Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 25 (2004) 1067–1074
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Segmentation of multispectral remote sensing images using active support vector machines

Pabitra Mitra *, B. Uma Shankar, Sankar K. Pal

Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

Received 21 October 2003; received in revised form 26 January 2004
Available online 15 April 2004

Abstract

The problem of scarcity of labeled pixels, required for segmentation of remotely sensed satellite images in supervised pixel classification framework, is addressed in this article. A support vector machine (SVM) is considered for classifying the pixels into different landcover types. It is initially designed using a small set of labeled points, and subsequently refined by actively querying for the labels of pixels from a pool of unlabeled data. The label of the most interesting/ambiguous unlabeled point is queried at each step. Here, active learning is exploited to minimize the number of labeled data used by the SVM classifier by several orders. These features are demonstrated on an IRS-1A four band multi-spectral image. Comparison with related methods is made in terms of number of data points used, computational time and a cluster quality measure.

Keywords: Image segmentation; Semi-supervised learning; Transductive learning; Query support vector machine

1. Introduction

Segmentation is a process of partitioning an image space into some nonoverlapping meaningful homogeneous regions. The success of an image analysis system depends on the quality of segmentation. Two broad approaches to segmentation of remotely sensed images are gray level thresholding and pixel classification (Richards, 1993). In thresholding (Pal et al., 2000) one tries to get a set of thresholds $\{T_1, T_2, \ldots, T_k\}$ such that all pixels with grey values in the range $[T_i, T_{i+1})$ constitute the ith region type. On the other hand in pixel classification, homogeneous regions are determined by clustering the feature space of multiple image bands. Multispectral nature of most remote sensing images make pixel classification the natural choice for segmentation.

In the unsupervised pixel classification framework, several clustering algorithms like split-and-merge (Laprade, 1988), fuzzy k-means (Pal et al., 2000; Cannon et al., 1986), neural networks based methods (Baraldi and Parmiggiani, 1995), scale space techniques (Wong and Posner, 1993) and statistical methods have been used for the purpose

* Corresponding author. Address: Department of Computer Science, Indian Institute of Technology and Engineering, Kanpur 208016, India. Tel.: +91-0512-259-7584.

E-mail addresses: pmitra@iitk.ac.in, pabitra_r@isical.ac.in (P. Mitra), uma@isical.ac.in (B. Uma Shankar), sankar@isical.ac.in (S.K. Pal).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2004.03.004

of segmentation. Statistical methods are widely used in unsupervised pixel classification framework because of their capability of handling uncertainties arising from both measurement error and the presence of mixed pixels which have certain degree of membership to more than one class. A general method of statistical clustering is by means of the expectation maximization (EM) algorithm (Dempster et al., 1977) and its variants (Pal and Mitra, 2002). However, the unsupervised pixel classification methods have many limitations. The number of clusters are often unknown, which results in region merging/splitting and also hinders the interpretation of the segmented images. Also, unsupervised methods mostly generate convex clusters, which leads to degradation in segmentation quality.

The aforesaid difficulties do not arise in supervised pixel classification, and several methods based on neural networks, genetic algorithms (Bandyopadhyay and Pal, 2001) has been developed in this framework. Recently, support vector machines are becoming popular for classification of multispectral remote sensing images (Brown et al., 2000; Huang et al., 2002).

The primary problem in supervised pixel classification is the pure availability of labeled data, which can be obtained only from ground truths and by costly manual labeling. Recently, active learning has become a popular paradigm for reducing the data requirement of large scale learning tasks (Angluin, 1988; Cohn et al., 1994). Here, instead of learning from 'random samples', the learner has the ability to select its own training data. This is done iteratively, and the output of a step is used to select the examples for the next step. Several active learning strategies exist in practice, e.g., error driven techniques, uncertainty sampling, version space reduction and adaptive resampling.

Support vector machines (SVM) are particularly suited for active learning since a SVM classifier is characterized by a small set of support vectors (SVs) which can be easily updated over successive learning steps. One of the most efficient active SVM learning strategy is to iteratively requests the label of the data point closest to the current separating hyperplane or which violates the margin constraint maximally (Mitra et al., 2000; Campbell et al., 2000). This accelerates the learning drastically compared to random data selection. The above technique is often referred to as active/query SVM. Besides active SVM, another active learning strategy based on version space splitting is presented in (Tong and Koller, 2001). The points which split the current version space into two halves having equal volumes are selected at each step, as they are likely to be the actual support vectors. Three heuristics for approximating the above criterion are described, the simplest among them selects the point closest to the current hyperplane as in (Campbell et al., 2000). A greedy optimal strategy for active SV learning is also described in (Schohn and Cohn, 2000). Here, logistic regression is used to compute the class probabilities, which is further used to estimate the expected error after adding an example. The example that minimizes this error is selected as a candidate SV.

The present article describes a pixel classification algorithm based on the query SVM algorithm. A conventional SVM is initially designed using a small set of points labeled manually. The SVM is subsequently refined by actively querying for the labels of pixels from a pool of unlabeled data. The most interesting/ambiguous unlabeled point is queried at each step and is labeled by an human expert. It is seen that the above active learning strategy reduces the number of labeled data used by the SVM classifier by several orders compared to conventional SVM, while providing comparable segmentation quality. These features are demonstrated on an IRS-1A four band image. Comparison with related methods is made in terms of the number of data points used, computational time and a cluster quality measure.

The article is organized as follows: the fundamentals of support vector machines are briefly mentioned in Section 2. The active SVM learning algorithm for pixel classification is described in Section 3. Experimental results are provided in Section 4, followed by conclusions in Section 5.

2. Support vector machines

Support vector machines are a general class of learning architecture inspired from statistical learning theory that performs *structural risk minimization* on a nested set structure of separating hyperplanes (Vapnik, 1998). Given a training data, the SVM training algorithm obtains the optimal separating hyperplane in terms of generalization error. We describe below the SVM design algorithm for a two class problem. Multiclass extensions can be done by designing a number of one-against-all on one-against-one two class SVMs.

Algorithm 1:

Suppose we are given a set of examples $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_l, y_l)$, $\boldsymbol{x} \in R^N$, $y_i \in \{-1, +1\}$. We consider functions of the form $\mathrm{sgn}((\boldsymbol{w} \cdot \boldsymbol{x}) + b)$, in addition we impose the condition

$$\inf_{i=1,\ldots,l} |(\boldsymbol{w} \cdot \boldsymbol{x}_i) + b| = 1. \tag{1}$$

We would like to find a decision function $f_{\boldsymbol{w},b}$ with the properties $f_{\boldsymbol{w},b}(x_i) = y_i; i = 1, \ldots, l$. If this function exists, condition (1) implies

$$y_i((\boldsymbol{w} \cdot \boldsymbol{x}_i) + b) \geqslant 1, \quad i = 1, \ldots, l. \tag{2}$$

In many practical situations, a separating hyperplane does not exist. To allow for possibilities of violating Eq. (2), slack variables are introduced like

$$\xi_i \geqslant 0, \; i = 1, \ldots, l, \tag{3}$$

to get

$$y_i((\boldsymbol{w} \cdot \boldsymbol{x}_i) + b) \geqslant 1 - \xi_i, \; i = 1, \ldots, l. \tag{4}$$

The support vector approach for minimizing the generalization error consists of the following:

$$\text{Minimize:} \quad \Phi(\boldsymbol{w}, \xi) = (\boldsymbol{w} \cdot \boldsymbol{w}) + C \sum_{i=1}^{l} \xi_i, \tag{5}$$

subject to the constraints (3) and (4).

It can be shown that minimizing the first term in Eq. (5), amounts to minimizing the VC-dimension, and minimizing the second term corresponds to minimizing the misclassification error (Burges, 1998). The above minimization problem can be posed as a constrained quadratic programming (QP) problem.

The solution gives rise to a decision function of the form:

$$f(\boldsymbol{x}) = \mathrm{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i (\boldsymbol{x} \cdot \boldsymbol{x}_i) + b\right].$$

Only a small fraction of the α_i coefficients are nonzero. The corresponding pairs of $\boldsymbol{x}_i$ entries are known as *support vectors* and they fully define the decision function. The support vectors are geometrically the points lying near the class boundaries.

The linear SVM was described above. However, nonlinear kernels like polynomial, sigmoidal and radial basis functions (RBF) may also be used. Here, the decision function is of the form:

$$f(\boldsymbol{x}) = \mathrm{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i) + b\right].$$

where $\kappa(\boldsymbol{x}, \boldsymbol{x}_i)$ is the corresponding nonlinear kernel function. In remote sensing images, classes are usually spherical shaped and the use of spherical RBF kernel is most appropriate. RBF kernels are of the form $\kappa(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathrm{e}^{-w|\boldsymbol{x}_1 - \boldsymbol{x}_2|^2}$. Again, the aforesaid two class SVM can easily be extended for multiclass classification by designing a number of one-against-all two class SVMs, e.g., a k-class problem is handled with k two class SVMs.

3. Active support vector learning for pixel classification

A limitation of the SVM design algorithm, described above, is the need to solve a quadratic programming (QP) problem involving a dense $l \times l$ matrix, where l is the number of points in the data set. Since most QP routines have quadratic complexity, SVM design requires huge memory and computational time for large data applications. Several approaches exist for circumventing the above shortcomings as well as to minimize the number of labeled points required to design the classifier. Many of them exploit the fact that the solution of the SVM problem remains the same if one removes the points that correspond to zero Lagrange multipliers of the QP problem (the nonSV points). The large QP problem can thus be broken down into a series of smaller QP problems, whose ultimate goal is to identify all of the nonzero Lagrange multipliers (SVs) while discarding the zero Lagrange multipliers (nonSVs). At every step, one solves a QP problem that consists of the nonzero Lagrange multiplier points from the previous step, and a number of other points queried. At the final step, the entire set of nonzero Lagrange multipliers has been identified; thereby solving the large QP problem. The active SVM design algorithm used here for pixel classification is based on the aforesaid principle. At each step the most informative point not belonging to the current SV set is queried along with its label; the goal is to minimize the total number of labeled points used by the learning algorithm. The method is described below and illustrated in Fig. 1. The steps need to be repeated k times for a k class problem with data from respective classes.

Algorithm 2:

Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_d]$ represent a pixel of a d-band multispectral image. Here, x_i is the grey value of the ith band for pixel $\boldsymbol{x}$. Let $A = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{l_1}\}$ denote the set of pixels for which class labels are known, and $B = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{l_2}\}$ the set of pixels for which class labels are unknown. Usually, $l_2 \gg l_1$. $SV(C)$ denotes the set of support vectors of the set C obtained using the methodology described in Section 2. $S_t = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ is the support vector set obtained after tth iteration, and $\langle \boldsymbol{w}_t, b_t \rangle$ is the corresponding separating hypersurface. Q_t is the point actively queried for at step t. The learning steps involved are given below:

Initial step: set $t = 0$ and $S_0 = \mathrm{SV}(A)$. Let the parameters of the corresponding RBF be $\langle \boldsymbol{w}_0, b_0 \rangle$.

While Stopping criterion is not satisfied:
  $Q_t = \{\boldsymbol{x} |\min_{\boldsymbol{x}\in B} \kappa(\boldsymbol{w}_t, \boldsymbol{x})\} + b$.
  Request label of Q_t.
  $S_t = \mathrm{SV}(S_t \cup Q_t)$.
  $B = B - Q_t$.
  $t = t + 1$.
End while

The set S_T, where T is the iteration at which the algorithm terminates, contains the final SV set representing the classifier.

Stopping criterion: $\min_{\boldsymbol{x}\in B} \kappa(\boldsymbol{w}_t \cdot \boldsymbol{x}) + b > 1$. In other words, training is stopped when none of the unlabeled points lie within the margin of the separating hypersurface.

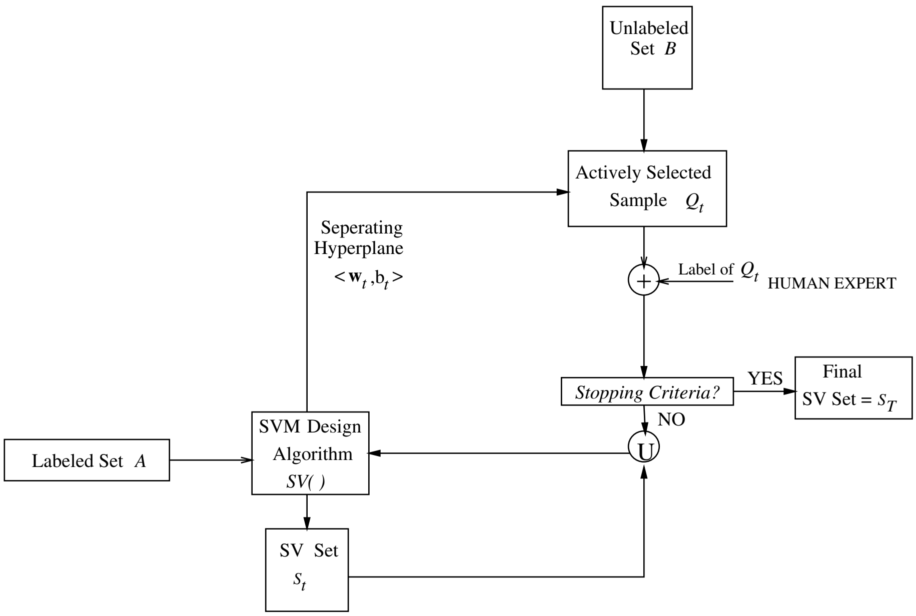

Fig. 1. Block diagram of the active SVM learning algorithm for pixel classification.

4. Experimental results and comparison

The multispectral image data, used in our experiment, contains observations of the Indian Remote Sensing (IRS) satellite for the city of Mumbai, India. The data contains images of four spectral bands, namely blue, green, red and infrared. The images contain 512×512 pixels and each pixel represents a 36.25 m $\times$ 36.25 m region.

Here the task is to segment the image into different landcover regions, using four features (spectral bands). The image mainly consists of six classes e.g., clear water (ponds), turbid water (sea), concrete (buildings, roads, airport tarmacs), habitation (concrete structures but less in density), vegetation (crop, forest areas) and open spaces (barren land, playgrounds). A labeled set (A) containing 198 points is initially used.

4.1. Algorithms compared

The performance of the active support vector learning algorithm (active SVM) is compared with the following multispectral image segmentation algorithms. Among them, methods SVM 1 and SVM 2 represent extreme conditions on the use of labeled samples. In SVM 1 the labeled set is very small in size but the labels are accurate, while in SVM 2 a large fraction of the entire data constitutes the labeled set, but the labels may be inaccurate. The k-means algorithm is a completely unsupervised scheme requiring no class labels.

(i) SVM 1: the conventional support vector machine, using only the initial labeled set as the entire design set.
(ii) k-means: the unsupervised k-means clustering algorithm.
(iii) SVM 2: the conventional support vector machine, using 10% of the entire set of pixels as the design set. The labels are supplied by the output of the k-means algorithm.

4.2. Evaluation criterion

The image segmentation algorithms are compared on the basis of the following quantities:

(i) Total number of labeled data points used in training (n_{labeled}).
(ii) Training time (t_{training}) on a Sun UltraSparc 350 MHz workstation.
(iii) *Quantitative cluster quality index* (β), β is defined as (Pal et al., 2000)

$$\beta = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^{\mathrm{T}}(X_{ij} - \bar{X})}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^{\mathrm{T}}(X_{ij} - \bar{X}_i)}, \tag{6}$$

where n_i is the number of points in the ith $(i = 1, \dots, k)$ cluster, X_{ij} is the feature vector of the jth pattern $(j = 1, \dots, n_i)$ in cluster i, $\bar{X}_i$ the mean of n_i patterns of the ith cluster, n is the total number of patterns, and $\bar{X}$ is the mean value of the entire set of patterns.

Note that the above measure is nothing but the ratio of the total variation and within-class variation. This type of measure is widely used for feature selection and cluster analysis (Richards, 1993; Pal et al., 2000). For a given image and k (number of clusters) value, the higher the homogeneity within the segmented regions higher would be the β-value.

4.3. Comparative results

The performances of different multispectral image segmentation methods are presented in Table 1. Among them, the proposed active SVM learning algorithm provides the best segmentation quality as measured by the β index. The SVM 1 algorithm provides the lowest β-value, which is expected since it uses a very small number of training samples. The unsupervised k-means algorithm also provides much lower β-value compared to the active SVM algorithm. The SVM 2 algorithm uses the labels generated by the k-means

Table 1
Comparative results for the IRS-1A image

Method	n_{labeled}	t_{training} (s)	β
active SVM	259	72.02 + (time for labeling 54 pixels)	6.35
SVM 1	198	28.15	3.45
k-means	0	1054.10	2.54
SVM 2	26,214	2.44×10^5	4.72

algorithm, but provides a relatively small improvement in performance compared to k-means. The visual quality of the classified images (Fig. 2) also reinforce these conclusion.

Among the supervised classification algorithms, namely, active SVM, SVM 1 and SVM 2, SVM 1 uses the least number of labeled samples and has minimum training time. However, the active SVM algorithm uses only 54 additional labeled points compared to SVM 1 with a substantial improvement in segmentation quality. This is due to the fact that the additional points queried by active SVM were the most informative ones and contributed to the increase in segmentation quality.

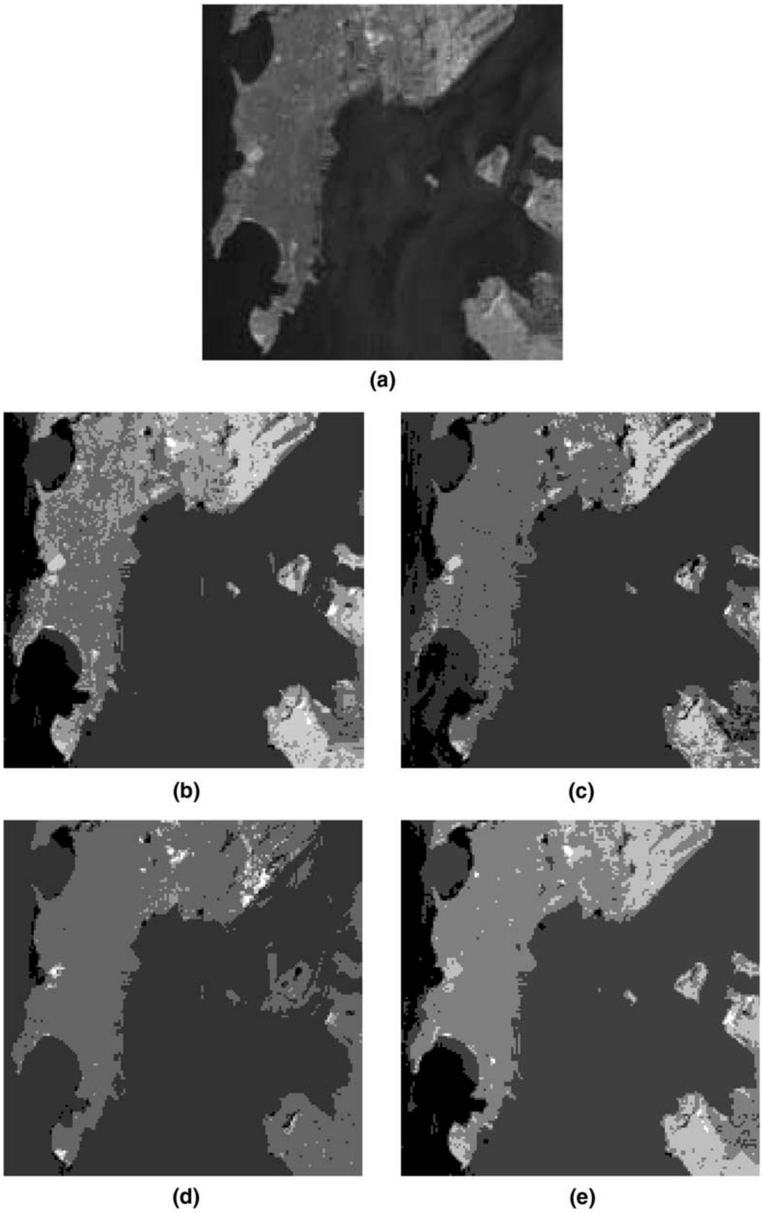

Fig. 2. IRS-1A: (a) original band four image; classified image using (b) active SVM, (c) SVM 1, (d) k-means, and (e) SVM 2.

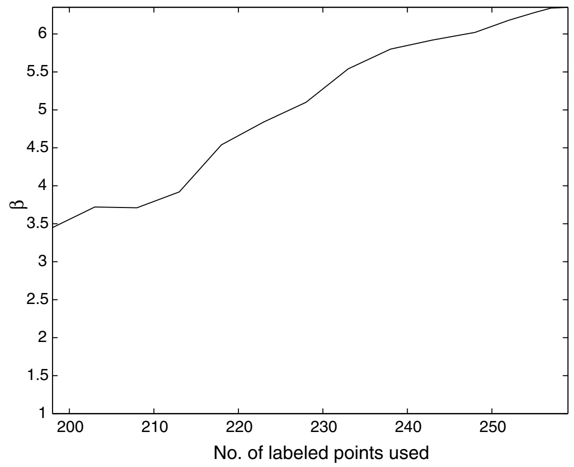

Fig. 3. Variation of β-value with the number of labeled data points used by the active SVM algorithm.

On the other hand, SVM 2 uses a large sized labeled set, consisting of randomly chosen points, for training. Since, accurate labels for the large training set used were not available, slightly inaccurate labels were used. The overall effect being: the performance of the SVM 2 algorithm is poorer compared to active SVM inspite of it requiring a much higher computation time.

The variation in segmentation quality (as measured by β index) with the number of labeled samples queried by the active SVM algorithm is shown in Fig. 3. It is seen that the initial SVM designed using the training set of SVM 1 provides a β-value of 3.45 which subsequently increases as more point are queried to a final value of 6.35.

5. Conclusions and discussion

We have presented an active support vector learning algorithm for supervised pixel classification in remote sensing images. The goal is to minimize the number of labeled points required to design the classifier. The algorithm uses an initial set of small number of labeled pixels to design a crude classifier, which is subsequently refined by using more number of points obtained by querying from a pool of unlabeled pixels. The class labels of the queried points are supplied by a human expert. It is seen that the number of labeled points required by the active learning algorithm is far less compared to the conventional support vector machine. It also provides better accuracy compared to completely unsupervised segmentation algorithms or a supervised algorithm having access to only inaccurate class labels of a large number of pixels.

The active learning strategy adopted in this article queries for the most interesting/ambiguous unlabeled point as measured by its distance from the current separating hypersurface. Other query strategies based on version space splitting, logistic regression may be used. Also, besides active learning, other semi-supervised learning techniques like transductive learning, co-training may also help in circumventing the problem arising from scarcity of labeled data in remote sensing image analysis.

The main goal of the active learning algorithm is to reduce the requirement of labeled pixels. Hence, an aggressive query strategy is adopted. However, the aggressive strategy is sensitive to wrong labeling by a human expert, resulting in performance degradation. If in some application, a higher number labeled pixels, with possibly few wrong labels, are available, a more conservative query strategy will provide better performance.

References

Angluin, D., 1988. Queries and concept learning. Machine Learning 2, 319–342.

Bandyopadhyay, S., Pal, S.K., 2001. Pixel classification using variable string genetic algorithms with chromosome differentiation. IEEE Transactions on Geoscience and Remote Sensing 39 (2), 303–308.

Baraldi, A., Parmiggiani, F., 1995. A neural network for unsupervised categorization of multivalued input pattern: An application to satellite image clustering. IEEE Transactions on Geoscience and remote Sensing 33, 305–316.

Brown, M., Lewis, H.G., Gunn, S.R., 2000. Linear spectral mixture models and support vector machines for remote sensing. IEEE Transactions on Geoscience and Remote Sensing 38 (5), 2346–2360.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2 (2), 1–47.

Campbell, C., Cristianini, N., Smola, A., 2000. Query learning with large margin classifiers. In: Proc. 17th Internat. Conf.

on Machine Learning. Morgan Kaufman, Stanford, CA, pp. 111–118.

Cannon, R.L., Dave, R., Bezdek, J.C., Trivedi, M., 1986. Segmentation of a thematic mapper image using fuzzy *c*-means clustering algorithm. IEEE Transactions on Geoscience and remote Sensing 24, 400–408.

Cohn, D., Atlas, L., Ladner, R., 1994. Improving generalization with active learning. Machine Learning 15, 201–221.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38.

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

Laprade, R.H., 1988. Split-and-merge segmentation of aerial photographs. Computer Vision Graphics and Image Processing 48, 77–86.

Mitra, P., Murthy, C.A., Pal, S.K., 2000. Data condensation in large databases by incremental learning with support vector machines. In: Proc. Internat. Conf. Pattern Recognition (ICPR2000), Barcelona, Spain. pp. 712–715.

Pal, S.K., Ghosh, A., Uma Shankar, B., 2000. Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. International Journal of Remote Sensing 21 (11), 2269–2300.

Pal, S.K., Mitra, P., 2002. Multispectral image segmentation using the rough set initialized EM algorithm. IEEE Transactions on Geoscience and Remote Sensing 40 (11), 2495–2501.

Richards, J.A., 1993. Remote Sensing Digital Image Analysis: An Introduction. Springer Verlag, New York.

Schohn, G., Cohn, D., 2000. Less is more: Active learning with support vector machines. In: Proc. 17th Internat. Conf. Machine Learning, Morgan Kaufman, Stanford, CA, pp. 839–846.

Tong, S., Koller, D., 2001. Support vector machine active learning with application to text classification. Journal of Machine Learning Research 2, 45–66.

Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.

Wong, Y.-F., Posner, E.C., 1993. A new clustering algorithm applicable to polarimetric and SAR images. IEEE Transactions on Geoscience and Remote Sensing 31 (3), 634–644.