

RESEARCH

Open Access

Remotely sensed image retrieval based on region-level semantic mining

Tingting Liu¹, Liangpei Zhang^{2*}, Pingxiang Li² and Hui Lin³

Abstract

As satellite images are widely used in a large number of applications in recent years, content-based image retrieval technique has become important tools for image exploration and information mining; however, their performances are limited by the semantic gap between low-level features and high-level concepts. To narrow this semantic gap, a region-level semantic mining approach is proposed in this article. Because it is easier for users to understand image content by region, images are segmented into several parts using an improved segmentation algorithm, each with homogeneous spectral and textural characteristics, and then a uniform region-based representation for each image is built. Once the probabilistic relationship among image, region, and hidden semantic is constructed, the Expectation Maximization method can be applied to mine the hidden semantic. We implement this approach on a dataset consisting of thousands of satellite images and obtain a high retrieval precision, as demonstrated through experiments.

Keywords: content-based image retrieval, segmentation, region-level image representation, semantic mining

1. Introduction

The information in remotely sensed images plays an important role in environmental monitoring, disaster forecasting, geological survey, and other applications. With the steadily expanding demand for remotely sensed images, many satellites have been launched, and thousands of images are acquired every day [1]. This leads to an exponential increase in the quantity of remotely sensed images in database. Therefore, how to retrieve useful images quickly and accurately from a huge and unstructured image database becomes a challenge.

Traditional image query techniques retrieve images by matching keywords [2], such as geographic location, sensor type, and time of acquisition. But, the content of the image which is much more important than attributes is not considered in these techniques [3]. In order to overcome this shortcoming of these techniques, image retrieval techniques are strongly focused on content-based image retrieval (CBIR). In a CBIR system, low-level features are used to represent image content and retrieve image from database, such as spectrum, texture, and

shape [4-6]. Although low-level features can accurately be extracted by various methods, they cannot easily be utilized to describe user's perception of an image [7,8]. Semantic feature is a high-level hidden concept, which is meaningful to user's perception. The difference between low-level feature and high-level semantic feature, caused by the absence of a direct relationship between low-level features and high-level concepts [9-13], is called the "semantic gap" [7]. To narrow this gap, a semantic-based image retrieval system should be built, in which high-level semantic features can automatically be extracted from low-level image features.

Semantic feature mining is essential to semantic-based image retrieval technique. The process of semantic feature mining can be divided into two steps: low-level feature extraction and high-level semantic feature extraction. At present, most studies on low-level feature extraction are based on pixel characteristics. Li and Narayanan [14] identified ground-cover information based on spectral characteristics using supervised classification and extracted textural features by characterizing spatial information using Gabor wavelet coefficients. Li et al. [15] developed an approach based on pixel-level textural information to extract global semantic features. But, pixel does not facilitate understanding of the image, thus it is often replaced

* Correspondence: zlp62@public.wh.hb.cn

²The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China

Full list of author information is available at the end of the article

by region. It can be assumed that region-level description of visual information is more comprehensible for users than pixel-level image description. However, most of the existing high-level semantic feature mining methods are based on pixel-level features. Datcu et al. [16] and Daschiel and Datcu [17] developed a Bayesian classifier to retrieve images from a remotely sensed image database by approximating the probabilities of images belonging to different classes using pixel-level probabilities. Aksoy et al. [1] proposed a pixel-based Bayesian framework for a visual grammar to narrow the gap between low-level features and high-level concepts. Therefore, in this article, a novel approach is proposed to achieve region-level semantic feature mining. First, a region-level image content description is developed to facilitate users' understanding of image. Based on region-level features, a probabilistic relationship among image, region, and hidden semantic is developed. Then, the Expectation Maximization (EM) method is used to mine the hidden semantic features. Finally, remotely sensed image retrieval is performed using region-level semantic features.

The rest of the article is organized as follows. In Section 2, details of the region-level image representation are provided. In Section 3, semantic mining using the EM method is discussed. In Section 4, experiments are presented to demonstrate the effectiveness of region-level semantic features. Finally, conclusions are presented in Section 5.

2. Region-level image representation

Region-level image representation includes the following components: image segmentation, regional information description, and codebook extraction. Figure 1 shows a flowchart of the region-level image representation process.

2.1 Image segmentation

The JSEG algorithm [18] is a region-based segmentation method that provides robust segmentation results for a large variety of images and videos [19-21]. In this article, the JSEG algorithm is improved to make it applicable to multi-spectral remotely sensed image segmentation.

The JSEG algorithm consists of two parts: color quantization and spatial segmentation. Figure 2 shows a schematic diagram of the original JSEG algorithm.

In color quantization step, the general Lloyd algorithm (GLA) [22] is used to quantize the image. In this algorithm, the distortion D can be represented as Equation (1)

$$D = \sum_i D_i \sum_i \sum_n v(n) \|x(n) - c_i\|^2, x(n) \in C_i \quad (1)$$

where C_i is the i th cluster in the image, c_i is the center pixel of cluster C_i , $x(n)$ and $v(n)$ are the color vector and the perceptual weight for pixel n , and D_i is the total distortion for cluster C_i .

Since multi-spectral Thematic Mapper (TM) images are used as experimental data, $x(n)$ is defined as $x(n) = \{a_{n1}, a_{n2}, \dots, a_{nj}\}$ in this algorithm, where j is the number of bands in the image and a_{nj} is the value of the n th pixel in the j th band of the image.

In spatial segmentation step, region growing method is used to segment image based on J-image, in which a threshold controls region growing result. In this research, 0.4 is chosen as an empirical value.

Remotely sensed images present complex spatial arrangement and spectral heterogeneity. It has been demonstrated that combining spatial and spectral information can improve land cover information extraction from satellite image data [23]. Therefore, in this research, Normalized Difference Vegetation Index (NDVI) [24], Normalized Difference Built-up Index (NDBI) [25], and textural features are substituted for the original spectral features to increase land cover separability. NDVI provides a standardized method of assessing whether the land cover being observed contains live green vegetation or not; it can be calculated as Equation (2)

$$\text{NDVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R})} \quad (2)$$

where R and NIR stand for the spectral reflectance measurements acquired in the visible (red) and near-infrared regions, respectively. NDBI serves to compare urban areas with built-up areas between satellite images;

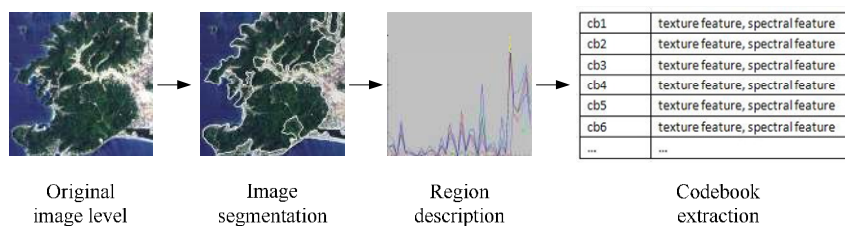
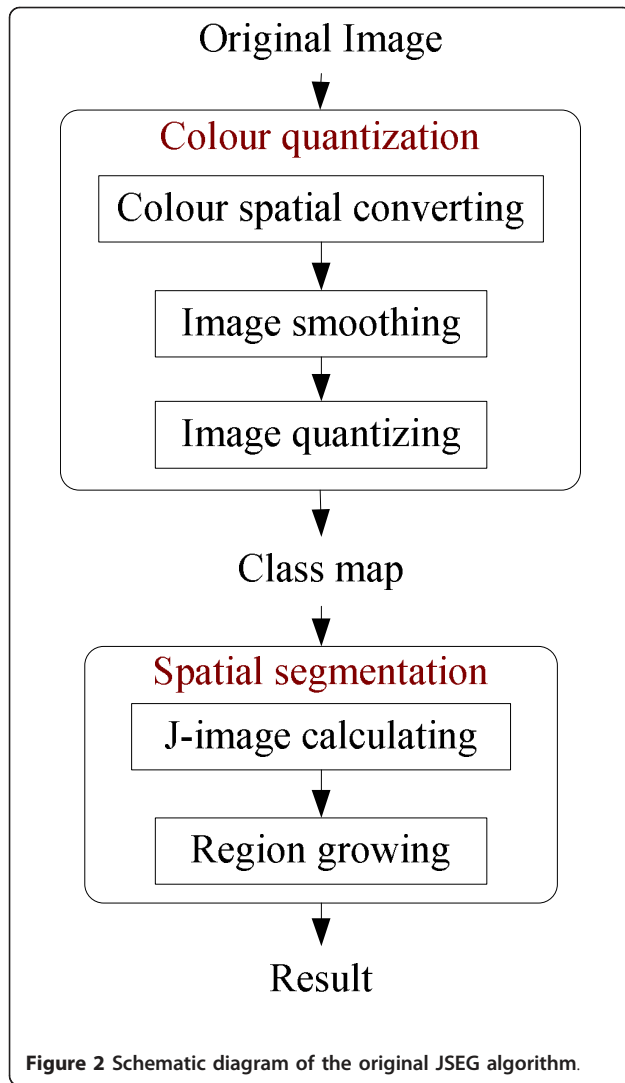


Figure 1 Flowchart of region-level image representation.



it can be calculated as Equation (3)

$$\text{NDBI} = \frac{\text{MIR} - \text{NIR}}{\text{MIR} + \text{NIR}} \quad (3)$$

where MIR and NIR stand for the spectral reflectance measurements acquired in the middle-infrared and near-infrared regions, respectively. Texture reflects the local variability of grey level in the spatial domain and reveals the information about the object structures in the natural environment [26]. In this research, the mean texture is used and extracted using Grey-Level Co-occurrence Matrix (GLCM), which is commonly applied in statistical procedure for interpreting texture. Finally, the pixels in original image can be represented as Equation (4)

$$f = \{f_{\text{NDVI}}, f_{\text{NDBI}}, f_{\text{texture}}\} \quad (4)$$

As described above, a flowchart for the improved segmentation algorithm is shown in Figure 3.

2.2 Regional information description

In this research, regional information is described using spectral and textural features. Spectral feature is the original pixel value, and textural feature is extracted using GLCM. These two features are extracted separately for each region in all images.

GLCM is a commonly used method in texture analysis. It describes the frequency at which one grey tone appears in a specified spatial linear relationship with another grey tone in the area under investigation. Fourteen statistical parameters [27] can be extracted using GLCM. However, in the retrieval system, the more features, the lower the efficiency is. The correlation matrix of the eight common parameters, namely mean, variance, homogeneity, contrast, dissimilarity, entropy, second angular moment, and correlation, is presented in Table 1.

The correlation matrix shows the correlation of the two parameters. The higher value indicates the higher correlation of the two parameters. In Table 1 there are two correlations over 0.8. The correlation value between contrast and variance is 0.80; the correlation value between dissimilarity and contrast is 0.87. These indicate that variance and dissimilarity are highly correlated with contrast. Therefore, variance and dissimilarity can be replaced by contrast, while the other six parameters describe textural features.

2.3 Codebook extraction

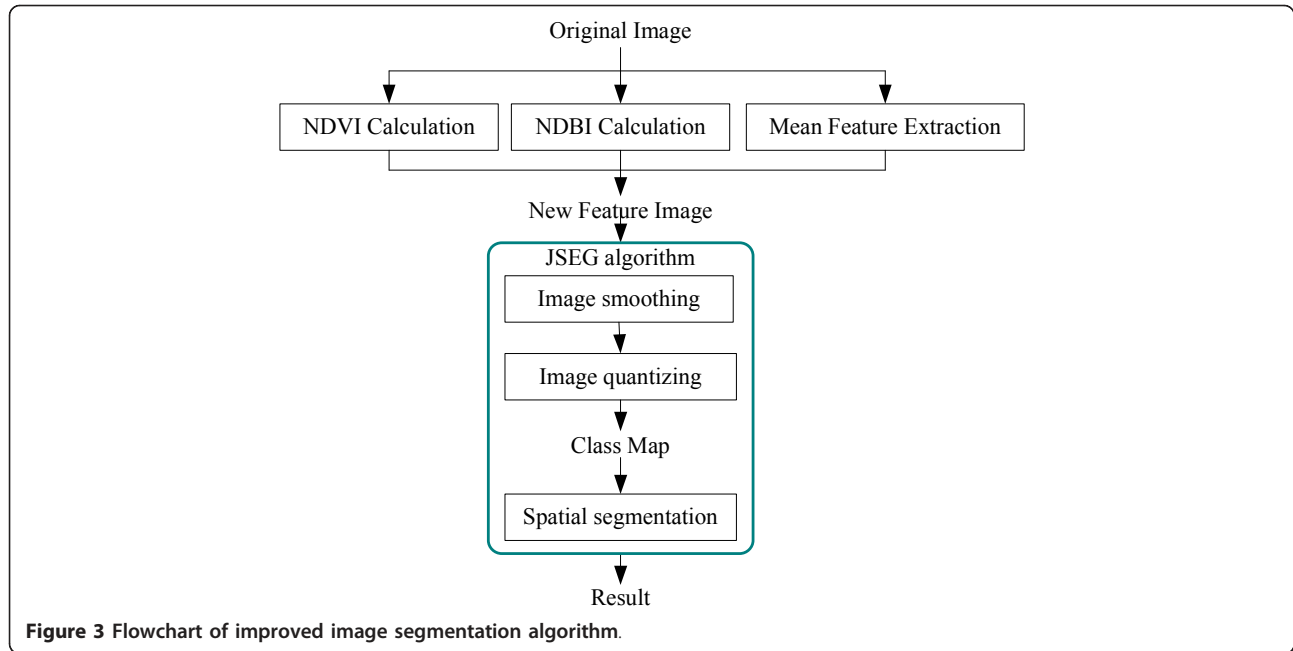
After the images have been segmented into several parts, a number of regions are generated and stored in the database. It will be time-consuming to calculate the similarity between two regional features for all pairs of regions.

However, many regions on different images are very similar in terms of spectral and textural features. Therefore, GLA is used to classify the low-level features into a set of codes based on which a codebook will be generated (as shown in Figure 4).

Figure 4 presents the principle of codebook extraction when image feature is two-dimensional feature space. In Figure 4, the blue point is a low-level feature, the black circle is a cluster, and the red point is the center of the cluster called code. Code_{*j*} is the mean of all features in corresponding cluster. All codes form a codebook. Then, each region can be represented by a code. For an image *I*, its *i*th region *R_i* can be represented by Code_{*j*}.

3. Semantic feature extraction

In this step, a probabilistic method is used to mine the relationship among semantic features, regions, and



images automatically. Then the EM method [28,29] is used to analyze the relationship and extract the latent semantic concepts.

First, various parameters are defined as follows:

- (a) *Image data*: d_j is an image in the database, $d_j \in \{d_1, \dots, d_M\}$; M is the total number of images.
- (b) *Regional feature data*: r_i is the i th region feature in the feature codebook, $r_i \in R = \{r_1, \dots, r_N\}$, where N is the total number of regional features.
- (c) *Hidden semantic features*: s_k is the hidden semantic feature, $s_k \in S = \{s_1, \dots, s_K\}$, where K is the total number of semantic features.

where j is the number of images, $j \in \{1, \dots, M\}$; i is the number of region features, $i \in \{1, \dots, N\}$; k is the number of semantic features, $k \in \{1, \dots, K\}$.

$P(d_j)$ denotes the probability that an image will occur in a particular image database. $P(r_i | s_k)$ denotes the class-conditional probability of region r_i given the hidden

semantic feature s_k . $P(s_k | d_j)$ denotes the class-conditional probability of the hidden semantic feature s_k given a particular image d_j . d_j and r_i are independently defined on the state of the associated hidden semantic feature. According to conditional probability formula, the joint probability of d_j and r_i can be described by Equation (5)

$$P(r_i, d_j) = P(d_j)P(r_i | d_j) \quad (5)$$

Then, applying total probability formula, Equation (5) can be transformed to Equation (6):

$$P(d_j)P(r_i | d_j) = P(d_j) \sum_{k=1}^K P(r_i | s_k)P(s_k | d_j) \quad (6)$$

The class-conditional probability of semantic feature s_k , $P(s_k | r_i, d_j)$, depends on image d_j and region feature r_i . Using the Bayesian formula, this class-conditional probability can be described by Equation (7)

Table 1 The correlation matrix of the eight common parameters

	Mean	Variance	Homogeneity	Contrast	Dissimilarity	Entropy	ASM	Correlation
Mean	1.00	0.62	-0.51	0.58	0.77	0.55	-0.29	-0.13
Variance	0.62	1.00	-0.44	0.80	0.76	0.32	-0.25	0.04
Homogeneity	-0.51	-0.44	1.00	-0.46	-0.70	-0.39	0.74	0.14
Contrast	0.58	0.80	-0.46	1.00	0.87	0.30	-0.24	0.03
Dissimilarity	0.77	0.76	-0.70	0.87	1.00	0.56	-0.41	-0.04
Entropy	0.55	0.32	-0.38	0.31	0.56	1.00	-0.57	-0.34
ASM	-0.29	-0.25	0.74	-0.24	-0.42	-0.57	1.00	0.28
Correlation	-0.13	0.04	0.14	0.03	-0.04	-0.34	0.28	1.00

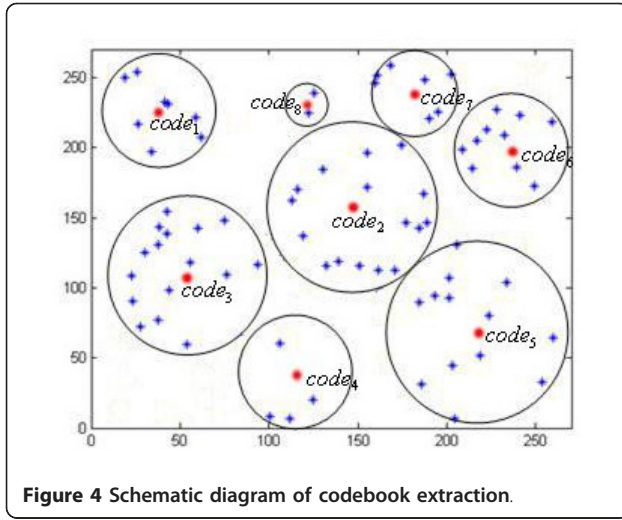


Figure 4 Schematic diagram of codebook extraction.

$$P(s_k | r_i, d_j) = \frac{P(r_i, d_j | s_k) P(s_k)}{P(r_i, d_j)} \quad (7)$$

Since d_j and r_i are independent, referring to Equation (6), Equation (7) can be transformed to

$$P(s_k | r_i, d_j) = \frac{P(d_j | s_k) P(r_i | s_k) P(s_k)}{P(d_j) \sum_{k=1}^K P(r_i | s_k) P(s_k | d_j)} \quad (8)$$

where l is the number of semantic features, $l \in \{1, \dots, K\}$.

Referring to Bayesian formula, Equation (8) can be transformed to

$$P(s_k | r_i, d_j) = \frac{P(d_j | s_k) P(r_i | s_k) P(s_k)}{P(d_j) \sum_{k=1}^K P(r_i | s_k) P(s_k | d_j)} = \frac{P(r_i | s_k) P(s_k | d_j)}{\sum_{k=1}^K P(r_i | s_k) P(s_k | d_j)} \quad (9)$$

Then, referring to the likelihood principle, $P(d_j)$, $P(r_i | s_k)$, and $P(s_k | d_j)$ can be determined by maximizing the log-likelihood function:

$$L = \log(P(R, D, S)) = \sum_{i=1}^N \sum_{j=1}^M n(r_i, d_j) \sum_{k=1}^K P(s_k | r_i, d_j) \log[P(s_k | d_j) P(r_i | s_k)] \quad (10)$$

where $n(r_i, d_j)$ indicates the number of occurrences of region r_i in image d_j .

The standard procedure for maximum likelihood estimation is the EM algorithm. This method has two steps: expectation step (E-step) and maximization step (M-step). The E-step can be interpreted as mining the relationship between current estimates of the parameters and the latent variables by computing posterior probabilities. The M-step can be interpreted as updating parameters based on the so-called expected complete-data log-likelihood.

According to the EM method, the process of obtaining Equation (8) can be considered as the E-step, and the process of obtaining Equation (9) can be considered as the process of log-likelihood estimation. Then, Equation (9) is maximized using Lagrange multipliers. Equations (11) and (12) can then be derived

$$P(r_i | s_k) = \frac{\sum_{j=1}^M n(r_i, d_j) P(s_k | r_i, d_j)}{\sum_{m=1}^N \sum_{j=1}^M n(r_m, d_j) P(s_k | r_m, d_j)}, \quad (11)$$

$$P(s_k | d_j) = \frac{\sum_{i=1}^N n(r_i, d_j) P(s_k | r_i, d_j)}{\sum_{n=1}^N n(r_n, d_j)} \quad (12)$$

where n is the number of regions and region features, $n \in \{1, \dots, N\}$.

The E-step and M-step equations are calculated alternately until a local maximum of the expectation in Equation (9) is found. Because the distributions of $P(R|S)$, $P(S|D)$, and $P(S|R,D)$ are uniform, their initial values can be set equal to $P(R|S)$. The number of iterations depends on experience; in this research, it is set to five.

Each image can then be represented by the posterior probability $P(s_k | d_j)$ instead of by the original image feature.

4. Experiments

In the experiments, TM images of Kii Peninsula (Japan), Wuhan (China), and Yancheng (China) are used. Each image is split into 256×256 subimages, and the total number of images is 2,000. Each image could manually be classified into eight land-cover types, namely sea, river, lake, farmland, urban area, cloud, forest, and bare soil.

4.1 Image segmentation

Experiments on TM image are performed to test this segmentation algorithm. For a comprehensible comparison, we use original JSEG algorithm and the well-established eCognition. The results are shown in Figure 5 (boundaries are highlighted by red lines).

Figure 5a shows two original TM images, both covering urban area and forest. The textural characteristics are clear in the forest area; because of the modest resolution of the TM sensor, spectral characteristics are more prominent than textural characteristics in the urban area. Figure 5b presents the results from the original JSEG method. Note that the JSEG method produces good results, but sometimes it cannot separate two different regions very well due to the complex spectral characteristics of remotely sensed image. Compared with the JSEG method, the proposed method (Figure 5c), which takes NDVI, NDBI, and textural features into consideration, makes the difference between the ground covers more obvious and generates much better segmentation results.

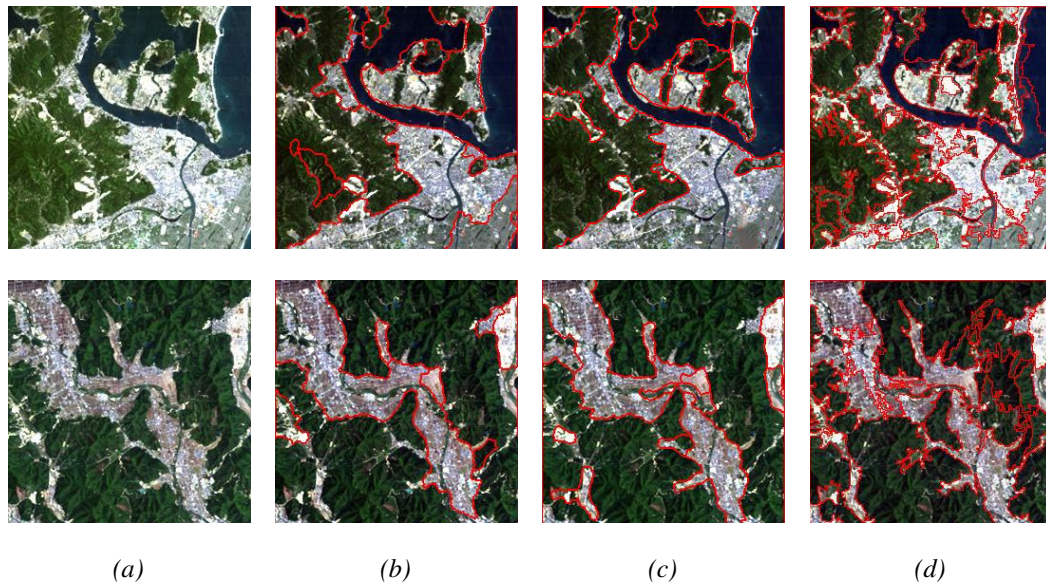


Figure 5 Images and corresponding segmentation results using different methods. (a) The original TM images with forest, urban area, and sea. **(b-d)** The results that are separately segmented by original JSEG algorithm, the proposed method, and eCognition, respectively.

Figure 5d presents the results obtained from eCognition. Although the boundary is clear in Figure 5d, the result contains many fragments and some oversegmentation.

These experimental results prove that the proposed method outperforms the other two methods in terms of visual evaluation. It not only produces a good segmentation boundary, but also avoids oversegmentation.

4.2 Semantic feature extraction

In this experiment, semantic features are extracted from spectral and textural features. To determine the optimal number of semantic features, different numbers of semantic features are used to retrieve images. Without considering time requirements, the retrieval precisions obtained for the initial 20(40) result images (denoted as Top (20(40))) are shown in Figure 6.

The result indicates the general trend that the larger the number of semantic features, the higher the retrieval precision is. This occurs because more semantic features are used to describe the image content, more details can be described. However, a larger number of semantic features will lead to greater computational complexity for hidden semantic feature extraction and time requirements for computing similarity for image retrieval. The first turning point in Figure 6 is 100; there is no change when the number of semantic features is larger than 100. Therefore, the number of semantic features is set to 100.

4.3 Differences between semantic features

In this experiment, two groups of original remotely sensed image and their corresponding semantic features are shown in Figures 7 and 8, respectively.

Figure 7a shows an image covering forest and urban area; Figure 7b shows the semantic features of these two kinds of ground cover. In Figure 7b, the cyan column presents semantic feature of mountain, while the red column presents semantic feature of urban area; each column indicates the value of corresponding semantic feature dimension. The higher the column, the larger the semantic feature value is. The total number of semantic features is 100.

In the original image (Figure 7a), it is clear that mountainous and urban areas present obviously different features; in Figure 7b, this difference is also clearly presented. For examples, in the 18th semantic feature,

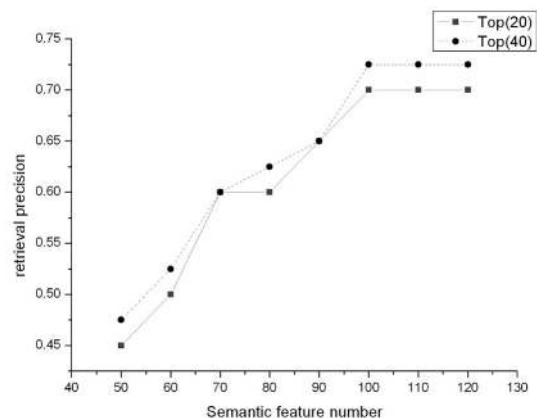


Figure 6 Retrieval precision for different numbers of semantic features.

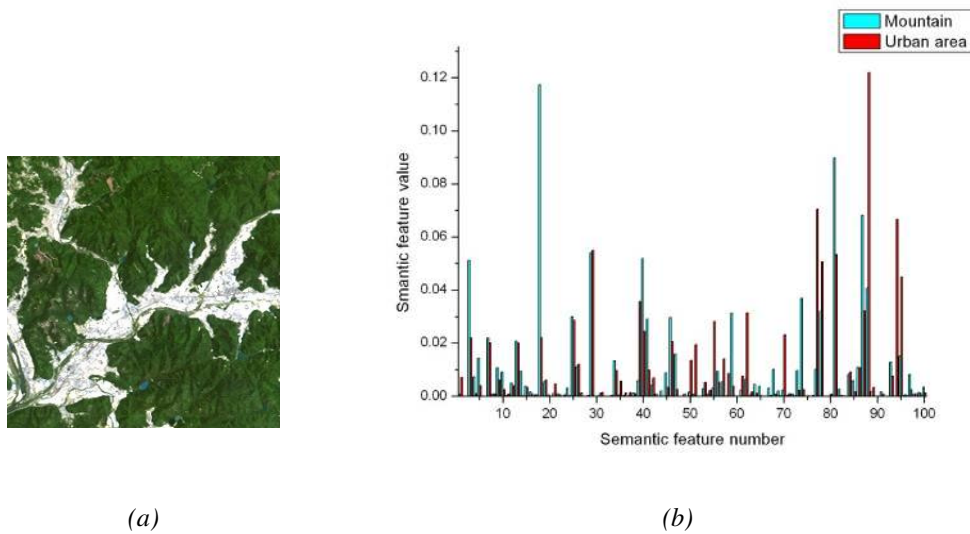


Figure 7 Different land cover types and corresponding semantic features. (a) Original image. (b) Semantic feature of mountain and urban areas.

the cyan column is extremely higher than the red, while in the 88th semantic feature, the red column is extremely higher than the cyan. This indicates that semantic features can be used as a replacement for low-level features to distinguish between different ground covers.

Figure 8a, b presents two images both covering sea area. Figure 8c shows the semantic features of these two sea images. In Figure 8c, the red column presents semantic features of the sea area of image a, while the cyan column presents semantic features of the sea area of image b; each column indicates the value of the

corresponding semantic feature dimension. The higher the column, the larger the semantic feature value is, and the total number of semantic features is 100.

Although the characteristics of the sea area in these two images are same, the concept is different because the area adjacent to sea in Figure 8a is urban while in Figure 8b it is mountain. This different concept between two sea areas is clearly presented in Figure 8c. For examples, in the 26th semantic feature, the red column is extremely higher than the cyan while in the 86th semantic feature, the cyan column is extremely higher

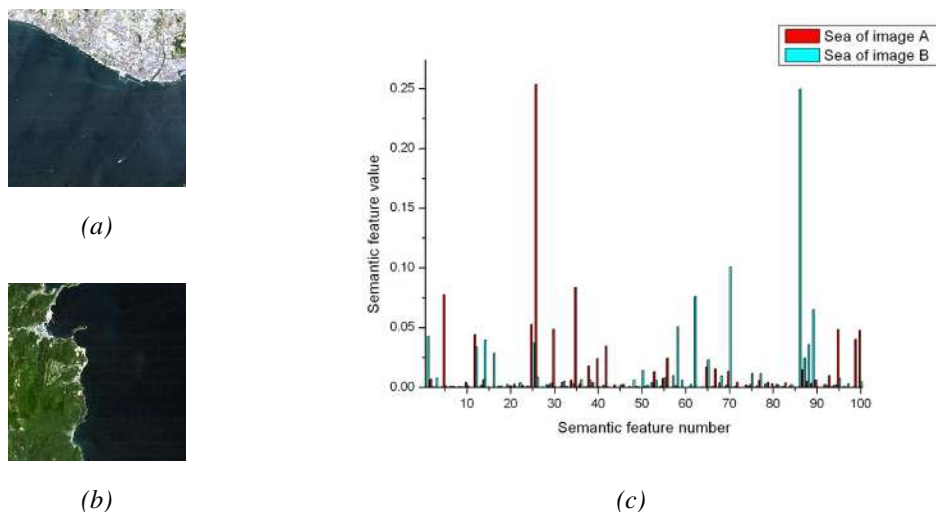


Figure 8 Similar land cover types and corresponding semantic features. (a) Original image. (b) Original image. (c) Semantic feature of sea of (a) and (b).

than the red. This indicates that these semantic features can also well describe the high-level semantic concepts hidden in images.

These two experiments lead to the conclusion that the extracted semantic features can well describe not only low-level image characteristics, but also high-level hidden concepts.

4.4 Image retrieval experiments

Once the semantic features have been extracted, each image can be represented. The Euclidean distance method is used to calculate the similarity between two images. The following two experiments present the different specimen images and the top 20 retrieved results.

4.4.1 Experiment A

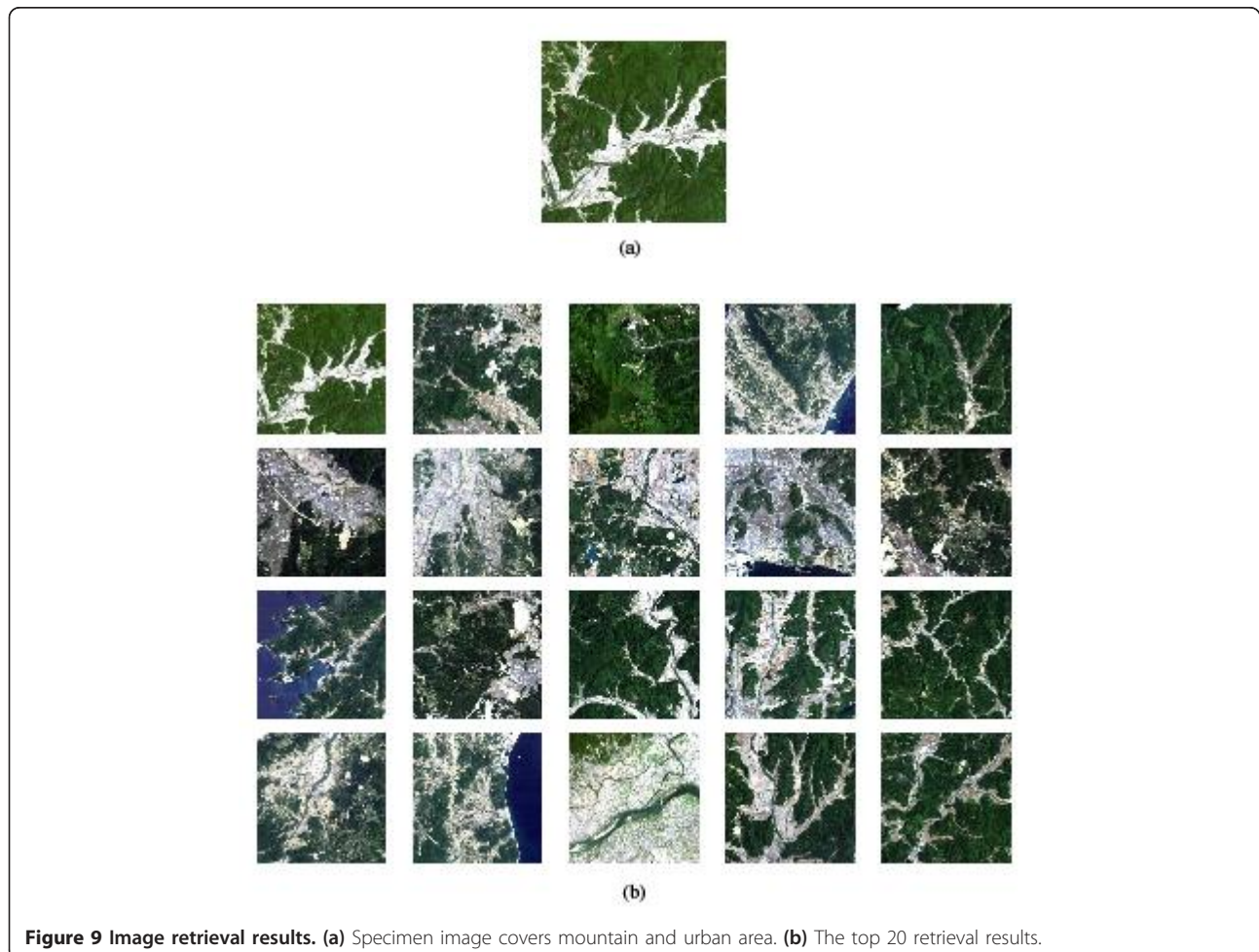
In this retrieval experiment, the specimen image is shown in Figure 9a, with the corresponding retrieval results consisting of the most similar images presented in Figure 9b. Specimen image covers mountain area and urban area, in which low-level features (spectral and

textural) are clearly seen, and semantic perception is that urban area is on the foot of a mountain.

The analysis of spectral and textural features shows that most of these 20 retrieval results are similar with the specimen image covering not only mountain, but also urban areas. The analysis of high-level semantic features shows that 14 images are similar with the specimen image in which urban area is surrounded by mountain, but others are different in which mountain and urban areas are adjacent.

4.4.2 Experiment B

In this experiment, an image covering cloud and mountain is chosen as the specimen image. The semantic perception in this image is that clouds float above mountain. The top 20 retrieval results, consisting of the most similar images, are shown in Figure 10b. The analysis of low-level features shows that most of results are similar to specimen image. However, according to semantic perception, 16 of them are similar to the specimen image.



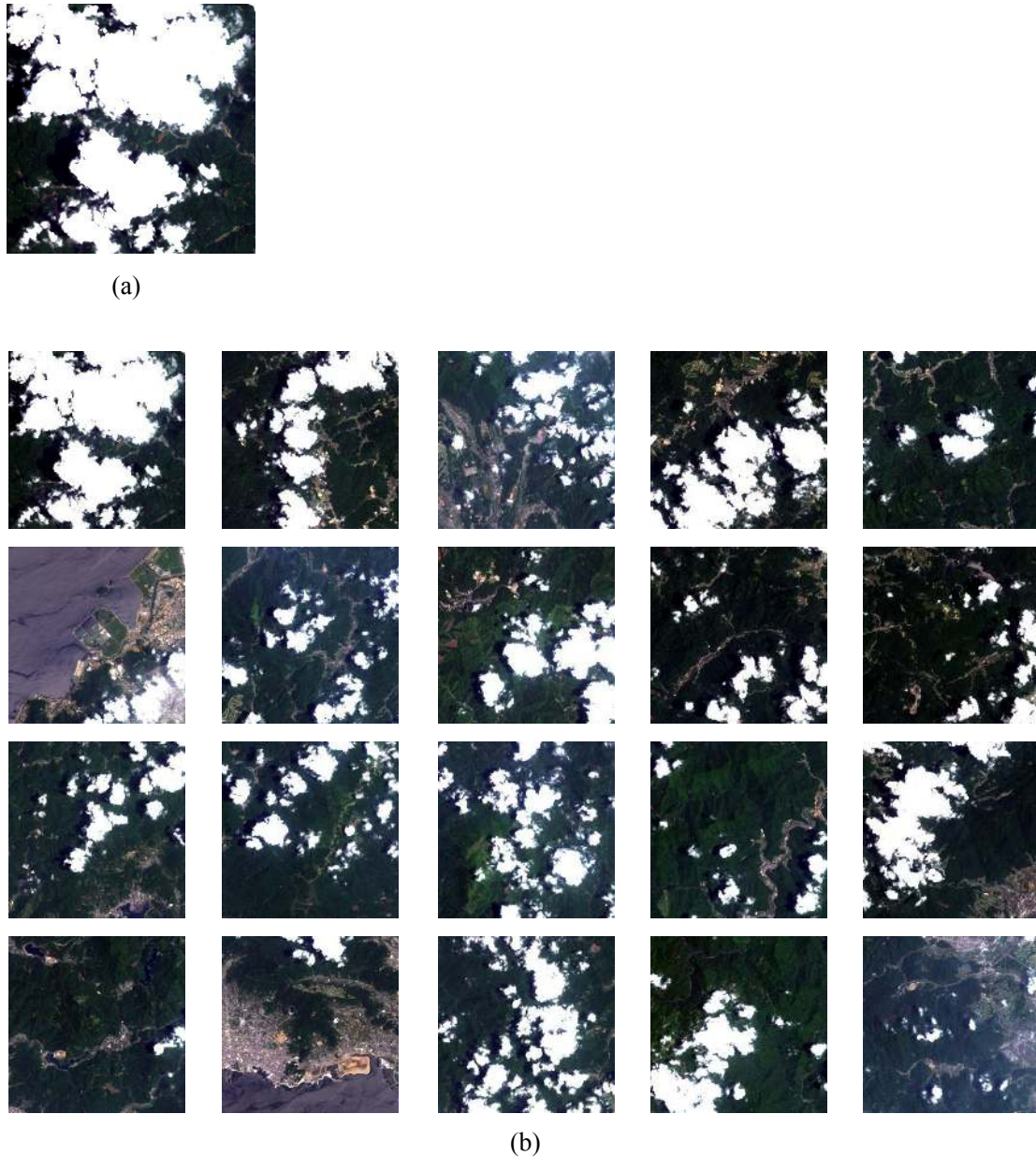


Figure 10 Image retrieval results. **(a)** Specimen image covers cloud and mountain. **(b)** The top 20 retrieval results.

4.4.3 Image retrieval precision and recall

According to the results as shown in Figures 9 and 10, retrieval precision and recall are calculated. The precision and recall used in this analysis are defined as Equations (13) and (14)

$$\text{Precision} = \frac{I_{\text{relevant}} \cap I_{\text{retrieved}}}{I_{\text{retrieved}}} \quad (13)$$

$$\text{Recall} = \frac{I_{\text{relevant}} \cap I_{\text{retrieved}}}{I_{\text{relevant}}}, \quad (14)$$

where I_{relevant} is the total number of relevant images, and $I_{\text{retrieved}}$ is the total number of retrieved images.

Figures 11 and 12 present the precision and recall results when the numbers of total retrieved images are 10, 20, 30, and 40, respectively. Although the precision and recall of experiment B are much higher than those of experiment A, both of them exhibit a slow drop in precision by increasing the number of retrieved images while the recall increases. It is consistent with normal trend of precision and recall. For a comprehensible comparison, low-level features of specimen images are

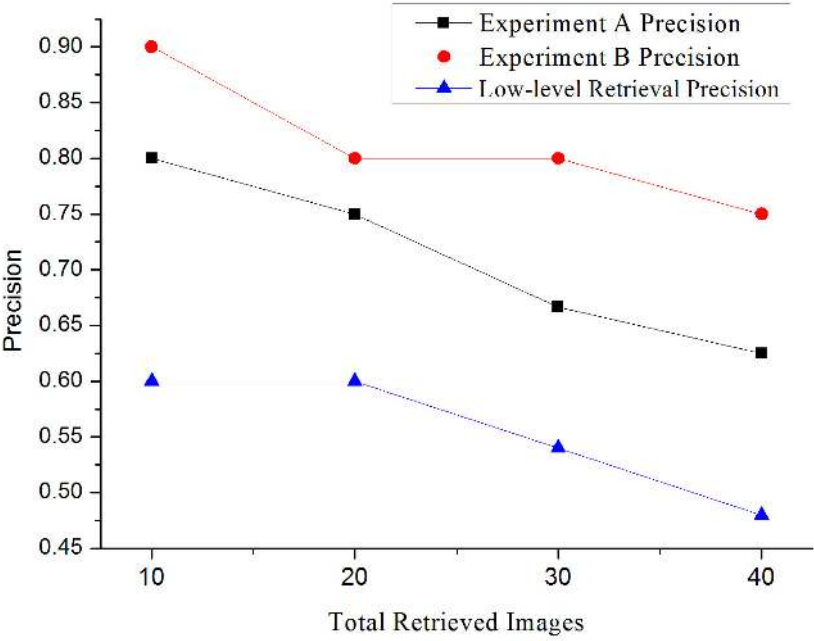


Figure 11 The precision results for different number of total retrieved images.

used to retrieve image. Precision and recall results are shown in each figure. It can be noticed that the proposed method obtains higher precision and recall.

5. Conclusions
In this article, a region-level semantic-based satellite image retrieval system is described between low-level

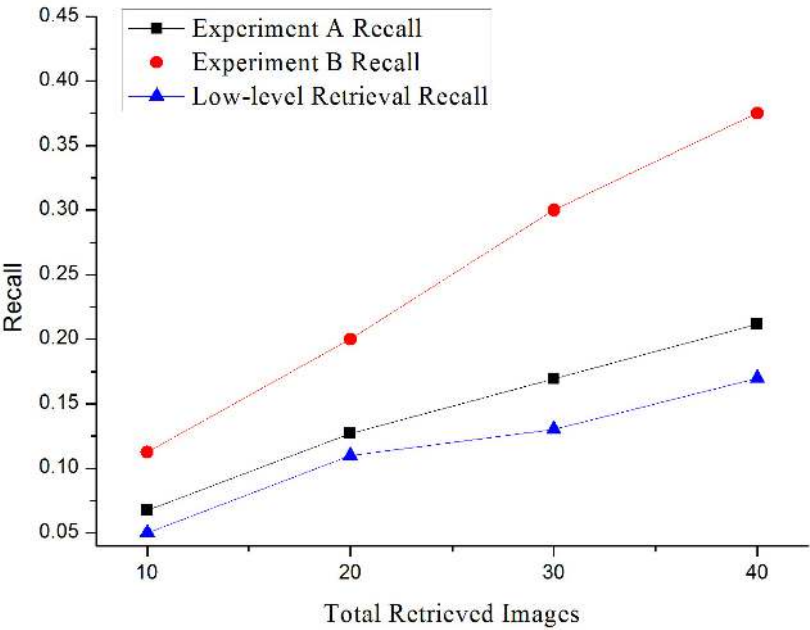


Figure 12 The recall results for different number of total retrieved images.

information and high-level concepts. Regional and semantic features are combined to narrow the semantic gap. An improved image segmentation algorithm is introduced which can obtain much better segmentation results than earlier algorithms. Semantic features are extracted using a probabilistic method, and experiments indicate that the new semantic features can represent not only low-level information, but also high-level concepts. Image retrieval experiments on two different specimens attain better retrieval precision and recall. The major limitation of this approach, however, is that a computationally expensive regeneration of the derived semantic model is required if new satellite images are added to the database.

Acknowledgements

This research was supported by the Chinese 863 program (No.2009AA12Z133), the National Nature Science Foundation of China (NSFC) (No. 41076126/D0611), and the National Basic Research Program of China (No. 2011CB7071).

Author details

¹Chinese Antarctic Center of Surveying and Mapping, Wuhan University, Wuhan, China ²The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China ³Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China

Competing interests

The authors declare that they have no competing interests.

Received: 15 February 2011 Accepted: 23 March 2012

Published: 23 March 2012

References

- Aksoy S, Koperski K, Tusk C, Marchisio G, Tilton JC: **Learning Bayesian classifiers for scene classification with a visual grammar.** *IEEE Trans Geosci Remote Sens* 2005, **43**(3):581-589.
- Bimbo AD: *Visual Information Retrieval* Morgan Kaufmann Publishers, Inc., San Francisco, CA; 1999.
- Vasconcelos N: **From pixels to semantic spaces: advances in content-based image retrieval.** *Computer* 2007, **40**(7):20-26.
- Flickner M, Sawhney H, Niblack W, Ashley J, Qian H, Dom B, Gorkani M, Hafner J, Lee D, Petkovic D, Steele D, Yanker P: **Query by image and video content: the QBIC system.** *Computer* 1995, **28**(9):23-32.
- Smith JR, Chang SF: **Automated binary texture feature sets for image retrieval.** In *IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 4.* Atlanta, GA, USA; 1996:2239-2242.
- Chun YD, Kim NC: **Content-based image retrieval using multiresolution color and texture features.** *IEEE Trans Multimedia* 2008, **10**(6):1073-1084.
- Ferecatu M, Boujemaa N: **Interactive remote-sensing image retrieval using active relevance feedback.** *IEEE Trans Geosci Remote Sens* 2007, **45**(4):818-826.
- Huang X, Zhang LP, Li PX: **Classification and extraction of spatial features in urban areas using high resolution multispectral imagery.** *IEEE Trans Geosci Remote Sens Lett* 2007, **4**(2):260-264.
- Diou C, Stephanopoulos G, Panagiotopoulos P, Papachristou C, Dimitriou N, Delopoulos A: **Large-scale concept detection in multimedia data using small training sets and cross-domain concept fusion.** *IEEE Trans Circ Syst Video Technol* 2010, **20**(12):1808-1821.
- Mylonas P, Spyrou E, Avrithis Y, Kollias S: **Using visual context and region semantics for high-level concept detection.** *IEEE Trans Multimedia* 2009, **11**(2):229-243.
- Li Y, Bretschneider TR: **Semantic-sensitive satellite image retrieval.** *IEEE Trans Geosci Remote Sens* 2007, **45**(4):853-860.
- Lee CS, Ma WY, Zhang HJ: **Information embedding based on user's relevance feedback for image retrieval.** *Proc of SPIE Photonics East Boston, MA;* 1999, 19-22.
- Rui Y, Huang TS, Ortega M, Mehrotra S: **Relevance feedback: a power tool for interactive content-based image retrieval.** *IEEE Trans Circ Syst Video Technol* 1998, **8**(5):644-655.
- Li J, Narayanan RM: **Integrated spectral and spatial information mining in remote sensing imagery.** *IEEE Trans Geosci Remote Sens* 2004, **42**(3):673-685.
- Li QY, Hu H, Shi ZZ: **Semantic feature extraction using genetic programming in image retrieval.** In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR). Volume 1.* Cambridge, England, UK; 2004:648-651.
- Datcu M, Daschiel H, Pelizzari A, Quartulli M, Galoppo A, Colapicchioni A, Pastori M, Seidel K, Marchetti PG, D'Elia S: **Information mining in remote sensing image archives: system concepts.** *IEEE Trans Geosci Remote Sens* 2003, **41**(12):2923-2936.
- Daschiel H, Datcu M: **Design and evaluation of human-machine communication for image information mining.** *IEEE Trans Multimedia* 2005, **7**(6):1036-1046.
- Deng YN, Manjunath BS: **Unsupervised segmentation of color-texture regions in images and video.** *IEEE Trans Pattern Anal Mach Intell* 2001, **23**(8):800-810.
- Wang ZY, Boesch R: **Color- and texture-based image segmentation for improved forest delineation.** *IEEE Trans Geosci Remote Sens* 2007, **45**(10):3055-3062.
- Chang Y, Lee D, Wang Y: **Color-texture segmentation of medical images based on local contrast information.** *Proc IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology* Honolulu, HI; 2007, 488-493.
- Han JW, Ngan KN: **Automatic segmentation of objects of interest in video: a unified framework.** *Proceeding of 2004 Intelligent Signal Processing and Communication Systems (ISPACS)* Seoul, South Korea; 2004, 375-378.
- Gersho A, Gray RM: *Vector Quantization and Signal Compression* Kluwer Academic: Norwell, MA; 1992.
- Dell'Acqua F, Gamba P, Ferrari A, Palmason JA, Benediktsson JA, Arnason K: **Exploiting spectral and spatial information in hyperspectral urban data with high resolution.** *IEEE Geosci Remote Sens Lett* 2004, **1**(4):322-326.
- Rouse JW, Haas RH, Schell JA, Deering DW: In *Monitoring vegetation systems in the great plains with ERTS*. Edited by: Freden SC, Mercanti EP, Becker MA. Third Earth Resources Technology Satellite-1 Symposium-Volume I: Technical Presentations, NASA SP-351, Washington, USA; 1974:310-317.
- Zha Y, Gao J, Ni S: **Use of normalized difference built-up index in automatically mapping urban areas from TM imagery.** *Int J Remote Sens* 2003, **24**(3):583-594.
- Zhao YD, Zhang LP, Li PX: **Texture feature fusion for high resolution satellite image classification.** *International Conference on Computer Graphics, Imaging and Vision: New Trends* Beijing, China; 2005, 19-23.
- Haralick RM: **Statistical and structural approaches to texture.** *Proc IEEE* 1979, **67**(5):786-804.
- Dempster AP, Laird NM, Rubin DB: **Maximum likelihood from incomplete data via the EM algorithm.** *J R Stat Soc Ser B* 1977, **39**(1):1-38.
- Moon TK: **The expectation maximization algorithm.** *IEEE Signal Process Mag* 1996, **13**:47-60.

doi:10.1186/1687-5281-2012-4

Cite this article as: Liu et al.: Remotely sensed image retrieval based on region-level semantic mining. *EURASIP Journal on Image and Video Processing* 2012 **2012**:4.