SALIENT TARGET DETECTION BASED ON THE COMBINATION OF SUPER-PIXEL AND STATISTICAL SALIENCY FEATURE ANALYSIS FOR REMOTE SENSING IMAGES

Libao Zhang*, Yue Wang and Yang Sun

College of Information Science and Technology, Beijing Normal University, Beijing 100875, China

ABSTRACT

The saliency analysis has become the important tool to detect the salient targets. However, due to complex target features and abundant background information interference, the traditional models are weak in salient target detection of remote sensing images. In this paper, a novel model based on the combination of super-pixel and statistical saliency feature analysis is proposed. The proposed model consists of three main steps. First, the statistical saliency feature map based on histogram statistical saliency analysis in the Lab color space is introduced. Then, information saliency feature map is obtained based on the combination of super-pixel segmentation and information entropy, and the statistical saliency feature map and the information saliency feature map are fused and enhanced to generate the final saliency map. Finally, the complete and accurate salient targets and regions of interest (ROIs) are obtained based on the improved Otsu segmentation method. Experimental evaluations show that the proposed model outperforms the state-of-the-art salient detection models.

Index Terms—Image processing, saliency analysis, histogram, super-pixel segmentation, thresholding

1. INTRODUCTION

Saliency analysis is a kind of quick and efficient method for researchers to complete the target detection and extraction. It is widely used in the field of object recognition, feature analysis, image understanding, and image compression [1-4]. Unlike conventional supervised models, it detects targets by a different strategy that is free from heavy training and manual prior information. Instead, the visual attention mechanism is utilized for detecting and recognizing meaningful regions and targets rapidly and efficiently [5-7].

Itti et al. [8] introduced a groundbreaking saliency model, which was based on the imitation of the selective mechanism of the human visual system. Harel et al. [9] draws on Itti’s approach, but it used a graph-based method (GBVS) when normalized. GBVS uses an idea from graph theory to concentrate mass on activation maps and form activation maps from raw features. Achanta et al. [10] proposed a frequency-tuned (FT) method based on a difference of Gaussian (DoG) band-pass filter. Goferman et al. [11] proposed the context-aware (CA) saliency detection model that analyzed saliency in different scales.

Traditional saliency models are efficient in detecting salient targets in natural scenes. However, due to different imaging mechanisms and image features, these approaches are not quite appropriate for remote sensing images [12-16]. As shown in Fig.1, in natural images, the salient targets as the flower and cross can be detected by CA and FT models efficiently. Nevertheless, in the remote sensing image, the residential areas as salient targets obtained by CA and FT models contain mass background interference.

In fact, natural pictures have a simple background, usually with only one salient target. However, most remote sensing images include more complex background information than natural images. In addition, the targets of a remote sensing image may include more than one, and not necessarily in the center, which makes traditional saliency models to be not accurate for remote sensing images.

In this paper, to suppress the background interference and obtain boundaries of targets more accurately, we propose a novel salient target detection model for remote sensing images. In our model, a combination of super-pixel and statistical feature analysis is introduced. Our model cannot only retain the target edge more accurately through super-pixel segmentation, but also ensure the integrity of salient goals, and finally get the precise targets and ROIs.

Fig.1. Salient target and ROI detected by CA and FT models in natural and remote sensing images respectively.


2. METHODOLOGY

In this section, we firstly show the framework of the proposed model. As illustrated in Fig. 2, the proposed model consists of four main steps: 1) Generate the statistical saliency feature map based on histogram statistical saliency analysis in the Lab color space. 2) Generate the information saliency feature map based on the combination of super-pixel segmentation and information entropy. 3) Fuse the statistical saliency feature map and the information saliency feature map and enhance the final saliency map. 4) Segment the final saliency map based on the improved Otsu method, and extracting the salient targets and ROIs.

![Fig.2. Framework of the proposed model.](image)

2.1. Statistical saliency feature analysis in Lab color space

In this paper, we present a new histogram statistical feature analysis strategy based on Lab color space to detect image saliency feature preliminary. We know that RGB color space corresponds to the visible features in color images. However, the mixing of chrominance and luminance data, the high correlation among three channels, and the perceptual non-uniformity make RGB color space susceptible to external interference. The color space transform of RGB to Lab is nonlinear, which is an attempt to correct the interference with dimension L channel in lightness and a channel and b channel in the color-opponent dimensions. The nonlinear relationships among L, a, and b are intended to mimic the nonlinear response of the eye [17].

Histogram represents the distribution of values of the image, and two-dimensional histogram can capture the emergence and coexistence of image pixels well, which can be used to calculate visual saliency. Moreover it has two good properties: having tolerance for change of the image scale and the size of neighborhood. It also can reflect global and local distribution of pixels values in the image and identify the image texture information. Thus, we use the histogram statistical analysis based on Lab space to detect saliency feature of remote sensing images.

Now we take the L channel as example. The input image is I, the size of the image is \( M \times N \), when convert the input image to Lab color space, the size of L channel is also \( M \times N \). Because usually the saliency and emergence probability are negatively correlated, so we use a logarithmic relationship to describe the initial saliency of L channel and define as follows:

\[
sa(m,n) = -\ln(p(m,n))
\]

(1)

Where \( p(m,n) \) actually captures the differences between statistical distribution and uniform distribution of L channel.

Similarly, we can get the saliency feature map of a channel and b channel. The final statistical saliency feature map of the input image is obtained through the weighted integration of L, a, b channels’ saliency feature map.

\[
S = g \times (S_L \omega_L + S_a \omega_a + S_b \omega_b)
\]

(2)

Where \( S \) represents the final statistical saliency feature map, \( g \) is a two-dimensional Gaussian filter. \( \omega_L, \omega_a \) and \( \omega_b \) are the weight values of L, a, b channels.

2.2. Saliency analysis based on super-pixel segmentation and saliency map enhancement

Super-pixel refers to a pixel block having a similar color, brightness, texture and other characteristics. The super-pixel segmentation can cluster the pixels, which have similar characteristics, and form the spatial and structural processing unit and greatly reduce the complexity of subsequent image processing. The output of super-pixel segmentation, namely super-pixel, has a certain image characterization, high data structure characterizations and atom region which has a certain sense of visual perception, particularly the natural boundary contour information in the target image, has certain semantic information that human understand images.

In this paper, we use the SLIC [18] super-pixel segmentation algorithm which can segment the image into a compact and uniform super-pixels quickly and accurately. It is based on clustering and calculates by the Lab space and \((x, y)\) pixel coordinates totally five-dimensional space.

After the super-pixel segmentation, we can get abundant information super-pixels, we use the information entropy to calculate the saliency of the super-pixel map. First, count the probability of each pixel intensity value occurrence and one-dimensional histogram for each band. Then, the information amount of each pixel intensity value is calculated based on the corresponding one-dimensional histogram for each band. Finally, according to the information amount, the information saliency feature map is generated based on the weighted fusion as follow.

\[
Grey(x, y) = \sum_{c=1}^{n} \omega_c \text{LOG}(x, y)
\]

(3)

Where \( Grey(x, y) \) is the final the information saliency feature map, \( \omega_c \) is the weight, \( \text{LOG}(x, y) \) is the information map, \( c \) represents the number of these bands.
Because the salient values within each super-pixel block are the average of all the pixels within the super-pixel block, the salient values of some salient areas are pulled low. So we need to enhance the saliency map. A point close to the center of attention has a higher saliency than a point far from the center of attention [11, 19]. Therefore, the saliency of the points near the most salient point is increased:

\[ G(x, y) = G'(x', y') (1 - d(x, y)) \]  

(4)

Where \( G(x, y) \) is the saliency value of the point \((x, y)\), \( G'(x', y') \) is the saliency value of the more salient points \((x', y')\). \( d(x, y) \) is the distance between the point \((x, y)\) and the point \((x', y')\). Apparently, the saliency of the points which are close to the salient region will be improved. The saliency value of points that are far from the saliency region will be reduced or unchanged.

In the section 2.1, we get the saliency map based on the statistical analysis histogram, in the section 2.2, we get the saliency map upon the superpixel segmentation and information entropy. When we segment the image into superpixel, a split cover can be get. Now we put this cover into the saliency map obtained in the section 2.1, which is equivalent to do superpixel segmentation on the saliency map. Thus, add salient value within each superpixel block and then take the average as the salient value of the corresponding superpixel blocks. Finally, fuse the two saliency maps to get the final saliency map. Through the superpixel segmentation on the input image and the saliency map based on the statistical analysis histogram, we can obtain the final saliency map completely and precisely.

2.3. Salient targets detection based on the improved Otsu method

The Otsu method proposed the maximum inter-class variance method and is widely used due to the simple calculation and strong self-adaptable [20]. However, the Otsu does not get the best result in all situations [21]. Actually, the threshold found by Otsu is the average of the mean values of the two classes divided by the threshold [22].

Suppose that a set of image gray level is \( L \), and the gray value of the pixel in the image is \([1, 2, \ldots, L]\). The image pixels are divided into two types according to the gray level threshold \( D \). That gray level \([1, \ldots, D]\), a class of pixels, is denoted by \( c1 \); the pixels with gray levels of \([D + 1, \ldots, L]\) constitute another class, denoted as \( c2 \).

\[
\sigma_1^2(D) = \sum_{i=1}^{D} (i - \mu_1(D))^2 \frac{p_1(D)}{p(D)}
\]

(5)

\[
\sigma_2^2(D) = \sum_{i=D+1}^{L} (i - \mu_2(D))^2 \frac{p_2(D)}{p(D)}
\]

(6)

Where the probability of occurrence of \( c1 \) and \( c2 \) is denoted respectively as \( p_1(D) \) and \( p_2(D) \). The gray scale mean of the two classes is denoted by \( \mu_1(D) \) and \( \mu_2(D) \). And \( \sigma_1^2(D) \) and \( \sigma_2^2(D) \) represent the variance of two types.

From the above analysis, we can see that the Otsu threshold will fall in the class of large variance. However, in remote sensing images, the background usually has a larger variance, and the variance of salient region is smaller. So when we use the Otsu method to extract salient regions, usually a portion of the non-salient regions are extracted. In this paper, we improved the Otsu threshold by shifting the threshold to the right, so we can obtain more accurate salient regions which could see clearly from Fig. 3.

![Fig.3. Framework of the improved Otsu method.](image)

3. EXPERIMENTAL RESULTS

In this section, we select more than 100 test images from SPOT5, Google Earth and WorldView-3 to evaluate the performance between our proposed model and other seven state-of-the-art models. The seven models include ITTI model by Itti et al [8], the GBVS model by Harel et al [9], the FT model by Achanta et al [10], the SR model by Hou and Zhang [23], the SACH model by Zhang and Li [24], the FDA model by Zhang and Yang [25], the MFF model by Zhang et al [26] and the proposed model. The experimental analysis includes qualitative and quantitative analysis.

3.1 Qualitative evaluation

In this section, the saliency maps and salient targets that are produced by our model and the seven competing models are shown in Fig. 4 and Fig. 5. The top two rows are the results of SPOT5 images and the salient targets are residential areas. The middle two rows are the results of Google Earth images and the salient targets are also residential areas. The bottom two rows are the results of WorldView-3 images and the salient targets are the buildings.

The proposed model presents a more accurate description of salient targets. The Itti model, GBVS model and SR model provide a saliency map of non-full resolution. When using the saliency map to extract the salient targets, it needs to be expanded to the saliency map of the full resolution. Therefore, these models cannot define target boundary accurately. And their detection results contain much background information, particularly the GBVS model.
and the Itti model. The FDA model and FT model fail to detect fully salient regions because they use low-pass filters to filter out low frequency information and enhance the high-frequency information respectively, which causes that the boundaries of salient targets is not complete and leads to an attenuated region interior. So they fail to detect the whole salient targets. The MFF model and SACH model give a good performance. Overall, the experimental results show that our model is better than other seven models. It can provide a clear boundary of the final salient maps, and detect the salient targets accurately and integrally.

3.2 Quantitative evaluation

We used receive operating characteristic curve (ROC) and PRF values to evaluate these saliency models objectively. PRF value contains comprehensive evaluation F-measure (F), precision (P), and recall (R). As illustrated in Fig. 6, our proposed model achieves the better ROC result and the higher F-Measure than other seven models.

3.3 Experimental results of the improved Otsu method

In Fig. 7, we can find that the experimental results with the Otsu method contain some obvious background interference. However, by using the improved Otsu method, we can obtain more accurate detection results. And the interference of the roads is eliminated and the vegetation is suppressed effectively.

4. CONCLUSION

The proposed model is a salient target detection model which is based on the combination of super-pixel and statistical saliency feature analysis. These main contributions of this paper include three points: 1) Combining the super-pixel segmentation and statistical feature analysis; 2) Generating the information saliency feature map based on the combination of super-pixel segmentation and information entropy; 3) Designing an improved Otsu method for extracting the salient targets. Comparing to the state-of-the-art salient detection models, our model has the better results of qualitative and quantitative evaluation and extracts salient targets from remote sensing images more effectively.
5. REFERENCES


