Asian Journal of Information Technology 6 (11): 1174-1180, 2007
ISSN: 1682-3915

ANN and SVM Based Image Classification Using Wavelet Decomposition

[1]V. Devendran, [2]Amitabh Wahi and [3]Hemalatha Thiagarajan
[1]Department of Computer Applications, [2]Department of Information Technology,
Bannari Amman Institute of Technology, Sathyamangalam, TamilNadu, India
[3]Department of Computer Applications, National Institute of Technology, Trichy, TamilNadu, India

Abstract: This study discusses the comparative study of ANN and SVM in the classification problem domain. We are classifying the static images into two categories i.e., images with cars and images with no cars. This is tested without any segmentation processes. Feed forward neural networks with backpropagation algorithm are used for paradigm of ANN. On the other hand, support vector machine with radial basis kernel function with p = 5 is used for the same image classification problem. Comparative results show that the classification rate of SVM is far better than the classification rate of ANN. Features are extracted from the static images using Multi-level wavelet decomposition method. The performance of ANN and SVM is tested in 5 different levels of decomposition. This research is carried out using real world data set.

Key words: Artificial neural networks, image classification, multi-level wavelet decomposition, support vector machines

INTRODUCTION

Intelligent Traffic Controlling System (ITCS) is an active computer vision research area because of its enormous demand in practice. Detection and classification of vehicles is a primary focus in the field of ITCS. ITCS may use static or moving cameras to monitor the vehicles for the smooth processing of traffic. Background subtraction method is often used to detect the vehicles, through the static camera. This method may only segment the vehicles which are in normal motion. Very slow or very fast moving vehicles may reduce detection rate when the background subtraction method is used. This method may not consider the standing (not moving) vehicles. Thus, ITCS may miss many vehicles from the monitoring processes. To overcome this problem, some methodologies should be introduced only to classify the images (frames) into two categories: i.e., images with vehicles and images with no vehicles. Vehicle covers every two and four wheelers irrespective of brands, models and classes. This study discusses the performance of classification for image classification problem using artificial neural networks and support vector machine classifiers. The detailed literature survey is discussed in the study, regarding the methodologies and importance of detection and classification of vehicles particularly car detection and classification.

Classification of vehicles is done (Harlow and Peng, 2001) in three types; Type I: Small vehicles such as motorcycles, Type II: Medium size vehicles such as automobiles and light trucks, Type III: Large vehicles such as large trucks and buses using range sensors. Vehicle type classification (Hsu and Lin, 2002) using visual-based dimension estimation method by extracting moving vehicles from traffic image sequences. This study (Gupte *et al*., 2002) presents an algorithm for vision-based detection and classification of vehicles in monocular image sequences of traffic scenes recorded by a stationary camera. Vehicle dimensions are used (Gupte *et al*., 2002) to classify vehicles into two categories: Cars (which constitute the majority of vehicles) and non-cars (vans, SUVs, Pickup trucks, tractor-trailors, semis and buses). Support vector machine (Yuan and Cheu, 2003; Singh and Markou, 2004) is applied for the pattern classification problem to classify the incident data with non-incident data. Polynomial kernel function with degree 2 and radial basis function is implemented and verified and used for novelty detection, respectively. In Foody and Mathur (2004) author proposes an approach for multi-class classification of airborne sensor data by single SVM analysis is evaluated against a series of classifiers that are widely used in remote sensing, with particular regard to the effect of training set size on classification accuracy. Learning-based approach is developed (Agarwal *et al*.,

Corresponding Author: V. Devendran, Department of Computer Applications, Bannari Amman Institute of Technology, Sathyamangalam, TamilNadu, India

2004) to detect the vehicle using sparse, part-based representation. A vocabulary of distinctive object parts is automatically constructed from a set of sample images of the object class of interest; images are then represented using parts from this vocabulary, together with special relations observed among the parts. New metric based (Tsai, 2005) on point-to-line segment distance is proposed to evaluate the similarity between an image region and an instantiation of a 3-D vehicle model under a given pose. An improved EKS is also proposed to track and to predict vehicle motion with a precise kinematics model. In Lou *et al.* (2005) focused on improving the performance of on-road vehicle detection by employing a set of Gabor filters specifically optimized for the task of vehicle detection. Sun *et al.* (2005) proposed a new method to detect the moving objects using the orthogonal Gaussian-Hermite moments in the process of traffic object detections and its action analysis. Two-level stacked generalization scheme composed (Wu *et al.*, 2006) of three generalizers of support vector machines for image classification. Hsien *et al.* (2006) presented an automatic traffic surveillance system to estimate important traffic parameters from video sequences using only one camera. This study (Dong *et al.*, 2006) investigates a procedure for detecting and identifying vehicles based on their RF emissions. In Ki and Baik (2006) suggested a new algorithm for inductive-loop detectors using back-propagation neural networks for vehicle classification.

ARTIFICIAL NEURAL NETWORKS

The Artificial Neural networks developed from the theories of how the human brain works. Many modern scientists believe the human brain is a large collection of interconnected neurons. These neurons are connected to both sensory and motor nerves. Scientists believe, that neurons in the brain fire by emitting an electrical impulse across the synapse to other neurons, which then fire or don't depending on certain conditions. The Artificial neural network is basically having three layers namely input layer, hidden layer and output layer. There will be one or more hidden layers depending upon the number of dimensions of the training samples. ANN is a machine learning algorithm that has been used for various pattern classification problems such as gender classification, face recognition and classification of facial expression and vehicle classification. ANN classifier has advantages for classification such as incredible generalization and good learning ability. The ANN takes the features vector as input and trains the network to learn a complex mapping for classification which will avoid the need for simplifying the classifier. The ANN paradigm that is used in this

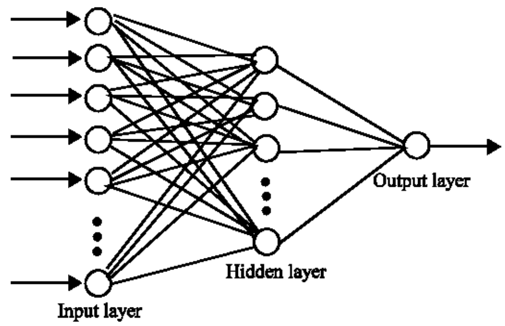

Fig. 1: Simple neural network structure

application is Multilayer Feed Forward Neural Networks (MFNN). MFNNs are a form of non-linear network consisting of a set of inputs (forming the input layer), followed by one or more hidden layers of non-linear neurons and an output layer of non-linear neurons as shown in Fig. 1. MFNN is an ideal means of tackling a whole range of difficult tasks in pattern recognition and regression because of its highly adaptable non-linear structure. In the order to train the network to perform a given tasks the individual weights (w_{ij}) for each neuron are the set using a supervised leaning algorithm known as the error-correction backpropagation algorithm, which involves repeatedly presenting the network with samples from a training set and adjusting the neural weights in order to achieve the required output. It is essentially a gradient descent method, where when adjusting the weight matrices, the direction is move to the greatest descent. The learning constant, η must be chosen with care. If it is too large, the algorithm may repeatedly overshoot the solution, which will lead to slow convergence or even no convergence at all. However, if it is too small, the algorithm will only approach the solution at a very slow rate, again leading to a slow convergence and increasing the chances of the algorithm becoming stuck in local minima. Two main methods of overcoming these problems are momentum and adaptive learning. For momentum method, if we are consistently moving in the same direction, then we want to build up some momentum in that direction. This will help us to go through any small local minima and hopefully speed up the convergence.

Standard : $\Delta w(t) = -\eta \nabla E(t)$
Momentum : $\Delta w(t) = -\eta \nabla E(t) + \alpha \nabla w(t - 1)$

Where α is the momentum term. For adaptive learning rate, adjusting the learning rate dynamically, usually starting with a large value and then decreasing it as we approach the solution in order to prevent overshoot.

The input data for training comes from the output data of feature extraction methods.

A learning problem with binary outputs (yes/no or 1/0) is referred to as binary classification problem whose output layer has only one neuron. A learning problem with finite number of outputs is referred to multi-class classification problem whose output layer has more than one neuron. The neural networks structure used in our experiment is consist of only three layers having 4000 neurons in the input layer, 150 neurons in the hidden layer and 1 neuron in the output layer as shown in Fig. 1.

SUPPORT VECTOR MACHINE

Support vector machine is a relatively new pattern classifier introduced by Vapnik (1998). A SVM classifies an input vector into one of two classes, with a decision boundary developed based on the concept of structural risk minimization (of classification error) using the statistical learning theory. The SVM learning algorithm directly seeks a separating hyperplane that is optimal by being a maximal margin classifier with respect to training data. For non-linearly separable data, the SVM uses kernel method to transform the original input space, where the data is non-linearly separable, into a higher dimensional feature space where an optimal linear separating hyperplane is constructed. On the basis of its learning approach, the SVM is believed to have good classification rate for high-dimensional data. Consider the problem of image classification where X is an input vector with 'n' dimensions. The SVM performs the following operation involving a vector $W = \{w_1 ... w_n\}$ and scalar b:

$$f(X) = \mathrm{sgn}(W \bullet X + b) \tag{1}$$

Positive sign of f(X) may be taken as images with cars and negative value of f(X) may be regarded as images with no cars. Consider a set of training data with l data points from two classes. Each data is denoted by (x_i, y_i), where I = 1, 2…l, $X_i = \{x_1, \dots x_n\}$ and $y_i \in \{+1, -1\}$. Note that yi is a binary value representing the two classes. The task of SVM learning algorithm is to find an optimal hyperplane (defined by W and b) that separates the two classes of data. The hyperplane is defined by the equation:

$$W \bullet X + b = 0 \tag{2}$$

Where, X is the input vector, W is the vector perpendicular to the hyperplane and b is a constant. The graphical representation for a simple case of two-dimensional input (n = 2) is illustrated in Fig. 2. According to this hyperplane, all the training data must satisfy the following constraints:

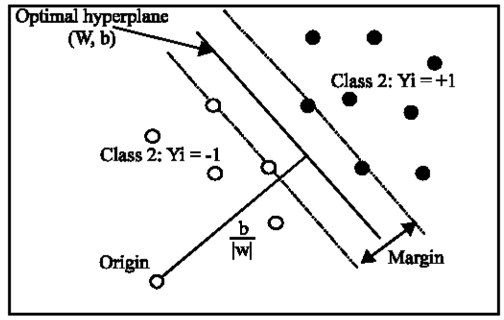

Fig. 2: Optimal separating hyperplane for a 2-dimensional 2-class problem

$$\begin{aligned} W \bullet X_i + b &\geq +1 \text{ for } \forall_i = +1 \\ W \bullet X_i + b &\geq -1 \text{ for } \forall_i = -1 \end{aligned} \tag{3}$$

which is equivalent to:

$$y_i (W \bullet x_i + b) \geq 1 \ \forall_i = 1, 2, ..., l \tag{4}$$

There are many possible hyperplane that separate the training data into the two classes. However, the optimal separating hyperplane is the unique one that not only separates the data without error but also maximizes the margin, i.e., maximizes the distance between the closest vectors in both classes to the hyperplane (Vapnik, 1998). As shown in Fig. 2, the margin, ρ, is the sum of the absolute distance between the hyperplane and the closest data points in each class. It is given by:

$$\rho = \min \frac{|W.X_i + b|}{\|W\|} + \min \frac{|W.X_i + b|}{\|W\|} = \frac{2}{\|W\|} \tag{5}$$

Therefore, the optimal separating hyperplane is the one that maximizes 2/||W||, subject to constraints (4). It is mathematically more convenient to replace maximization of 2 / ||W|| with the equivalent minimization of $\|W\|^2/2$ subject to constraints (4) can be solved by the Lagrangian formulation:

$$\text{Min } L = \frac{1}{2} \|W\|^2 - \sum_{i=1}^{l} \alpha_i \left[y_i (W.X_i + b) - 1 \right] \tag{6}$$

Where α_i is the Lagrange multiplier ($\alpha_i >= 0$, i=1,2,..,l). The Lagrangian has to be minimized with respect to W and b and maximized with respect to α_i. The minimum of the Lagrangian with respect to W and b is given by:

$$\frac{\partial L}{\partial W} = 0 \Rightarrow W = \sum_{i=1}^{l} \alpha_i X_i y_i \tag{7}$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \quad (8)$$

Substituting (7) and (8) into (6), the primal minimization problem is transformed into its dual optimization problem of maximizing the dual Lagrangian L_D with respect to α_i:

$$\text{Max } L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (X_i.X_j) \quad (10)$$

subject to

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \quad (11)$$

$$\alpha_i \geq 0 \ \forall_i = 1,\ldots,l \quad (12)$$

Thus, the optimal separating hyperplane is constructed by solving the above quadric programming problem defined by (10)-(12). In this solution, those points have non-zero Lagrangian multipliers ($\alpha_i > 0$) are termed support vectors. Support vectors satisfy the equality in the constraint (4) and lie closest to the decision boundary (they are circles in Fig. 2). Consequently, the optimal hyperplane is only determined by the support vectors in the training data. Based on the α_i values obtained, W can be calculated from (7). b can be obtained by using the Karush-Kuhn-Tucker (KKT) complementary condition for the primal Lagrangian optimization problem:

$$\alpha_i \left[y_i(W.X_i + b) - 1\right] = 0 \ \ \forall_i = 1,\ldots,l \quad (13)$$

One b value may be obtained for every support vector (with $\alpha_i > 0$). Burges recommends that the average value of b be used in the classification (Burges, 1998). With the solution, the SVM classifiers becomes

$$f(X) = \text{sgn}(W.X + b) = \text{sgn}\Big(\sum_{\forall i, \alpha_i > 0} y_i \alpha_i (X_i.X) + b\Big) \quad (14)$$

Note that, in (14), one only needs to make use of X_i, y_i and α_i of the support vectors, while X is the input vector to be classified. When a linear boundary is inappropriate (i.e., no hyperplane exists to separate the two classes of data), the extension of above method to a more complex decision boundary is accomplished by mapping the input vectors $X \in R_n$ into a higher dimensional feature space H through a non-linear function

$$\phi : R^n \rightarrow H.$$

In H, an optimal separating hyperplane is then constructed using training data in the form of dot products $\phi(X_i)$. $\phi(X_j)$ instead of the $X_i.X_j$ term in (10). To

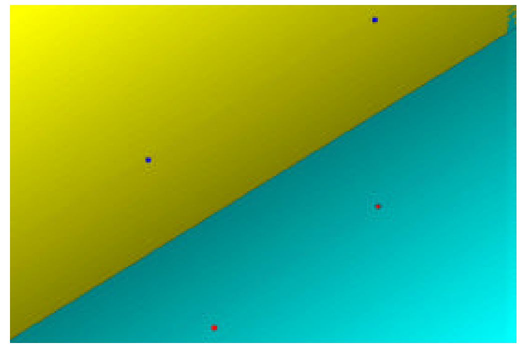

Fig. 3: SVM classification of linearly separable data

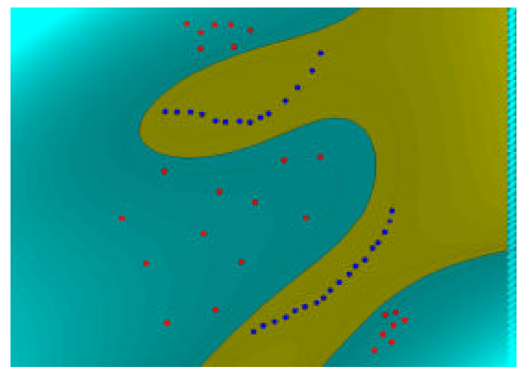

Fig. 4: SVM classification of non-linearly separable data

avoid the expensive computations of $\phi(X_i)$. $\phi(X_j)$ in the feature space, it is simpler to employ a kernel function such that

$$K(X_i.X_j) = \phi(X_i) \bullet \phi(X_j) \quad (15)$$

Thus, only the kernel function is used in the training algorithm and one does not need to know the explicit form of ϕ. The computation in (15) results in some restrictions on the form and parameter values of non-linear functions that can be used as the kernel functions. Detailed discussions can be found in Berges (1998) and Vapnik (1998). Some commonly used kernel functions are:

$$\text{Polynomial function: } K(X_i, X_j) = \left(X_i.X_j + 1\right)^d \quad (16)$$

$$\text{Radial basis function: } K(X_i, X_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad (17)$$

$$\text{Sigmoid function } K(X_i, X_j) = \frac{1}{1 + e^{[v(X_i.X_j) - \delta]}} \quad (18)$$

Where d is a positive integer and σ, v and δ are real constants. These four parameters must be defined by the user prior to SVM training. With the use of a kernel function, the SVM capable of performing non-linear classification of input X becomes

$$f(X) = \operatorname{sgn}\left(\sum_{\forall_i, \alpha_i > 0} y_i \alpha_i K(X_i, X) + b \right) \qquad (19)$$

The hyperplane and support vectors used to separate the linearly separable data are shown in Fig. 3. And the hyperplane and support vectors used to separate the non-linearly separable data are shown in Fig. 4. Radial basis kernel function with p = 5 used for this non-linear classification. Individual colors represents particular each class of data.

WAVELET FEATURE EXTRACTION

In this study, we utilize the 1-Dimensional Multi-level wavelet analysis to decompose an image using the Daubechies (db1) wavelet. We have tested the image classification problem in all the five levels of decomposed coefficients. The five levels of decomposition structure used for our problem are shown in Fig. 5.

Each decomposition gives two categories of coefficients, i.e., approximate coefficients and detailed coefficients. Though the grayscale input image is of size 40×100, the decomposed approximate coefficients size is as 4000×1 in all the levels. Approximate coefficients are considered as the features for our problem in all the five levels of decomposition. cA1, cA2, cA3, cA4 and cA5 represent first, second, third, fourth and fifth level of approximate coefficients. cD1, cD2, cD3, cD4 and cD5 represent first, second, third, fourth and fifth level of detailed coefficients. Approximate coefficients are considered as the features for our problem in all the five levels of decomposition.

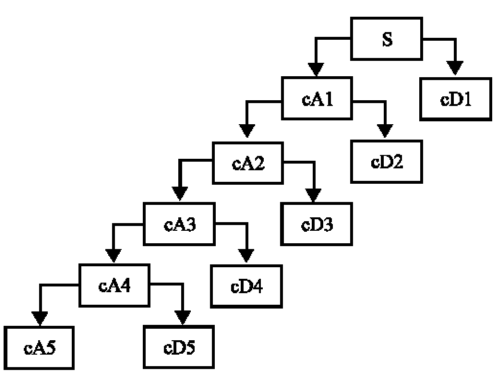

Fig. 5: Structure of 1-D Multi-level decomposition

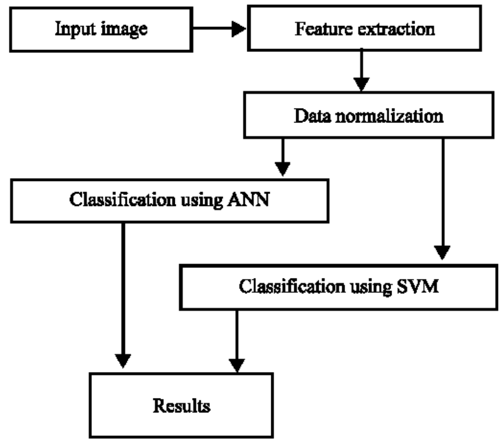

Fig. 6: Detailed description of proposed work

PROPOSED WORK

This study is two folded. First, classifies the given images into two broad categories; images with cars and images with no cars. Second, classification rates are evaluated using Artificial Neural Networks and Support Vector Machine. Multi-level 1-Dimensional wavelet analysis is used to obtain the feature data in all the 5 levels of decomposition. As the input image is of size 40×100, the decomposed approximate coefficients are of size 4000×1. Zero-mean normalization is applied in the features for normalizing the feature data. The normalized feature is given as input to ANN and SVM individually and the results are compared. The detailed description of our proposed work is shown in Fig. 6.

IMPLEMENTATION

Implementation using ANN: Feed-forward neural network with backpropagation algorithm is used for image classification. Neural network has an input layer, one hidden layer and an output layer. Input layer has 4000 neurons, hidden layer has 150 neurons and output layer has only one neuron. Tansigmoidal and logsigmoidal is used in hidden layer and output layer, respectively for transfer functions. One hundred samples are used for training phase and 1000 samples are used for testing phase. Training includes 50 positive and 50 negative samples and testing includes 500 positive and 500 negative samples. The results obtained from all the 5 levels are shown in Table 1.

Image classification using SVM: Radial basis kernel function with d = 5 is used or training and testing phase.

Table 1: Overall performances of artificial neural networks

Decomposition stage	Classification rate True positive	False positive	Overall performance
Level 1	92.4	52.6	72.5
Level 2	76.8	72.2	74.5
Level 3	84.4	66.4	75.4
Level 4	65.4	87.6	76.5
Level 5	86.8	64.0	75.4

Table 2: Overall performances of support vector machine

Decomposition stage	Classification rate True positive	False positive	Overall performance
Level 1	92.8	84.6	88.7
Level 2	93.0	84.0	88.5
Level 3	93.0	83.8	88.4
Level 4	93.0	83.8	88.4
Level 5	93.0	83.8	88.4

100 samples are used for training phase and 1000 samples are used for testing phase. Training includes 50 positive and 50 negative samples and testing includes 500 positive and 500 negative samples. The results obtained in all the five levels are shown in Table 2.

DISCUSSION

This study discusses the image classification problem and studies the performance of classification using ANN and SVM classifiers. The above experiments are carried out on the standard real images taken from UIUC Image database (http://l2r.cs.uiuc.edu/~cogcomp/Data/Car). Multi-level 1-Dimensional wavelet (db1) is applied in all the five levels without any segmentation processes. Zero-mean data normalization technique is applied in the extracted decomposed coefficients for normalizing the data. Normalized feature data are given as input to ANN and SVM separately. The classification results obtained from ANN and SVM are shown in Table 1 and 2, respectively. The comparative study on ANN and SVM is shown in Table 3 which figures out the overall performances of ANN and SVM classifiers over the image classification implemented on UIUC Image Databases. The implementation results proves that classification rate of SVM is much higher than ANN in the image classification problem. And SVM gives more than 90% classification rate for the true positive and more than 80% classification rate for false positive. But ANN is not stable over true and false positives classification as shown in Table 1.

The comparative results are given in graphical representation is shown in Fig. 7.

Table 3: Comparative results of ANN and SVM classifiers

Decomposition stages	Classification rate ANN (%)	SVM (%)
Level 1	72.5	88.7
Level 2	74.5	88.5
Level 3	75.4	88.4
Level 4	76.5	88.4
Level 5	75.4	88.4

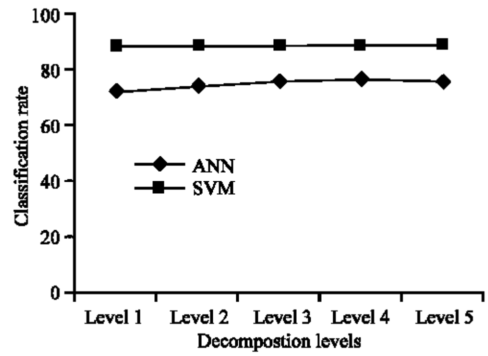

Fig. 7: Performance graphs of ANN and SVM classifiers

CONCLUSION

This study concentrates on the classification of images as images with cars and images with no cars. Various levels of multi-level decomposition wavelet are applied using db1 and classified the images using ANN and SVM classifiers. Classification rate of ANN is ranging between 72 and 77%. But classification rate of SVM is of almost 88% in all the levels. The results are proving that SVM is working well towards the classification problems than ANN. This complete work is implemented in Matlab 6.5 with the standard UIUC Image databases (http://l2r.cs.uiuc.edu/~cogcomp/Data/Car). The SVM is implemented using the SVM-Matlab toolbox.

REFERENCES

Agarwal, S., A. Awan and D. Roth, 2004. Learning to Detect Objects in Images via a Sparse, Part-Based Representation. IEEE. Trans. Pattern Analysis and Machine Intelligence, 26: 11.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition, data mining and knowledge discovery Available from http://svm.first.gmd.de, 2: 121-167.

Dong, X., H. Weng, D.G. Beetner, T.H. Hubing, D.C. Wunsch, M. Noll, H. Goksu and B. Moss, 2006. Detection and identification of vehicles based on their unintended electromagnetic emissions. IEEE. Trans. Electromagnetic Compatibility, 48: 4.

Giles M. Foody and A. Mathur, 2004. A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. IEEE. Trans. Geoscience and Remote Sensing, 42: 6.

Gupte, S., O. Masoud, Robert F.K. Martin and Nikolaos P. Papanikolopoulos, 2002. Detection and Classification of Vehicles. IEEE. Trans. Intelligent Transport. Sys., 3: 1.

Harlow, C. and S. Peng, 2001. Automatic vehicle classification system with range sensors. Transport. Res., pp: 231-247.

Hsieh, J.W., S.H. Yu, Y.S. Chen and W.F. Hu, 2006. Automatic Traffic Surveillance System for Vehicle Tracking and Classification. IEEE. Trans. Intelligent Transport. Sys., 7: 2.

Hsu, C.W. and C.J. Lin, 2002. A Comparision of Methods for Multiclass Support Vector Machines. IEEE .Trans. Neural Networks, 13: 2.

http://l2r.cs.uiuc.edu/~cogcomp/Data/Car.

Ki, Y.K. and D.K. Baik, 2006. Vehicle-Classification Algorithm for Single-Loop Detectors Using Neural Networks. IEEE. Trans. Vehicular Technol., 55: 6.

Lou, J., T. Tan, W. Hu, H. Yang and S.J. Maybank, 2005. 3-D Model-Based Vehicle Tracking. IEEE. Trans. Image Processing, 14: 10.

Singh, S.and M. Markou, 2004. An Approach to Novelty Detection Applied to the Classification of Image Regions. IEEE. Trans. Knowledge and Data Eng., 16: 4.

Sun, Z., G. Bebis and R. Miller, 2005. On-Road Vehicle Detection Using Evolutionary Gabor Filter Optimization. IEEE. Trans. Intelligent Transport. Sys., 6: 2.

Tsai, C.F., 2005. Training support vector machines based on stacked generalization for image classification. Neurocomputing, 64: 497-503.

Vapnik, V.N., 1998. The support vector method of function estimation. In: Generalization in Neural Network and Machine Learning. Springer-Verlag, New York, pp: 239-268.

Wu, Y., J. Shen and M. Dai, 2005. Traffic Object detections and its action analysis. Pattern Recognition Letters, 26: 1963-1984.

Yuan, F. and R.L. Cheu, 2003. Incident detection using support vector machines. Transport. Res., 11: 309-328.