals and proceedings, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE Conference on Computer Vision and Pattern Recognition. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

ZHIGUO JIANG received the B.E., M.S., and Ph.D. degrees from Beihang University, Beijing, China, in 1987, 1990, and 2005, respectively. He was appointed as a Professor in image processing and pattern recognition in 2005. He currently serves as a Standing Member of the Executive Council of China Society of Image and Graphics. His research interests include remote sensing image analysis, medical imaging and analysis, target classification, detection, and recognition.

CAI MENG received the B.S. and Ph.D. degrees in 1999 and 2004, respectively. From 2004 to 2006, he held a Postdoctoral position with the Robotics Institute, Beihang University. Since 2006, he joined the IPC, Beihang University, as a Teacher. He was a Research Fellow with the Social Robotics Lab, National University of Singapore, from 2010 to 2011, and a Visitor with the Medical Robotics Lab, JHU, from 2016 to 2017. He is currently an Associate Professor of the Image Processing Center, School of Astronautics, Beihang University. His current research interest includes machine vision technique and its application in robotics.