INTEGRATING OBJECT-BASED CLASSIFICATION WITH ONE-CLASS SUPPORT VECTOR MACHINES IN MAPPING A SPECIFIC LAND CLASS FROM HIGH SPATIAL RESOLUTION IMAGES

Qinghua Guo[a], Guoming Du[a,b], Yu Liu[a, c], Desheng Liu[d]

[a]School of Engineering, University of California Merced, P.O. Box 2039, Merced, CA, USA 95344
[b]School of Geography Sciences and Planning, Sun Yat-sen University, Guangzhou, China, 510275
[c]Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing, China 100871
[d]Department of Geography, Ohio State University, 154 North Oval Mall, Columbus, OH, USA 43210

Commission WGs WG IV/9

KEY WORDS: Object-Based Classification, One-Class Support Vector Machines, High Spatial Resolution Imagery

ABSTRACT:

Remote sensing techniques have been commonly used to map land cover and land use types. For many applications, users may only be interested in a specific land class in an image such as extracting urban areas from an image, or retrieving dead trees from a forest. This could be referred to as a one-class classification problem. In addition, with the increasing availability of high spatial resolution imagery, earth objects can be mapped in detail, which enable us to quickly update and monitor the change of a specific class. However, conventional pixel-based classification methods have difficulty in dealing with high spatial resolution remote sensing data. In this study, we use urban house extraction as an example, and propose to classify houses from high spatial resolution images by integrating one-class Support Vector Machines (SVMs) and object-based classifiers. We also compared the performance from the proposed method with the one-class SVMs and pixel-based method. The results indicate that the proposed method outperforms the pixel based method, and could be a promising way to provide relatively quick and efficient way in extracting a specific land class from high spatial resolution images.

1. INTRODUCTION

With the increasing availability of remote sensing data over the past few decades, remote sensing data have been commonly used in a wide variety of urban and environmental applications such as: monitoring land use change, mapping suitable habitat, and detecting invasive species. Traditionally, all the land types in an image were completely mapped via remote sensing classification methods. However, for some applications, we may only be interested in a specific class without considering other land types (Foody et al. 2006). For example, if the objective of the project is to extract roads from remote sensing data, we may not be interested in classifying forests, and agricultural lands. Another reason for mapping a specific class of interest is to reduce significant amount of efforts in collecting, training, and testing ground truth data; as it is very time consuming to collect ground-truth data for all the land classes. Therefore, it is needed to develop methods to retrieve only one land type from the remote sensing data. The idea is to separate a specific class from the rest of the land classes. This question could be referred to as a one-class classification problem. Other disciplines also have similar issues. For example, species collection data from natural museums often contain presence-only data (i.e. absence data are often not collected, or unreliable to collect such an animals or invasive species), scientists are interested in classifying the habitat based on presence-only data in order to find suitable or potential habitat for species. Another example in a handwritten number recognition problem is to classify the handwriting number such as "8" when we only have a sample a set of handwritten "8"s. One common solution to deal with one-class classification problem is based on similarity matching. Numerous machine learning and non machine learning approaches can be applied to this problem. Among many methods, support vector machines (SVMs), originally developed by Vapnik (1995), are considered to be a new generation of learning algorithms. SVM have several appealing characteristics for modellers, including: they are statistically based models rather than loose analogies with natural learning systems, and they theoretically guarantee performance (Cristianini and Scholkopf, 2002). SVM have been applied successfully to a range of remote sensing classification applications (Huang et al., 2002). Recently, Scholkopf et al. (1999) developed one-class SVM to deal with the one-class problem. This method has proved useful in document classification, texture segmentation, and image retrieval.

Moreover, with the advance of sensor techniques, high spatial resolution remotely sensed images have become commercially available and increasingly used in various aspects of environmental monitoring and management (Mumby and Edwards 2002). Conventional pixel-based classifiers such as maximum likelihood classification (MLC) and Iterative Self-Organizing Data Analysis Technique (ISODATA), which label unknown areas pixel by pixel based on spectral similarity, do not perform well with high spatial resolution images (Xia 1996). This is because the inherent spectral variability in specific ground targets increases as resolution becomes finer (Martin and Howarth 1989). Therefore, retrieving one land class from high spatial resolution imagery based on pixel-based method may result in significant misclassification. In recent years, object-based methods have gained much attention as alternative methods for classifying high resolution images. An "object" is defined here as a group of spectrally similar contiguous pixels, and ideally, it should represent a physically or ecologically homogeneous land class. One reason that the object-based methods perform well in classifying high resolution images are because once the object is created by a segmentation approach, many more features such geometrical (e.g. shape and area) and topological properties (e.g. relationship between objects) can be extracted from the segmented image. This feature is particularly useful in classifying high spatial resolution images since high spatial resolution images often contain relatively fewer spectral bands (e.g. IKNONOS, QIUCKBIRD) compared to coarser images (e.g. MOIDS, Landsat TM). Consequently, methods that rely on only spectral information could have difficulty in distinguishing spectrally similar classes such as buildings and roads). Yet, it will be much easier to differentiate a building from a road if we can incorporate the object shape into the classification process.

Hence, in this study, we propose to develop an integrated approach based on one-class SVMs and object-based methods to classify one land class from high spatial resolution images. We first segmented an image by a segmentation approach; both spectral and spatial properties were then extracted from objects, the one-class SVMs were then applied to extract one land type based on properties extracted from the objects. We also performed the comparisons among the proposed method and the one-class SVM with pixel-based classification. The overall accuracy and Kappa coefficient were calculated and used in the comparison (Congalton and Mead, 1983).

2. DATA AND METHODS

2.1 Data

The high resolution remote sensing data used in this research are from aerial photos with 0.3 meter spatial resolution. The study area is located in Oakland, California (Figure 1).

Figure 1. Aerial photograph of the study area (color image with 0.3 meter spatial resolution)

2.2 Methods

2.2.1 Segmentation method: Segmentation methods are used to generate image objects for classification and image retrieval, the object is defined as a group of spectrally similar contiguous pixels. Numerous algorithms have been proposed to segment an image. In this study, we used the segmentation method from Definies software, which is based on a multi-resolution segmentation algorithm. The segmentation results are tuned based on scale parameters, color, smoothness, and compactness. The final segmentation results are shown in Figure 2.

2.2.2 Features extraction: After image segmentation, we then extract features to be used for image classification. One advantage of the object-based method is the ability to extract a wealth of features that could aid in classifying the imagery. Fourteen features are chosen in this study, detailed descriptions are described as follows:

(1) Mean value (MV), which represents mean brightness value of every image object. Since the aerial photo includes three bands (i.e. green, blue, and red), we will have three mean values as features for image classification. The formula is as follows:

$$L_k = \sum_{i=1}^{n} B_{ik} \ /n$$

Where, L_k is the mean brightness value; n is the number of pixels in the image object; B_{ik} is brightness value of ith pixel contained in the image object in band k.

(2) Mean difference to scene (MDS), which is the difference between mean brightness value of an image object and mean brightness value of the whole scene in band k. The formula is as follows:

$$S_k = \sum_{i=1}^{n} B_{ik} \ /n - \sum_{j=1}^{m} B_{jk} \ /m$$

Where, S_k is mean difference to the band k; m is the number of pixels of the whole scene. Similarly, there are three features which exist in the mean difference to the scene.

Figure 2. The result of image segmentation

(3)Mean difference to neighbour (MDN), which is the difference between mean brightness value of an image object and mean brightness value of its direct neighbours in band k. The formula is defined as follows:

$$N_k = \sum_{i=1}^{n} B_{ik} \ /n - \sum_{l=1}^{p} \sum_{j=1}^{m_l} B_{ljk} \ / \sum_{l=1}^{p} m_l$$

Where, p is the number of direct neighbours; m_l is the number of pixels of the neighbour l. There are three MDNs.

(4)Standard Deviation, which represents the standard deviation of brightness value of all the pixels contained in an image object in bank k. There are also three standard deviations.

(5)Area, which represents the number of all the pixels contained in an image object.

(6)Shape Index (SI) describes the smoothness of the image object borders, which is useful in differentiating houses and other man-made objects such as road. The definition of SI is:

$$SI_i = \frac{P_i}{4\sqrt{A_i}}$$

Where, SI_i is the shape index; P_i is the perimeter; A_i is the area of image object *i*.

The number of features used in this study is summarized in Table 1:

Features	The number of features
Mean value	3
Mean difference to scene	3
Mean difference to neighbour	3
Standard Deviation	3
Area	1
Shape Index	1

Table 1. Features and the number of features used in one-class SVMs

The values of features vary significantly. For example, the area of objects ranges from 1 to 10650 while the shape index ranges from 1 to 8.73. Therefore, before we implemented support vector machines to classify the objects, we firstly normalized the features by using a maximum and minim method, and the final values for each feature range from 0 – 1.

2.2.3 Feature selection: The feature selection aims to select a subset of the features that could perform well in the classification algorithms. It could reduce the dimensionality of the feature space while still maintain high classification accuracy and avoid over-fitting. In this study, we used the brute-force search algorithm (also referred to as the exhaustive search), which searches for all of the possible combinations of features and finds the optimum solution for the classifier. Assuming the number of features is *n*, the number of possible combination of features will be:

$$C_n^1 + C_n^2 + C_n^3 + \ldots + C_n^n = 2^n - 1$$

Since there are 14 features in this study, which results in 16383 possible combinations. The computational calculation is acceptable with current PC (the computer used in this study is Intel CPU 2.4, and takes approximately 2.6 hours to search all possible combinations as well as implemented one-class SVMs). Although the brute-force algorithm is very time-consuming, it solves the optimization problem directly and always finds the solution if it exists. However, it should be noted that when the number of features increase, the brutal search method could become too time-consuming as its cost is proportional to the number of candidate features. In these cases, other feature selection methods can be more appropriate (e.g. the stepwise method). The Kappa values and cross validation methods are used to evaluate model performance with difference feature combinations.

2.2.4 Accuracy assessment: Ground truth data were acquired by manual photo interpretation. Totally, 100 house objects and 20 non-house objects were selected. For one class support vector machines, we only used house data for training, and both house and non-housing data to evaluate the classification results. We used the Kappa value to measure the classification accuracy, Kappa values take into account an agreement that can occur by change (expected agreement). In General, Kappa values of 0.75 and higher are considered good classification results (Eric et al 2004). A five-fold cross-validation method is used to evaluate the model performance based on the training data. The cross validation method is implemented as follows (Guo et al. 2007):

(1) Select 100 houses as set H and 20 non houses as set H_n as training set for cross validation.

(2) Randomly split set H into five subsets, i.e., every subset contains 20 houses. Every subset H in combination with H_n in turn composes of set T_e. Other four subsets in set H compose of set T_r.

(3) The set T_r is used as training samples; the set T_e is used as testing samples to evaluate the model performance of one-class SVMs.

(4) The Kappa value is calculated for each iteration. The average Kappa value is reported based on five implements.

2.2.5. Implement SVMs for object-based classification

Assuming we have l training points x_i (i =1, 2… l), we want to find a hypersphere as small as possible to contain the training points in multidimensional space. Meanwhile, we also allow a small portion of outliers to exist using a slack variable (ξ_i):

$$\text{Min } R^2 + \frac{1}{vl}\sum_i \xi_i \qquad (1)$$

$$\text{Subject to: } (x_i - c)^T (x_i - c) \le R^2 + \xi_i,\ \xi_i \ge 0 \text{ for all } i \in [l] \qquad (2)$$

Where c and R are the center and radius of the sphere, T is the transpose of a matrix, and $v \in (0, 1]$ is the trade-off between the volume of the sphere and the number of training points rejected. When v is large, the volume of the sphere is small thus more training points will be rejected than when v is small, where more training points will be contained within the sphere. v can be roughly explained as the percentage of outliers in the training dataset (Scholkopf et al., 2001). This optimization problem can be solved by the Lagrangian:

$$L(R,\xi,c,a_i,\beta_i) = R^2 + \frac{1}{vl}\sum_{i=1}^{l}\zeta_i - \sum_{i=1}^{l} a_i\left\{R^2 + \zeta_i - (x_i^2 - 2cx_i + c^2)\right\} - \sum_{i=1}^{l}\beta_i\xi_i \qquad (3)$$

Where $a_i \ge 0$ and $\beta_i \ge 0$. Setting the partial derivative of L with respect to R, a_i, c equal to 0, we get:

$$\sum_{i=1}^{l} a_i = 1 \qquad (4)$$

$$0 \le a_i \le \frac{1}{vl} \qquad (5)$$

$$c = \sum_{i=1}^{l} a_i x_i \qquad (6)$$

Substituting equation (4)-(6) to equation (3), we have the dual problem:

$$\min_a \sum_{i,j} a_i a_j (x_i \cdot x_j) - \sum_i a_i (x_i \cdot x_i) \qquad (7)$$

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

To determine whether or not a test point (x) is within the sphere, we can calculate the distance between the test point and the center C. It can be expressed as:

$$(x \cdot x) - 2\sum_i a_i (x \cdot x_i) + \sum a_i a_j (x_i \cdot x_j) \le R^2 \qquad (8)$$

So far, we have assumed that the data are spherically distributed. In reality, the data are often not spherically distributed. To make the method more flexible and to capture the non-linearity such as multi-mode distribution, the kernel function $K(x_i, x_j)$ can be introduced. We can express the inner product in equation 8 as the kernel function:

$$K(x,x) - 2\sum_i a_i k(x,x_i) + \sum_{i,j} a_i a_j K(x_i,x_j) \le R^2 \quad (9)$$

Two types of kernels are often used: polynomial and Gaussian kernels, however, the former usually does not produce a tight description of the data and is sensitive to outliers when the polynomial degree is high (Tax and Duin, 1999a). A more robust way is to construct the Gaussian kernel, which has been commonly used for one-class SVMs (Tax and Duin, 1999a; Scholkopf et al., 2001):

$$K(x_i,x_j) = e^{-(x_i-x_j)^2/S^2} \quad (10)$$

where S is the kernel width. The Gaussian kernel was applied in this study. It should be noted that the method above was proposed by Tax and Duin (1999), another approach proposed by Scholkopf et al (1999) is to find some hyperplane to separate the training data from the origin with the maximum margin. For the Gaussian kernel, these two methods are equivalent (Scholkopf et al., 2001). We implemented the one-class SVMs by the modified version of LIBSVM-a library for support vector machines developed by Chang and Lin (2001). A more detailed mathematical derivation of one-class SVMs can be found in Scholkopf et al. (1999), and Tax and Duin (1999a). In this study, both Gaussian kernel width (S) and ν are estimated from the cross validation method that maximizes the classification accuracy (i.e. Kappa value).

2.2.6 One class SVMS with pixel-based classification: as a comparison, we implemented the one class SVMs with pixel-based classification. Features used in pixel based classification include digital values of three bands for each pixel. In order to catch the texture information in the classification, we also used the variance of digital value for each pixel with a 5×5 processing window in each band. As a result, six features are used in one-class SVMs with pixel-based classification. The training samples and testing samples are used similar to the object-based method. The major difference is from the perspective of data storage and processing: in the object-based method, the segmented objects as well as training and testing samples are stored as polygons and hence processed in the vector format, while, in the pixel-based method, the data are stored and processed in the raster format only.

3. RESULTS

For the one-class SVMs with object-based method, we found that the combination of shape index and mean difference to neighbour provided the highest Kappa score based on the five-fold cross validation method. Kappa value is 0.79. The kernel width and the ν for one-class SVMs are 0.03 and 0.06 respectively. The final classification result is shown in Figure 3.

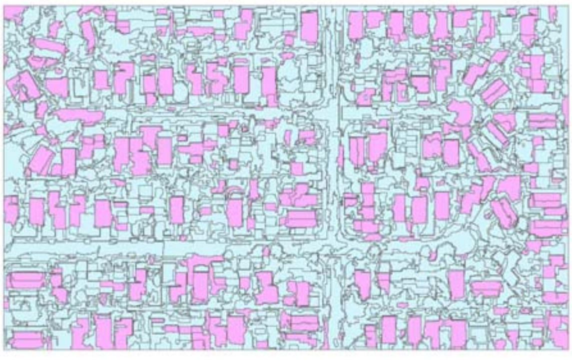

Figure 3. one-class SVMs with object-based classification. The pink objects are houses, and blue objects are non-houses.

For the one-class SVMs with pixel-based method, the kernel width and the ν for one-class SVMs are 0.07 and 0.02 respectively by using the brute-force search method. The optimal Kappa value for the classification is 0.30. Figure 4 shows the final classification result based on one-class SVMs with the pixel-based method. This method can effectively distinguish houses from grasslands, but it is difficult to distinguish houses from roads.

Figure 4. Classification results based on one-class SVMs with the pixel-based classification. The pink objects represent houses and the gray objects represent non-houses.

4. DISCUSSION

The result indicates that one-class SVMs with the object-based classification (kappa = 0.79) outperformed the pixel-based counterpart (kappa = 0.30). Two reasons may contribute to the differences:

1) The object-based classification can provide more meaningful candidate features such as shape and mean difference to neighbour, which is very useful in differentiating houses from roads. Because of the spectral similarity between houses and roads, conventional pixel-based methods are difficult to distinguish between them as shown in Figure 4.

(2) Since the pixel-based classifies the image pixel by pixel, the final classification is very fragmented, particularly in classifying high spatial resolution images. As shown in Figure 4, many pixels are misclassified as houses. Based on the shape and size of those pixels, we could easily tell that they are not houses. Therefore, in order to make use of the classification result from pixel-based methods for extracting houses from high resolution images, users normally need to manually or semi-manually post process the classified image. On the other hand, the object-based method seeks to first segment the images based on spectral similarity among pixels, and then classify the images based on features extracted from the segmented objects. The classification results are much smoother and almost ready for environmental or urban applications (Figure 3).

In sum, with the increasingly available high spatial resolution imagery, one-class SVMs together with object-based classification methods provide a promising way in extracting a specific land class type from high resolution images.

REFERENCES

Congalton, R.G. and Mead, R.A., 1983, A quantitative method to test for consistency and correctness in photointerpretation, Photogrammetric Engineering and Remote Sensing, 49: 69-74.

Cristianini, N. and Scholkopf, B., 2002, Support vector machines and kernel methods - the new generation of learning machines, Ai Magazine, 23: 31-41.

Eric A. Whitsel, Kathryn M. Rose, Joy L. Wood et al. 2004 Accuracy and Repeatability of Commercial Geocoding. American Journal of Epidemiology, 160(10):1023-1029.

Foody, G.M., Mathur, A., Sanchez-Hernandez, C. and Boyd, D., 2006, Training set size requirements for the classification of a specific class, Remote Sensing of Environment, 104: 1 - 14.

Guo, Q., Kelly, M., Gong, P. and Liu, D., 2007, An object-based classification approach in mapping tree mortality using high spatial resolution imagery, GIScience & Remote Sensing, 44: 24 - 47.

Huang, C., Davis, L.S. and Townshend, J.R.G., 2002, An assessment of support vector machines for land cover classification, International Journal of Remote Sensing, 23: 725-749.

Martin, L.R.G. and Howarth, P.J., 1989, Change-detection accuracy assessment using spot multispectral imagery of the rural urban fringe, Remote Sensing of Environment, 30: 55-66.

Mumby, P.J. and Edwards, A.J., 2002, Mapping marine environments with ikonos imagery: Enhanced spatial resolution can deliver greater thematic accuracy, Remote Sensing of Environment, 82: 248-257.

Scholkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C., 1999, Estimation the support of a high-dimensional distribution, Technical Report MSR-TR-99-87, Microsoft Research.

Vapnik, V., 1995, The nature of statistical learning theory, New York: Springer-Verlag.

Xia, L., 1996, A method to improve classification with shape information, International Journal of Remote Sensing, 17: 1473-1481.