《**Note**》

Region-based Image Retrieval Using Semantic Mining

Tingting Liu
(Wuhan University, China, Nara Women's University)

Kanako Muramatsu
(Department of Advanced Information and Computer Sciences,
School of Interdisciplinary Scientific Phenomena and Information,
Nara Women's University/ Kyousei Science Center for Life and Nature, Nara)

Motomasa Daigo
(Doshisha University)

Pingxiang Li
(Wuhan University, China)

Liangpei Zhang
(Wuhan University, China)

Abstract

In this paper, a multi-level image representation model is developed and used to mine semantic feature hidden in the original remote sensing image. This model is consisted of three levels : region level, region feature level and semantic level. The first two levels aim at represent image content by using region feature. Semantic level aims at extracting hidden semantic feature. At last, interested part and uninterested part method is used to improve the retrieval precision. Experiment shows that this method can well improve the accuracy of the retrieval result.

Key words : Remote sensing, image retrieval, semantic

INTRODUCTION

With the development of the aviation technology, sensor technology, network technology and database technology, enormous amounts of remote sensing images are continuously collected. How to index or retrieve them is a challenge which is the focus of many research projects.

The first generation image retrieval system is a kind of text-based image retrieval system. This kind of system only retrieves scenes based on geographical location, spatial characteristic of the imaging instrument, acquisition date and etc. However, queries can not process this type of information, as the text-based method focuses on the search for a scene that shares a similar ground cover characteristic with certain query scene. After near one decade of research, it was found that content based image retrieval (CBIR) is a practical and satisfactory solution to this challenge. The concept of CBIR system is to describe the image content by a priori automatically extracted descriptors, called feature vectors, which express spectrum feature, texture feature and spatial feature[1]. Then the feature vector is compared with a correspondingly vector extracted from the query image. However, it is well known that the performance of CBIR is mainly limited by the gap between low-level features and high-level semantic concepts of user queries. This gap is caused by the missing direct relationship between low-level features and high-level semantic concepts[2, 3]. In order to reduce this gap, region based features are used instead of raw features of whole image to represent the visual content of an image[4–6], and semantic features are used to replace low-level features to describe concepts in user's mind. A common technique to provide regions with semantic meaning is the manual annotation. Combined with a powerful segmentation method this technique can produce in a good meaningful classification. However, hand-annotation images are tedious and human expensive, so methods of learning image representations directly from data are required.

Bayesian networks are the method for uncertainty reasoning and knowledge representation and were advocated at the end of the 20 th Century. It is a probabilistic graphical model, which has been used for probabilistic reasoning in expert systems. Because the novel method has a powerful ability for reasoning and a flexible mechanism to learn, it provides an effective way to deal with causality or uncertainty. Bayesian networks are proved to have surprisingly broad applications, such as medical diagnoses, image classification and understanding, prediction and forecasting. It is proved to be particularly useful in knowledge discovery and data mining. Datcu[7] used a simple Bayesian network to establish the stochastic linkages between the joint space of signal classes and semantic concepts, and partially bridged the semantic gap. However, an important element of image understanding is the spatial information because complex land cover structures usually contain many pixels and regions that have different feature characteristics. Furthermore, two scenes with similar regions can have very different interpretations if the regions have different spatial arrangements. So, this drawback of simple Bayesian network may significantly impact on the retrieval performance.

In this paper, a multi-level image representation model is developed to extract the semantic concepts hidden in the images, and then use these semantic features to represent and retrieve image. This model is consisted of three levels : region level, region feature level and semantic level. In region level, improved JSEG algorithm is used to divide image into several regions. In region feature level, spectral feature and texture feature are extracted and used to represent region content. In semantic level, Bayesian method and Expectation Maximization (EM) are used to mine hidden semantic feature. At last, in order to analyze similarity between each image, interested part and uninterested part method is used to improve the retrieval precision.

IMAGE REPRESENTATION MODEL

Image representation model is built to mine hidden semantic feature, and use these features to represent image content. In this model : firstly use improved JSEG algorithm to segment image ; secondly use spatial feature and texture feature to represent region content ; lastly, use Bayesian method and Expectation Maximization (EM)[8] method to mine hidden semantic.

1. 1 Region segmentation

Region-based JSEG[9] algorithm is a new image segmentation algorithm which not only considers color information but also considers spatial information. Due to the large band number and the high correlation between different bands, we improve JSEG algorithm and make it applicable for remote sensing image segmentation. At last, a good segmentation of remote sensing image is acquired.

1. 2 Region Feature Extraction

There are various kinds of image features : color, texture, shape and so on. The different feature plays a different role in image analysis. In this paper, according to the characteristics of multispectral image, we choose spectrum feature and texture feature as remote sensing image's feature.

2. 2. 1 Spectral feature

Color feature is one of the most important features of pictures, and it is the first feature used in image retrieval. In remote sensing image, spectral feature has the same importance of color feature in the picture. In this paper, we use original image pixel's value as spectral feature.

2. 2. 2 Texture feature

Texture is often seen as a kind of local characters of image, and also a measure of the relationship between pixels in local region. It is a very useful property of spatial structure information in images. Incorporating texture information, heterogeneous objects and different objects with same spectrum could be much more detected effectively. According to this, Gauss Markov random field (GMRF) model[9] is used to extract texture feature of image in this paper.

1. 3 Semantic mining

In this step, Bayesian method is used to mine the relationship between region, image and semantic. Firstly, we suppose region $r_i \in R = \{r_1, \cdots, r_N\}$ and image $d_j \in \{d_1, \cdots, d_M\}$ are independent to each other, N and M is the total number of the region feature and image. Secondly, we suppose that the hidden semantic feature $s \in S = \{s_1, \cdots, s_K\}$exists, K is the number of semantic feature. A joint probability of and is

$$P(r_i,d_j) = P(d_j)P(r_j \mid d_j) = P(d_j)\sum_{k=1}^{K} P(r_i \mid s_k)P(s_k \mid d_j) \tag{1}$$

When region r_i and image d_j exist, the probability of S_k existing is $P(S_k \mid r_i,d_j)$. Formula (1) combines Bayesian formula can get :

$$P(s_k \mid r_i,d_j) = P(r_i,d_j \mid s_k)/P(r_i \mid s_k)P(s_k \mid d_j)/\sum_{l=1}^{K} P(r_i \mid s_l)P(s_l \mid d_j) \tag{2}$$

Then, EM method is used for hidden semantic estimation. From formula (2) can get :

$$E(\log P(R, D, S)) = \sum_{i=1}^{N}\sum_{j=1}^{M} n(r_i,d_j)\sum_{k=1}^{K} P(s_k \mid r_i,d_j)\log[P(s_k \mid d_j)P(r_i \mid s_k)] \tag{3}$$

After maximizing (3) with Lagrange multipliers, obtains :

$$P(r_i \mid s_k) = \sum_{j=1}^{M} n(r_i,d_j)P(s_k \mid r_i,d_j) / \sum_{m=1}^{N} \sum_{j=1}^{M} n(r_m,d_j)P(s_k \mid r_m,d_j) \tag{4}$$

$$P(s_k \mid d_j) = \sum_{i=1}^{N} n(r_i,d_j)P(s_k \mid r_i,d_j) / \sum_{n=1}^{N} n(r_m,d_j) \tag{5}$$

The (2) and (4), (5) formulas are alternated until a termination condition met.

SEMANTIC-BASED IMAGE RETRIEVAL

Based on the multi-level image representation model, each image in the database can be represented by the posterior probability $P(s_k \mid d_j)$ instead of original image feature. It is not only decrease the dimension of the feature but also improves the image retrieval precision because that semantic feature is much more closed to user's thought.

In image retrieval process, we can find that there are some images which the user is interested in and some images which the user is not interested in. So, we think that all the examples can be divided into two groups : interested examples and uninterested examples. However, most of the image retrieval systems only care about interested examples but neglect uninterested ones. When interested examples are mixed with uninterested examples, system precision will be highly impacted. Then, they have to use relevant feedback to improve the retrieval precision, but that also increase the complexity of retrieval.

In this paper, we suppose each image include both interested part and uninterested part. So, the original posterior probability can be replaced by the new one to represent images. The new posterior probability is $aP(s_k|d_j)+bP(-s_k|d_j)$. In this formula, $P(s_k|d_j)$ is the interested part and a is the weight, $P(-s_k|d_j)$ is the uninterested part and b is the weight. In the process of image retrieval, a and b can be calculated by comparing image's similarity with interested example and uninterested example.

IMAGE RETRIEVAL EXPERIMENT

1. 4 Experiment data and experiment platform

In this paper, different scenes of TM images are utilized. Each image is split into small size of 256 *256, and total number of small size image is 500. After preprocessing, each image and its features are stores in the image database. In this image database, images can be classified into seven concepts : cloud, sea, river, mountain, urban area, farm land, bare soil.

1. 5 Experiment and result

4. 2. 1 Semantic feature experiment

In this experiment, two images are compared by each other from segmentation result and semantic extraction result.

Fig. 1 is two original images and their segmentation results : a picture and c picture are the original images, b picture and d picture are the segmentation results by using improved JSEG algorithm. In b, image A is roughly divided into three parts : sea, mountain and urban area. In d, image B is accurately divided into two parts : sea and urban area. From the segmentation result we can find that the difference from image A and image B is the mountain region.

Fig. 2 is two images semantic features, x-axis is semantic feature number, y-axis is feature value, black line is image A's sematicfeature and gray line is image B's sematicfeature Because of comparing only two images' semantic feature, we use 10 semantic features in this experiment. From the

a : Original Image A b : Image A's result c : Original Image B d : Image B's result

Fig. 1 Two images and their segmentation results

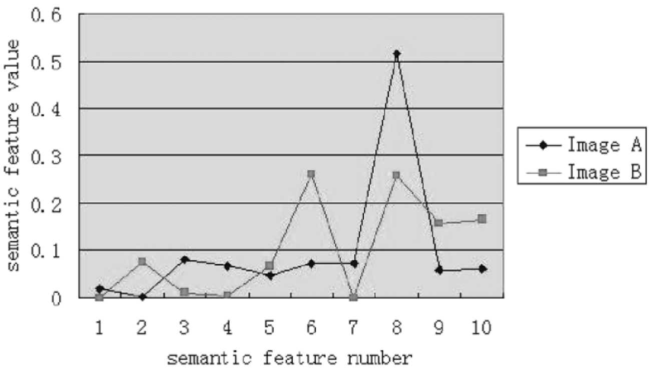


Fig. 2　Semantic of image A and Image B

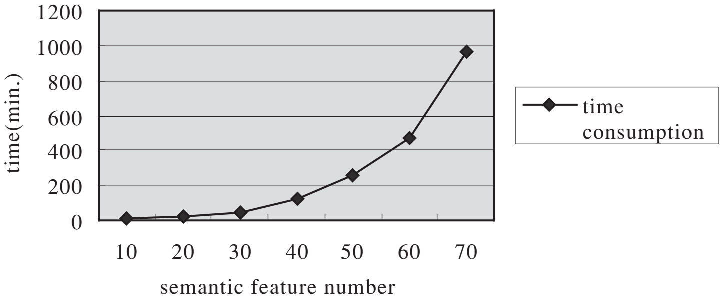


Fig. 3　Time consumption of semantic feature extraction

graph of semantic feature result we can find that image A and image B can be well separated by using the sixth and the eighth semantic feature. From the above, we can get that semantic feature can well represent image.

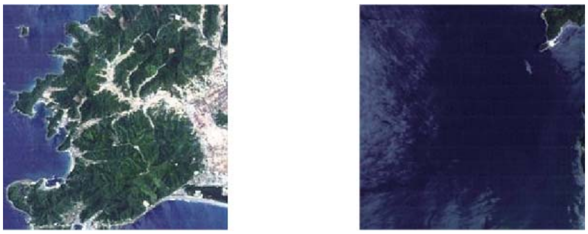
4. 2. 2 EM algorithm's computational complexity experiment

In this experiment, different semantic feature number is chosen for testing EM algorithm's computational complexity.

Fig. 3 is the graph of semantic feature extraction time consumption. X axis is the number of semantic feature and Y axis is the time consumption of extracting corresponding semantic feature. From the graph we can find, with the feature number adding, time consumption rapidly increases. When the feature number is 70, the time consumption is nearly 1000 minutes. Simultaneously, there are only seven concepts in image database. So, caring about various aspects, 50 is decided as semantic feature number.

4. 2. 3 Image retrieval experiment

In this experiment, we choose one image as interested example and one image as uninterested example. Suppose user want an image same as interested example but don't need big area of sea same as uninterested example.



Interested example   Uninterested example

Fig. 4 interested example and uninterested example of image retrieval

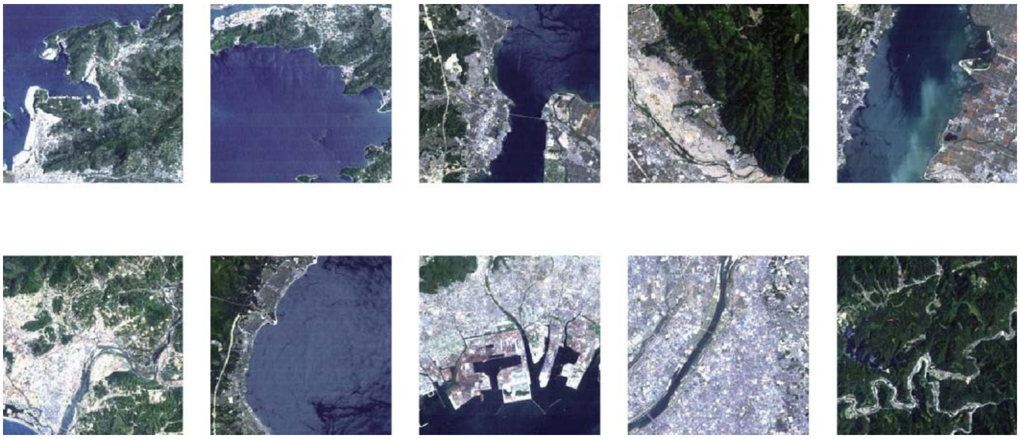


Fig. 5 The first 10 retrieval results by only using interested example

Fig. 6 The first 10 retrieval results by using interested example and uninterested example

In Fig. 5, there are 7 images same as interested image, both contain mountain, sea and urban region. So, we can find that semantic features play an important role in image retrieval and also get a good result. However, in user's opinion, they need an image which is same as interested image and different from uninterested image. Although the results are same as interested image, are not close to user's requirement. After separate image to interested part and uninterested part, similarity is recalculated again. In Fig. 6, the first five images only contain mountain and urban area, the last five images although contain region of sea but not so large. Comparing Fig. 5 and Fig. 6 can obviously find that by using interested example and uninterested example make result more similar with user's requirement.

CONCLUSION

In this paper, a multi-level image representation model is developed to extract semantic feature hidden in remote sensing image. In this model, Bayesian method and Expectation Maximization (EM) method are utilized to actualize semantic extraction. At the image retrieval step, concept of interested example and uninterested example is used for improving retrieval accurate. The experiments show that this model can well replace original feature and narrow the semantic gap and also show that concept of interested example and uninterested example used in image retrieval can get a good result.

However, because of high computational complexity of Bayesian method and EM method, time assumption of semantic extraction is a serious problem. So, the future work will be focus on improve the efficiency of semantic feature extraction.

ACKNOWLEDGEMENTS

This research was supported by Academic Frontier Project (1999–2010).

REFERENCES

1 Yikun Li, TimoR. Semantic-sensitive satellite image retrieval. IEEE transaction on Geoscience and Remote Sensing, Volume : 45, Issue : 4, April 2007, page(s) : 853–860.

2 C. Lee, W. Y. Ma, and H. J. Zhang, “Information embedding based on user's relevance feedback for image retrieval,” in Proc. SPIE Conf. Multimedia Storage and Arch. Syst., 1999, vol. 5, pp. 20–22.

3 Y. Rui, T. S. Huang, M. Ortega, and S. Mehrota, “Relevance feedback : A power tool for interactive content-based image retrieval,” IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp. 644–655, Sep. 1998.

4 W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. IEEE Trans. on PAMI, 22 : 1349–1380, 2000.

5 C. Carson, M. Thomas, and e. a. S. Belongie. Blobworld : A system for regionbased image indexing and retrieval. In the 3rd Int'l Conf. on Visual Information System Proceedings, pages 509–516, Amsterdam, Netherlands, June 1999.

6 Y. Chen and J. Z. Wang. A region-based fuzzy feature matching approach to content-based image retrieval. IEEE Trans. on PAMI, 24(9) : 1252–1267, 2002.

7 Mihai Datcu. A Bayesian multi-class image content retrieval. IGARSS 2007. 23–28 July 2007 Page(s) : 326–329

8 A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the em algorithm. J. Roy. Statist. Soc., vol 39, no. 1, pp. 1–38, 1977.

9 Deng YN, Manjunath BS. Unsupervised Segmentation of Color -Texture Regions in Images and Video [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 23, Issue 8, Aug. 2001 Page(s) : 800–810.