

## REMOTE SENSING IMAGE RETRIEVAL WITH COMBINED FEATURES OF SALIENT REGION

Zhenfeng Shao<sup>1</sup>, Weixun Zhou<sup>1\*</sup>, Qimin Cheng<sup>2</sup>

1-State key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China

2-The Department of Electronics and Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China  
weixunzhou1990@whu.edu.cn

**KEY WORDS:** Visual Attention Model, Object Saliency, Salient Regions, Combined Features, Remote Sensing Image Retrieval

### ABSTRACT:

Low-level features tend to achieve unsatisfactory retrieval results in remote sensing image retrieval community because of the existence of semantic gap. In order to improve retrieval precision, visual attention model is used to extract salient objects from image according to their saliency. Then color and texture features are extracted from salient objects and regarded as feature vectors for image retrieval. Experimental results demonstrate that our method improves retrieval results and obtains higher precision.

### 1. INTRODUCTION

With the rapid development of remote sensing technology, more and more remote sensing data become available. However, issues such as lack of effective organization and management methods on the massive remote sensing data have impeded its effective applications. Content-based image retrieval (CBIR) technology is an appropriate choice to alleviate such issues and has become a hot topic in remote sensing field (Wang M et al., 2013).

In general, current researches primarily focus on image retrieval based on low-level features (color, texture and shape, etc.) (Wang M et al., 2013). Color features tend to be closely associated with objects or scenes in images and depend less on image size, direction and viewpoint. Texture features can well express abundant visual information of images and are appreciated in machine vision field. Thus color and texture features are widely used in CBIR community because of above characteristics. In addition, retrieval method using combination of color and texture features is increasingly drawing more attention. Nevertheless, because of the existence of semantic gap between low-level features and high-level semantics, aforementioned strategies tend to obtain unsatisfactory results, particularly for remote sensing images with multiple objects.

To improve image retrieval results using low-level features and bridge the distance between low-level features and high-level semantics, methods that concentrate on salient objects rather than whole image may be functional from the perspective of Human Visual System (HVS) because semantics are related to specific objects. In the field of computer vision, salient region detection and salient object extraction are popular topics among researchers (Mayer H, 1999). Inspired by such studies, segment methods have been studied and applied into CBIR in order to extract salient objects from remote sensing images and then low-level features extracted from these salient objects are used for retrieval. To some extent, features from salient regions can describe image content much better than those from entire image. Unfortunately, segmentation itself still remains an open question in computer field, and salient objects can hardly be extracted from remote sensing images directly. To solve these problems and improve retrieval results of remote sensing images, visual attention model has been introduced in this paper.

There are some famous visual attention models, such as Itti model (Itti et al., 1998), GBVS model (Jonathan Harel et al, 2007), etc., which have been widely used in many fields. In this paper the GBVS attention model is chosen due to its better performance and adaptability. Furthermore the GBVS model has better results than Itti model in many cases, it follows the architecture of Itti model and the greatest difference may be GBVS defines Markov Chains on activation maps. The following are brief procedure of Itti model. Given one remote sensing image, color, intensity and direction features are extracted firstly through Gaussian pyramid with various scales. Then saliency maps with respect to three features are obtained by using center-surround algorithm. Finally, color saliency map, intensity saliency map and direction saliency map are fused together and final saliency map is received.

The paper is organized as follows. In section 2, we briefly introduce Itti visual attention model and then present the extraction of color and texture features from salient objects. In section 3, some high resolution remote sensing images from one public image database are chosen to constitute the retrieval database. In addition, experiments with combined color and texture features and analyses are shown. Conclusions are given in section 4.

### 2. METHODOLOGY

#### 2.1 Framework of methodology

In order to describe the content of remote sensing images better, GBVS model is introduced to extract salient objects from images. Figure 1 illustrates the architecture of our methodology.

From the flowchart, we can see two main parts are included in this methodology, salient region extraction and feature extraction. To achieve best results, several groups of saliency threshold are tried according to saliency value of each image class. Saliency values ranging from 0.4 to 0.8 are appropriate empirically.

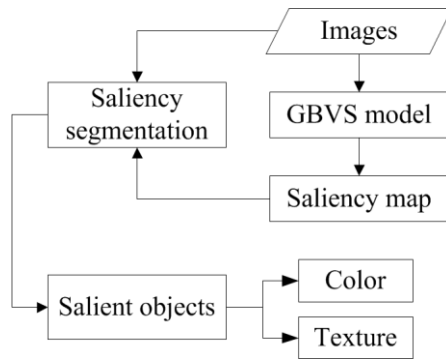


Figure1. Architecture of proposed method

## 2.2 Feature extraction

Color and texture features are chosen as low-level features used in our methodology since they are among the more expressive visual features (Manjunath et al., 2001). Certainly, all features are extracted from salient regions as well as whole images. Because many color and texture descriptors are existed (Datta et al., 2008), thus commonly used descriptors color histogram and edge histogram are adopted in this paper.

### 2.2.1 Color feature extraction

When it comes to color feature, one indispensable ingredient may be color space. Existed color spaces, such as RGB, HSV, YCbCr, et al., have more or less weakness to some extent. In this paper, we choose one uniform and simple space HSV. Three color components of HSV are quantized to 16, 4, 4 respectively. And then, color feature can be defined by

$$C = 16H + 4S + 4V \quad (1)$$

Finally, a total of 256 color histogram is gained.

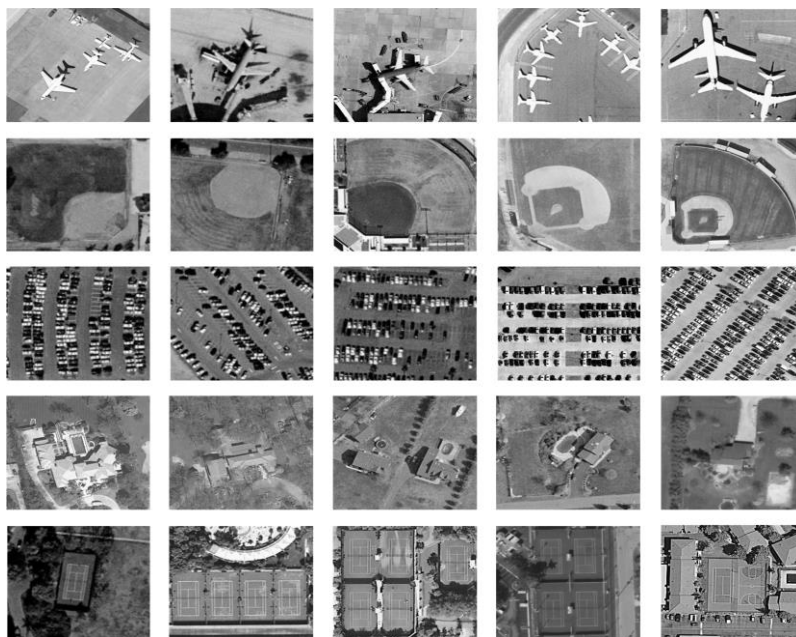


Figure2. Some image patches of retrieval database

### 2.2.2 Texture feature extraction

Edge histogram (EH) is one widely used texture descriptor and, as MPEG-7 recommended texture descriptor, it defines 5 kinds of edges and tends to achieve better results than the majority of other texture descriptors. However, original edge histogram is not utilized in our method and one simpler method similar to edge histogram called simpler edge histogram (SEH) is used. SEH has not cropped the image into 16 patches, which is the only difference between EH and SHE. Thus, a total of 5 texture feature is gained.

### 2.3 Combination of color and texture

As a result of semantic gap, low-level features cannot describe the content of images well, and image retrieval with multiple features seems to be an effective method to improve the retrieval results. There are some popular retrieval systems that hold this principle. For example, in QBIC project (see Niblack et al., 1993), retrieval methods based on color, texture and shape are studied. In (Iqbal et al., 2002), structure, color and texture are combined for efficient image retrieval. In this paper, we extract color and texture features of salient objects from images and conduct combined features for image retrieval.

### 2.4 Normalizations

For image retrieval based on the combined features, normalization is a necessary procedure before similarity measure because great differences of mean and standard deviation may exist between different features. Gaussian normalization is one appropriate method in image retrieval, and the details can be found in (Ortega et al., 1997).

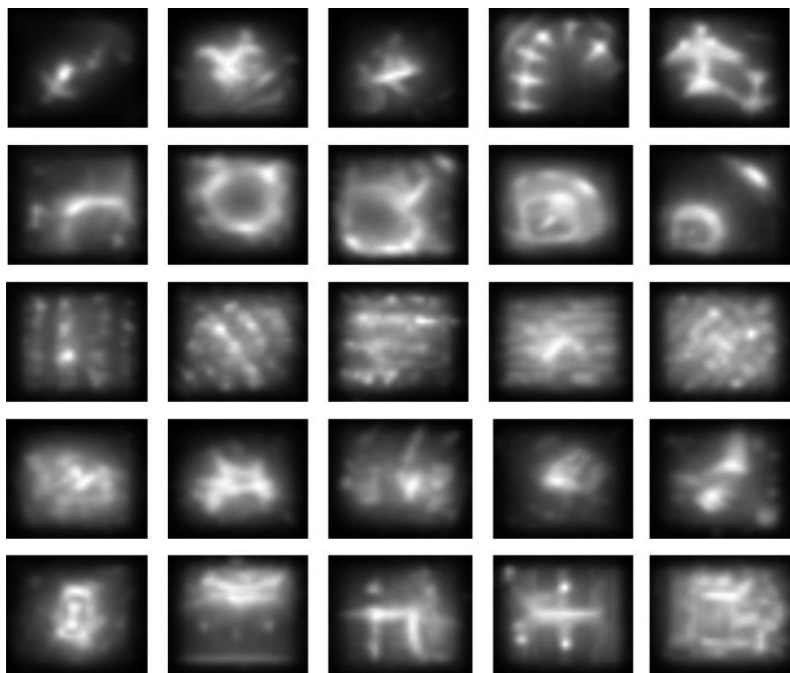


Figure3. Respective saliency maps of images in Figure 2

### 3. EXPERIMENT AND ANALYSIS

#### 3.1 Data Set

Retrieval image database is of great importance in image retrieval field; however, there has no public image databases for remote sensing image retrieval, which is different from natural image retrieval. In remote sensing image retrieval community, one common strategy is dividing the whole remote sensing image into many image patches and relevant patches are regarded as similar images. In this paper, the image database used in (Yang et al., 2013) is chosen, which includes 21 image classes and each class has 100 RGB images with the resolution of 30cm and size of  $256 \times 256$  pixels. 5 class images, i.e. airplane, baseball diamond, parking lot, residential and tennis court are used. Consequently, retrieval database composed of 500 images are prepared.

Figure 2 shows some example image patches of 5 class images, which include are airplane, baseball diamond, parking lot, residential and tennis court respectively from left to right. Figure 3 shows saliency maps of images in Figure 2 by GBVS visual attention model respectively. Note that the lighter the pixels are, the more conspicuous they are, which means that the objects are more likely to attract our attention compared with background pixels from the perspective of HVS.

#### 3.2 Experiment

In this section, some comparison experiments are conducted to demonstrate the validity of proposed retrieval method based on combined color and texture features extracted from salient regions. Meanwhile, some retrieval results are illustrated as well by means of query-by-example.

#### Experiment A.

Table1 exhibits retrieval precision by using color and texture features extracted from the whole images and salient regions of images respectively. Note that the numbers in Table 1 are precision calculated by

$$p = \frac{n}{m} \quad (2)$$

where m means the number of returned images and n means the number of similar images in m.

Image Class	Color Feature		Texture Feature	
	whole image	salient region	whole image	salient region
airplane	0.4281	1.0000	0.4291	0.3719
baseball diamond	0.2360	0.2858	0.4290	0.3650
parking lot	0.4148	0.6606	0.6290	0.6431
residential	0.3707	0.5198	0.3565	0.3191
tennis court	0.4849	0.8278	0.3003	0.2832
<b>Average</b>	0.3869	<b>0.6588</b>	0.4288	<b>0.3965</b>

Table1. Retrieval precision of features extracted from whole image and salient region respectively

Some conclusions can be drawn from Table 1:

- From average precision of each feature, color features extracted from salient regions perform best and improve retrieval results dramatically.
- For each image class, color features extracted from salient regions increase precision obviously, especially for airplanes.
- Texture features extracted from salient regions achieve lower average precision than features extracted from whole image, which may attribute to the extraction of salient objects destroys texture pattern of images to some degree.
- Image retrieval strategy based on salient region is a good solution to improve retrieval results.

### Experiment B.

According to experimental results in Experiment A, image retrieval using color and texture features extracted from salient regions simultaneously is discussed. Among many existed strategies, assigning relevant weights to different features and image retrieval method based on the combined features is commonly accepted. Then the similarity of two images is denoted by

$$s = \alpha s_1 + \beta s_2 \quad (3)$$

where  $\alpha$  and  $\beta$  mean weights of features respectively, while  $s_1$  and  $s_2$  mean similarities of different features respectively.

In our method, weight parameters are established on the retrieval results of features respectively, which means features that have better results will be assigned larger weights. Thus in following experiments, different weights can be assigned to each image class.

Figure 4 illustrates retrieval results of images in our database using combined color and texture features extracted from salient regions with different numbers of returned images. In order to decrease calculation complexity, not all images in database are chosen as queried images. Conversely, 20 images are selected randomly from each images class as query images. In addition, weights used in similarity measure for each image class are given in Table 2. Certainly, these weights parameters achieve better results than any other parameters.

Based on the illustration of Figure 4, it is clear that each image class has good retrieval results except for baseball diamond images, whose retrieval precision is under 0.4 when the number of returned images is 20 or more. For the other four image classes, even if 100 images, the number of images in each class, are returned, the precision is still above 0.4.

Figure 5 illustrates the performance of proposed image retrieval method based on combined features. We can see clearly that the combined features perform best and achieve highest precision from the bin group "Average". In addition, image retrieval using combined features also achieves better results than other features with respect to each image class. And in Figure 6, some retrieval examples of each image class are shown. Note that the upper-left corner image is query image and the first twenty similar images are returned.

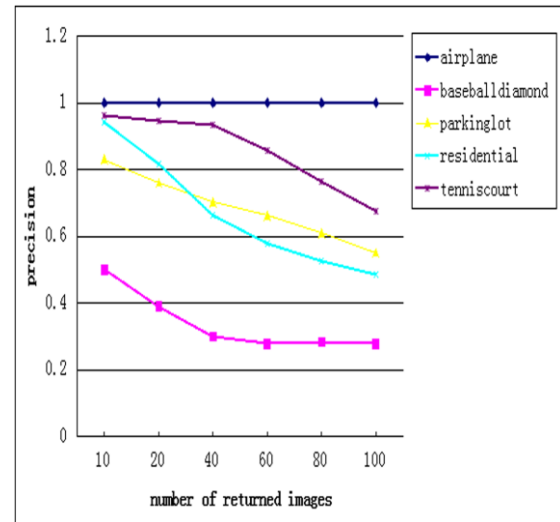


Figure4. Retrieval results of five class images using combined features with different numbers of returned images

	$\alpha$	$\beta$
airplane	1	0
baseball diamond	0.1	0.9
parking lot	0.6	0.4
residential	0.9	0.1
tennis court	0.9	0.1

Table2. Weights used in similarity measure for each images class

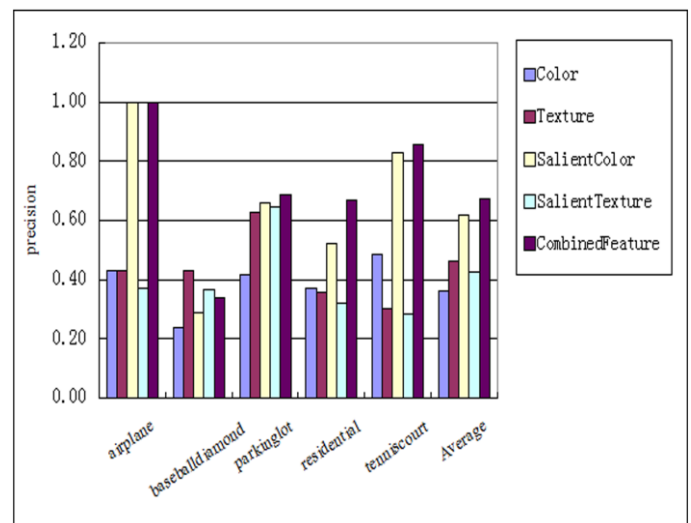


Figure5. Comparisons of proposed method with other methods

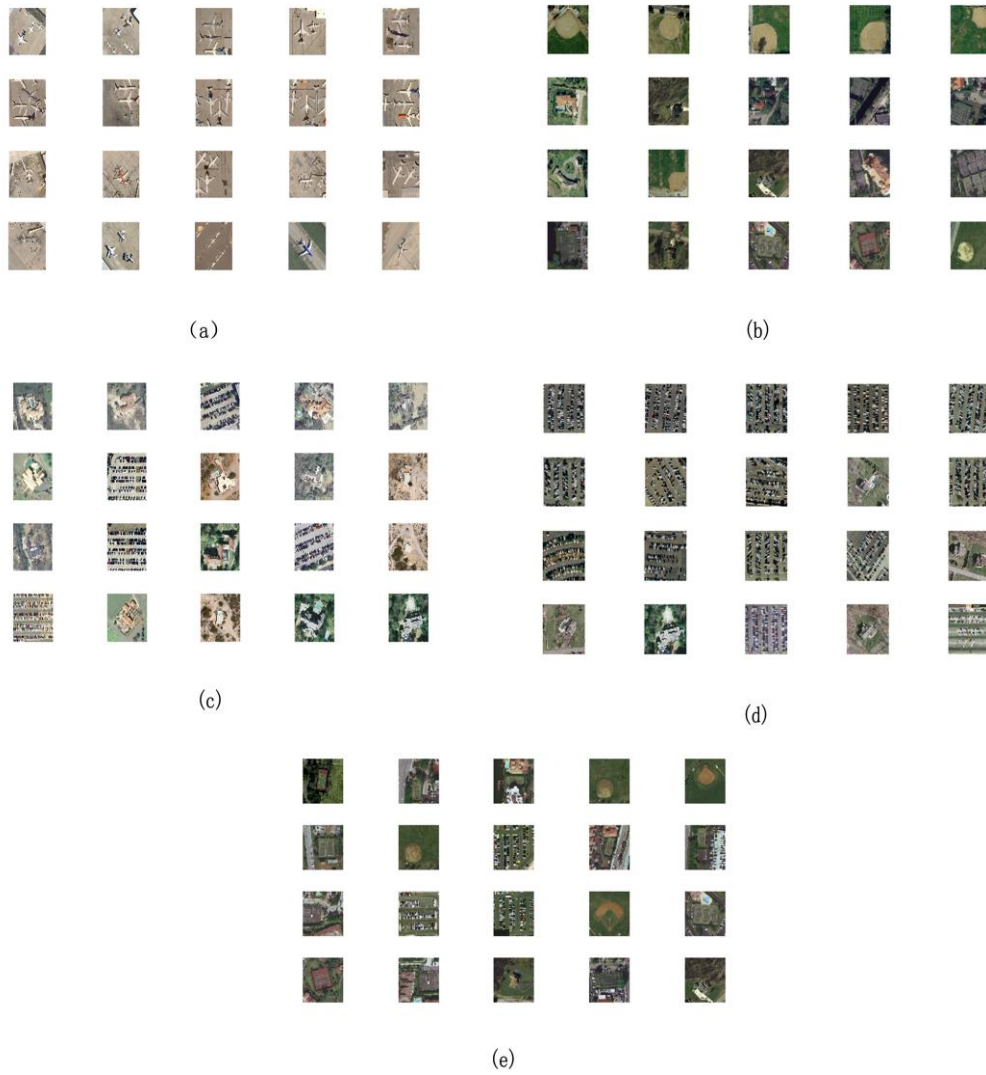


Figure6. Examples of 5 class images retrieval based on combined features

#### 4. CONCLUSIONS

In this paper, a retrieval method with combined features based on salient region for remote sensing is proposed. Firstly saliency maps of images in database are calculated based on GBVS attention model; then salient regions are extracted from images according to their saliencies, where color and texture features are extracted from. Finally, the validity of proposed method is demonstrated by several experiments, which has improved retrieval results greatly compared with that of by using features extracted from the whole images.

#### REFERENCES

- Datta, R., Joshi, D., Li, J., & Wang, J. Z., 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)* 40(2), 5.
- Itti, L., Koch, C., & Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20(11), pp.1254-1259.
- Iqbal Q, Aggarwal J K. 2002. Combining structure, color and texture for image retrieval: A performance evaluation. 16th International Conference on Pattern Recognition, pp.438-443.
- Jonathan Harel, Christof Koch, Pietro Perona, 2007. Graph-Based Visual Saliency, pp.545-552.
- Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., & Yamada, A., 2001. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology* 11(6), pp.703-715.
- Mayer H. 1999. Automatic object extraction from aerial imagery—a survey focusing on buildings. *Computer vision and image understanding*, 74(2), pp.138-149.
- Niblack C W, Barber R, Equitz W, et al., 1993. QBIC project: querying images by content, using color, texture, and shape. *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, pp.173-187.
- Ortega M, Rui Y, Chakrabarti K, et al. 1997. Supporting similarity queries in MARS, *Proceedings of the fifth ACM international conference on Multimedia*. pp.403-413.
- Wang M, Song T. 2013. Remote Sensing Image Retrieval by Scene Semantic Matching. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5), pp.2874-2886.
- Yang, Y., Newsam, S. 2013. Geographic image retrieval using local invariant features. *IEEE Transactions on Geoscience and Remote Sensing* 51(2), pp.818-832.