A CLASSIFICATION APPROACH TO FINDING BUILDINGS IN LARGE SCALE AERIAL PHOTOGRAPHS

C.J. Bellman[1] , M.R. Shortis[2]

[1]Department of Geospatial Science, RMIT University, Melbourne 3000, Australia - Chris.Bellman@rmit.edu.au
[2]Department of Geomatics, University of Melbourne, Parkville 3052, Australia – Mark.Shortis@rmit.edu.au

Commission III, WG III/4

KEY WORDS: Building Recognition, Classification, Learning, Neural Networks

ABSTRACT:

Automatic building extraction remains an open research problem in digital photogrammetry. While many algorithms are proposed for building extraction, none of these solve the problem completely. One of their limitations is in the initial detection of the presence or absence of a building in the image region. One approach to the initial detection of buildings is to cast the problem as one of classification, where the image is divided into patches that either contain or do not contain a building. Support Vector Machines (SVMs) are a relatively new classification tool that appear well suited to this task. They are closely related to other machine learning techniques such as neural networks but have a stronger base in statistical theory and produce a generalised solution to the classification problem, using the principles of structural risk minimisation. They have been used successfully in other image classification and object recognition problems. Due to the high resolution of digital aerial photographs, compression and characterization of the image content is an essential part of the process. While many methods are available for this, an over-sampled, multi-resolution form of the Haar wavelet is a simple and convenient method for establishing coefficients that can be used in the machine learning phase. A Support Vector Machine (SVM) was trained on a large sample of building and non-building examples and achieved very high training accuracy. The trained SVM was then used to classify previously unseen image patches from the same photography and achieved an accuracy of more than 80%. Images from a variety of other sources were also tested and the classification accuracy was again consistently high. The relative simplicity of the method and the high success rates suggest that these techniques are quite promising for use in the initial detection of buildings in aerial images and may be a useful adjunct to the suite of algorithmic tools employed for building recognition and extraction.

1. INTRODUCTION

The use of digital imagery has enabled the automation of many traditional photogrammetric tasks, such as the measurement of fiducial marks, terrain extraction, relative orientation and aerotriangulation. One task that has proved difficult to automate is the extraction of buildings. At first glance, this seems a simple task, due to the distinct characteristics of building features such as parallelism and orthogonality. In practice, despite extensive research effort, the problem remains poorly understood (Schenk, 2000).

Many algorithms have been developed for building extraction from aerial imagery, largely as a result of several important research projects. The topic of building extraction has also been central to three research workshops held at Ascona, Switzerland in 1995 (Gruen *et. al.*, 1995), 1997 (Gruen *et. al.*, 1997) and 2001 (Baltsavias *et. al.*, 2001) on the automatic extraction of man-made features from aerial and space images.

Object extraction from digital images requires the initial identification of an object, usually through image interpretation and classification, and the precise tracking of the boundary to determine its outline (Agouris *et. al.*, 1998). Much of the research in photogrammetry has focused on the second of these tasks and left the first to a combination of low-level processing routines, such as edge detection, and human operator involvement, where the operator identifies suitable image patches over which to apply the algorithms. In these semi-automated systems, the human operator performs a task that is analogous to the pre-attentive stage of vision and 'finds' the object within the image space. The photogrammetric algorithms are then applied to determine the exact boundary of the object and extract it. Examples of these systems can be found in Henricsson (1996) and Michel *et. al*. (1998).

Other approaches to finding suitable candidate patches make use of ancillary data such as digital surface models (Zimmermann, 2000), multi-sensor and multi-spectral data (Schenk, 2000) and geographic information system databases (Agouris *et. al.*, 1998). Simple image segmentation methods are not sufficiently robust for building extraction (Nevatia *et. al.*, 1997), although some recent work based on splitting and merging appears to lead to good image segmentation (Elaksher *et. al.*, 2003).

An alternative to the algorithmic approach to finding candidate patches is to treat the problem as one of machine learning and 'train' the system recognize patches of image that contain a building. The research presented in this paper explores the use of image processing and machine learning techniques to identify candidate image patches for building extraction.

The approach presented is fairly simple. The candidate patches are preprocessed using wavelet techniques to obtain a set of image coefficients. These coefficients are then used train a classifier to distinguish between coefficients associated with a building patch and those of a non-building patch.

2. MACHINE LEARNING

Machine learning techniques are popular strategies in many image analysis and object recognition applications. They are often based on connectionist systems such as neural networks or support vector machines (SVM). In photogrammetry, machine learning techniques have been applied to road extraction (Sing & Sowmya, 1998), knowledge acquisition for building extraction (Englert, 1998) and for landuse classification (Sester, 1992). Neural techniques have been used in feature extraction (Li *et. al.*, 1998), stereo matching (Loung & Tan, 1992) and image classification (Israel & Kasabov, 1997).

Where recognition is involved, the task is generally treated as a problem of classification, with the correct classifications being learnt from a number of training examples. When the images are small (i.e. have few pixels), a direct connection approach can be employed, with each image pixel directly connected to a node in the network. For digital aerial photographs, such an approach is not feasible due to the large number of pixels involved and the combinatorial explosion that would result. To overcome this, a preprocessing stage is required to extract key characteristics from the image. Many such strategies for preprocessing are available, such as edge detection (Canny, 1986), log-polar forms (Grossberg, 1988) and texture segmentation (Lee & Schenk, 1998).

2.1 Wavelet Processing

One approach to preprocessing that has some attractive properties is wavelet processing. Although often associated with image compression, wavelets also have useful properties for the characterization of images. Of particular interest are the multi-resolution representations that highlight both strong edges and patterns of texture (Mallat, 1989). Some psycho-physical experiments support the idea that mammalian vision systems incorporate many of the characteristics of wavelet transforms (Field, 1994).

A combination of wavelet processing (for the preprocessing phase) and support vector machines (for the learning phase) has been used successfully in system to recognize the presence of a pedestrian in a video image (Papageorgiou *et. al.*, 1998; Poggio & Shelton, 1999) and for face recognition (Osuna *et. al.*, 1997). The images used in these studies have many of the characteristics found in aerial images such as clutter, noise and occlusions and so this approach seems worthy of further exploration.

Although many complex forms of wavelet processing are available, for this research a simple Haar transform was used. This is a step function in the range of 0-1 where the wavelet function $\psi(x)$ is expressed as:

$$\psi(x) := \begin{cases} 1 \text{ for } 0 \le x < 1/2 \\ -1 \text{ for } 1/2 \le x < 1 \\ 0 \quad \text{otherwise} \end{cases} \qquad (1)$$

The wavelet transform is computed by recursively averaging and differencing the wavelet coefficients at each resolution (Figure 1).

The Haar basis is a discrete wavelet transform. Despite having the advantage of being easy to compute, it is not well suited to many image analysis problems because it does not produce a dense representation of the image content and is sensitive to translations of the image content. To overcome these limitations, an extension of the Haar wavelet can be applied that introduces a quadruple density transform (Papageorgiou *et. al.*, 1998; Poggio & Shelton, 1999). In a conventional application of the discrete wavelet transform, the width of the support for the wavelet at level n is 2^n and this distance separates adjacent wavelets. In the quadruple density transform, this separation is reduced to ¼ 2^n (Figure 2(c)). This oversamples the image to create a rich set of basis functions that can be used to define object patterns.

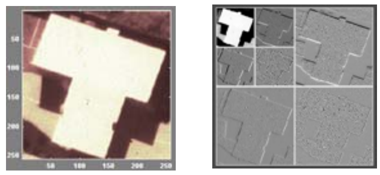
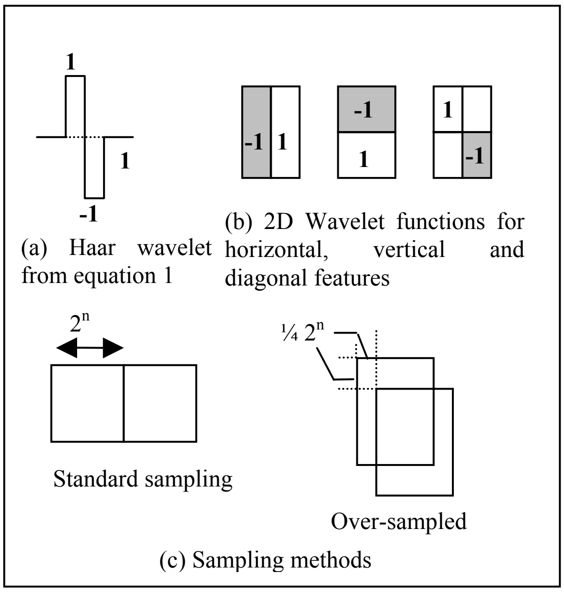


Figure 1. (a) Building image and (b) Haar wavelet compressed image and coefficients

2.2 Support Vector Machines



(a) Haar wavelet from equation 1

(b) 2D Wavelet functions for horizontal, vertical and diagonal features

(c) Sampling methods

Figure 2. The Haar wavelet characteristics (after (Papageorgiou *et. al.*, 1998))

The Support Vector Machine (SVM) is a relatively new tool for classification and regression problems. It is based on the principles of structural risk minimization (Vapnik, 1995) and has the attractive property that it minimizes the bound on the generalisation error of the learning solution rather than minimizing the training error. It is therefore not subject to problems of local minima that may occur with many neural network classifiers such as multilayer perceptrons.

SVMs work by finding a separating hyperplane between two classes. In a binary classification problem, there could be many hyperplanes that separate the data. As shown in figure 3, the optimal hyperplane occurs when the margin between the classes is maximised. In addition, only a subset of the data points will be critical in defining the hyperplane. These points are the support vectors.

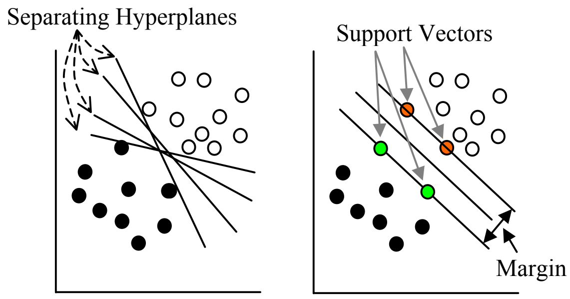


Figure 3. (a) Possible separating hyperplanes; (b) Optimal separating hyperplane

Another attractive property of the SVM is that its decision surface depends only on the inner product of the feature vectors. As a result, the inner product can be replaced by any symmetric positive-definite kernel (Cristianini & Shawe-Taylor, 2000). The use of a kernel function means that the mapping of the data into a higher dimensional feature space does not need to be determined as part of the solution, enabling the use of high dimensional space for the learning task without needing to address the mathematical complexity of such spaces. This offers the prospect of being able to separate data in high dimensional feature space and find classifications that were not possible in simple, lower dimensional spaces (Figure 4).

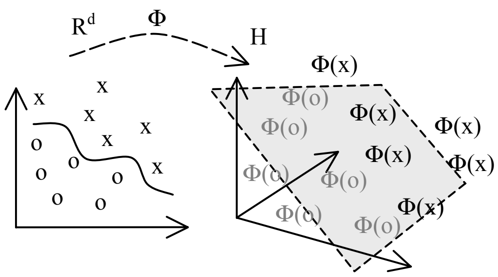


Figure 4. Mapping data into a higher feature space can make the data easier to separate

3. TEST DATA

Several datasets exist in the public domain for use in building extraction. Two of the most commonly used are the Avenches dataset (Henricsson, 1996) and the Fort Hood dataset (1999). Another dataset from the town of Zurich Hoengg was added to the public domain for the Ascona 2001 workshop (Hoengg dataset, 2001). The small number of buildings in these images makes these datasets unsuitable as the basis for research using a learning machine like SVM. As the learning machine is trained by example, a large number of examples of each object class must be presented to the learning machine to ensure valid learning. These public domain datasets simply do not contain enough data for this purpose.

To generate sufficient training data, a new database of images was created for the purposes of this research. Several large-scale aerial photographs of the city of Ballarat, in central Victoria, were available for this task. The images were acquired at a scale of 1: 4000, originally for the purpose of asset mapping within the Ballarat city centre. As such, they were taken in a standard stereo-mapping configuration, with a near vertical orientation and a 60% forward overlap.

Three images from this set were scanned from colour diapositives on a Zeiss Photoscan™ 1 at a resolution of 15 microns. The resultant ground sample distance for the images was 6 cm. This compares well to a ground sample distance of 7.5cm for the Avenches dataset and 7 cm for the Zurich Hoengg data. The Zeiss scanner produces a separate file for each colour band of the image (red-green-blue (RGB)). These files are produced in a proprietary format and were converted into uncompressed three-colour 24-bit Tagged Image Format File (TIFF) files for ease of use with other systems.

3.1 Image patches

In order to train the classifier and test whether effective class discrimination was possible, the classification problem was simplified by producing discrete image patches of a regular size. Each patch was 256 pixels by 256 pixels and contained either a single building or non-building piece of the image. The recognition problem was simplified further by limiting the building patches to those containing single, detached residential houses, where the extent of the house fitted completely within the 256 x 256 pixel area. This may seem extremely restrictive but the problem of building extraction has proven to be very difficult and a generalised solution appears unlikely at this stage. In a classification approach, it is likely that there will be a class for each category or type of building i.e. residential detached, residential semi-detached, commercial, industrial and so on. As this area appears largely unexplored, the scope of the classification was limited to a very specific case to increase the chances of success.

The aerial image TIFF files were used to create a collection of image patches, where each patch was stored in a separate TIFF file. As the area is predominantly urban residential in character, many of the non-building image patches contained residential street detail, usually kerb and channel bitumen roadways (Figure 5).

4. INITIAL TESTS

Initial classification tests were based on a balanced test set of 100 building images and 100 non-building images. Image coefficients were extracted using the quadruple sampled wavelet process described earlier. A public domain support vector machine, SVM^{Lite} (Joachims, 1998) was used to classify the image patches into building or non-building categories. Results of these tests have been reported previously (Bellman & Shortis, 2002) and showed that although the classification had a predicted success rate of 73%, the actual success on a small independent test set was only 40%. As the success rate is strongly dependent on the sample size, the low rate of detection is most probably due to the small size of the training set.

Although preprocessing is done using the wavelet transform, there are many variables that can influence the preprocessing stage. In studies by others, some attempt had been made to identify the optimum set of parameters but it was found that

(a)

(b)

Figure 5. Examples of image patches used for initial tests (a) building examples, (b) patches not containing buildings

these varied from case to case (Papageorgiou, 2000). As an extension to the initial testing, a further series of tests were performed using the small test set to determine which preprocessing methods produced the best results. The issues investigated included:

- The resolution level of the wavelet coefficients (32 x 32 pixels, 16 x 16 pixels or 8 x 8 pixels)
- The use of over-sampled or standard wavelet coefficients
- The use of normalised image or raw images
- The use of wavelet coefficients or standard colour values (or a combination)
- The use of single resolution or multi-resolution data

The various combinations of these parameters, together with both a linear and polynomial kernel in the SVM classifier, resulted in 216 separate tests. As expected, many of these tests produced poor results. Those that produced successful results were ranked according to the predicted generalization error, the number of training errors, the number of iterations and kernel evaluations taken to reach a solution and characteristics of the high dimensional feature space used in the solution. This resulted in 22 parameter sets that warranted further investigation with a larger training set.

5. LARGE TEST SET

To expand the training data, a new data set was created from the same photography. This dataset contained 1624 examples, with 974 building patches and 650 non-building patches. To validate the training, this data was split into a training set of 452 building and 354 non-building patches and a testing set of 522 building and 296 non-building patches. To generate a richer set of data and to incorporate different building orientations into the training, new image patches were generated from the original set by rotating each patch through 90, 180 and 270 degrees and by mirror reversing the images horizontally and vertically. This generated five additional images for each patch and increased the training set to 4836 images and the testing set to 4908 images.

The SVM was trained using the training data and then the test set was classified using this training model. This process was undertaken separately for all 22 parameter sets identified in the earlier tests. Several tests achieved good results, while five of the tests failed to reach a solution. The results of the successful tests are shown in Table 1.

To further evaluate the result of this training, 57 additional building examples (342 test cases) were produced from a range of public domain sources (Figure 6). These included the Avenches and Hoengg datasets, colour infrared photographs (courtesy of ISTAR Corporation), large scale photographs of a nearby country town and screen copies of a photomosaic of Sydney, Australia. The quality, resolution and scale of these images varied considerably. To meet the requirements of the software, the image patches were re-sampled to 256 x 256 pixels. These images were then used as additional test data for the best of the classifications derived earlier. Although no additional training was undertaken, the classifier identified more than 65-80% of the patches correctly, depending on the classification method used. The majority of the errors occurred with the Sydney images, which were of poor quality compared to the others.

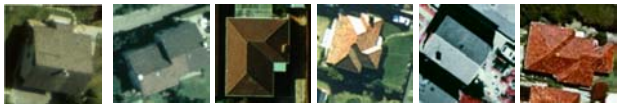


Figure 6. Examples of additional test images

6. DISCUSSION

All methods that established a classification were able to produce quite good results on the out-of-sample data and showed that the predicted generalization error from training is somewhat pessimistic. This is consistent with other work that has shown these estimators generally underestimate the true accuracy (Joachims, 2000; Duan *et. al.*, 2003).

From Table 1, it is difficult to determine a parameter set that is clearly superior to all others. However, some general trends emerge. The tests with suffix 'b' used a polynomial kernel and generally produced better results than those with the linear kernel (suffix 'a'). Test 2_7b, 3_7b and 4_7b all produced quite good results. The only parameter to vary between these tests was the method of normalization of the image content. The first was normalized in the wavelet domain, the second in the image domain and for the third, no normalization was performed. These tests were all at the mid-range resolution (16 x 16 pixels) and used multi-resolution data.

Tests with the prefix '6' were all at the coarsest image resolution of 8 x 8 pixels and although some of these tests produced good results, they generally required many more kernel evaluations and are therefore more computationally intensive. Test 6_8a produced the best results in terms of correct building classifications but this was at the cost of more errors in the non-building patches (false-positives) and a very large number of kernel evaluations.

It is clear from Table 1 that good classification results are possible using the polynomial kernel. The method of pre-processing appears to be less important but does influence the efficiency of the computations. One factor that is not apparent from the table is the size of the coefficient files. The training and testing data sets varied in size from about 10 Mbytes up to several hundred Mbytes, depending on the resolution level, the over-sampling strategy and whether colour image coefficients were included in the output.

		Accuracy estimates after training			Accuracy for Out-of-sample testing			Number Misclassified		
Test Number	No. of kernel evaluations	Error (<=%)	Recall (<=%)	Precision (<=%)	Accuracy (%)	Recall (%)	Precision (%)	Buildings	Non-buildings	Total
2-3a	13621468	14.7	87.2	86.7	82.9	81.5	90.8	579	258	837
2-7a	1326497	24.4	76.7	79.1	81.7	84.8	86.2	476	424	900
2-7b	8279892	25.3	81.5	75.4	87.4	88.6	91.5	358	258	616
3b	9431103	27.1	78.3	74.6	84.7	90.7	86.1	290	459	749
3-2a	275218851	33.7	66.1	66.1	84.8	89.1	87.3	341	406	747
3-4b	29534759	26.0	76.1	77.2	84.7	82.4	92.9	551	198	749
3-7a	1132296	24.1	77.6	79.0	82.0	86.9	85.2	410	474	884
3-7b	8575675	29.9	77.2	71.7	85.7	92.5	86.1	234	466	700
4-3a	14108050	14.7	87.1	86.7	83.0	81.5	90.8	578	258	836
4-7a	1326937	24.4	76.7	79.1	81.7	84.8	86.2	476	424	900
4-7b	7690877	25.3	81.5	75.4	87.4	88.6	91.5	358	258	616
6-1b	11131518	23.9	77.8	79.3	83.5	85.3	88.4	460	350	810
6-3b	18612448	21.5	80.0	81.3	87.4	89.0	91.0	345	274	619
6-4b	3002719641	28.4	74.1	75.0	87.1	85.9	93.3	442	193	635
6-7a	21478593	30.9	74.9	71.4	84.0	89.7	85.9	324	462	786
6-7b	36570891	23.5	78.5	79.3	84.2	90.3	85.7	305	471	776
6-8a	117950052	18.9	83.4	82.9	85.0	92.7	85.1	230	508	738

Table 1. Results of classification and testing using large training sample

A general set of optimal parameters for the pre-processing of image data and the training of the SVM is difficult to determine. It is likely that while some general principles can be established, fine tuning of the classification approach is data dependent and must be reviewed on a case-by-case basis.

Based on the tests in this research, over-sampled wavelet coefficients at a resolution of 16 x 16 appear to offer the best trade off between classification accuracy and computational efficiency. Combined with normalisation in the image domain (test 3-7b), this set of parameters produced fewer errors in classifying the buildings but at the expense of a higher false positive rate. The classifier produced by this test also achieved the highest recognition rates with the additional test data.

7. CONCLUSION

Machine learning methods have been used successfully in several image processing and machine vision domains. The research presented here extends this to building recognition for photogrammetric applications.

An important aspect of machine learning in vision applications is to extract a representative set of characteristics from the image. The multi-resolution approach of wavelets achieves this effectively and leads to a solution that is computationally feasible. One potential limitation of the wavelet approach is that for large training sets, the coefficient files can become very large and unwieldy.

With sufficient training data, an effective classification model can be obtained using a polynomial kernel with the support vector machine. This classification model performs well in out-of-sample testing and has a success rate of more than 80% in correctly recognizing building image patches.

While these techniques cannot satisfy the metric requirements of photogrammetry, they can provide useful starting points and heuristic filters in the area of automated object extraction. With some refinement, this method could be incorporated into a building extraction system as a heuristic filter and be used to ensure that only image patches with a high probability of containing a building were passed to the algorithms that performed the extraction.

REFERENCES

Agouris, P., Gyftakis, S. & Stefanidis, A. 1998. Using A Fuzzy Supervisor for Object Extraction within an Integrated Geospatial Environment. In: *International Archives of Photogrammetry and Remote Sensing*. Ohio, USA, XXXII(III/1), pp. 191-195.

Baltsavias, E. P., Gruen, A. & Van Gool, L., Eds. 2001. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*. Zurich, A. A. Balkema.

Bellman, C. J. & Shortis, M. R., 2002. A Machine Learning Approach to Building Recognition in Aerial Photographs. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Graz, Austria. Vol. XXXIV (3A), pp 50-54.

Canny, J. F., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6), pp. 679-686.

Cristianini, N. & Shawe-Taylor, J. 2000. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge, UK, Cambridge University Press.

Duan, K., Keerthi, S. S. & Poo, A. N., 2003. Evaluation of Simple Performance Measures for Tuning SVM Hyperparameters. *Neurocomputing*, Vol. 51, pp. 41-59.

Elaksher, A. F., Bethel, J. S. & Mikhail, E. M., 2003. Roof Boundary Extraction Using Multiple Images. *Photogrammetric Record*, 18(101): 27-40.

Englert, R., 1998. Acquisition of Complex Model Knowledge by Domain Theory-Controlled Generalization. In: *International Archives of Photogrammetry and Remote Sensing*, Ohio, USA, Vol. XXXII(3/1), pp. 317-324.

Field, D., 1994. What is the Goal of Sensory Coding?, *Neural Computation*, 6(4), pp. 559-601.

Grossberg, S., 1988., Nonlinear Neural Networks. Principles, Mechanisms, And Architectures. *Neural Networks* 1, pp. 17-61.

Gruen, A., Baltsavias, E. P. & O., H., Eds., 1997. *Automatic Extraction of Man_Made Objects from Aerial and Space Images (II)*. Basel, Switzerland, Birkhauser Verlag.

Gruen, A., Kuebler, O. & Agouris, P., Eds., 1995. *Automatic Extraction of Man-Made Objects from Aerial and Space Images.* Basel, Switzerland, Birkhauser Verlag.

Henricsson, O., Bigone, F., Willuhn, W., Ade, F., Kubler, O., Baltsavias, E. P., Mason, S. & Gruen, A., 1996. Project Amobe: Strategies, Current Status and Future Work. *International Archives of Photogrammetry and Remote Sensing,* Vienna, Austria, Vol. XXXI(Part B3): 321-330.

Henricsson, O., 1996. Ascona Workshop Test Dataset. http://www.photogrammetry.ethz.ch/news/events/ascona/dataset/dataset.html, (Accessed June, 2001).

Israel, S. & Kasabov, N. , 1997. Statistical, connectionist and fuzzy inference techniques for image classification. *Journal of Electronic Imaging*, 6(3), pp. 1-11.

Joachims, T., 1998. Text Categorisation with Support Vector Machines: Learning with Many Relevant Features. In: *European Conference on Machine Learning*, Springer.

Joachims, T., 2000. Estimating the Generalization Performance of a SVM Efficiently. In: *International Conference on Machine Learning*, Morgan Kaufman.

Lee, D. & Schenk, T., 1998. An Adaptive Approach for Extracting Texture Information and Segmentation. In: *International Archives of Photogrammetry and Remote Sensing,* Ohio, USA, Vol. XXXII(3/1).

Li, R., Wang, W. & Tseng, H.-Z., 1998. Object Recognition and Measurement from Mobile Mapping Image Sequences using Hopfield Neural Networks: Part 1. In: *ASPRS Annual Conference*, Tampa, Florida, USA, American Society of Photogrammetry and Remote Sensing.

Loung, G. & Tan, Z., 1992. Stereo Matching Using Artificial Neural Networks. In: *International Archives of Photogrammetry and Remote Sensing*, Washington, Vol. XXIX(B3/III), pp. 417-422.

Mallat, S. G. 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, II(7), pp. 674-693.

Michel, A., Oriot, H. & Goretta, O., 1998. Extraction of Rectangular Roofs on Stereoscopic Images - An Interactive Approach. *International Archives of Photogrammetry and Remote Sensing*, Ohio, USA, Vol. XXXII(3/1).

Nevatia, R., Lin, C. & Huertas, A., 1997. A System for Building Detection from Aerial Images. In: *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Switzerland, Birkhauser.

Osuna, E., Freund, R. & Girosi, F., 1997. Training Support Vector Machines: An Application to Face Detection. In: *IEEE Computer Vision and Pattern Recognition*, Puerto Rico.

Papageorgiou, C. P., 2000. A Trainable Object Detection System: Car Detection in Static Images. Unpublished PhD Thesis, Centre for Biological and Computational Learning Artificial Intelligence Laboratory. Massachusetts Institute of Technology. Cambridge, MA.

Papageorgiou, C. P., Oren, M. & Poggio, T., 1998. A General Framework for Object Detection. In: *Sixth International Conference on Computer Vision*, Bombay, India.

Hoengg Dataset, 2001. Institute of Geodesy and Photogrammetry, ETH Zurich, http://www.photogrammetry.ethz.ch/news/events/ascona/dataset_new/, (Accessed March 2001).

Poggio, T. & Shelton, C. R., 1999. Machine Learning, Machine Vision, and the Brain. *AI Magazine*, 20, pp. 37-55.

Schenk, T., 2000. Object Recognition in Digital Photogrammetry. *The Photogrammetric Record*, XVI(95), pp. 743-759.

Sester, M., 1992. Automatic Model Acquisition by Learning. *International Archives of Photogrammetry and Remote Sensing*, Washington, Vol. XXIX, B3.

Sing, S. & Sowmya, A., 1998. RAIL: Road Recognition from Aerial Images Using Inductive Learning. *International Archives of Photogrammetry and Remote Sensing*, Ohio, USA, XXXII(3/1), pp. 367-378.

Fort Hood Dataset, 1999, Sensor Data Management Systems, http://www.mbvlab.wpafb.af.mil/public/sdms/datasets/fthood/, (accessed April 2001).

Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. New York, Springer Verlag.

Zimmermann, P., 2000. A New Framework for Building Detection Analysing Multiple Cue Data. *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, The Netherlands, Vol. XXXIII(3/2), pp. 1063-1070.