Content-based remote sensing image retrieval using co-training of multiple classifiers

LI Shijin, TAO Jian, WAN Dingsheng, FENGJun

School of Computer & Information Engineering, Hohai University, Jiangsu Nanjing 210098, *China*

Abstract: There are usually few training samples in the tasks of content-based remote sensing image retrieval, which will lead to over-learning problem while using this small data set for training. In this paper a novel approach using co-training in multiple classifier systems is presented, which can label the unclassified samples automatically by using the cooperative determination of the classifiers which are created on several different feature sets, so that the small sample problem can be raveled out. Compared with the technique of relevance feedback, the experiments indicate that they have their own strengths and can obtain almost the same results. However, the proposed approach of co-training in multiple classifier systems is superior in regard of avoiding the needs of human intervention through relevance feedback.

Key words: remote sensing, content-based image retrieval (CBIR), co-training, multiple classifier systems, semi-supervised learning

CLC number: TP751.1 **Document code:** A

Citation format: Li S J, Tao J, Wan D S and Feng J. 2010. Content-based remote sensing image retrieval using co-training of multiple classifiers. *Journal of Remote Sensing*. **14**(3): 493—506

1 INTRODUCTION

During the last decades, the imaging satellite sensors have acquired huge quantities of data, e.g., NASA can collect more than 1 000 GB remote sensing imageries every day. With the fast increasing of these data, it is helpful to environment monitoring, hazards prevention, agricultural management and urban planning. However, the access to image archives will become more difficult due to the enormous data quantity acquired by the new generation of high-resolution satellite sensors. The state-of-the-art systems for accessing remote sensing data and images, in particular, allow only queries by geographical coordinates, time of acquisition, and sensor type. This information is often less relevant than the content of the scene, e.g., structures, patterns, or objects. The text annotation of image content is impractical due to the tedious manual labor and human subjectivity. Content-based image retrieval (CBIR) has emerged in the 1990s and achieved great advancements recently. The underlying rational of CBIR is the automatic extraction of image features, such as color, texture, shapes and their combination. Since these features represent the intrinsic characteristics of the image, the subjectivity of the above-mentioned manual annotation can be alleviated to some extent.

There are many efforts to utilize CBIR in remote sensing imageries. Zhu *et al*. (2000) proposed to use Gabor-filter based texture features to retrieve aerial photos over a test-bed of 800 aerial photographs. Luo *et al*. (2001) focused on visual feature based CBIR problem of remote sensing image, and the feature selection, extraction, similarity model and their implementation were thoroughly investigated. Both the spectrum information and wavelet-based features were utilized to describe the content of the image. In the work by Lu *et al*. (2004), a remote sensing image retrieval approach using the linear weighted fusion of color and texture features was presented, in which the image was decomposed using quin-tree. Zeng *et al*. (2005) proposed an algorithm for the retrieval of large-scale remote sensing images. After the partition of the original image, some statistics of the texture features were calculated using the method like Hu's invariant moment, and a group of texture features connected with the texture's position was extracted. Bao and Guo (2006) experimentally investigated two kinds of similarity measures for remote sensing CBIR, i.e., feature vector based measure and the probabilistic one. It was found that χ^2 statistical distance and cosine of the angle performed well for the first similarity measure, while the K-nearest neighbor-based rule was effective for the second one.

During the past few years, relevance feedback (Rui *et al*., 1998; Ferecatu & Boujemaa, 2007) has been put forward to tackle the problem of semantic gap, which is caused by the inconsistency between low-level features and the high-level semantics of the user's query. The user is requested to give some hints on the relevance or irrelevance of the query result.

Received: 2009-05-06; **Accepted**: 2009-08-24

Foundation: Natural Science Foundation of China (No.60673141).

First author biography: LI Shijin (1973—), male, Associate Professor. He graduated from the Department of Computer Science, Nanjing University of Science & Technology, with the degree of Ph.D majoring in pattern recognition and intelligent system in 2000. His current research interests focus on image analysis and pattern recognition. Dr. Li has published more than 30 papers. E-mail: lishijin@hhu.edu.cn

Specifically, a query task is established by several rounds of feedback, during which the target concept is approached by some strategies, such as query shifting and relevance estimation. With the increasing of the tagged samples by the user, the performance will be boosted rapidly. But in the retrieval of remote sensing imagery, the number of the tagged samples is scarce and thus can result in over-fitting phenomenon. In real applications, it is easy to obtain large number of unlabeled samples, while it is very difficult to get many labeled ground-truth samples. Semi-supervised learning(SSL) has been proposed to deal with this problem (Blum & Mitchell, 1998).

In this paper a novel approach using co-training in multiple classifier systems is presented, which can label the unclassified samples automatically by using the cooperative determination of the classifiers which are created on several different feature sets, so that the small sample problem can be raveled out. Compared with the technique of relevance feedback, the experiments indicate that they have their own strengths and can obtain almost the same results. However, the proposed approach of co-training in multiple classifier systems is superior in view of avoiding the needs of human intervention through relevance feedback.

2 RELATED WORK ON SSL IN CBIR

Semi-Supervised Learning (SSL) aims to tackle the problem of insufficient training samples and it can be defined as: Given some labeled samples, and abundant unlabeled samples, how to predict the labels of the test samples, during which the issue is to utilizing the hidden information in the unlabeled samples.

Recently, several researchers have proposed to use SSL in CBIR. Zhou *et al*. (2007) put forward to use canonical correlation analysis (CCA) to find several correlated projections of the two views in the co-training algorithm, before the semi-supervised learning. Dong and Bahnu (2003) had put forward a new EM algorithm for image retrieval, where the image distribution in feature space was modeled as a mixture of Gaussian densities. Tian *et al*. (2004) provided a new analysis on the value of unlabeled data by considering different distributions of the labeled and unlabeled data and showing the migrating effect for semi-supervised learning. Extensive experiments have been performed in the context of image retrieval applications. Yu (2006) studied graph based semi-supervised learning to the problem of image categorization and proposed a novel neighborhood preserving graph-based semi-supervised learning method. Hoi *et al*. (2008) proposed a novel idea of learning with historical relevance feedback log data, and adopt a new paradigm called Collaborative Image Retrieval. By integrating both log data and unlabeled data information through an effective graph regularization framework, they proposed a novel semi-supervised distance metric learning technique, called “Laplacian Regularized Metric Learning”, for learning robust distance metrics. Yao *et al*. (2008) addressed the medical image retrieval problem by utilizing the semi-supervised method and semantic error-correcting output codes model. Sharma *et al*. (2008) developed a weakly semi-supervised ensemble of decision tree classifiers trained based on a co-training framework for the tagging of the images over the Web.

Most of the current semi-supervised learning methods can be categorized into three main paradigms: the generative model, the graph-based methods and the co-training methods (Zhou, 2007). In this paper we focus on the co-training methods (Blum & Mitchell, 1998). In the co-training paradigm, it is assumed that features can be divided into two independent sets, each of which contains sufficient information to construct two classifiers, C_1, and C_2. These two classifiers are initially trained on a small labeled set L, after which the training set is augmented by the classification on the unlabeled set U and the samples with high confidence are added into L. This training process continues until there is no more sample can be labeled. The underlying rational of co-training method is that, one classifier maybe can not label the unknown sample; however, the other classifier can label it correctly. Hence, each classifier can add some labeled samples into the training set, which are informative to the other classifier.

The co-training paradigm can also easily be generalized to multiple classifiers system (MCS), whose precondition requires that the base classifiers of the MCS are independent and consistent (Didaci & Roli, 2006). This requirement can not be satisfied in many real applications and Goldman and Zhou (2004) have put forward a new algorithm for co-training without sufficient and redundant views. They run different algorithms of decision tree on the same set of features and each algorithm can divide the training samples into several equivalent classes. The newly labeled samples are tagged by the average confidence of every classifiers. It can be found that co-training paradigm can also suit the situation of two classifiers trained on the same feature set.

Didaci and Roli (2006) have extended the co-training of two classifiers to co-training multiple classifiers. But the algorithm in (Didaci & Roli, 2006) is a general framework and is not specific for CBIR, without considering the problems induced by error samples, while the final retrieval results will be unsatisfactory with the adding of error unlabelled samples. During the procedure of CBIR, the focus is concentrated on the similar images or patches in a large remote sensing image. Different query may need different feature sets. When adding the unlabelled samples, some improvements are proposed in this paper, and different classifiers are required to maintain coherence and collaboration in selecting samples. Previous work in CBIR has seldom considered this issue and we propose to co-train multiple classifiers on different features and select the most appropriate feature sets for a specific query session by taking consideration of the cooperation of multiple classifiers. In addition, most of the previous work on SSL in CBIR employed the relevance feedback technique, while the approach proposed in this paper has better degree of automation without it. In the next section, we will present the proposed algorithm in detail.

3 CONTENT-BASED REMOTE SENSING IMAGE RETRIEVAL ALGORITHM USING CO-TRAINING OF MULTIPLE CLASSIFIERS

The principle of content-based remote sensing image retrieval using co-training of multiple classifiers is to train classifiers respectively on some different feature sets. In this paper, four feature sets are utilized to conduct co-training, which are divided into groups (color group and texture group), and jointly determine the unlabelled samples in a manner of group decision. All the classifiers used in this paper are k-nearest neighbor classifiers, and some other kind of classifiers can also be employed, such as support vector machines (SVM), or the Bayes classifier; however, it may increase the time costs of training.

The details of the algorithm are as follows:

(1) Divide the original remote sensing image into patches. For content-based remote sensing image retrieval, it is necessary to divide an image into patches reasonably and efficiently. To avoid partitioning a target region into different patches, the strategy of overlapped sub-block is adopted (Li & Ning, 2006). The size of each patch is the same as the query sample, not larger than 128, and the overlapped area between two patches is width/2 by height/2 pixels. The purpose of doing so is that we think when the user selects a sample image patch for query, the size of this patch represents some typical land cover primitives in the scene and by respectively constraining the width and the height within the range of 128 pixels, we can avoid too rough retrieval results induced by too large sub-images.

(2) Extract color features. Here we adopt the dominant color histogram in the HSI space and Lab space respectively, since these two color models can approximate the perception of human beings and in the Lab color space, the Euclidean distance of different colors coincide with our human perceptions (Gonzalez & Woods, 2007).

(3) Conduct color quantization. In the HSI space, we quantize H, S, and I into 12, 4, 4 intervals respectively, so there are totally 192 bins in the HSI histogram. While in the Lab space, we quantize the L, a, and b components into 4, 8, 8 intervals, so that there are totally 256 bins in the Lab histogram.

(4) In order to remove the influence of noises, here we adopt the dominant color descriptor (Manjunath *et al*., 2001). We count the number of pixels of each bin respectively in HSI histogram and Lab histogram, and zero the bins whose number is less than a given threshold T_1 (in HSI histogram) or T_2 (in Lab histogram). The thresholds of T_1 and T_2 are determined adaptively on the basis of their own 0.9TAP (Total Amount of Pixels).

(5) Compute the texture features of each patch, which include the features derived from the co-occurrence matrix of grey level (Haralick *et al*., 1973) and the texture features based on Gabor filter (Manjunath & Ma, 1996). In the first category of texture features, four of the features corresponding to our human perceptions, i.e., consistency, entropy, contrast, and correlation, are extracted. And the Gabor filter-based texture feature is another set of widely adopted descriptor for texture images and has been recommended for the MPEG-7 standard (Manjunath *et al*., 2001). When computing the texture features based on Gabor filter, the resolution has two scales and four orientations, which are the orientations of 0°, 45°, 90° and 135°, so the Gabor texture features are 16-dimensional (mean and variance).

(6) Suppose that the positive training sets are "Labeled" *(K)* (K is the size of the set, and initially it is constituted by the examples selected by the user). Using the set of "Labeled" *(K)* to construct four nearest neighbor classifiers respectively on the four feature sets described in the last step.

(7) For all classifiers, each patch image for classification is assigned a score which is one or zero. An image for classification is given a score of one, if the distance between the positive training set and its corresponding feature is smaller than a threshold T; otherwise it is given a score of zero. During the process of assigning a score, firstly, we divide the classifiers into two groups, one is the color group (its feature sets are HSI dominant color histogram and Lab color histogram), another one is the texture group (its feature sets are co-occurrence matrix of grey level and texture features based on Gabor filter). We consider an unlabeled image as a positive example, if it gains one score at least respectively in texture group and color group, then we label it and put it into the set "Labeled" *(K)*. If the size of "Labeled" *(K)* stops increasing, the iteration stops. At this moment, the samples in the set "Labeled" *(K)* are the resulting positive sample images, otherwise the process goes to the 6th step and continues learning till the size of "Labeled" *(K)* stops expanding.

The distance D is defined as follows:

$$D=\min D_i \, (i\in[1, K]) \tag{1}$$

where D_i is the distance between the patch image for classification and a training sample i ($i\in[1, K]$) in the set "Labeled" *(k)*.

As the distance threshold T for each classifier is concerned, it is dynamically determined as follows. First, we calculate the distances DD_j, $j\in[1, M-K]$, between the image patches to be classified and those positive samples in "Labeled" *(K)*, M is the total image patches in the image. Second, $\{DD_j\}$ are sorted in increasing order and is denoted as $\{NDD_j\}$. T is initialized as NDD_1. If the number of the selected image patches K_0 is in $[K_l, K_h]$, then the current T is returned, otherwise $T=NDD_{1+\text{step}}$, where "step" is an increasing variable until K_0 is in the range of $[K_l, K_h]$. K_l and K_h is the minimum (maximum) number of samples permitted to be added to "Labeled" *(K)* in each round of iteration. In this paper, K_l =1 and K_h=4. During this process, it can be guaranteed that only few positive image patches are added into "Labeled" *(K)*, which are the most similar with the samples in "Labeled" *(K)* in the color and texture feature space.

4 EXPERIMENTAL RESULTS AND ANALYSIS

In the experiment, we have selected four scenes of remote sensing images to verify the validness of our algorithm, whose target areas are different type of land covers, such as regions of soil erosion, fishing ponds, and residential areas. Table 1 gives the details on the statistics of selected features by each classifier in the iterations. Fig. 1, Fig. 2, Fig. 3 and Fig. 4 give the retrieval results.

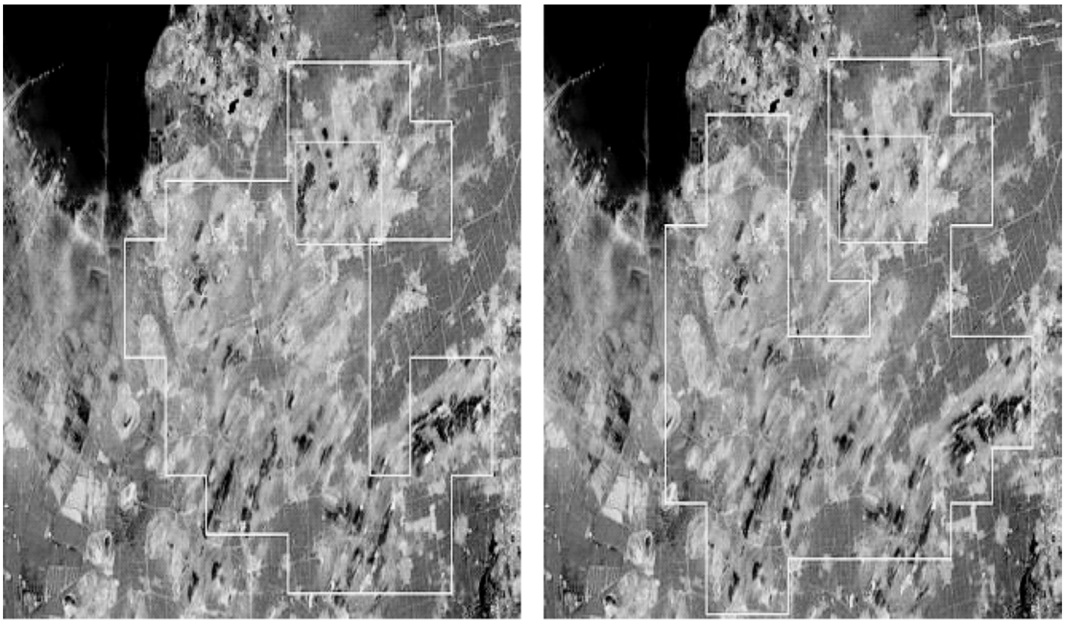
As shown in Table 1, we can see that:

(1) For Fig. 1, its color scores reach the number of 33 (19+14), which are 7 points higher than the number (26) of its texture scores; while for Fig. 2, the scores of color and texture

Table 1 Iteration number of retrieval using co-training of multiple classifiers and statistics of selected features

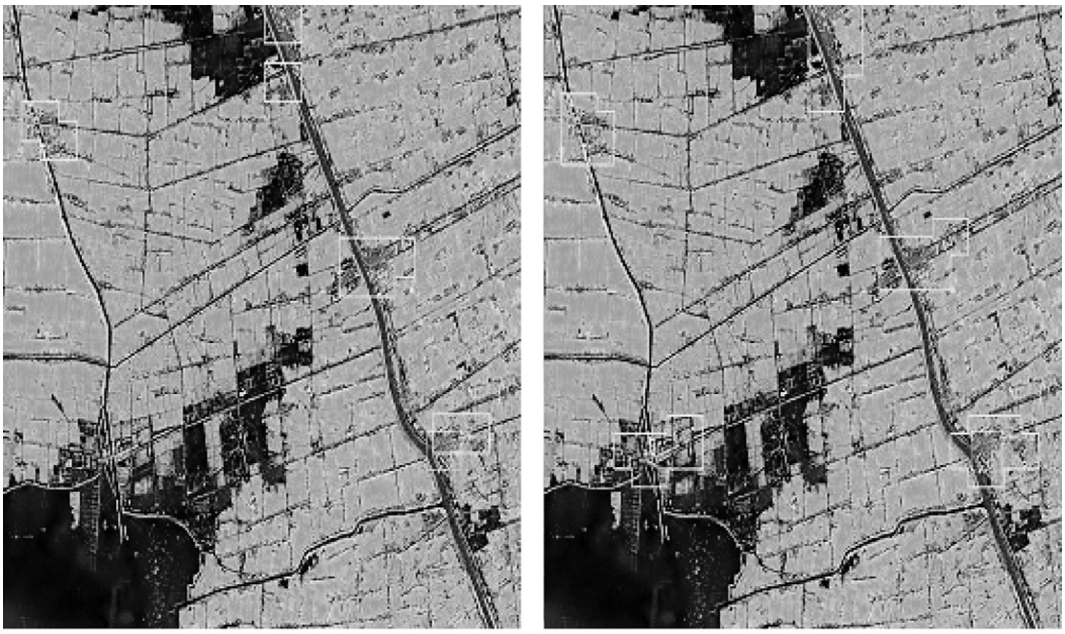
/Times

	Itertions	#(HSI) selected	#(Lab) selected	#(GLCM) selected	#(Gabor) selected
Fig.1	7	14	19	10	16
Fig.2	5	7	5	9	4
Fig.3	6	5	6	4	5
Fig.4	6	4	4	5	2



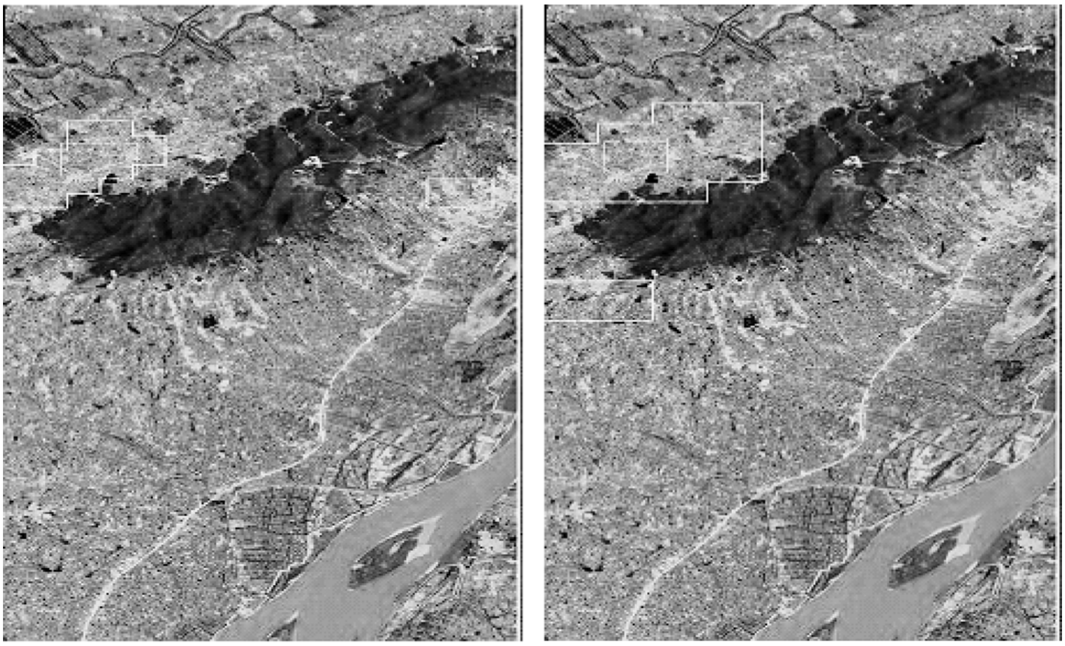
(a) (b)

Fig. 1 Retrieval results of soil erosion area by co-training and CTRF
(The small rectangle with white lines on the top-right indicates the initial query, while the large area enclosed by white lines represents the final retrieval results.)
(a) Co-training; (b) CTRF



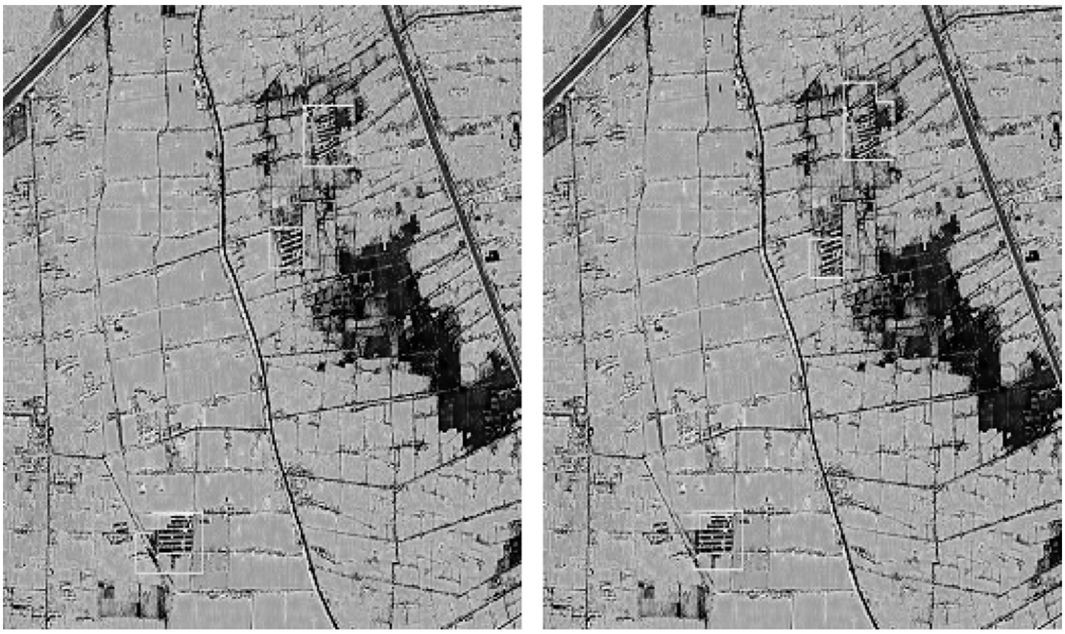
(a) (b)

Fig. 2 Retrieval results of residential areas by co-training and CTRF
(The small rectangle with white lines on the bottom-right indicates the initial query, while the large area enclosed by white lines represents the final retrieval results.)
(a) Co-training; (b) CTRF



(a) (b)

Fig. 3 Retrieval results of soil erosion area by co-training and CTRF
(The small rectangle with white lines on the left-top indicates the initial query, while the large area enclosed by white lines represents the final retrieval results.)
(a) Co-training; (b) CTRF



(a) (b)

Fig. 4 Retrieval results of fishing ponds area by co-training and CTRF
(The small rectangle with white lines on the bottom-left indicates the initial query, while the large area enclosed by white lines represents the final retrieval results.)
(a) Co-training; (b) CTRF

are almost the same. So, we can see that in Fig. 1 the role of color information is more important than the texture, while in Fig. 2 the role of color and texture are similar, which is consistent with the visual perception of the images. Previous work on the combination of color and texture features need to set the weights of each other, however, in our method it naturally resolves or bypasses this issue.

(2) Through the process of iteration, the unlabeled samples score respectively in four features, and there are no patches getting a score of zero. And it also shows that the features are complementary for each other, which is an advantage of using co-training: for a specific sample, one classifier perhaps can not label it correctly, but another one can.

(3) In remote sensing CBIR, different feature sets are required for different scenes and land covers. In Fig.1, Fig.2 and Fig. 4, the selection of texture features are different. For example, during the 7 iterations, the Gabor texture feature has been selected 16 times in Fig.1, while it is only selected 4 and 2 times in

Fig.2 and Fig.4 respectively. Hence, it can be observed that the proposed method in this paper can select different feature sets during the retrieval process by the co-training of multiple classifiers.

From Fig. 1—Fig. 4, we can see that the retrieval results are satisfactory by using co-training of multiple classifiers (the images have been scaled properly due to space limit of the paper). Compared with the results of relevance feedback (CTRF), which integrates color and texture features, the retrieval results based on co-training in multiple classifiers are satisfying. For Fig. 1 the retrieval performance of co-training is better than the performance of CTRF, it is easy to see the purity of soil erosion area in Fig. 1(a) is better than that in Fig. 1(b). For Fig. 2, the retrieval results of CTRF is a little better than that of co-training in multiple classifiers, and in Fig. 2(b) one more residential area at the left bottom is retrieved by CTRF. In Fig. 3, the retrieved area of soil erosion is smaller than CTRF, but has a higher purity.

In fact, the two methods have their own advantages. The user's feedback is the greatest advantage of CTRF, and it is guided by the user to a learning process and narrows "semantic gap" through the help of human-computer interaction, as to make two images similar to each other by decreasing the distance between the feature vectors, while the two images represented by initial low-lever features are not similar. As is shown in Fig. 2, the urban residential area in the lower-left corner is not visually consistent with the user's initial training samples, but it can be retrieved through the feedback of the user. However, there are some disadvantages of taking the user into the learning process, the greatest of which is that human beings have certain subjectivity, and different users may have different understandings about the same image and then give different relevance feedback. Meanwhile, the content-based remote sensing images are more complex and versatile than the natural images, so sometimes it is difficult for the user to determine if a sample to feedback is a positive example or not. The approach of co-training in the multiple classifiers is different, whose greatest advantage is no need for human intervention, and it exploits the inherent nature of low-level image feature vectors fully, making the information of different features complementary for each other. Certainly, it also has some shortcomings, such as it is less flexible than CTRF in certain situations. However, the retrieval performance of co-training of multiple classifiers is satisfactory, and it has a higher degree of automation.

5 CONCLUSIONS

In this paper we propose a novel approach to remote sensing CBIR using co-training of multiple classifiers, which aims at tackling the small sample size problem in the target application. We compare it with the retrieval method of CTRF through experiments, and the CTRF method is the integration of color, texture and relevance feedback. The results show that the two methods have their own advantages and disadvantages, and their retrieval performance is almost the same, but the method of co-training of multiple classifiers avoids the need of human intervention, and it exploits the inherent nature of low-lever image feature vectors fully. Further, it makes the information of different features complementary for each other, as to avoid the tedious human feedback in the retrieval process and enhance the degree of automation.

Currently, the features are divided into the color and the texture group. Our future work includes extracting new features for content-based remote sensing image retrieval and employing image segmentation technique to sketch out the target area in the query sample roughly before submitting the query, as to further improve the recall rate and the precision rate. During our experiments, it is observed that the partition of the original image is of great importance to the success of remote sensing image retrieval. If the target land cover region is divided into several small patches, it is of high probability that the target will be missed as the ratio is too small in an image patch. So a better partition scheme is also one of the issues to be resolved in the future.

REFERENCES

Bao Q and Guo P. 2006. Comparative studies on similarity measures for remote sensing image retrieval based on histogram. *Journal of Remote Sensing*, **10**(6): 893—900

Blum A and Mitchell T. 1998. Combining labeled and unlabeled data with co-training. Proceedings of 11th annual Conference on Computational Learning Theory. New York, NY, United States: ACM Press

Didaci L and Roli F. 2006. Using co-training and self-training in semi-supervised multiple classifier systems. Proceedings of SSPR&SPR. Berlin: Springer Press

Dong A and Bhanu B. 2003. A new semi-supervised EM algorithm for image retrieval. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, WI, United states: IEEE Press

Ferecatu M and Boujemaa N. 2007. Interactive remote sensing image retrieval using active relevance feedback *IEEE Transactions on Geoscience and Remote Sensing*, **45**(4): 818—826

Gonzalez R C and Woods R E. 2007. Digital Image Processing(2nd edition). Beijing: Publishing House of Electronics Industry

Haralick R, Shanmugam K and Dinstein I. 1973. Textural features for image classification. *IEEE Transactions. System, Man & Cybernetics*, **3**(6): 610—621

Hoi S C ,Liu W and Chang S F. 2008. Semi-supervised distance metric learning for collaborative image retrieval. Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition. Madison, WI, United states: IEEE Press

Li D R and Ning X G. 2006. A new image decomposition method for content-based remote sensing image retrieval, *Transactions on Geomatics and Information Science of Wuhan University*, **31**(8): 659—662

Lu L Z, Liu R Y and Liu N. 2004. Remote sensing image retrieval using color and texture fused features. *Journal of Image and Graphics*, **9**(3):328—332

Luo R, Zhang Y S, Fan Y D and Deng X Q. 2001. Research on content-based remote sensing image retrieval: the strategy for visual feature selection, extraction, description and similarity measurement. Proceedings of 2001 International Conferences on Info-tech and Info-net. Beijing: IEEE Press

Manjunath B S and Ma W Y. 1996. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis and Machine Intelligence*,**18**(8): 837—842

Manjunath B S, Ohm J R, Vasudevan V V and Yamada A. 2001. Color and texture descriptors. *IEEE Trans. Circuits and Systems for Video Technology*, **11**(6): 703—715

Rui Y, Huang T S, Ortega M and Mehrota S. 1998. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions. on Circuits and Systems for Video Technology*, **8**(5): 644—655

Sharma A, Hua G, Liu Z and Zhang Z. 2008. Meta-tag propagation by co-training an ensemble classifier for improving image search relevance, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Anchorage, AK, United states: IEEE Press

Tian Q, Yu J, Xue Q and Sebe N. 2004. A new analysis of the value of unlabeled data in semi-supervised learning for image retrieval. Proceedings of the IEEE International Conference on Multimedia Exposition. IEEE Press

Yao J, Zhang Z, Antani S, Long R and Thoma G. 2006. Automatic medical image annotation and retrieval using SEMI-SECC. Proceedings of the IEEE International Conference on Mutilmedia and Exposition. Piscataway, NJ, United States: IEEE Press

Yu Z H. 2006. Image Categorization with semi-supervised learning. Proceedings of the IEEE International Conference on Image Processing, Atlanta, GA, United States: IEEE Press

Zeng Z M, Li F, Fu K and Ding C B. 2005. A method for extracting texture characters from large-scale remote sensing image. *Transactions on Geomatics and Information Science of Wuhan University*, **30**(12): 1080—1083

Zhou Y and Goldman S. 2004. Democratic co-learning. Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence. Boca Raton: IEEE Press

Zhou Z H. 2007. Co-training paradigm in semi-supervised learning. Zhou Z H and Wang J. Machine learning and its application, Beijing: Tsinghua University Press

Zhou Z H, Zhan D and Yang Q. 2007. Semi-supervised learning with very few labeled training examples. Proceeding. 22nd AAAI Conference on Artificial Intelligence. Menlo Park, CA, United States: American Association for Artificial Intelligence

Zhu B, Ramsey M and Hsinchun C. 2000. Creating a large-scale content-based airphoto image digitial library, *IEEE Transactions on Image Processing*, **9**(1): 163—167

多分类器实例协同训练遥感图像检索

李士进，陶 剑，万定生，冯 钧

河海大学 计算机及信息工程学院，江苏 南京 210098

摘 要： 提出一种基于多分类器协同训练的遥感图像检索方法，该方法在不同特征集上分别建立分类器，利用不同分类器的协同性自动标记未知样本，从而有效解决了小样本问题。通过与相关反馈方法进行实验比较分析，结果表明，这两种方法各有优劣，检索结果基本相当，然而多分类器协同训练方法避免了相关反馈过程中人工的多次反馈，自动化程度更高。

关键词： 遥感，基于内容的图像检索， 协同训练， 多分类器，半监督学习

中图分类号： TP751.1 **文献标识码：** A

引用格式： 李士进，陶剑，万定生，冯 钧. 2010. 多分类器实例协同训练遥感图像检索. 遥感学报, **14**(3): 493—506

Li S J, Tao J, Wan D S and Feng J. 2010. Content-based remote sensing image retrieval using co-training of multiple classifiers. *Journal of Remote Sensing*. **14**(3): 493—506

1 引 言

随着航空航天技术，传感器技术以及网络技术的迅速发展，每天获取的遥感图像数据日渐增多，如美国宇航局(NASA)每天获得的遥感图像数据量可达 1000 GB。遥感图像数据的快速增长为环境监测、灾害管理、农情监测和城市规划等领域的应用创造了有利的条件。然而对于海量的遥感数据，快速浏览及高效检索感兴趣的区域成为一项非常繁重而又困难的工作。传统的基于文本的图像检索面临诸多问题,如大量的人力工作量，人工标注难以描述图像中包涵的丰富内容，个人对图像内容理解上的主观差异也会降低文本表述的精度等。基于内容的图像检索(content-based image retrieval, CBIR)则利用视觉特征，如颜色、纹理及形状特征等作为图像索引，可用图像处理方法自动获得。将 CBIR 技术运用于遥感图像领域，其思想是根据遥感图像(或局部区域)的颜色、纹理、形状等特征以及这些特征的组合来表达图像，从而克服了以上问题。

自遥感图像 CBIR 提出以来，研究者们已经取得了许多研究成果，如 Zhu 等(2000)使用 Gabor 纹理特征作为图像特征来检索航空图像。Luo 等(2001)对基于内容的遥感影像检索中视觉特征选择、提取、描述和相似性度量进行了研究，采用光谱和基于小波变换的纹理特征描述遥感图像内容，并建立了相应的相似性模型。陆丽珍等(2004)提出一种基于五叉树分解的线性加权颜色和纹理特征的遥感图像检索方法。曾志明等(2005)提出了一种大尺寸遥感图像的纹理特征提取算法，将遥感图像分块后，用类似 Hu 不变矩的方法对各子块的纹理特征进行统计，从而得到各子块中和空间位置相关的一组纹理特征。包倩和郭平(2006)针对单波段遥感图像检索，分别研究了基于特征向量的相似性度量和基于概率的相似性度量，发现χ^2 统计距离和相似夹角余弦度量对第一种相似性度量较有效，而基于 K-近邻法则的计算方法对第二种相似性度量较有效。

相关反馈(relevance feedback)是 CBIR 中最常用的学习策略(Rui 1998; Ferecatu & Boujemaa, 2007)，在人机交互过程中，用户对系统返回的图像进行标记，然后系统利用一些监督学习算法来逼近用户的查询概念，从而优化查询结果，其性能随着训练样本集的增大而提高。然而在基于内容的遥感图像检索中，训练样本的数目往往非常少，用这小部分样本训练分类器，容易造成模型的过学习现象，这种

收稿日期：2009-05-06；修订日期：2009-08-24

基金项目：国家自然科学基金(编号：60673141)。

第一作者简介：李士进(1973—)，男，江苏泰县人，毕业于南京理工大学计算机系，获模式识别与智能系统专业博士学位，副教授，主要研究领域为图像分析，模式识别。 E-mail: lishijin@hhu.edu.cn。

现象对于高维特征空间更加严重。在实际应用中，获得大量已标记的训练样本比较困难，而获得不带类别标记的样本则要容易得多，为此，利用少量已标记样本和大量未标记样本构造分类器的半监督学习(semi-supervised learning，SSL)，成为近年的研究热点之一(Blum & Mitchell, 1998)。

本文针对图像检索中示例样本较少情况，提出一种基于多分类器协同训练(co-training)的遥感图像检索方法。遥感图像检索中，不同场景地表覆盖具有不同特点，因此需要采用不同的特征。该方法在不同特征集上分别建立分类器，自适应地采用不同的特征，利用不同分类器的协同性来自动标记未知样本，从而有效解决了图像检索中的小样本问题。

2 半监督学习与 CBIR

半监督学习可以定义为：给定部分已标记类别的样本数据集，如何预测其他未标记样本的类别问题。其关键是：除了已标记类别样本的信息外，如何有效利用数据集中大量未标记类别样本的信息来提高分类器的性能。

近年来，研究者们在利用半监督学习技术进行 CBIR 研究取得了较多的成果。Zhou 等(2007)针对训练样本少的情况，提出了基于经典相关分析(canonical correlation analysis，CCA)的半监督学习方法。Dong 和 Bahnu (2003) 将数据的混合分布模型，相关反馈和长期学习策略相结合，提出了一种新的基于 EM 算法的半监督分类方法，它可以处理在相关反馈中因用户对某些样本误标记而产生的错误。Tian 等(2004)指出，只有当未标记样本集与已标记样本集服从同样的分布时，半监督学习中未标记样本才能帮助提高精度。Yu(2006)提出了一种新的基于图论的半监督学习算法。与一般基于图论方法所不同的是，该方法只考虑某个节点(图像)邻近的 k 个节点，而不考虑其他距离较远的节点，其依据是考虑了局部一致性。由于传统的欧氏距离会受到“语义鸿沟”(semantic gap)的影响，Hoi 等(2008)提出一种新的半监督距离度量学习方案，该方法通过分析相关反馈历史日志信息为检索寻找最优的相似性度量。Yao 等(2008)提出了一种新的基于半监督学习的医学图像检索方法(SEMI-SECC)，该方法对图像集进行语义标注，采用错误纠正输出编码的方法来作为分类器的输出，未知样本标记情况由上一次迭代后的分类结果决定。Sharma 等(2008)在对 Web 图像进行语义标注时，提出了融合决策树和半监督学习的框架来标记未知图像。

按照使用的方法，半监督方法可以分为生成模型(Generative model)估计、直推学习(Transductive learning)和协同训练(Co-training)方法(周志华，2007)。协同训练方法最早是由 Blum 和 Mitchell(1998)提出，它假设样本特征可以被分成两个独立的子集，每个子集都包含足够的信息进行分类学习，在每个特征子集上建立各自的分类器 C_1 和 C_2。这两个分类器分别在一个小的已标记初始样本集 L 上进行训练。然后每个分类器被用来对未标记样本集 U 进行分类，将 U 中置信度高的若干样本添加到 L 中，从而增大已标记样本集 L。这两个分类器将在该持续增大的已标记样本集中不断迭代训练，直到没有更多合适的未标记样本加入。在协同训练的方法背后，实际上存在着这样的一个推断：对于一个特定的样本，一个分类器可能不能将其正确地标记，但它可能被另一个分类器正确地标记。因此，每个分类器可以增加一些特定的样本到训练集中，其中这些特定样本对于其他的分类器是“富信息” 的(informative)。

Co-training 方法也可以很容易推广到多分类器系统，其前提是构成该多分类器系统的基分类器满足一致性和独立性(Didaci & Roli, 2006)。然而，在许多实际应用中这种假设很难满足。Goldman 和 Zhou(2004) 提出了一种不需要充分冗余视图的协同训练算法。他们使用不同的决策树算法，在同一个属性集上训练两个不同的分类器，每个分类器都可以把示例空间划分为若干个等价类。在协同训练过程中，每个分类器通过平均置信度来标记新的正例样本。Zhou 和 Goldman(2004)的研究表明 Co-training 同样也适用于两个采用相同特征集的分类器。

由于 Didaci 和 Roli(2006)中多分类器 Co-training 并非针对图像检索，未考虑到错误样本引入问题，而在图像检索中，如果有错误未标记样本加入，则最终检索结果将不能令人满意。因此本文对此进行了改进，在增加未标记样本时，要求不同分类器在选择样本时必须具有一定的协同一致性，因此本文方法更加能够体现 Co-training 中的协同性。另外，大多数的半监督学习方法结合了相关反馈技术，而本文的方法不需要引入相关反馈技术，自动化程度更高。

3 基于多分类器 Co-training 方法的遥感图像检索算法

基于多分类器 Co-training 方法遥感图像检索的原理是在一些不同特征集上分别建立分类器，利用

不同分类器的协同性来自动标记未知样本。本文针对遥感图像检索，对分类器进行了分组(颜色组和纹理组)，以组为单位对未标记样本进行共同判定。本文所选用的分类器均为最近邻分类器，虽然也可以选择支持向量机、贝叶斯分类器等，但这可能会由于样本数目有限而导致过学习，并且同时增加训练的时间。

下面给出算法具体步骤:

(1) 将原图像进行分块。对于遥感图像检索，合理有效的图像分块是必须的。为了避免将同一目标分入不同的小块之中，本文采取了重叠分块策略(李德仁 & 宁晓刚, 2006)。每块大小为, width=min(128, 样本图像 width), height=min(128,样本图像 height), 块与块之间重叠 width/2×height/2 像元。这样做的目的是用户在勾选查询示例图像块时，该图像块的大小代表了一定的模式基元，而将子图像的长和宽限制在 128 像元内是为了避免子图像过大而导致检索结果太粗糙。

(2) 提取颜色特征。此处选用了 HSI 主颜色直方图和 Lab 主颜色直方图，分别将图像从 RGB 空间转换到 HSI 空间和 Lab 空间。之所以选择 HSI 和 Lab 颜色空间，是因为它们和人的视觉颜色感知比较接近，并且在 Lab 颜色空间，不同颜色之间的欧氏距离和人的颜色差异感觉很吻合(Gonzalez & Woods, 2007)。

(3) 对颜色进行量化。在 HSI 空间中，将 H 量化为 12 等份，S 和 I 分别量化为 4 等份，这样 HSI 直方图就有 192 个间隔(bin)。而在 Lab 空间中，将 L 量化为 4 等份，a 和 b 分别量化为 8 等份，从而 Lab 颜色直方图就有 256 个间隔(bin)。

(4) 为了去除噪声，此处采用主色特征(Manjunath 等, 2001)。分别在 HSI 直方图中和 Lab 直方图中统计每个 bin 的像元数 N，对于 N 小于给定阈值 T_1(HSI 直方图中)或 T_2(Lab 直方图中)的 bin 将其清零。其中 T_1 和 T_2 的确定是根据各自 90%的像元来自适应确定。

(5) 计算每幅图像的纹理特征，包括基于灰度共生矩阵的特征(Haralick 等, 1973)和 Gabor 纹理特征(Manjunath & Ma, 1996)。纹理特征中，图像的灰度共生矩阵已经被理论证明并且实验显示它在纹理分析中是一个很好的方法，广泛用于从灰度图像中提取纹理特征。本文提取了与人类视觉感知特性有明确对应关系的 4 个特征，包括纹理的一致性、熵、对比度和相关性。基于 Gabor 滤波器的纹理特征(本文简称 Gabor 纹理特征)是另一类比较广泛采用的纹理图像描述特征，并被推荐为 MPEG-7 标准中的纹理图像描述子之一(Manjunath 等, 2001)。本文采用 Manjunath 和 Ma(1996)中的方法提取了 2 个尺度下 4 个方向的均值和方差特征，即 Gabor 纹理有 16 维特征。

(6) 设正例样本集为 Labeled(K)(K 为样本集的大小，初始正例样本集即为用户所选示例图像块)。使用 Labeled(K)集在步骤(2)—(5)中描述的 4 个特征集上分别构造 4 个最近邻分类器。

(7) 对每一个分类器，根据一个阈值 T 为待分类图像块赋分，分值为 1 分或 0 分。若待分类图像块在该特征上与正例样本集的距离 D 小于对应的阈值时，则赋予 1 分，否则 0 分。在最终判断过程中，将 4 个分类器分为两组，一组为颜色组(特征集为 HSI 主颜色直方图和 Lab 主颜色直方图)，另一组为纹理组(特征集为灰度共生矩阵特征和 Gabor 纹理特征)。若待分类图像块在颜色组和纹理组分类器至少各得 1 分，则判定该图像块为正例，并将之加入 Labeled(K)集。若 Labeled(K)不再增大，则算法结束，此时 Labeled(K)集中的样本即为最后检索所得的相似图像块，否则返回步骤(6)，继续学习。

其中距离 D 和阈值 T 分别定义如下:

距离 D 的定义: 若某一待分类图像与 Labeled(K)集中某一元素 $i(i\in[1, K])$的距离为 D_i，则 $D=\min D_i$ $(i\in[1, K])$。

阈值 T 的定义: 首先计算所有待分类图像与正例样本集的距离 DD_j $(j\in[1, M-K])$，其中 M 为图像库中图像总数。然后将这些距离升序排列得到新距离 NDD_j，令初始的 $T=NDD_1$，若此时满足 $K_0\in[K_l, K_h]$，则返回该 T 值，否则增大 T，使得 $T=NDD_{1+\text{step}}$，step 为递增步长，直到满足 $K_0\in[K_l, K_h]$的条件。其中 K_0 为本次迭代新添加的正例数目，K_l, K_h 分别为用户设置的每次新添加的正例数目的下限和上限，典型的可以取 1, 4 或 4, 8 等，本文取[1, 4]。需要指出的是，条件 $K_0\in[K_l, K_h]$是由 4 个分类器共同决定的。可以看出，阈值 T 在每次迭代时会自适应发生改变。这种阈值的设定方法保证了在每次迭代中只加入一定数量的正例，并且保证新添加的正例在低层特征上与正例集最相似。

4 实验结果及分析

为了验证本文方法的有效性，我们选取了 4 幅典型遥感图像进行检索实验，其目标检索区域分别为不同类型的地表覆盖，如土壤侵蚀区域、鱼塘和居民地等。表 1 给出了在不同图像中多分类器协同

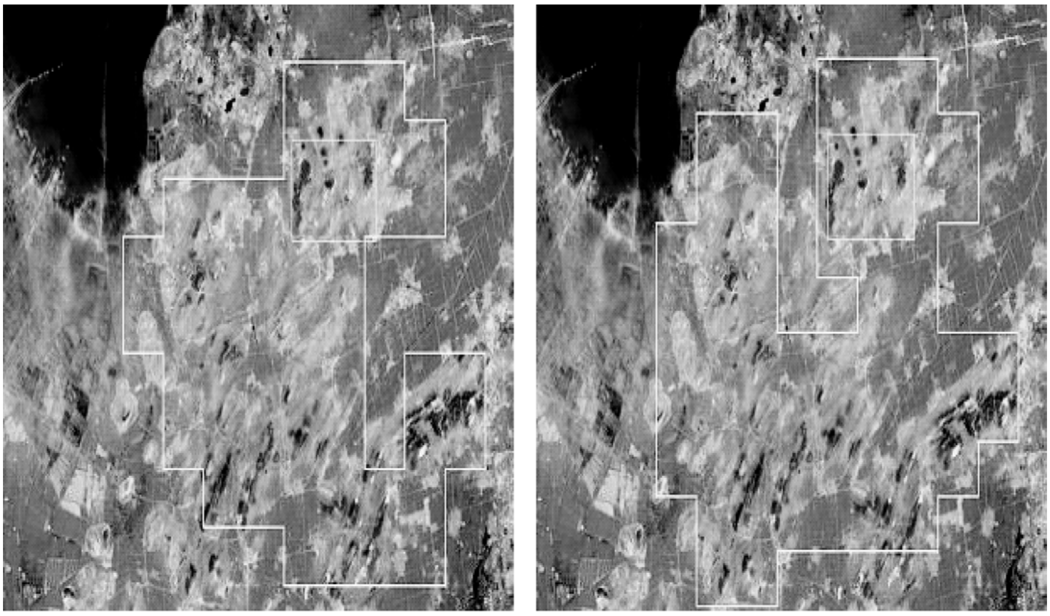
训练迭代次数以及所选择特征的统计数据。图 1—图4给出了部分检索结果，其中包含了多分类器Co-training 检索结果、融合颜色和纹理及相关反馈检索结果。图中的检索目标分别是土壤侵蚀区域、城镇居民地以及鱼塘(图像已经过缩放处理)。

从表 1 可以发现:

(1) 对于图1，其颜色得分总次数达到了 33(14+

表 1 多分类器 **Co-training** 检索迭代次数及所选特征统计 /次

图像检索过程	迭代次数	HSI 次数	Lab 次数	Glcm 次数	Gabor 次数
图 1	7	14	19	10	16
图 2	5	7	5	9	4
图 3	6	5	6	4	5
图 4	6	4	4	5	2

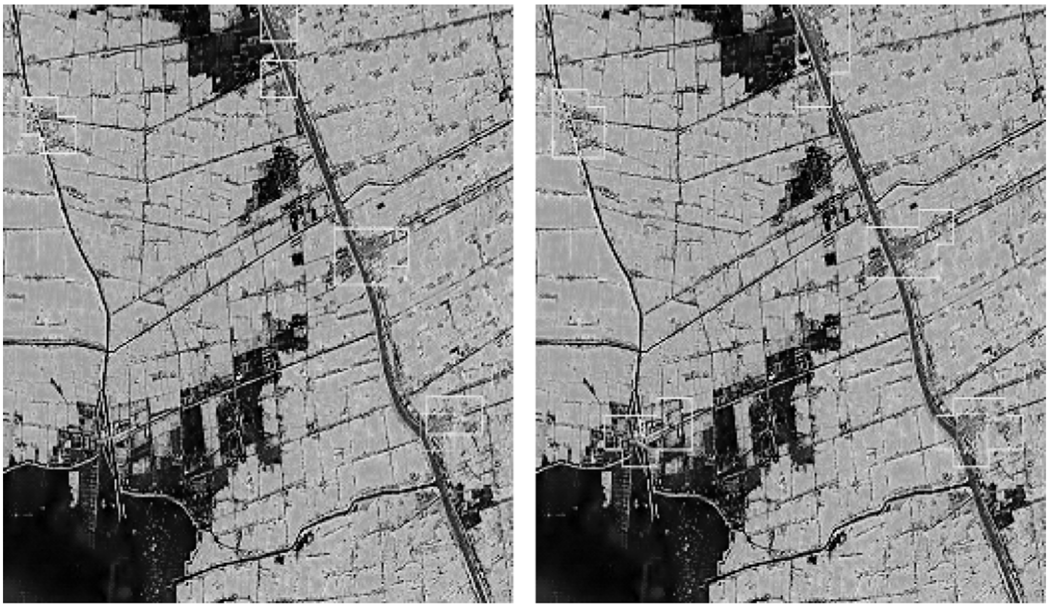


(a) (b)

图 1 Co-training 方法和融合颜色和纹理及相关反馈在土壤侵蚀范围检索结果比较

(图中右上角小矩形白线框为检索示例块，大范围的白线框为最后检索结果)

(a) Co-training 方法检索结果; (b) CTRF 方法检索结果

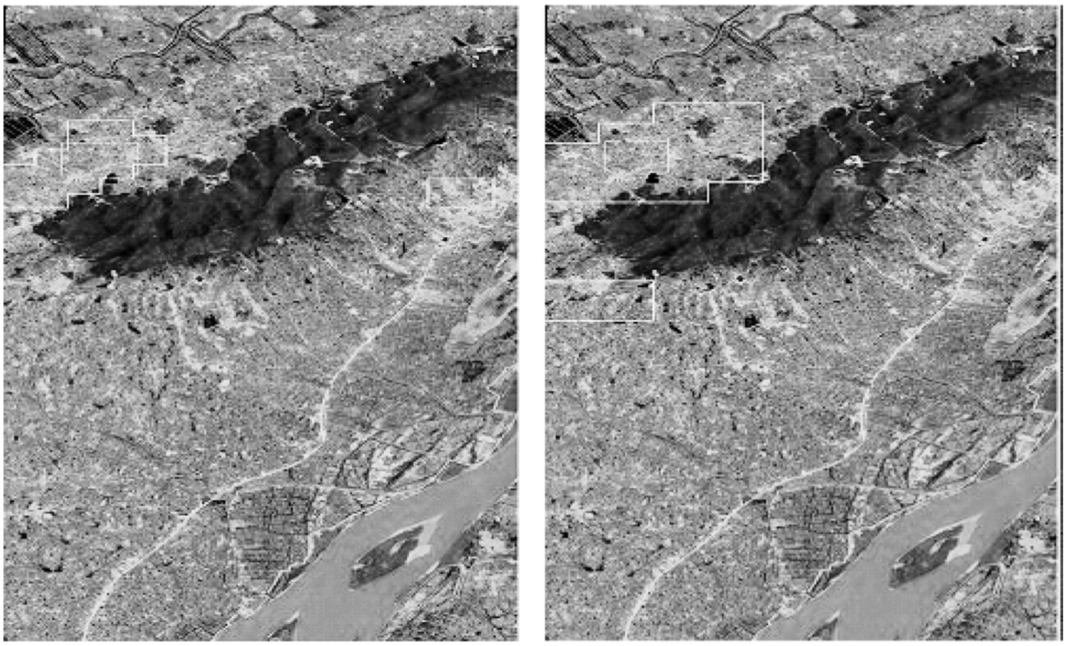


(a) (b)

图 2 Co-training 方法和融合颜色和纹理及相关反馈在城镇居民地检索结果比较

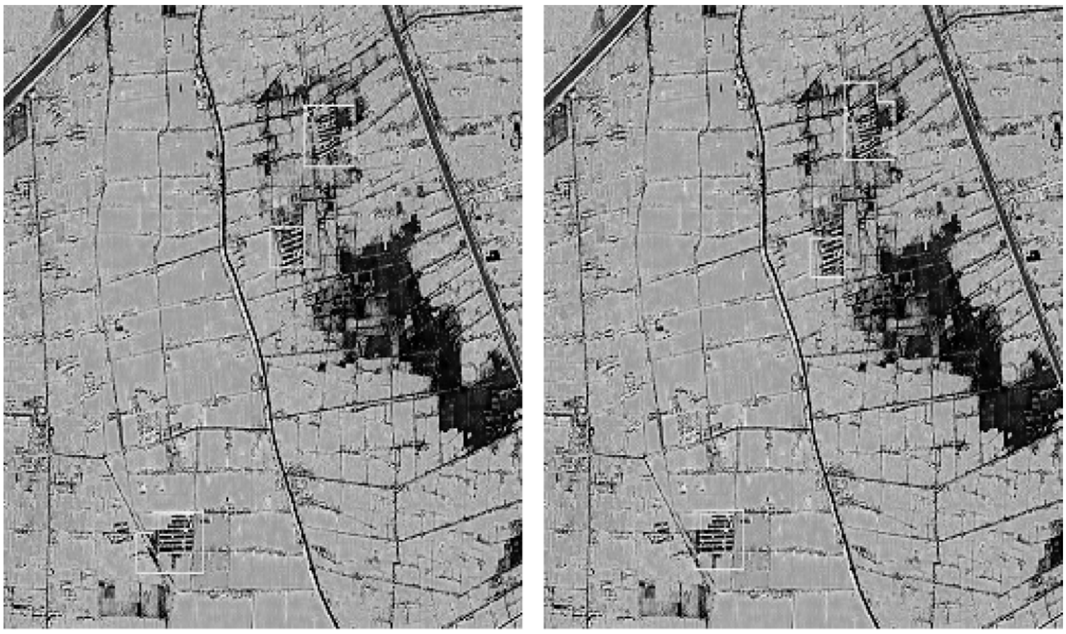
(图中右下角小矩形白线框为检索示例块，其余的白线框为检索结果)

(a) Co-training 方法检索结果; (b) CTRF 方法检索结果



(a) (b)

图 3 Co-training 方法和融合颜色和纹理及相关反馈在土壤侵蚀范围检索结果比较(南京浦口地区老山附近)
(图中左上角小矩形白线框为检索示例块，大范围的白线框为最后检索结果)
(a) Co-training 方法检索结果; (b) CTRF 方法检索结果



(a) (b)

图 4 Co-training 方法和融合颜色和纹理及相关反馈方法在苏北里下河某地区鱼塘检索结果比较
(图中左下角小矩形白线框为检索示例块，其余的白线框为检索结果)
(a) Co-training 方法检索结果; (b) CTRF 方法检索结果

19)，比纹理得分总次数(26)多了 7 次，而在其他的图像中，颜色得分总次数和纹理得分总次数基本持平，这表明该图中颜色信息的作用要强于纹理信息，而图 2 中颜色和纹理的作用基本相当，这与图像本身的视觉感知特征一致。

(2) 在整个迭代过程中，未知样本在 4 种特征中均有得分，即并没有出现其中一个赋分为 0 的情况，这表明特征之间存在互补作用，这也是 Cotraining 方法的优点：对于一个特定的样本，一个分类器可能不能将其正确地标记，但它可能被另一个分类器

正确地标记。

(3) 遥感图像检索中，不同场景地表覆盖具有不同特点，因此需要不同的特征。如图 1、图 2 和图 4 在纹理特征的选择上就有很大不同，图 2 在 Gabor 纹理特征上得分较多，7 轮迭代中有 16 次，而图 2 和图 4 在 Gabor 纹理特征上得分很少，分别只有 4 次和 2 次，这就说明适合图 1、图 2 和图 4 的纹理特征是不同的，本文利用多特征多分类器协同学习，能够自适应地选用不同的特征。

从图 1—图 4 可以看出，与基于颜色和纹理特征融合及相关反馈的遥感图像检索方法 CTRF(color and texture-based relevance feedback)相比，多分类器 Co-training 方法的检索效果比较令人满意。对于图 1 Co-training 方法的检索效果要略好于 CTRF 的效果，不难看出图 1(a)中的土壤侵蚀区域的纯度要大于图 1(b)中的。图 2 中 CTRF 方法检索出了多分类器 Co-training 方法没有检索到的左下角的一块城镇区域。图 3 中基于多分类器 Co-training 方法检索得到的土壤侵蚀区域范围要小于 CTRF 的，但其纯度较高。

其实这两种方法各有优点，CTRF 检索方法的最大优势是相关反馈，它通过人机交互，将用户引入学习过程，缩小了语义鸿沟，使得两幅当使用原始低层特征表达的图像不相似时，通过改变不同特征权重而拉近图像向量间的距离从而变得相似，如图 2 中左下角的那块城镇区域与用户初始训练样本在视觉上并不是非常一致，通过相关反馈技术可以将其检索出来。然而将用户引入学习也有一些弊端，最大的弊端就是人具有一定的主观性，不同的用户可能会对同一幅图像的理解不同，从而得到不同的相关反馈结果。并且，遥感图像相比于常规自然图像更加复杂多样，有时候用户很难判断一个待反馈样本是否为正例。而多分类器 Co-training 方法就不同，它的最大优势是不需要人为的反馈干预，充分利用了图像低层特征向量的固有性质，并且将不同特征间的信息进行互补。当然，它也有一些缺点，如在某些场合不如 CTRF 灵活。尽管如此，多分类器 Co-training 方法的检索效果还是令人满意的，并且自动化程度更高。

5 结论与讨论

本文针对基于内容的遥感图像检索中出现的小样本问题，提出了一种基于多特征多分类器 Cotraining 的遥感图像检索方法。通过与融合颜色和纹理及相关反馈遥感图像检索方法 CTRF 进行实验比较分析，结果表明，CTRF 方法与多分类器 Cotraining 方法各有优劣，在检索效果上基本相当，不过本文的基于多分类器 Co-training 方法不需要人为反馈干预，它充分利用了图像间低层特征向量的固有性质，并且将不同特征间的信息进行互补，避免了相关反馈过程中人工的多次反馈，自动化程度更高。

本文在检索过程中，将特征分为颜色和纹理两组，下一步研究将针对遥感图像特点，提取融合颜色和纹理特性的新的彩色图像特征，进一步提高检索的查全率和查准率。另外，不同图像分块对检索结果影响比较大。我们在研究中发现，如果图像块分得不好，会出现漏检索的现象，即有可能将一块相似目标区域分到不同子块中去，而目标区域在这些子块中占的比例较小，因此不能和示例图像块匹配，从而发生漏检。更好的分块策略也是我们今后需要解决的问题之一。

REFERENCES

Bao Q and Guo P. 2006. Comparative studies on similarity measures for remote sensing image retrieval based on histogram. *Journal of Remote Sensing*, **10**(6): 893—900

Blum A and Mitchell T. 1998. Combining labeled and unlabeled data with co-training. Proceedings of 11th annual Conference on Computational Learning Theory. New York, NY, United States: ACM Press

Didaci L and Roli F. 2006. Using co-training and self-training in semi-supervised multiple classifier systems. Proceedings of SSPR&SPR. Berlin: Springer Press

Dong A and Bhanu B. 2003. A new semi-supervised EM algorithm for image retrieval. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, WI, United states: IEEE Press

Ferecatu M and Boujemaa N. 2007. Interactive remote sensing image retrieval using active relevance feedback *IEEE Transactions on Geoscience and Remote Sensing*, **45**(4): 818—826

Gonzalez R C and Woods R E. 2007. Digital Image Processing(2nd edition). Beijing: Publishing House of Electronics Industry

Haralick R, Shanmugam K and Dinstein I. 1973. Textural features for image classification. *IEEE Transactions. System, Man & Cybernetics*, **3**(6): 610—621

Hoi S C ,Liu W and Chang S F. 2008. Semi-supervised distance metric learning for collaborative image retrieval. Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition. Madison, WI, United states: IEEE Press

Li D R and Ning X G. 2006. A new image decomposition method for content-based remote sensing image retrieval, *Transactions on Geomatics and Information Science of Wuhan University*, **31**(8): 659—662

Lu L Z, Liu R Y and Liu N. 2004. Remote sensing image retrieval using color and texture fused features. *Journal of Image and Graphics*, **9**(3):328—332

Luo R, Zhang Y S, Fan Y D and Deng X Q. 2001. Research on content-based remote sensing image retrieval: the strategy for visual feature selection, extraction, description and similarity measurement. Proceedings of 2001 International Conferences on Info-tech and Info-net. Beijing: IEEE Press

Manjunath B S and Ma W Y. 1996. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis and Machine Intelligence*,**18**(8): 837—842

Manjunath B S, Ohm J R, Vasudevan V V and Yamada A. 2001. Color and texture descriptors. *IEEE Trans. Circuits and Systems for Video Technology*, **11**(6): 703—715

Rui Y, Huang T S, Ortega M and Mehrota S. 1998. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions. on Circuits and Systems for Video Technology*, **8**(5): 644—655

Sharma A, Hua G, Liu Z and Zhang Z. 2008. Meta-tag propagation by co-training an ensemble classifier for improving image search relevance, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Anchorage, AK, United states: IEEE Press

Tian Q, Yu J, Xue Q and Sebe N. 2004. A new analysis of the value of unlabeled data in semi-supervised learning for image retrieval. Proceedings of the IEEE International Conference on Multimedia Exposition. IEEE Press

Yao J, Zhang Z, Antani S, Long R and Thoma G. 2006. Automatic medical image annotation and retrieval using SEMI-SECC. Proceedings of the IEEE International Conference on Mutilmedia and Exposition. Piscataway, NJ, United States: IEEE Press

Yu Z H. 2006. Image Categorization with semi-supervised learning. Proceedings of the IEEE International Conference on Image Processing, Atlanta, GA, United States: IEEE Press

Zeng Z M, Li F, Fu K and Ding C B. 2005. A method for extracting texture characters from large-scale remote sensing image. *Transactions on Geomatics and Information Science of Wuhan University*, **30**(12): 1080—1083

Zhou Y and Goldman S. 2004. Democratic co-learning. Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence. Boca Raton: IEEE Press

Zhou Z H. 2007. Co-training paradigm in semi-supervised learning. Zhou Z H and Wang J. Machine learning and its application, Beijing: Tsinghua University Press

Zhou Z H, Zhan D and Yang Q. 2007. Semi-supervised learning with very few labeled training examples. Proceeding. 22nd AAAI Conference on Artificial Intelligence. Menlo Park, CA, United States: American Association for Artificial Intelligence

Zhu B, Ramsey M and Hsinchun C. 2000. Creating a large-scale content-based airphoto image digitial library, *IEEE Transactions on Image Processing*, **9**(1): 163—167

附中文参考文献

包倩，郭平．2006．基于直方图的遥感图像相似性检索方法比较，遥感学报，**10**(6):893—900

李德仁，宁晓刚．2006．一种新的基于内容遥感图像检索的图像分块策略，武汉大学学报(信息科学版)，**31**(8):659—662

陆丽珍，刘仁义，刘南．2004．一种融合颜色和纹理特征的遥感图像检索方法，中国图象图形学报(A)，**9**(3):328—332

曾志明，李峰，傅琨，丁赤飚．2005．一种大尺寸遥感图像基于内容检索的纹理特征提取算法，武汉大学学报(信息科学版)，**30**(12): 1080—1083

周志华．2007．半监督学习中的协同训练风范．周志华，王珏，机器学习及其应用．北京：清华大学出版社