Support Vector Machines: Optimization and Validation for Land Cover Mapping Using Aerial Images and Lidar Data

J.C.Trinder [a], M. Salah [b, *]

[a] School of Surveying and Spatial Information Systems, The University of New South Wales,
UNSW SYDNEY NSW 2052, Australia - j.trinder@unsw.edu.au
[b] Dept. of Surveying, Faculty of Engineering Shoubra, Benha University, 108 Shoubra Street, Cairo, Egypt – engmod2000@ yahoo.com

Abstract − **This work investigates the optimization and validation of Support Vector Machines (SVMs) for land cover classification from multispectral aerial imagery and lidar data. For the optimization step, a new method based on a curve fitting technique was applied to minimize the grid search for the Gaussian Radius Basis Function (RBF) parameters. The validation step was based on two experiments. In the first, four SVM kernel models (Gaussian Radius Basis Function; Linear; Polynomial; and Sigmoid) were tested and compared to each other. In the second, SVMs were compared against two classifiers of different characteristics (the Self-Organizing Map (SOM) and the Classification Trees (CTs)) based on four study areas with different sensor and scene characteristics. The comparison is based on two criteria: overall classification accuracy; and commission and omission errors per class. The results demonstrate: the higher overall classification accuracy; the lower range of commission and omission errors per class of the SVMs as compared to other classifiers.**

Keywords: Aerial Images, Lidar, GLCM, Attributes, SVMs, Optimization, Validation.

1. INTRODUCTION

Research on land cover classifcation from aerial images and lidar data has been fuelled in recent years by the need for data acquisition and updating. The high dimensionality of aerial and satellite imagery presents a challenge for traditional classification methods based on statistical assumptions. On the other hand Support Vector Machines (SVMs), as one of the more recent developments in the field of machine learning, have proved reliable and accurate in many pattern classification and nonlinear regression tasks. There have been few applications of SVMs for the analysis of aerial imagery and laser scanner data. Secord and Zakhor (2006) proposed a two step method for detecting trees from aerial imagery and lidar data consisting of segmentation followed by SVM-based classification. The results showed that the segmentation followed by classification outperforms the point-wise method. Haitao et al. (2007) applied an object-oriented SVM for land cover classification by fusing high-resolution aerial imagery, Digital Surface Model (DSM) and the calculated textural features (variance, contrast) of aerial imagery using Gray Level Co-occurrence Matrix (GLCM). The results showed that SVMs provided greater classification details and accuracy and the computational performance for classification was improved.

The motivation behind this study is that there are several issues requiring consideration in respect of the application of SVM for classifying aerial images and lidar data: (1) how can we limit the range of the inherent grid search of SVM parameters using an adequate approach in terms of complexity and time of computations; (2) which SVM kernel performs the best for image and lidar data fusion; (3) how accurate, in terms of overall classification accuracy and class-specific accuracy, are SVMs compared to statistical and neural classifiers?. The objective of this paper is to answer these questions, based on an intensive investigation of four SVM models using four datasets with different sensor and scene characteristics. The rest of the paper is organized as follows. The next section describes the study areas and data sources, followed by a brief description of the methods used. After that, the results are presented and evaluated. We summarise our results in the last section.

2. STUDY AREAS AND DATA SOURCES

2.1 Test Zones and Input Data

Four test datasets of different sensor and scene characteristics were used in this study as summarized in Table 1 and 2. Test area 1 is a part of the region surrounding the University of New South Wales campus, Sydney Australia, which is a largely urban area. The colour imagery was captured by film camera at a scale of 1:6000. The film was scanned in three colour bands (red, green and blue) in TIFF format, with 15µm pixel size (GSD of 0.09m) and radiometric resolution of 16-bit as shown in figure 1(*a*). Test area 2 is a part of Bathurst city, NSW Australia, which is a largely rural area. The colour (red, green and blue) images were captured by a Leica ADS40 line scanner sensor and supplied as an ortho image as shown in figure 1(*b*). Test area 3 is over suburban Fairfield, NSW Australia covering low density development, and large industrial buildings as shown in figure 1(*c*). The image data was acquired by a film camera at a scale of 1:10,000 which was scan digitized and supplied as an ortho image. Test area 4 is over Memmingen Germany, featuring a densely developed historic centre and industrial areas as shown in figure 1(*d*). Multispectral images (CIR), including an infrared image with the same resolution as the colour bands, were acquired by a line scanner sensor and supplied as an ortho image.
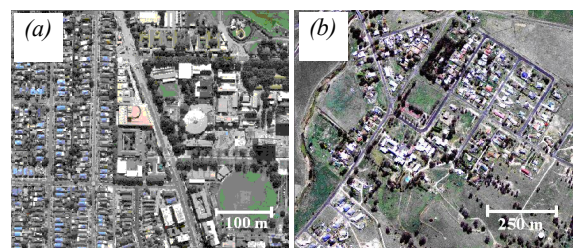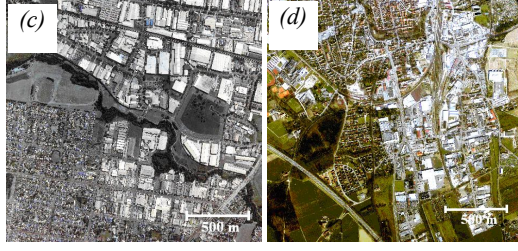

Figure 1. Orthophotos for: *(a)* UNSW; *(b)* Bathurst; *(c)* Fairfield; and *(d)* Memmingen.

Table 1. Characteristics of image datasets.

Test area	Size (Km)	bands	pixel size (cm)	Camera
UNSW	0.5 x 0.5	RGB	9	LMK1000
Bathurst	1 x 1	RGB	50	ADS40 Line scanner
Fairfield	2 x 2	RGB	15	LMK1000
Memmingen	2 x 2	CIR	50	TopoSys Falcon II line scanner

Table 2. Characteristics of lidar datasets.

	UNSW	Bathurst	Fairfield	Memmingen
	Optech ALTM 1225	Leica ALS50	Optech ALTM 3025	TopoSys
Spacing across track (m)	1.15	0.85	1.2	0.15
Spacing along track (m)	1.15	1.48	1.2	1.5
Vertical accuracy (m)	0.10	0.10	0.15	0.15
Horizontal accuracy (m)	0.5	0.5	0.5	0.5
Density (Points/m^2)	1	2.5	1	4
Sampling intensity (mHz)	11	150	167	125
Wavelength (μm)	1.047	1.064	1.047	1.56
Laser swath width (m)	800	777.5	700	750
Recorded pulse	1st and last	1st and last	1st and last	1st and last

2.2 Training and Reference Datasets

All tests were conducted using identical training sets. The training data for each test area consists of 1644, 1264, 1395 and 1305 training pixels for buildings, trees, roads and ground respectively for each band of the input data. Class "ground" mainly corresponds to grass, parking lots and bare fields.

In order to evaluate the accuracy of the results, reference data were captured by digitising buildings, trees, roads and ground in the orthophotos. In order to overcome the horizontal layover problem of tall objects such as buildings, roofs were first digitized and then each roof polygon was shifted if possible so that at least one point of the polygon coincided with the corresponding point on the ground. For Fairfield, the orthophoto and the lidar data correspond to different dates. Thus, we excluded from the analysis 41 building polygons that were only available in one data set. Larger areas covered by trees were digitised as one polygon. Information on single trees was captured where possible.

3. METHODOLOGY

The classification process was implemented in several stages as follow:

3.1 Filtering of lidar point clouds

First the original lidar point clouds were filtered to separate on-terrain points from points falling onto natural and human made objects. A filtering technique based on a linear first-order equation which describes a tilted plane surface has been used (Salah et al., 2009). Data from both the first and the last pulse echoes were used in order to obtain denser terrain data. After that, the filtered lidar points were converted into an image DTM, and the DSM was generated from the original lidar point clouds. Then, the nDSM was generated by subtracting the DTM from the DSM.

3.2 Generation of Attributes

Our experiments were carried out characterizing each pixel by a 32-element feature vector which comprises: 25 generated attributes, 3 image bands (R, G and B), intensity image, DTM, DSM and nDSM. The 25 attributes include those derived from the Grey-Level Co-occurrence Matrix (GLCM), Normalized Difference Vegetation Indices (NDVI), slope and the polymorphic texture strength based on the Förstner operator. The NDVI values for the UNSW, Bathurst and Fairfield test areas were derived from the red image and the lidar reflectance values, since the radiation emitted by the lidars is in the IR wavelengths. Since the images derived for the Memmingen dataset include an IR channel, the NDVI was derived from the image data only. The attributes were calculated for pixels as input data for the three classifiers. Table 3 shows the attributes and the images for which they have been derived. These attributes have been selected to be uncorrelated based on the problem of correlation between feature attributes. All the presented attributes were used for every test area. A detailed description of the filtering and generation of attributes process can be found in Salah et al. (2009).

Table 3. The full set of the possible attributes from aerial images and lidar data. √ and x indicate whether or not the attribute has been generated for the image. PTS refers to polymorphic texture strength; HMGT refers to GLCM/homogeneity; Mean refers to GLCM/ Mean; entropy refers to GLCM/ entropy.

attribute	Red Band	Green Band	Blue Band	Intensity	DSM	nDSM
PTS	√	√	√	√	√	√
HMGT	√	√	√	√	√	√
Mean	√	√	√	√	√	√
entropy	√	√	√	√	√	√
Slope	x	x	x	x	x	√

3.3 SVM Classification

SVMs are based on the principles of statistical learning theory (Vapnik, 1979). SVMs delineate two classes by fitting an optimal separating hyperplane (OSH) to those training samples that describe the edges of the class distribution. As a consequence they generalize well and often outperform other algorithms in terms of classification accuracies. Furthermore, the misclassification errors are minimized by maximizing the margin between the data points and the decision boundary. Since the One-Against-One (1A1) technique usually results in a larger number of binary SVMs and then in subsequently

intensive computations, the One-Against-All (1AA) technique was used to solve for the binary classification problem that exists with the SVMs and to handle the multi-class problems in aerial and lidar data. The Gaussian radial basis function (RBF) kernel has been used, since it has proved to be effective with reasonable processing times in remote sensing applications. Two parameters should be specified while using RBF kernels: (1) C, the penalty parameter that controls the trade-off between the maximization of the margin between the training data vectors and the decision boundary plus the penalization of training errors; (2) γ, the width of the kernel function. In order to estimate these values and to avoid making exhaustive parameter searches by approximations or heuristics, we used a grid-search on C and γ using a 10-fold cross-validation. SVMs were compared against two classifiers of different characteristics: the Self-Organizing Map (SOM) (Kohonen, 2001); and the Classification Trees (CTs) (Breiman et al. 1984).

3.4 Model Selection for SVMs

The problem is that there is no rule for the selection of the kernel's parameters and it is not known beforehand which are the best values of C and γ for the current problem. Both parameters C and γ depend on the data range and distribution and they differ from one classification problem to another. As a result there are an infinite number of possible pairs of parameters that could be taken into consideration. The most effective method to optimize the C and γ parameters is a grid-search using the cross-validation error as a measure of quality. Based on these facts, we used a grid-search on C and γ based on a 10-fold cross-validation. The cross-validation procedure can prevent overfitting problems and results in better accuracy (Hsu et al., 2009). Basically pairs of (C, γ) were tested and the one with the best cross-validation accuracy was selected. First we applied a coarse grid with ranges of values of [0.001, 0.01, 1,……, 10 000] for both C and γ. Then we applied a finer grid search in the neighbourhood of the best C and γ, obtained from the coarse grid, with ranges of values [(C or γ)-10, (C or γ) +10] and with interval of 0.01 to obtain a better cross-validation. Once C and γ have been specified, they were used with the entire training set, to construct the optimal hyperplane.

4. RESULTS AND DISCUSSIONS

4.1 Grid-search optimization

Although the grid-search approach achieves good model selection, resolution searches can result in a large number of evaluations. An intensive literature search in remote sensing, machine learning and pattern recognition fields showed that the issue of limiting the range of the SVM inherent grid-search has not been studied. In this section, we introduce a new approach, based on a curve fitting technique, for optimizing the grid-search for the kernel parameters. Since both parameters, C and γ, depend on the data range and distribution and in order to minimize the grid search range, a set of 22 statistics have been calculated for each group of attributes. Then, the correlation coefficients between each statistic and both of C and γ were computed to examine the strength and direction of the relationship between each statistic and C and γ, which demonstrates a relatively strong negative relationship between the sum of the input data and both of C and γ. We therefore minimized the range used for the grid search based on the inverse proportionality relationship between the sum of the input data and both of C and γ. First, three values of C and γ were selected (representing minimum sum, maximum sum and

the closest value to the mean sum). Then, a single-term exponential curve was fitted to these three values of C and γ as shown in Figure 2. Then the confidence bounds have been calculated for the fitted curve with a 99% level of certainty. This means that we will have a 99% chance that all the three values of C or γ are contained within the lower and upper confidence bounds. Figure 2 demonstrates that C and γ for the remaining 4 groups of attributes, shown by the crosses in the figure, are still also contained within the lower and upper prediction bounds produced by the three values of C and γ because of the correlation between C, γ and the sum. Since both C and γ are positive values, then the grid searches will be in the positive range of the confidence bounds. These results show that the sum can be an adequate indicator to minimize the grid search for both C and γ.
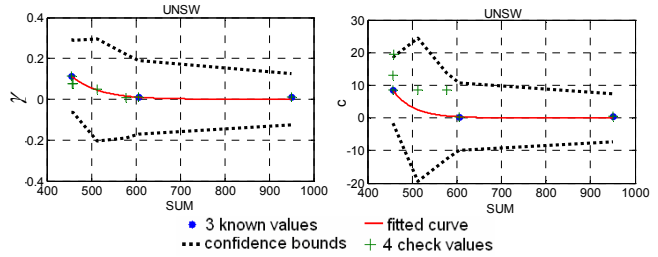

Figure 2. Minimizing the bounds of the grid search for UNSW test area.

4.2 Performance evaluation of SVMs kernels

Before comparing SVMs with SOM and CT, the four SVMs kernels were compared to each other in terms of overall classification accuracy to select the kernel with the best performance as a representative of SVMs. The previously described attributes were applied in seven separate *Groups* which include those from: *Red*, *Green* and *Blue* bands of the aerial image; *Intensity/IR* image; *DSM*; *nDSM* and the *Total Group* of attributes. Then, SVMs were performed 28 times for each test area (7 groups of attributes for 4 kernel functions). Classification results for the UNSW test area using the SVMs and the *Total Group* of attributes are shown in Figure 3.
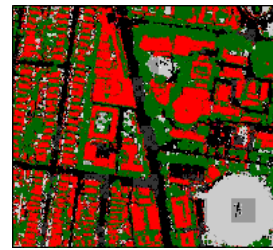

Figure 3. Classification results using the SVMs and the *Total Group* of attributes for UNSW test area. Red: buildings, green: trees, black: roads and grey: ground.

By comparing the classification results with the reference data, the contributions of each *Group* of attributes to the overall classification accuracy were computed and plotted against the *Groups* of attributes for the four test areas. Figure 4 is a typical example obtained for UNSW test area. Figure 4 shows that, using attributes generated from *Red Group*, *Green Group*, *Blue Group*, resulted in lower classification accuracies. In this case, RBF, Linear, Sigmoid and Polynomial kernels produced classification accuracies comparable to each other. On the other hand, incorporating attributes generated from lidar data, such as

Intensity/IR Group, DSM Group and *nDSM Group* of attributes into the classification process, has significantly improved the classification accuracies. In these cases, the classification accuracies produced by the four kernels are slightly different with RBF performing the best. The *Total Group* of attributes resulted in the highest classification accuracy, and the four kernels performed comparable to each other. These results provide evidence that, in some cases of land cover classification by aerial images and lidar data fusion, the selection of the kernel in SVMs is not a critical factor for overall classification accuracy. If only attributes from lidar data are used, the RBF kernel would be the best choice.
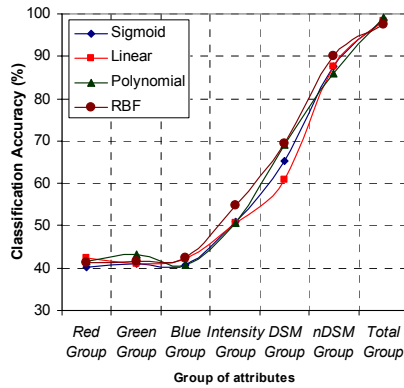

Figure 4. Contribution of each *Group* of attributes to the accuracy of the classification results in case of UNSW test area.

4.3 SVMs versus neural and statistical classifiers

In order to evaluate the ability of the SVMs to classify the test data sets, the classification accuracies derived by SOM and by CTs were compared to the accuracies derived by the SVMs, and based on the same reference data. For this comparison, the *Total Group* of attributes was used and the overall classification accuracies were computed for each classifier using the four data sets as shown in Table 4. Results show that classification results were slightly more accurate with the SVMs, 96.8%, than with the SOM, 95.9%, or with the CTs, 93.7%.

Table 4. Performance evaluation of all classifiers for the four test areas.

Test area	Classification accuracy (%)		
	SOM	CT	SVM
UNSW	96.8	95.05	96.9
Bathurst	95	92.85	96.5
Fairfield	96.8	96.15	97
Memmingen	95	90.75	96.6

Two additional measures were used to evaluate the performance of SVMs, namely: commission and omission errors. Unlike overall classification accuracy, commission and omission errors clearly show how the performance of the proposed method improves or deteriorates for each individual class. Commission errors are the percent of incorrectly identified pixels associated with a class, and omission errors are the percent of unrecognized pixels that should have been identified as belonging to a particular class. An assessment of the mean commission and omission errors for the four test areas confirms that SVMs and SOM produced commission and omission errors comparable to each other. On the other hand, CTs performed the worst in terms of commission and omission errors. One advantage of the SVMs over SOM and CTs is that the achieved commission and omission errors are less variable: whereas the

application of SOM resulted in standard deviation of 3.69 % and 8.55 % for commission and omission errors respectively, and the application of CTs resulted in standard deviation of 7.43 % and 5.70 % for commission and omission errors respectively, the application of SVM reduces these values to 3.22 % and 5.25 %.

5. CONCLUSION

The optimization and validation of SVMs for land cover classification based on the fusion of lidar, multispectral aerial images and 26 auxiliary attributes were presented. Four test areas in different urban environments, based on lidar data derived from different sensors and different vegetation types were used. A new method to minimize the grid search for the RBF kernel parameters, *C* and *γ*, based on a curve fitting technique was successfully applied. On the other hand, the results show that in some cases the selection of the kernel in SVMs is not a critical factor for overall classification accuracy. If only attributes from lidar data are used, the RBF kernel would be the best choice. The results also demonstrate that SVMs outperform both SOM and CT in terms of overall classification accuracy, and commission and omission errors.

REFERENCES

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (Editors). 1984. *Classification and Regression Trees*, Chapman & Hall, New York, 358 p.

Haitao, LI., Haiyan, GU., Yanshun, HAN., and Jinghui, YANG. 2007. Fusion of high-resolution aerial imagery and lidar data for object-oriented urban land-cover classification based on svm. *In Proceedings of the ISPRS Workshop on Updating Geo-spatial Databases with Imagery & The 5th ISPRS Workshop on Dynamic and Multi-dimensional GIS*, 28 - 29 August 2007, Urumchi, China. Edited by J. Jie and Z. Renliang, International Society for Photogrammetry and Remote Sensing. pp. 179-183.

Hsu, C.W., Chang, C.C., and Lin, C.J. 2009. A Practical Guide to Support Vector Classification. Department of Computer Science, National Taiwan University [online]. Available from < http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf > [cited (July 23, 2009)].

Kohonen, T. 2001. *Self-Organizing Maps*. Third Edition, Springer-Verlag, Berlin.

M. Salah, J. Trinder, and A. Shaker, "Evaluation of the self-organizing map classifier for building detection from lidar data and multispectral aerial images", *Journal of Spatial Science*, Vol. 54 No. 2, pp 15-34, 2009.

Secord, J., and Zakhor, A. 2006. Tree detection in aerial lidar and image data. *In Proceedings of the International Conference on Image Processing, ICIP 2006, October 8-11, Atlanta, Georgia, USA*. Edited by M. Hayes, Y. Altunbasak and G. AlRegib, Georgia Institute of Technology. pp. 2317-2320.

Vapnik, V. 1979. *Estimation of Dependences Based on Empirical Data* [in Russian]. Nauka, Moscow, 1979. (English translation: Springer Verlag, New York, 1982).