Support Vector Machine Experiments for Road Recognition in High Resolution Images

J. Y. Lai[a], A. Sowmya[a], J. Trinder[b]

[a]School of Computer Science and Engineering - (jlai, sowmya)@cse.unsw.edu.au
[b]School of Surveying and Spatial Information Systems - j.trinder@unsw.edu.au
University of New South Wales, Sydney, NSW 2052, Australia

Abstract. Support Vector Machines have received considerable attention from the pattern recognition community in recent years. They have been applied to various classical recognition problems achieving comparable or even superior results to classifiers such as neural networks. We investigate the application of Support Vector Machines (SVMs) to the problem of road recognition from remotely sensed images using edge-based features. We present very encouraging results from our experiments, which are comparable to decision tree and neural network classifiers.

1 INTRODUCTION

Road extraction from remotely sensed images is an important process in the acquisition and updating of Geographical Information Systems. Automatic and semi-automatic road recognition is an active area of research [7]. RAIL is a road recognition system that has been under development by our group for a number of years. It serves as a framework to research new directions in applying machine learning to image understanding

Support Vector Machines (SVMs) provide a relatively new classification technique that has grown from the field of statistical learning theory [11]. SVMs construct a hyper plane in the feature space that separates the positive and negative training samples. SVMs have been applied to many classic pattern recognition problems with great success including face recognition, hand-written character recognition and speech recognition [1]. In the domain of remote sensing, SVMs have been applied mostly to land cover classification. Camps-Valls et al. [2] use hyper spectral data of 128 bands to classify 6 types of crops. SVM yielded better outcome than neural networks. SVMs also performed reasonably well in situations where feature selection was not used. Pal and Mather [8] report that SVMs performed better than maximum likelihood, univariate decision tree and back-propagation neural network classifier, even with small training data sets. Both groups used pixel-based features.

There are two main motivations to incorporate SVMs into RAIL. First of all, SVMs have been successful in other application domains. However, there have been no results (prior to [13]) published on applying SVMs to the problem of road recognition. Therefore, our experiment will be of interest to pattern recognition communities as well as remote sensing researchers. Secondly, RAIL uses a meta-learning framework that facilitates model selection for classifiers, amongst other types of learning. Incorporating SVMs into RAIL expands the base algorithm sets to promote meta-learning research.

This paper is organised as follows. Section 2 describes implementation improvements on RAIL. Section 3 describes the experiment and the results are presented in Section 4. We summarise our results in Section 5.

2 RAIL

RAIL is an adaptive and trainable multi-level edge-based road extraction system which has been developed within our group for a number of years [10]. Starting with low-level objects (edges), RAIL incrementally builds higher-level objects (road network). The levels of classification are

1. Road Edge Pairs - pairs of edges that enclose a segment of road
2. Linked Road Edge Pairs - adjacent road edge pairs that form continuous roads
3. Intersections - road edge pairs that meet to form intersections
4. Road Network - linked roads and intersections.

SVM was applied to the preprocessing stage (edge extraction) and Level 1 of RAIL with encouraging results [13]. This paper extends the use of SVM to Level 2 while removing SVM use in the preprocessing stage. Several implementation improvements have been made to RAIL that affected the previous SVM experimentation. These include the image processing stage, the reference model, feature extraction and feature selection stages.

Image Processing: The parameters used in Vista's Canny edge-detector were tuned to produce outputs with less noise. This was accomplished by adding noise to the original image prior to a Gaussian smoothing function with a large standard deviation. Adding artificial noise to our images before blurring removes very small features such as noise that are present in high resolution images. The improvement was a dramatic decrease in the number of extracted edges, up to 90% less in several images, which meant that SVM could be used to learn Level 1 data without an additional SVM preprocessing. Removing this preprocessing stage gives results that can be compared to other algorithms in RAIL which also do not use any additional preprocessing stage. Another advantage is the reduction in misclassification during the SVM preprocessing stage (approximately 14%) so that a more complete road network can be recovered at higher levels.

Reference Model: RAIL has recently adopted a centreline reference model based on Wiedemann et al. [12] which can assess the learned outputs more correctly by checking that the extracted pairs have edges that do in fact lie opposite each other near the reference model. Previously we used an edge based model which produced a slightly more modest value in assessing the correctness of the outputs.

Feature Extraction: Additional features have been added to Level 1 and Level 2 (see Table 1) and a relevant subset from each level was selected by using feature subset selection (FSS) methods, which is described in section 3.4. The highlighted entries are the feature subsets that were discovered. Descriptions of the features may be found in [6]. *Selected Level 1 Features: Pair width, enclosed intensity (mean), bearing and projection* form an intuitive feature subset that describes road segments, i.e. roads have similar width and intensity and their opposite sides are almost parallel. Pair length is a good feature because in our preprocessing stage we have set a maximum length for

Table 1. Extracted Features

Level 1	Level 2
Width (mean)	Width (mean)
Enclosed Intensity (mean)	Width (var)
Enclosed Intensity (var)	Width Difference
Pair Length (centreline)	**Enclosed Intensity (mean)**
Length Difference	Enclosed Intensity (var)
Bearing Difference	**Enclosed Intensity Difference**
Intensity Gradient Difference	**Gap Intensity (mean)**
Projection	**Gap Intensity (var)**
	Length Combined
	Length Difference
	Minimum Gap Separation
	Maximum Gap Separation
	Gap Separation (mean)
	Bearing Difference
	Intensity Gradient Difference (left)
	Intensity Gradient Difference (right)

edges. Generally road sides are long and continuous and get split into smaller segments after preprocessing. When road pairs are formed their lengths do not vary too much. This is because non-road edges are usually of shorter length.

Enclosed intensity variance did not prove to be a good feature since the area enclosed by an edge pair is small and the intensity is fairly similar. Length difference between edges was also discarded by FSS. We expect road pairs to have similar edge length but non-road pairs maybe also have similar edge lengths, thus it does not convey much information. Intensity gradient difference between the two edges do not show consistencies between road pairs and non-road pairs. The assumption that the intensity levels are the same on both the external sides of the road is invalid.

Selected Level 2 Features: At Level 2, linked road pairs should have similar *enclosed intensity* with little *difference*. Ideally linked pairs should be minimally separated and have no gap, thus *gap intensity* and *gap separation* are excellent features to distinguish between linked road pairs and other linked edge pairs. Roads generally have smooth curves except at an intersection, therefore the *bearing difference* between linked road pairs should not be very large.

Width features are not good attributes for Level 2 because Level 1 outputs all have similar widths. The same argument applies to length attributes. *Enclosed intensity variance* and *gap intensity variance* are not very good features for the same reason discussed earlier, i.e. intensity level do not change much in enclosed edge pair or in a road gap. Again, intensity levels across edges cannot be assumed to be the same on both sides of the linked edge pairs.

Feature Subset Selection: The goal of FSS is to choose the most relevant features for classification, in other words, removing irrelevant attributes that may distract a machine learning algorithm. We compiled 9 sets of data from our images. 7 were from individual images and 2 were random selections from all the images. The sample size ranges from 130 to 360 examples in each set. We did not use one large test set since we had different road types and having one set of data might cause the result to be biased towards the most frequent road type.

The Weka data mining suite (version 3.4) [1] was used to conduct the FSS experiments. The FSS algorithms are grouped into *Type I*, consisting of correlation-based, classifier and wrapper algorithms, and *Type 2*, consisting of Chi squared, Relief, Information Gain, Gain Ratio and Symmetrical uncertainty algorithms. Type I algorithms select the 'best' subset of features. The frequency of each attribute was recorded and averaged. Type II algorithms rank the individual attributes by assigning them a weighting. These were normalised and averaged.

We wanted to select a subset of features which have a high frequency score in Type I and a high weighting in Type II. We ranked the Type I and Type II results and picked the smallest subset where the features are the same in each type. For example, if the top 4 attributes in Type I and Type II are the same disregarding their relative ranking position, then we would have a subset of 4 features. This has produced good classification results.

Features Versus Heuristic Preprocessing: Although we are using new image processing parameters to produce less noisy outputs, we are still dealing with fairly large datasets for Level 1 (C_2^n), as each edge can be paired up with every other edge and the ordering is irrelevant). Thus we use heuristic preprocessing to reduce the data size so that it becomes more manageable. We do not use heuristic rules for Level 2 since the data size is comparatively smaller than Level 1.

The heuristic rules throw away cases where an expert would agree that a positive classification is impossible. For example, in Level 1 we used the heuristic that if edges in an edge pair do not project onto each other, then they cannot be classified as an edge pair, since they are not opposite each other. Because this feature has a binary output, by using this attribute as a heuristic filter we have effectively removed projection from the feature space, since the heuristic rule outputs only those edge pairs that do project on to each other. We also have a heuristic rule that leaves out any edge pairs that are wider than twice the maximum road width in the images. We have effectively reduced the feature space that SVM would need to learn from.

Theoretically this should not make any difference to machine learning algorithms because the data we are leaving out have no influence on how the classes are separated. For SVMs, the data points discarded are distant from the class separation region and the support vectors, thus the construction of the separation hyper plane is independent of them.

Dataset: Seven high resolution aerial images were used in the experiment. Image A and B are from a rural area in France.These images have a ground resolution of 0.45m/pixel. The other 5 images are from a rural area in Morpeth in Australia. These images have a ground resolution of 0.6m/pixel. The image properties are given in Table 2.

A total of 333 and 227 positive and negative examples were selected from the images (some images contain more examples) for Level 1 and Level 2 respectively. The heuristic preprocessing outputs serve as inputs data for Level 1, and the Level 1 outputs

[1] Software available at `http://www.cs.waikato.ac.nz/ml/weka`

feed into Level 2. The size of the test data ranges from 2400 to 11200 instances for Level 1 and between 1500 to 18200 instances for Level 2.

Since we only had seven images to experiment with, we used 7-fold cross validation technique (leave-one-out) for evaluating the learned output, i.e. we train using six images and test on the unseen image. Note however that at the edge pair and twin linked edge pair level where the learning takes place, we have thousands of instances in each image.

Table 2. Image Properties

Image	Dimensions	No. of Edges
A	776*289	1530
B	757*821	3055
C	1500*848	2912
D	1700*1300	3290
E	1400*1300	1858
F	1400*1200	3893
G	1600*1100	3204

3 EXPERIMENTAL DESIGN

SVM experiments have been conducted on Level 1 and Level 2 of RAIL (the level references are different to those in [13]). The SVM implementation used was changed to LIBSVM [2] (version 2.4) which offers more in terms of tools and programming interfaces.

The training data and test data were scaled to [-1,1] to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during SVM calculations [3].

We used five different kernels for training SVMs for Support Vector Classification (C-SVC). They can be separated into two categories: Polynomial and Radial Basis Function (RBF). The polynomial kernels are of the first, second and third degree (with default C=1). The RBF kernels are standard (RBFs, C=1, γ=1) and optimised (RBFo, C, γ picked by a grid search function provided by LIBSVM). C is the penalty parameter that controls the margin and hence the over fitting of data, and γ is an internal variable for RBF.

The SVM kernels are compared to two well known classifiers within Weka, namely decision tree (DT) and neural network (NN), with default settings.

[2] Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

4 EXPERIMENTAL RESULTS

The metrics used to evaluate the results are taken from Wiedemann et al. [12]. They address two questions: 1. How complete is the extracted road network, and 2. How correct is the classification. They are calculated to percentage values, given by:

$$completeness = \frac{length_{TP}}{length_{reference}} \tag{1}$$

$$correctness = \frac{length_{TP}}{length_{classifed}} \tag{2}$$

Completeness measures the percentage of the road reference as detected by SVM. Correctness measures the percentage of the SVM classification that are actual road pairs. A high completeness means that SVM has extracted most of the road network, whereas high correctness implies that SVM has not classified too many incorrect road pairs.

We combine the two measures above into a more general measure of the quality. We call this *cxc* which is expressed as:

$$cxc = completeness^2 * correctness \tag{3}$$

Clearly, this measure is biased towards completeness. RAIL uses the output of Level 1 as the input of Level 2, so it is more important to have high completeness at the lower levels for input to higher levels. For example, Level 2 will only be as complete as its input (Level 1 output). Higher correctness value will result as higher levels discard non-road pairs.

Tables 3 shows the classification results (rounded to nearest percent) for Level 1 and Level 2. The completeness (cmp), correctness (cor) and *cxc* are shown for each classifier on each image. The entry with the highest *cxc* for each image in Level 1 is used as input to Level 2. The highest *cxc* obtained by SVM classifier has been highlighted for each image. Fig. 1 to Fig. 4 shows the SVM results visually for Image F. The images consist of the road reference, the low-level edges as inputs and the best Level 1 and Level 2 outputs.

Some of the completeness values are a little over 100%, this is because the centreline reference model uses a buffer zone both to the left and to the right of the road reference. Although the buffer width is only set to 3 pixels on either side, on some noisy road sections, two or more edges maybe measured as true positives for that same section. However, this is only true in a few cases. In all images with completeness greater than 100%, detailed analysis show that more than 98% of the reference road network is recognised.

Level 1 SVM classifiers have an average of 97% completeness and 35% correctness. Level 2 SVM classifiers have an average of 90% completeness and 49% correctness. These results are very encouraging because high completeness values are obtained. Clearly, the polynomial kernel of degree 3 and the optimized RBF kernel outperform the other kernels except for Image E. Additionally, the SVM classifiers compare well to DT and NN classifiers. In most cases, the results are very similar, though on images containing dirt roads in Level 1 (Image E and F), SVM classifiers appear to outperform both DT and NN, see Table 3.

Fig. 1. Image F - Road Reference

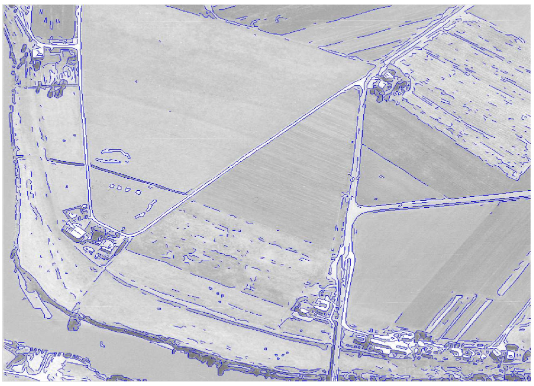

Fig. 2. Image F - Input

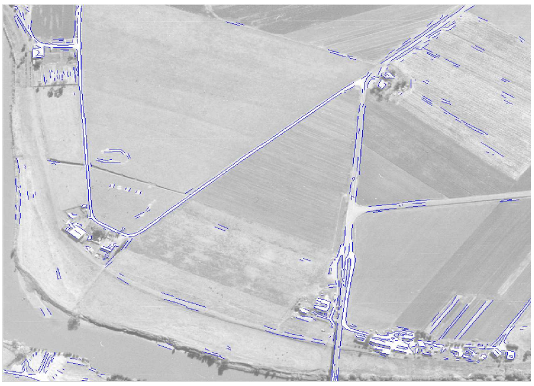

Fig. 3. Image F - Level 1 output

Table 3. Classification Results

	Classifier	L1 Cmp	L1 Cor	L1 cxc	L2 Cmp	L2 Cor	L2 cxc
A	Poly. 1	101	32	34	98	53	50
A	Poly. 2	100	32	33	98	54	51
A	**Poly. 3**	**101**	**37**	**38**	98	53	51
A	RBFs	101	35	36	98	54	51
A	**RBFo**	101	35	36	**98**	**54**	**52**
A	DT	100	31	31	98	54	51
A	NN	101	36	37	98	54	51
B	Poly. 1	107	34	39	105	51	57
B	Poly. 2	96	18	17	105	51	58
B	**Poly. 3**	94	39	34	**105**	**54**	**60**
B	RBFs	103	36	38	105	52	57
B	**RBFo**	**108**	**36**	**42**	105	53	59
B	DT	102	29	30	105	52	57
B	NN	107	39	44	105	51	56
C	Poly. 1	92	26	23	81	41	27
C	Poly. 2	89	21	17	81	41	27
C	Poly. 3	87	33	25	81	41	27
C	RBFs	94	29	25	81	41	27
C	**RBFo**	**94**	**30**	**27**	**81**	**42**	**27**
C	DT	92	29	25	81	41	27
C	NN	92	31	26	81	41	27
D	Poly. 1	100	22	21	98	30	28
D	Poly. 2	98	23	23	98	30	29
D	**Poly. 3**	97	23	22	**98**	**30**	**29**
D	RBFs	99	23	22	98	30	28
D	**RBFo**	**99**	**24**	**24**	98	30	29
D	DT	97	27	26	98	30	29
D	NN	98	26	25	98	30	28
E	Poly. 1	85	47	34	84	70	49
E	Poly. 2	80	43	28	84	69	48
E	Poly. 3	81	56	37	84	66	46
E	**RBFs**	**91**	**56**	**46**	**84**	**70**	**49**
E	RBFo	87	57	43	84	70	49
E	DT	37	41	6	84	70	49
E	NN	51	55	15	84	70	49
F	Poly. 1	83	32	22	68	54	25
F	Poly. 2	91	27	22	68	55	25
F	**Poly. 3**	**88**	**37**	**29**	**68**	**55**	**25**
F	RBFs	88	35	27	68	55	25
F	RBFo	84	33	24	67	55	25
F	DT	70	36	18	67	55	25
F	NN	73	43	23	68	54	25
G	Poly. 1	98	35	34	99	42	40
G	Poly. 2	97	30	28	98	43	40
G	**Poly. 3**	**96**	**39**	**36**	**99**	**43**	**41**
G	RBFs	100	38	38	99	42	40
G	RBFo	100	38	37	99	42	41
G	DT	92	31	27	96	42	39
G	NN	95	34	30	99	42	41

Fig. 4. Image F - Level 2 output

The low correctness value in Level 1 does not worry us. One of the major causes of the large number of false positives is that SVM classified road pairs have similar road properties, but only a fraction of them are represented by the centreline road reference. The others appear to be driveways and crop separations (perhaps for tractors) which are non-roads, but picked up well by the classifiers. The other main reason is that road properties may vary slightly between different images. SVMs learn these variations to a certain degree and thus the classified output may contain a range of road properties, some of which might be non-roads depending on the images.

Some images had lower completeness in Level 2, particularly Images C, E and F. The main causes of this are 1) because the road is very similar to its surroundings (especially roads with lower intensity), which means edges are not extracted well, and 2) dirt roads have been misclassified in Level 2 since the edge pairs are not closely linked. Fig. 4 is a good example where narrower roads with high intensity have been detected while wider and lower intensity roads have been missed. This problem can be fixed by applying a further preprocessing stage before edge extraction, e.g. multilevel thresholding/segmentation or by using an ensemble of SVMs and combining the results.

5 CONCLUSIONS

In this paper we have experimented with SVM and road extraction using edge-based features, which is significantly different from other SVM experiments in the remote sensing domain. The results for Level 1 and Level 2 are very encouraging and comparable to decision trees and neural networks. We plan to extend SVM to level 3 of RAIL which currently uses a relational learning algorithm to recognise the attributes of junctions [9].

The current experiments also show that it is feasible to select a suitable kernel that is best for the data. In the future we plan to experiment with other kernel functions and apply meta-learning techniques to find the best kernel and the parameters that are associated with them (Gold and Sollich, 2003).

References

1. Byun, H., Lee, S., 2003. A survey on pattern recognition applications of support vector machines. *IJPRAI*, 17(3), pp. 459-486.
2. Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J. Soria-Olivas, E., Martin-Guerrero, J. D., Moreno, J., 2003. Support Vector Machines for Crop Classification Using Hyper spectral Data. *Proc. IBPRIA 2003*, pp. 134-141.
3. Chang, C., Lin, C., 2001. LIBSVM: A library for support vector machines. Available at `http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf` (accessed 28 Feb. 2004).
4. Chen, A., Donovan, G., Sowmya, A., Trinder, J., 2002. Inductive Clustering: Automating low level segmentation in high resolution images. *Proc. ISPRS Commission III Symp. Photogrammetric Computer Vision*, pp. 73-78.
5. Gold, C., Sollich, P., 2003. Model selection for support vector machine classification. *Neurocomputing*, pp. 221-249.
6. Lai, J., Sowmya, A. and Trinder, J., 2004. Support Vector Machine Experiments for Road Recognition in High Resolution Images, Tech. Report, School of Comp. Sci. and Engg, University of New South Wales, UNSW CSE TR 0413.
7. Mena, J. B., 2003. State of the art on automatic road extraction for GIS update: A novel classification. *Pattern Recognition Letters*, 24(16), pp. 3037-3058.
8. Pal, M., Mather, P. M., 2004. Assessment of the effectiveness of support vector machines for hyper spectral data. *Future Generation Computer Systems*, Article in Press.
9. Nguyen, H. S., Sowmya, A., Trinder, J., 2005. Improved Road Junction Recognition in High Resolution Images using Relational Learning. (personal coomunication).
10. Sowmya, A., Singh, S., 1999. RAIL: Extracting road segments from aerial images using machine learning. *Proc. ICML 99 Workshop on Learning in Vision*, pp. 8-19.
11. Vapnik, V., 1995. The Nature of Statistical Learning Theory. Spring-Verlag, New York.
12. Wiedemann, C., Heipke, C., Mayer, H., Jamet, O., 1998. Empirical Evaluation of Automatically Extracted Road Axes. In *CVPR Workshop on Empirical Evaluation Methods in Computer Vision*, pp. 172-187.
13. Yager, N., Sowmya, A., 2003. Support Vector Machines for Road Extraction from Remotely Sensed Images. *Proc. CAIP 2003*, pp. 285-292.