

---

# APLSSVM: Hybrid Entropy Models for Image Retrieval

Li Jun-yi<sup>1</sup>, Li Jian-hua<sup>1</sup>, Zhu Jin-hua<sup>2</sup>, Chen Xiao-hui<sup>3</sup>

<sup>1</sup>School of Electronic Information and Electrical engineering, Shanghai JiaoTong University, Shanghai, China

<sup>2</sup>College of Network Communication Zhejiang Yuexiu University of Foreign Languages, Zhe Jiang, China

<sup>3</sup>Information Engineering School, Yulin University, Yulin, Shanxi, China

## Email address:

leejy2006@163.com (Li Jun-yi)

## To cite this article:

Li Jun-yi, Li Jian-hua, Zhu Jin-hua, Chen Xiao-hui. APLSSVM: Hybrid Entropy Models for Image Retrieval. *International Journal of Intelligent Information Systems*. Special Issue: C Content-based Image Retrieval And Machine Learning. Vol. 4, No. 2-2, 2014, pp. 9-14. doi: 10.11648/j.ijis.s.2015040202.13

---

**Abstract:** Aiming at properties of remote sensing image data such as high-dimension, nonlinearity and massive unlabeled samples, a kind of probability least squares support vector machine (PLSSVM) classification method based on hybrid entropy and  $L_1$  norm was proposed. Firstly, hybrid entropy was designed by combining quasi-entropy with entropy difference, which was used to select the most “valuable” samples to be labeled from massive unlabeled sample set. Secondly, a  $L_1$  norm distance measuring was used to further select and remove outliers and redundant data from the sample set to be labeled. Finally, based on originally labeled samples and screened samples, PLSSVM was gained through training. Experimental results on classification of ROSIS hyperspectral remote sensing images show that the overall accuracy and Kappa coefficient of the proposed classification method reach 89.90% and 0.8685 respectively. The proposed method can obtain higher classification accuracy with few training samples, which is much applicable to classification problem of remote sensing images.

**Keywords:** Remote Sensing Image,  $L_1$  Norm, Active Learning, PLSSVM (Probability Least Squares Support Vector Machine), Hybrid Entropy

---

## 1. Introduction

Classification of remote sensing images means to make each pixel point region in the image belong to a category in several categories or one among several special elements. The classification results is to divide image space into several sub-regions, and each sub-region presents a practical land object<sup>[1-2]</sup>. In actual classification of remote sensing images, there are usually massive unlabeled samples, while the proportion of labeled samples is very small. Thus, it is very difficult to look for the information in need of labeling from these massive unlabeled samples. Besides, the cost used to label these samples is very high. Active learning algorithm is a new method for sample training. It is different from passive learning algorithm where samples are selected randomly<sup>[3-4]</sup>. In the process of machine learning, learners can actively choose the data most beneficial to improving properties of a classifier, automatically mark and add them in training samples for learning so as to effectively avoid excessive manual intervention and reduce the number of labeled samples.

The core of active learning algorithm is that which strategic

selection function is used to select the most “valuable” sample for labeling from unlabeled samples. Since the evaluation criteria for “value” are different, multiple active learning algorithms appear. Literature [4] selects the samples for labeling which current classifier cannot confirm the category mostly. Generally, this is called uncertain sampling. This method can fully select the samples beneficial to the classifier, and gain better results than random algorithm. But it has large randomness, so only sub-superior samples set can be picked out. In addition, outliers and redundant data may be easily chosen<sup>[7]</sup>. The introduction of quasi-entropy can reduce sampling randomness to some extent. Literature<sup>[8]</sup> proposes a heuristic active learning algorithm which selects the most possible misclassified samples based on committee. This algorithm chooses the most possible misclassified samples of current classifier during every sampling and eliminates the samples more than a half in the space so as to gain faster convergence speed than mainstream selection algorithm. Literature [9] randomly selects unlabeled samples from uncertain misclassified samples on the verification set for labeling. This algorithm owns better accuracy rate than standard algorithm. But, these algorithms still probably

choose outliers and redundant data, and calculation complexity is high. The introduction of entropy difference can help pick up misclassified samples more conveniently. In order to get more refined sample set, hybrid entropy is gained through fusing quasi-entropy and entropy difference. Since the algorithm may result in selecting outliers and redundant data,  $L_1$  norm distance measurement is used to choose these data and eliminate them.

This paper proposes an active learning algorithm based on hybrid entropy and  $L_1$  norm. This algorithm improves selection function from two aspects: 1) the most "valuable" samples are selected with hybrid entropy, and a rough sample set to be labeled is gained;  $L_1$  norm distance measurement is used to choose and eliminate possible outliers and redundant data; 2) remote sensing image data usually own such features as high dimension, nonlinearity and massive data, so support vector machine can be used to analyze and treat them. But traditional support vector machine classification method only takes into account of two extreme cases during deciding sample classification, i.e. the label for the sample belonging to the category is +1 and the label for the sample which does not belong to the category is -1. However, in practical application, due to the existence of uncertainty and influence of external factors, every sample has different division methods. Especially for some problems, due to sample randomness and fuzziness, they cannot be classified into a class explicitly, but can only be classified into a class according to certain probability or certain membership degree. So, it is improper to express class information only with  $\{-1, +1\}$  [10]. Thus, for the samples selected on the basis of active learning algorithm, PLSSVM is adopted as the classifier to classify and identify hyperspectral remote sensing images.

## 2. Plssvm

Aiming at classification inaccuracy and uncertainty of traditional support vector machine as well as defects of interference samples, Literature [10] designs PLSSVM to classify the samples which cannot be explicitly classified into a class according to certain probability. In this way, sample classification has qualitative interpretation and quantitative evaluation. Posterior probability of sample  $x$  belonging to each class is:

$$\begin{bmatrix} 1 & -p(1|2) & \cdots & -p(1|c) \\ -p(2|1) & 1 & \cdots & -p(2|c) \\ \vdots & \vdots & & \vdots \\ -p(c|1) & -p(c|2) & \cdots & 1 \end{bmatrix} \quad (1)$$

$$(p_1, p_2, \dots, p_c)^T = (0, 0, \dots, 0)^T,$$

Where,  $c$  is the number of classes;  $p(c|1)$  is posterior probability that sample  $x$  belongs to the  $c$ th class under the condition where sample  $x$  belongs to the first class. Similarly,  $p(1|c)$ ;  $p_m$  is posterior probability that sample  $x$  belongs to the  $m$ th class ( $m = 1, 2, \dots, c$ ).

Formula (1) can be regarded as  $c$  equation sets used to solve  $c$  unknown variables  $p_m$ . Through solving Formula (1), in output probability modeling of multi-classification problem, decision function of  $p_m$  of sample  $x$  in each class can be gained, i.e. take the class with the largest posterior probability as the sample. The class that  $x$  belongs to is as follows:

$$y(x) = \arg \max_{m=1,2,\dots,c} (p_m). \quad (2)$$

## 3. Active Learning Based on Hybrid Entropy and $L_1$ Norm

Labeled sample set  $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$  from unknown distribution and an unlabelled sample set  $U = \{x_{l+1}, x_{l+2}, \dots, x_n\}$  are given. Overall sample set is  $\mathcal{X} = L \cup U$ . There are  $c$  classes.  $x_i \in \mathbb{R}^d$  ( $i = 1, 2, \dots, n$ ;  $d$  refers to the number of dimensions of samples) and  $y_i \in \{1, 2, \dots, c\}$  is the label of sample  $x_i$ . The system adopts labeled sample set  $L$  as the training set to gain initial PLSSVM classifier, and actively selects some samples with large information quantity from unlabelled sample set  $U$  according to a strategy. Then, experts label them and add them in the training set. Thus, new PLSSVM classifier is obtained. After repeated cycles, classification results will finally reach the threshold value of an evaluation index or specified cycle times.

A. Sample selection strategy based on hybrid entropy

The classifier may easily make mistakes during judging the most uncertain sample classification, thus leading to low classification accuracy rate. Therefore, uncertainty is an important factor that experts should consider when selecting the samples to be labeled. Sample uncertainty algorithms can be based on Shannon entropy, posterior probability and the nearest boundary etc. The algorithm based on Shannon entropy has gained good results in many applications, but it cannot select the optimal samples so that calculation complexity is high during training the set. Thus, optimization selection standard (i.e. quasi-entropy with high quality factor) is needed to measure sample uncertainty. Literature [11] points out that quality factor of  $-p^a$  ( $0 < a < 1$ ) convex function is higher than that of  $p \log p$ . If the quality factor is larger, quasi-entropy is more sensitive to probability distribution evenness near the minimum value, and the shape of minimum value of quasi-entropy is shaper. So, quasi-entropy surpasses Shannon entropy in terms of significance index of minimum value. Therefore, quasi-entropy with high quality factor replaces Shannon entropy. Assuming posterior probabilities that sample  $x_i$

belongs to every class are  $p_1, p_2, \dots, p_c$ , and  $\sum_{m=1}^c p_{im} = 1$  is met, uncertainty measure of sample  $x_i$  can be expressed as

$$\lambda_i = \sum_{m=1}^c f(p_{im}) \quad (3)$$

Where,  $f(p_{im}) = -p_{im}^a; \lambda_i$  has the following properties:

Property 1: when posterior probability distribution is most even (i.e. all  $p_m$  are equal),  $\lambda_i$  is the minimum and equal to  $cf(1/c)$ . This is also the situation where uncertainty is the largest.

It can be known from Property 1 that when posterior probabilities  $p_m$  that sample  $x_i$  belongs to every class are equal, sample uncertainty is the largest, and the value of quasi-entropy  $\lambda_i$  is the smallest. So, quasi-entropy can be used to figure out uncertainty measurement value of each sample. If quasi-entropy value of samples is smaller, the information quantity is larger.

In information entropy, the samples which may be easily misclassified can be expressed with the absolute value of differences of two absolute values:

$$\begin{aligned} d_i &= |H(p_{\max}) - H(p_{\sec})|, \\ H(p_{\max}) &= -p_{\max} \log p_{\max}, \\ H(p_{\sec}) &= -p_{\sec} \log p_{\sec}, \end{aligned} \quad (4)$$

Where,  $p_{\max}$  is the maximum posterior probability that sample  $x_i$  belongs to every class;  $p_{\sec}$  is the second largest posterior probability that sample  $x_i$  belongs to every class.

Entropy difference distance metric function of density functions  $p_{\max}$  and  $p_{\sec}$  of the two posterior probabilities have the following characteristic [12]

$$\|p_{\max} - p_{\sec}\|_{L_1} = \sum |p_{\max} - p_{\sec}| \leq \frac{1}{2} \quad (5)$$

Where,  $\|p_{\max} - p_{\sec}\|_{L_1}$  is standard Minkowski L1 norm distance measurement, then

$$|H(p_{\max}) - H(p_{\sec})| \leq \|p_{\max} - p_{\sec}\|_{L_1} \quad (6)$$

This characteristic shows retrieval results of Entropy difference distance metric is included in retrieval results of L1 norm distance measurement, and the retrieval range narrows.

It can be seen from Formula (4), when posterior probability of samples changes slightly, and the change in entropy value will also be small. When entropy difference value is smaller, the possibility that sample  $x_i$  belongs to some two classes is close, i.e. this sample may be misclassified most easily, and the information quantity is also the largest.

According to analysis of quasi-entropy and entropy difference, the following conclusions can be drawn: if quasi-entropy value is smaller, sample uncertainty is larger; entropy difference value is smaller, the sample may be misclassified more easily. If the values of quasi-entropy and entropy difference are smaller, information quantity is larger and there are larger impacts in classification effects. In massive data sets, the sample size selected purely by quasi-entropy or entropy difference strategy is also large. In order to pick out more refined samples and reduce labeling cost, quasi-entropy and entropy difference are fused to gain a

new sample selection measurement strategy - hybrid entropy.

$$u_i = \lambda_i d_i. \quad (7)$$

M samples with the highest information quantity are worked out according to Formula (7), i.e. M samples with the smallest  $u_i$  value.

#### B. Sample similarity measurement based on L1 norm

The samples selected by hybrid entropy may have outliers and redundant data. These data make little contributions to classification accuracy of the classifier and even will affect its classification accuracy. Therefore, L1 norm distance measurement will be adopted to work out similarity among samples. Outliers and redundant data will be removed according to similarity value.

Literature [13] adopts L1 norm, L2 norm and quadric expression to compare data retrieval properties. The testing results show these distance measurement methods differ little in retrieval property. L1 norm distance measurement is more robust than L2 norm distance measurement, and L1 norm distance measurement is the most simplest in calculation. So, L1 norm distance measurement is adopted to calculate similarity among samples to be labeled.

$$\begin{aligned} s_{hj} &= \sum_{k=1}^d |x_{hk} - x_{jk}|, \\ (h, j &= 1, 2, \dots, v), \end{aligned} \quad (8)$$

Where,  $x_{hk}$  and  $x_{jk}$  are the kth attribute in the hth and jth samples; v is the number of samples.

Assuming mean space distance of samples of the same class is  $\theta$ ,  $a = \theta/100$ ,  $\beta = \theta$ . If  $s_{hj} < a$ , sample  $x_j$  is judged to be redundant information and eliminated; if  $\min s_{hj} > \beta$ , sample  $x_j$  is judged to be an outlier and deleted. Then, the remaining samples are selected and submitted to experts for labeling. This deletes outliers, eliminates redundant data, further narrows scale of sample set to be labeled and reduces cost of manual labeling.

## 4. Algorithm Steps

Input: labeled sample set is expressed as L and unlabelled sample set is expressed as U; the number of samples is expressed as M; ending condition is expressed as S; the parameter is expressed as a.

*Algorithm process:*

- 1) Train classifier PLSSVM with labeled sample set;
- 2) Carry out a~g repeatedly until ending condition S is met;
  - a) Posterior probability that unlabeled sample set U belongs to each class is calculated with classifier PLSSVM;
  - b) Calculate quasi-entropy  $\lambda_i$  and entropy difference  $d_i$  of unlabeled samples according to posterior probability gained, Formula (3) and (4);
  - c) Calculate hybrid entropy  $u_i$  according to Formula (7);

- d) Select  $m$  samples with the smallest  $u_i$  value and add them in the sample set to be labeled;
- e) Calculate similarity of  $M$  samples according to Formula (8), eliminate the samples meeting  $s_{hj} < \alpha$  and  $\min s_{hj} > \beta$ , and make the remaining samples form new sample subset A;
- f) Submit A to experts for labeling and add labeled samples in L;
- g) PLSSVM. Utilize L to train classifier PLSSVM again.

Output: train sample set L finally labeled and gain classifier PLSSVM.

## 5. Experiment and Analysis

### A. ROSIS hyperspectral experimental data

ROSIIS hyperspectral experimental data come from Literature [14]. Spectral region is 0.43~0.86  $\mu\text{m}$ , with 610×340 pixel, 103 wave bands and 1.3 spatial resolution. Besides, training region and testing region actually measured synchronously are provided. The training samples include 9 classes of land objects: bituminous pavement (548 pixel), tree (524 pixel), brick (514 pixel), shadow (231 pixel), pitch roof (375 pixel), bare land (532 pixel), metal plate (265 pixel), grit (392 pixel) and grassland (540 pixel). Testing samples include 9 classes of land objects: bituminous pavement (6592 pixel), tree (3064 pixel), brick (3682 pixel), shadow (942 pixel), pitch roof (1330 pixel), bare land (5029 pixel), metal plate (1345 pixel), grit (2099 pixel) and grassland (18675 pixel). ENVI4.7 software is utilized to transform original data corresponding to the regions ROSIS hyperspectral image training sample and testing sample are interested in to ASCII data so as to process data in Matlab 7.8 environment.

### B. Calculation results of remote sensing image and analysis of results

Active learning algorithm is adopted to select training samples for the classifier and to construct two types of APLSSVM, expressed as APLSSVM1 and APLSSVM2 in this paper. In the experiment process, parameter setting is as follows: kernel function of PLSSVM adopts polynomial kernel function; the optimal values of penalty parameter  $C$  and kernel parameter  $\gamma$  are confirmed with cross validation method,  $\alpha=0.6$  and  $M=100$ .

- 1) Based on the same initial sample set, change the number of newly-added training samples, evaluate effects of the number of newly-added training samples on classification accuracy of two type of APLSSVM; the ending condition S is that the difference between adjacent two classification accuracies is less than 0.002 or the number of iteration times reaches 15. This

indicates high classification accuracy can be gained when PLSSVM is used to process remote sensing images; when the number of newly-added training samples is less than 300, classification accuracy of APLSSVM1 boots rapidly with the rise in the number of labeled samples; when the number of newly-added training samples exceeds 300, classification accuracy of APLSSVM1 basically tends to be stable and maintains about 90% with the rise in the number of labeled samples; for APLSSVM2, its classification accuracy increases slowly with the rise in the number of labeled samples; to reach the same classification accuracy with APLSSVM1, APLSSVM2 needs more labeled samples, which will consumes more time and energy of experts. So, the cost is expensive.

- 2) In the experiment, given training samples serve as the initial sample set. Under the condition where the number of newly-added training samples is the same, classification effects of two APLSSVM classifiers and passive PLSSVM classifier are compared. APLSSVM1 and APLSSVM2 selects newly-added training samples for labeling through iteration of active learning algorithm; passive PLSSVM directly selects samples of the same number as newly-added samples for training. The number of training samples the three classifiers select is: original sample set + 300 newly-added samples. The ending condition S is that the number of iterations reaches 3. Table 1, Table 2 and Table 3 are confusion matrix and Kappa coefficient corresponding to each figure.

It can be seen that APLSSVM2 and passive PLSSVM classify most grassland into bare land, and the misclassification phenomenon is serious; APLSSVM1 performs relatively well in this aspect and can well classify the two types of land objects; misclassification accuracy of other land objects approaches for the three classifiers.

The following can be gained according to Table 1-3:

User's accuracy: among all kinds of land objects, user's accuracy differs mostly for bare land. User's accuracy of APLSSVM1 is 80.04%, up over 30% compared with user's accuracy of APLSSVM2 and passive PLSSVM. According to confusion matrix in Table 2 and Table 3, APLSSVM2 and passive PLSSVM misclassify most grassland into bare land. Thus, the proportion of grassland in bare land samples exceeds a half. For pitch roof, the largest user's accuracy of APLSSVM2 is 83.90%, followed by APLSSVM1 (70.29%), and passive PLSSVM has the smallest user's accuracy (65.58%). For the three classifiers, user's accuracy differs little among other land objects.

Table 1. Confusion matrix obtained by APLSSVM 1

	Bituminous pavement	Tree	Brick	Shadow	Pitch roof	Bare land	Metal plate	Grit	Grassland	User's accuracy/%
Bituminous pavement	5416	5	166	0	115	9	0	25	11	94.24
Tree	0	2747	3	0	0	13	0	0	465	85.10
Brick	273	0	3196	0	11	50	5	379	41	80.81
Shadow	26	1	0	942	0	0	2	0	0	97.01

	Bituminous pavement	Tree	Brick	Shadow	Pitch roof	Bare land	Metal plate	Grit	Grassland	User's accuracy/%
Pitch roof	418	0	36	0	1201	2	53	5	0	70.29
Bare land	23	201	0	0	0	4799	0	0	973	80.04
Metal plate	0	2	0	0	0	35	1281	0	0	97.19
Grit	405	0	251	0	3	0	2	1687	0	71.85
Grassland	31	94	28	0	0	121	2	3	16901	98.38
Producer's accuracy/%	82.16	90.07	86.85	100	90.30	95.43	95.24	80.37	91.90	Overall accuracy=89.90% Kappa=0.8685

Table 2. Confusion matrix obtained by APLSSVM 2

	Bituminous pavement	Tree	Brick	Shadow	Pitch roof	Bare land	Metal plate	Grit	Grassland	User's accuracy/%
Bituminous pavement	5717	0	240	0	129	4	0	31	13	93.20
Tree	0	2889	0	0	0	12	0	0	703	80.16
Brick	135	0	3172	0	3	18	0	380	30	84.86
Shadow	27	0	0	942	0	0	1	0	0	97.11
Pitch roof	209	0	17	0	1193	0	0	3	0	83.90
Bare land	10	59	3	0	0	4958	1	0	5115	48.87
Metal plate	0	1	0	0	0	0	1287	0	0	99.92
Grit	345	0	206	0	1	0	0	1675	0	75.21
Grassland	19	76	24	0	0	9	1	4	12796	98.97
Producer's accuracy/%	88.47	95.50	86.62	100	89.97	99.14	99.77	80.03	68.59	Overall accuracy=81.56% Kappa=0.7691

Table 3. Confusion matrix obtained by passive PLSSVM

	Bituminous pavement	Tree	Brick	Shadow	Pitch roof	Bare land	Metal plate	Grit	Grassland	User's accuracy/%
Bituminous pavement	5341	5	146	1	111	5	0	24	21	94.46
Tree	0	2805	2	0	0	15	0	1	981	73.74
Brick	284	0	3174	0	9	51	9	373	79	79.77
Shadow	3	1	0	841	0	0	2	0	100	88.81
Pitch roof	411	0	24	0	1107	4	51	3	88	65.58
Bare land	19	84	2	0	0	4801	0	0	5253	47.26
Metal plate	0	2	0	0	0	35	1181	0	98	89.74
Grit	402	0	208	0	3	0	1	1596	78	69.76
Grassland	32	67	26	0	0	18	1	2	12477	98.84
Producer's accuracy/%	82.27	94.63	88.61	99.88	90	97.40	94.86	79.84	65.07	Overall accuracy=78.48% Kappa=0.7305

Producer's accuracy: producer's accuracy of grassland differs most greatly. Producer's accuracy of APLSSVM1 is 91.90%, up over 20% compared with producer's accuracy of APLSSVM2 and passive PLSSVM. According to confusion matrixes in Table 2 and Table 3, nearly 1/3 grassland samples are misclassified into bare land. Producer's accuracy of other land objects approaches for the three classifiers. Overall accuracy and Kappa coefficient: since overall accuracy takes into account of corresponding weight relationship of each class, it is relatively objective; since Kappa coefficient considers the prelatship between user's accuracy and producer's accuracy, it has become classification accuracy evaluation index of remote sensing images together with overall accuracy. Based on analysis of Table 1-3, overall

accuracy and Kappa coefficient of APLSSVM1 are the highest, followed by APLSSVM2. Passive PLSSVM performs most poorly.

Experiment results show, APLSSVM1 over considers sample uncertainty and samples which may be easily misclassified, and eliminates outliers and redundant data from samples to be selected. Finally, more refined training sample set is gained. Therefore, under the same number of training samples, APLSSVM1 has higher classification accuracy than other classifiers.

## 6. Conclusions

a) Hybrid entropy gained through fusing quasi-entropy and

entropy difference can measure sample uncertainty and avoid sample misclassification. Sample selection strategy based on hybrid entropy can choose more refined samples and reduce the cost of manual labeling.

- b) Sample similarity measurement method based on  $L_1$  norm can screen out outliers and redundant data, which further reduces the scale of sample set to be labeled and cost of manual labeling.
- c) Compared with heuristic active learning algorithm which selects the most possible misclassified samples based on committee, active learning algorithm based on hybrid entropy and  $L_1$  norm can pick out more valuable samples to be labeled and gain high classification accuracy with few training samples.
- d) PLSSVM owns both qualitative explanation and quantitative evaluation during classifying uncertain samples, suitable for classifying remote sensing image data.
- e) For remote sensing image data with massive unlabelled samples, active learning can help find out the most valuable information from massive unlabeled samples. Compared with passive PLSSVM which selects samples randomly, APLSSVM owns higher classification accuracy.

---

## References

- [1] WANG Yuan-yuan, CHEN Yun-hao, LI Jing. Application of model tree and support vector regression in the hyperspectral remote sensing [J]. *Journal of China University of Mining & Technology*, 2006.35(6):818-823.
- [2] CHEN Shao-jie, LI Guang-li, ZHANG Wei, et al. Land use classification in coal mining area using remote sensing images based on multiple classifier combination [J]. *Journal of China University of Mining & Technology*, 2011, 40(2):273-278.
- [3] HAMANAKA Y, SHINODA K, TSUTAOKA T, et al. Committee-based active learning for speech recognition [J]. *IEICE Transactions on Information and Systems*, 2011, 94(10):2015-2023.
- [4] ZHANG L J, CHEN C, BU J J, et al. Active learning based on locally linear reconstruction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(10): 2026-2038.
- [5] SUN Z C, LIU Z G, LIU S H, et al. Active learning with support vector machines in remotely sensed image classification[C]//QIU P H, YIU C, ZHANG H, et al. *Proceedings of the 2nd International Congress on Image and Signal Processing*. Piscataway: IEEE Computer Society, 2009: 1-6.
- [6] TUIA D, RATLE F, PACIFICI F, et al. Active learning methods for remote sensing image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(7): 2218-2232.
- [7] CHEN Y, HE Z. Blind separation using a class of new independence measures[C]//IEEE. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. Piscataway: IEEE Signal Process, 2003: 309-312.
- [8] LONG J, YIN J P, ZHU E. An active learning method based on most possible misclassification sampling using committee [J]. *Lecture Notes in Computer Science*. 2007. 4617: 104-113.
- [9] BRUZZONE L, PERSELLO C. Active learning for classification of remote sensing images[C]//IEEE. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*. Piscataway: IEEE In-corporated, 2009: 693-696.
- [10] GAO Y, WANG X S, CHENG Y H, et al. Fault diagnosis using a probability least squares support vector classification machine [J]. *Mining Science and Technology*, 2010, 20(6) : 917-921.
- [11] CHEN Yang. Properties of quasi-entropy and their application [J]. *Journal of Southeast University: Natural Science Edition*, 2006, 36(2): 221-225.
- [12] COX I J, MILLER M L, MINKA T P, et al. The bayesian image retrieval system, PieHunter: theory, implementation, and psychophysical experiments [J]. *IEEE Transactions on Image Processing*, 2000, 9(1): 20-37.
- [13] JOHN R S. *Integrated spatial and feature image systems: retrieval, analysis, and compression* [D]. New York: Columbia University, 1997.
- [14] WEI L, SAURABH P, JAMES E F, et al. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012, 50(5): 1185-1198.