Sparse and Stratified Sampling Method for One-Class Classification of Remote Sensing Images

Xin Pan[1*], Hongbin Sun [1] and Yuan Yue [2]

[1] *School of Computer & Information Technology, Changchun Institute of Technology, Changchun 130012, China*
[1] *Otago Business School, University of Otago, Dunedin 9054, New Zealand*

Abstract

One-class classification for model generation using only one sample category plays an important role in particular object recognition, missing category treatment and prediction of geographic event distributions. However, traditional one-class classification algorithm is difficult to achieve high accuracy in remote sensing because of existing larger amounts of spatial bands in remote sensing image as well as uncertainty or inconsistency information in relatively few training samples. To solve these problems, this study proposed a sparse and stratified sampling method for one-class classification of remote sensing images (SSSOCC). Spatial features were first divided into multiple groups which have different sample subspaces through sparse description and stratified sampling. Each group trained a one-class support vector machine as a sub-classifier, and then made joint decisions according to multiple sub-classifiers. Experiments show that our method's classification accuracy and stability are better than those of traditional one-class support vector machine. Thus, this study could meet the demands for one category land use or cover information extraction from remote sensing images. It could also solve problems such as larger amounts of spatial bands, uncertainty or inconsistency information in relatively few training samples.

Keywords: One-class classification, Ensemble, Sparse description, Stratified sampling, Remote Sensing

---

1. Introduction

Supervised classification algorithm for image classification in remote sensing can automatically obtain land use/covering information from remote-sensing images timely and accurately using a classification model obtained by training samples [1]. Traditional methods usually require inputted samples with complete category structure and containing all categories of the entire image [2]. However, the following situations are encountered in practice: users are only interested in a particular type of land coverage; data of several categories of sample may be lacking or difficult to collect; an image contains too much categories and they cannot be all listed; data contain only a category of information (such as location of a specific geographic incident). Traditional methods cannot adapt to the above situations, thus cannot conduct effective classification [3,4]. By contrast, one-class classifier (OCC) can distinguish objects by studying only one category of samples [5].

2. State of the Art

At present, OCC mainly uses probability density, boundary, super-sphere , distance, and some other technologies, which

have attracted attention from various fields. [6] distinguished oil spills and non-oil spills area in the synthetic aperture radar image with OCC. [7] separated specific types of wetlands with OCC. [8] analyzed the function of several types of OCCs in the prediction geographic spatial events. [9] solved the problem too much categories in the changes detection of high-resolution image with OCSVM. [10] proposed an interactive method to improve the quality of one-class remote-sensing samples. [11] presented a one-class classification method based on maximum information entropy method. [12, 13] improved the accuracy of one-class classification by introducing unlabeled samples. Although the above studies have achieved good results, following problems are still encountered in remote sensing image classification:

Most one-class algorithms are derived from outlier detection technology, which is very sensitive to sparse data, and the increase in dimension will directly cause sparsification of the data, and higher dimension produces more sparse data [14]. One-class algorithm needs more samples to determine the boundary of the category because of absence of an opposition category [15]. The accuracy is lower when samples are insufficient. Remote sensing images include more spatial bands or spatial feathers (such as texture information of gray-level co-occurrence matrix), and the type boundaries are generally not definite. At the same time, the proportion of sample quantity in the entire image is relatively low; some single sample is next to nothing (such as, the place of rare wild animals showing up). All of these problems greatly affect the accuracy of one-class algorithm,

---

* E-mail address: 101103991@qq.com

thereby limiting their function for image classification in remote sensing.

The remainder of this paper is organized as follows. Section 3 proposed the stratified sampling algorithm and construction process of remote sensing image one-class classifier. Section 4 analyzes the method and compares the classification result with traditional OCSVM. Section 5 summarizes the conclusions.

3. Methodology

One-class classifier is different from the traditional dual-class and multi-class classifier: It can obtain the model based on probability density, boundary, super sphere or distance by studying the sample with only one category; this model can judge on an input data and get the conclusion that whether the data belong to the input category.

3.1 Sample and feather sparseness description

The sparseness of the sample determines difficulty of the one-class classification, so feather sparseness can be used as an important index to distinguish the classification ability of features on the behavior of one-class classifier [12]. The sample of remote sensing images can be described with the triple $<x, y, s>$, where, x is the spatial feather of the sample $x=\{b_1, b_2,..., b_{FN}\}$ contains FN spatial properties. y is category label $y=\{1, -1\}$, there, 1 means that it belongs to a certain category, while -1 means the opposite; s is class label and $s=\{1,0\}$, there, 1 means there is class label, while 0 means the opposite. All data for one-class classification problems when $s=1$ contain only samples when $y=1$, while samples when $s=0$, the situation can be $y=1$ or $y=-1$. For a sample set, although for samples without category labels, their category may be 1 or -1, they can still show the rule of sparseness with labeled samples [16, 17]. For the set of N samples containing M labeled samples, a certain spatial feature can be divided into K subsections in the way of equal depth (with same numbers), each subsection contains N/K samples; the subsection of i contains NS labeled samples and NN unlabeled sample, $NS+NN=N/K$, and the probability of containing labeled samples is:

$$p(s = 1 \mid fraction = i) = \frac{NS}{N/K} \tag{1}$$

The sparseness of this attribute containing labeled sample can be described as:

$$\varphi = \sum_{i=1}^{K} (1 - p(s = 1 \mid fraction = i))^2 \tag{2}$$

The value of this formula is a non-negative number, the lower the result is, the sparser the labeled samples distribute, and vice versa. With this formula, we can describe and distinguish effects of spatial features on the single-class distinction.

3.2 Division of stratified samples

Random forest or ensemble classifier algorithm improve its classification precision and prevent over fitting through the co-decision of multi sub-classifiers, one of key principles of this classifier is that there is sufficient differences between sub-classifiers. Stratified sampling is a method for extracting samples from each group after grouping samples, studies show that this method can better ensure the differences

between sub-spaces of samples than simple random sampling [19]. By using this method, the features of remote sensing image can be divided into several subsets, and sufficient differences between subsets can be maintained. An feature set $A=\{A_1, A_2,..., A_N\}$ describes a sample space, the characteristic φ_i of every A_i can be calculated with Formula (2), the normalization of the attribute can be described as:

$$\theta_i = \frac{\varphi_i}{\sum_{k=1}^{N} \varphi_i} \tag{3}$$

With the normalization formula, the attribute description can be mapped to 0 and 1, and all attributes are grouped according to Formula (3):

(1) All attributes are ranked in descending order according to Formula (3).

(2) A is divided into two groups: A_s and A_w according to the specified threshold value α. A_s has a higher sparseness, while A_w has a lower one, $A = A_S \cup A_W$ and $A_S \cap A_W = \phi$.

At this point, the attribute space is divided into two groups with different characteristics: A_s and A_w, the stratified sub-space is divided into T groups, NUM spatial attributes of each group can be calculated with the following algorithms:

Spatial partitioning algorithms of stratified samples
(A_S, A_w, T, NUM)

Begin
For $i=1:T$
    $N1=Random(NUM-1)+1;$
    $A1= RandomSelect\ (As, N1)$; Selecting N1 attributes
    from As randomly
    $A2= RandomSelect\ (Aw, Num-N1)$; Selecting the rest
    attributes from Aw
    Partition[i]= $A_1 \cup A_2$;
End
return Partition;
End

The method can partition A into T groups of attribute subspaces, and each contains NUM spatial attributes. At least one of them comes from the group with higher sparseness, the rest part come from the lower one. We can obtain several groups of sub spaces with great differences between groups, yet less inter-group features

3.3 One-class support vector machine

The training objective of OCSVM is to find a smallest super sphere containing all the samples for the given sample l $<x_i, y=1, s=1>(i=1,2,...,l)$, the calculation formula is as follows:

$$Min \sum_{i,j} a_i a_j (x_i \cdot x_j) - \sum_i a_i (x_i \cdot x_i) \tag{4}$$

$$Subject\ to: 0 \le a_i \le \frac{1}{vl}, \sum_{i=1}^{l} a_i = 1$$

where, $v \in [0, 1]$ is used to adjust the volume of the sphere and the number of sample, which are not included in the sphere. With this optimized formula, a spatial super sphere constituted by a_i can be obtained, the formula for

judging whether an unknown sample is within the sphere is as follows:

$$(x \cdot x) - 2\sum_i a_i(x \cdot x_i) + \sum_{i,j} a_i a_j(x_i \cdot x_j) \leq R^2 \qquad (5)$$

The inner product $(x_i.x_j)$ can be replaced by the kernel function $K(x_i, x_j)$. With OCSVM, we can study one-class samples and obtain the classification model of the spatial super-sphere. In the classification prediction, a sample belongs to a specific category if the sample is inside the sphere.

3.4 Construction process of SSSOCC for image classification in remote sensing

The construction process of SSSOCC for remote sensing image classification is shown as Figure 1:
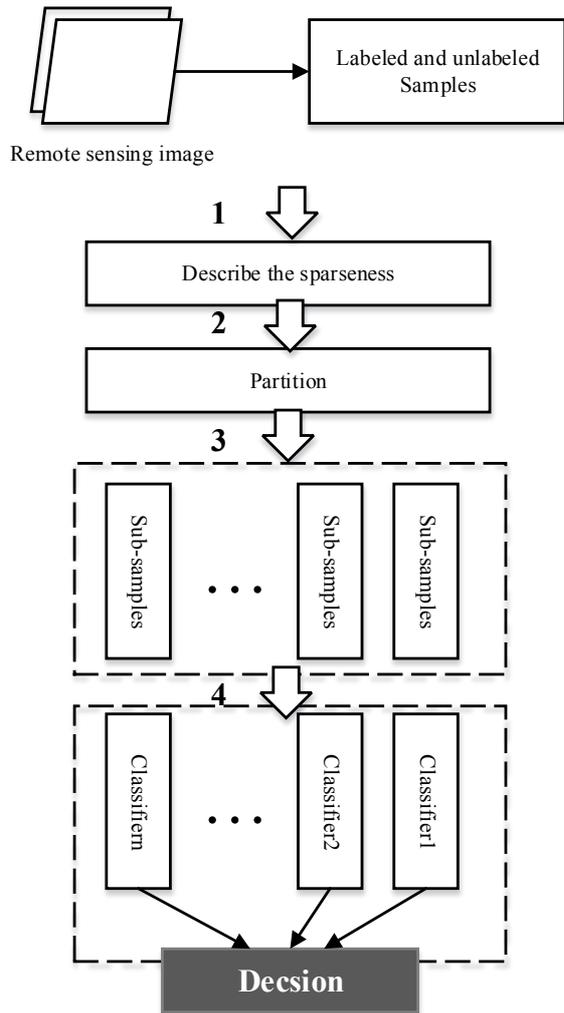
Fig. 1. Construction process of SSSOCC classifier

The construction process of the algorithm is as follows:
(1) For a remote-sensing image containing N features, M one-class training samples are introduced. Each feature is specified to be divided into K sections with the same depth, (K−1)*M unlabeled samples are drawn randomly from this image, and the sparseness of each feature can be calculated using Formula 2.
(2) A normalized description of the sparseness is obtained using Formula 3 and the band is ranked and specified in a descending order:

$$\alpha = \frac{\sum_{i=1}^{N} \theta_i}{N} \qquad (6)$$

All the spatial attributes are divided into two groups: A_s and A_w. The number of spatial attributes NUM of each subattribute space is specified, and the results obtained after stratified sampling is partitioned with the spatial portioning algorithm of the stratified samples.
(3) For each set of features in the partition, the corresponding features of the training sample are selected to construct subsamples; the subsamples of each group train an OCSVM as a subclassifier using Formula 4.
(4) All the subclassifiers are combined as an ensemble classifier; each OCSVM subclassifier can make judgments using Formula 5. If Formula 5 is satisfied and the output result is 1 or −1, then the final output of the ensemble classifier is as follows:

$$Sr = \sum_{i=1}^{T} result_i \qquad (7)$$

$$Output = \begin{cases} 1 & Sr \geq 0 \\ -1 & Sr < 0 \end{cases} \qquad (8)$$

If the output of the ensemble classifier is 1, then the input data belong to the same category with the sample data; if the output is −1, then the opposite is true. According to the above steps, SSSOCC classifier can be constructed, which contains the T subclassifiers. Each subclassifier has corresponding *NUM* spatial features; as such, the spatial features of the subclassifier are reduced. Meanwhile, the accuracy and generalization ability are improved because of the integrated decisions of multiple classifiers.

4. Analysis of Results and Discussion

This paper selects Landsat-8 as test data. The image was obtained on August 9, 2013 in the northern part of Xiang Hai National Reserves in Jilin Province. The image size was truncated to 1024×1024, has 11 spatial bands in total, and a spatial resolution of 30m×30m. The land of this region is mainly covered by waters, saline and alkaline lands, farmlands, and grasslands, as shown in Figure 2:
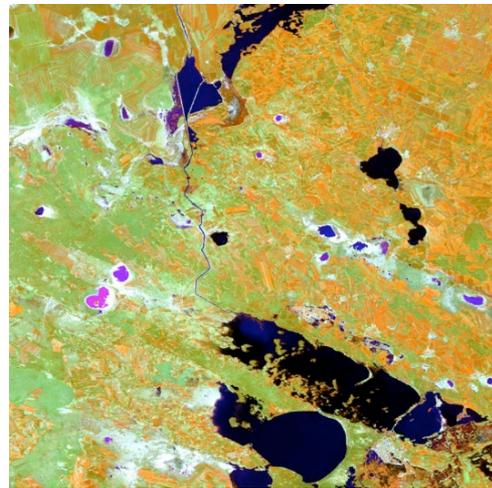

Fig. 2. Image of the study area (composite of bands 5, 6, 4)

This research uses Matlab to implement all algorithms. Grassland is chosen as the test object of one-class classification because of the absence of certain boundaries at the junctions of grasslands, saline and alkaline lands and cultivated lands. Through artificial selection, 700 grassland samples and 500 non-grassland samples are selected randomly from 700 grassland samples; 200 of them are selected randomly as the one-class training samples, the rest 500 samples are taken as testing samples. The 200 samples selected are further divided into 20 groups of samples. The SSSOCC and OCSVM algorithm proposed in this paper are then trained. When the number of subclassifier *T*=5 in SSSOCC algorithm and the number of attribute *NUM*=3 in each group, the accuracies of these two algorithms are shown in the following table and figures:

Table 1. Classification accuracy of the two methods

The number of samples	SSSOCC（%）		OCSVM(%)	
	Identification of grasslands	Identification of non-grasslands	Identification of grasslands	Identification non-grasslands
10	81.2	98.2	63.6	99.8
20	92.2	94.6	78.8	98.8
30	91.4	95.2	77.2	98.2
40	94.8	92.8	82.3	98
50	92.2	93	87	97.2
60	93	95.5	82.6	98
70	95.4	94.4	84.2	97.2
80	93.4	95	84.2	98.2
90	94	95.8	91.5	95.8
100	94.2	95.2	87.6	97.4
110	94.2	96.4	90	95.8
120	95	95.2	90.2	96
130	94.8	95.2	90.2	96
140	94.8	95.8	90.6	95.8
150	94	95.8	91.2	96
160	95	95.6	91.8	95.8
170	95	95.6	91.8	95.8
180	94.2	95.8	92	95.8
190	94.2	95.8	91.8	95.8
200	95	95.8	91.8	95.8
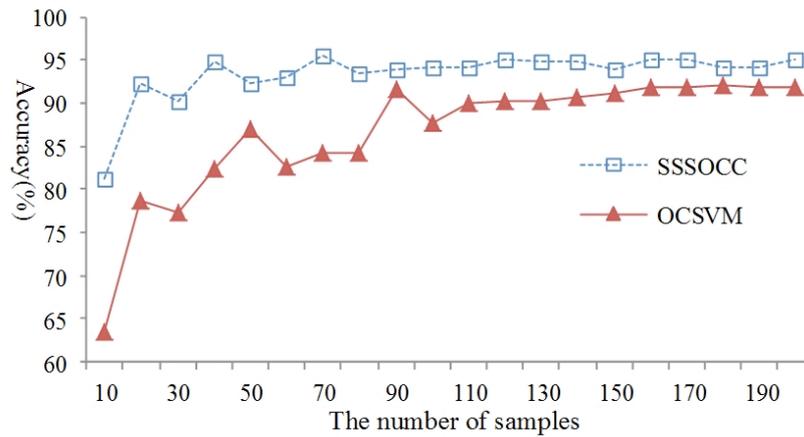
Fig.3. Comparison of identification of grasslands
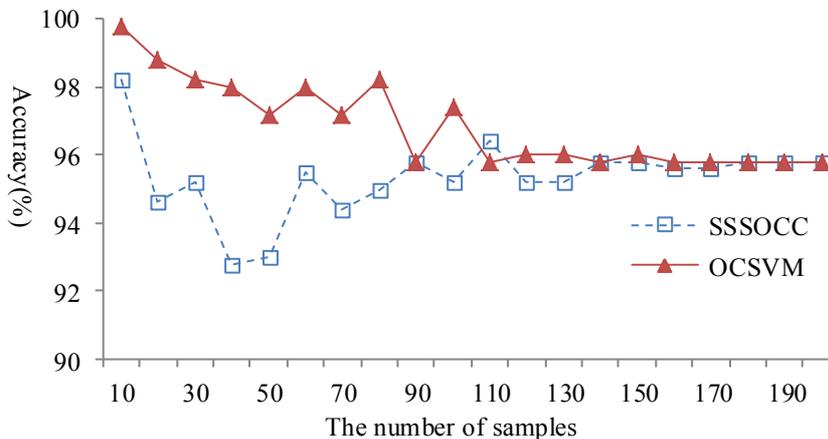
Fig.4. Comparison of identification of non-grasslands

Table 1 and Figures 3 and 4 show the comparisons of the two methods. In the identification of grasslands, the classification accuracy of SSSOCC in all sample set is better than that of OCSVM; the highest accuracy of SSSOCCC reaches to 95.4%, whereas that of OCSVM is 91.8%. When there are only 10 samples, the accuracy of SSSOCCC reaches 81.2%, whereas that of OCSVM is 63.6%. The main reason is that samples with more than 11 bands have shown obvious sparsification, and only 3 input attributes are present in each subsamples of SSSOCC; thus, this problem is avoided. At the same time, we can see the classification accuracy of SSSOCC is stable all the time (Figure 3), whereas that of OCSVM fluctuates greatly when the number of samples is small. In addition, the accuracies of SSSOCC are higher than that of OCSVM in all 20 groups. In the identification of non-grassland, the accuracy of SSSOCC reaches 94.4% to 98.2%, and that of OCSVM reaches 95.8% to 99.9%. Thus, both algorithms achieve higher accuracy.

Furthermore, both algorithms show higher accuracy in the grassland identification accuracy than in the non-grassland identification accuracy, which is determined by the characteristics of super sphere. That is, lower accuracy of the grassland identification indicates smaller super sphere and lower possibility of non-grassland misidentification. The non-grassland identification accuracy of OCSVM is higher before the 100th sample, because there are fewer samples showing sparsification at this moment; thus, the corresponding super sphere is smaller. The non-grassland identification accuracy of SSSOCC is lower from the 10th to 100th sample, because its model is constituted by multiple subclassifiers. However, the increase in grassland identification accuracy of OCSVM cannot make up for its overall accuracy; the accuracy of OCSVM is lower than that of SSSOCC in general. Table 1 and Figure 1 show that the classification accuracy and stability of SSSOCC are higher than that of traditional OCSVM algorithm. The classification results are shown in Fig. 4.
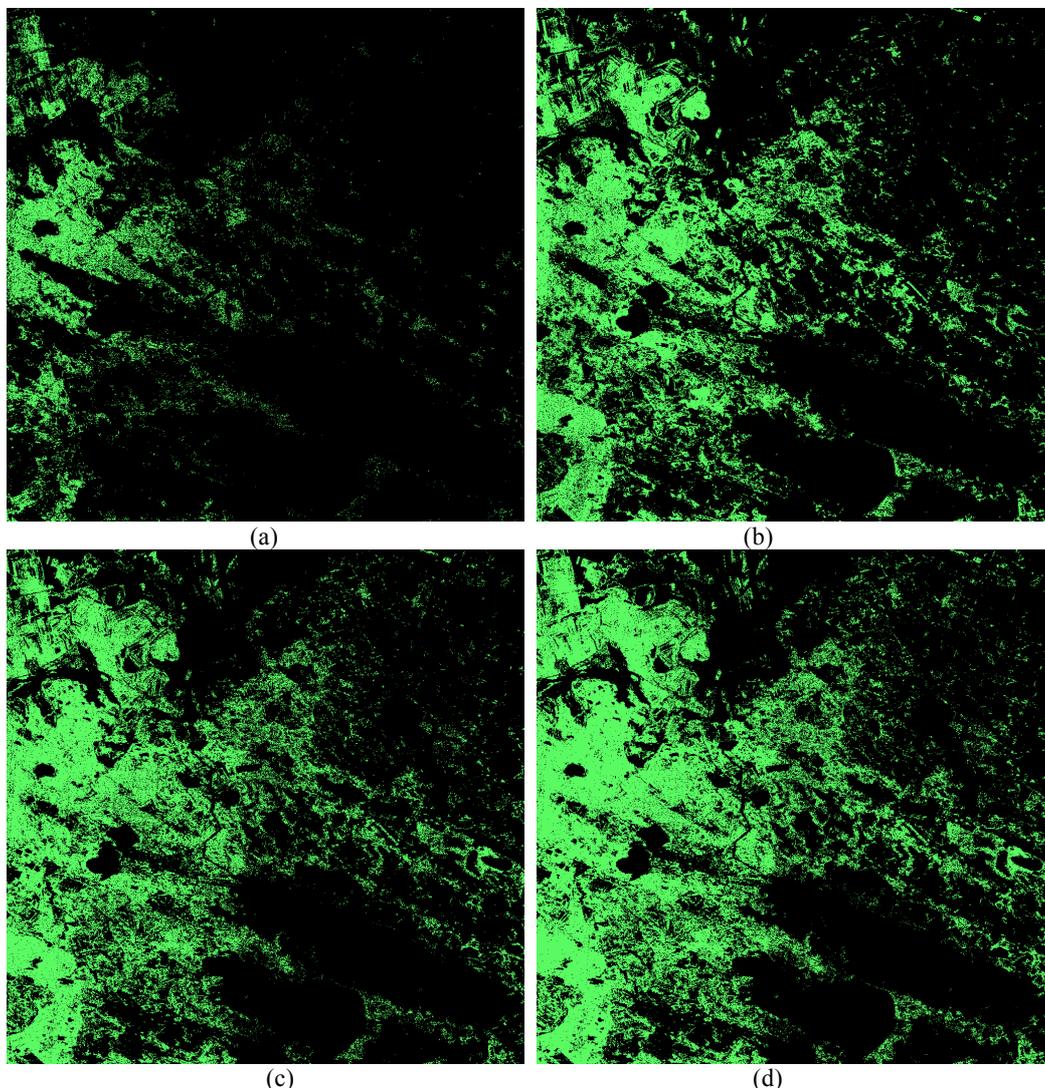

Fig. 5. One-class classification result of the two methods for the remote-sensing image. (a) OCSVM-10. (b) SSSOCC -10. (c) OCSVM-200. (d) SSSOCC -200.

Figure 5(a) shows the classification results of 10 samples in OCSVM algorithm, which shows that sparsification considerably affect the model of super-sphere, and thus a large amount of grassland has not been identified. Figure 5(b)

shows the classification results of 10 samples in SSSOCC algorithm, in which most of the grasslands have been identified. In addition, when the sample is small, SSSOCC algorithm is obviously better than OCSVM. Figures 5(c) and

5(d) are respectively the classification result of OCSVM and SSSOCCC in the 200th sample. The figures show that SSSOCC algorithm has better classification results on the boundaries of grasslands, farmlands, saline and alkaline lands. Moreover, SSSOCC algorithm is also more continuous inside the grasslands with less miscellaneous points, which indicates that the identification ability of SSSOCC is better than that of OCSVM.

5. Conclusions

Sparse and stratified sampling method was introduced into one-class classification of remote sensing images to solve problems on larger amounts of spatial bands, uncertainty or inconsistency information in relatively few training samples.

(1) Sparseness description can describe distinguishable degree of a spatial feature for remote sensing images classification. Through sparseness description, stratified sampling divides spatial features of a remote sensing image into several subsets having sufficient differences. Each subset can train a sub-classifier. Combining the sub-classifiers can create a decision model, and multi sub-classifiers can give a powerful classification ability to this model.

(2) With the help of sparse and stratified sampling method, SSSOCC's classification accuracy and stability are higher than traditional OCSVM algorithm; even in the case of the number of training samples is relatively small. SSSOCC is more appropriate to deal with the one-class classification problems in remote sensing images.

This study can meet the demands for one-class classification of remote sensing images and obtain land use or cover information accurately. However, the method is limited by special image resolution conditions, such as moderate or low resolution remote sensing images. Further study is needed to solve the problem in high resolution remote sensing images.

References

1. Xie, L., Li, G., Xiao, M., Peng, L., "Novel classification method for remote sensing images based on information entropy discretization algorithm and vector space model". *Computers & Geosciences*, 89(1), 2016, pp.252-259.
2. Munoz-Marf, J., Bruzzone, L., Capmps-Vails, G., "A support vector domain description approach to supervised classification of remote sensing images". *IEEE Transactions on Geoscience and Remote Sensing*, 45(8), 2007, pp.2683–2692.
3. Foody, G. M., "Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes". *International Journal of Remote Sensing*, 25(1), 2004, pp. 3091–3104.
4. Foody, G. M., Mathur, A., Sanchez-Hernandez, C., Boyd, D. S., "Training set size requirements for the classification of a specific class". *Remote Sensing of Environment*, 104(1), 2006, pp.1–14.
5. Khan, S. S., Madden, M. G.,"A Survey of Recent Trends in One Class Classification". *Lecture Notes in Computer Science*, 6206, 2010, pp.188-197.
6. Gambardella, A., Giacinto, G., Migliaccio, M., Montali, A., "One-class classification for oil spill detection". *Pattern Analysis and Applications*, 13(3), 2010, pp.349-366.
7. Sanchez-Hernandez, C., Boyd, D. S., Foody, G. M., "One-class classification for monitoring a specific land cover class: SVDD classification of fenland". *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), 2007, pp.1061-1073.
8. Guo, Q., Li, W., Tong, D., "Predicting potential distributions of geographic events using one-class data: concepts and methods". *International Journal of Geographical Information Science*, 25(10) , 2011,1697-1715.
9. Li, P., Xu, H., Guo, J., "Urban building damage detection from very high resolution imagery using OCSVM and spatial features". *International Journal of Remote Sensing*, 31(13), 2010, pp.3393-3409.
10. Fang, G., Rao, M., "SVM-Based Data Editing for Enhanced One-Class Classification of Remotely Sensed Imagery". *IEEE Transactions on Geoscience and Remote Sensing*, 5(3), 2008, pp.189-913.
11. Stenzel, S., Feilhauer, H., Mack, B., Metz, A., Schmidtlein, S., "Remote sensing of scattered Natura 2000 habitats using a one-class classifier". *International Journal of Applied Earth Observation and Geoinformation*, 33(1), 2014, pp.211-217.
12. Li, W., Guo, Q., Elkan, C., "A Positive and Unlabeled Learning Algorithm for One-Class Classification of Remote-Sensing Data". *IEEE Transactions on Geoscience and Remote Sensing*, 49(2), 2011, pp. 717 – 725.
13. Muñoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L., Camps-Valls, G., "Semisupervised One-Class Support Vector Machines for Classification of Remote Sensing Data". *IEEE Transactions on Geoscience and Remote Sensing*, 48(8), 2010, pp.3188 – 3197.
14. Bouguess, M., "A practical outlier detection approach for mixed-attribute data". *Expert Systems with Applications*, 42(22), 2015, pp.8637-8649.
15. Tax, D., Duin, R., "Uniform Object Generation for Optimizing One-class Classifiers". *Journal of Machine Learning Research*, 2(1), 2002, pp.155-173.
16. Elkan, C., Noto, K., "Learning classifiers from only positive and unlabeled data". *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1 , 2008, pp.213-220.
17. Kleinberg, E, M., "On the Algorithmic Implementation of Stochastic Discrimination". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5), 2000, pp.473-490.
18. Zhong, S., Xie, X., Lin, L., "Two-layer random forests model for case reuse in case-based reasoning". *Expert Systems with Applications*, 42(24), 2015, pp.9412-9425.
19. Ye, Y., Wu, Q., Huang, Z, J., Ng, M, K., Li, X., "Stratified sampling for feature sub space selection in random forests for high dimensional data". *Pattern Recognition*, 46(3), 2013, pp.769–787.
20. Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., Williamson, R. C., "Estimating the Support of a High-Dimensional Distribution". *Neural Computation*, 13(7), 2001, pp.1443-1471.