International Journal of Transportation
Vol.3, No.3 (2015), pp.41-48
http://dx.doi.org/10.14257/ijt.2015.3.3.04

Prediction Performance of Support Vector Machines with Fused Data in Road Scene Analysis

Daehyon Kim

Department of Marine & Civil Engineering,
Chonnam National University, Dundeok-Dong, Yeosu-Si, Jeollanam-Do, Korea
daehyon@chonnam.ac.kr

Abstract

Automatic video-based vehicle detection is one of the main research topics in Intelligent Transportation Systems (ITS) and is a key element for automatic traffic surveillance systems. Support Vector Machines (SVMs) have been increasingly applied to an automatic video-based vehicle detection and a road scene analysis because of their remarkable performance in prediction accuracy. The property of input data for learning on SVMs determines the predictive performance. It is important task to choose the best input vectors in order to improve the predictive performance. It is normal to use a single property of input vectors in the application of learning models. However, the composition of different input vectors may affect predictive performance and a new input vector will be created by combining two raw data. In this paper, two information sources of edge information and pixel gray value have been combined to detect vehicle in road scene images and see how the fused data affect the predictive performance in SVMs. The experimental results of this study show that the fused data may provide better performance in predictive accuracy than a raw input data. Moreover, the results show that SVMs could provide much better performance than the Backpropagation model which is the best known neural network.

Keywords: *Intelligent Transportation Systems (ITS), Support Vector Machines (SVMs), vehicle detection, fused data*

1. Introduction

Automatic video-based vehicle detection which is one of the main research topics in Intelligent Transportation Systems (ITS) is a key element for integrated automatic traffic surveillance systems. This technology is used to solve the transportation problems of road safety and congestion.

Even though there has been growing interest in vehicle detection in traffic scene images (Oh, 2003; Tsai, *et al.*, 2007; O'Malley *et al.*, 2011), detection accuracy is still low in complicated circumstances, and more efficient algorithms should be developed to improve prediction accuracy in vehicle detection.

Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) which are best known algorithms in image processing, pattern recognition problems and applications of real-time systems have been used as superior tools in video-based vehicle detection systems.

In machine learning systems such as ANNs and SVMs, the choice of input vectors is as important as the choice of learning algorithm since the predictive performance of the model might be highly dependent on the input vectors. The predictive performance of learning models can be improved by choosing more efficient data and informative input vectors (Kim, 2010).

ISSN: 2287-7940 IJT

In video-based computer vision, the gray value of each pixel in the image has been used as common inputs to learning models. However, the edge segmentation which is an example of discontinuity based on abrupt changes in pixel properties can also be an efficient source for pattern recognition. Kim (2010) showed that edge information could be an important source of inputs to leaning models and produce a good performance with high detection accuracy in traffic scene analysis.

Even though a single input property such as gray intensity values or edge information has been used as a good source of input vectors, a fused data might be valuable additional information in video based vehicle detection systems. Data fusion is the integration of data or information from different sources and the data fusion techniques have been extensively employed on multi-sensor environments with the aim of fusing and aggregating data from different sensors (Castanedo, 2013). However, data fusion has also been applied to the learning models to obtain a lower detection error and a higher reliability because the fused data is more synthetic and informative than the original raw inputs (Kim, 2010; Lie and Chi, 2014).

In order to improve the predictive performance of a neural network model in traffic scene analysis, the data fusion technique was proposed and used and in Kim (2010). The data fusion technique was based on the combination of two sources of raw data which are edge detection image and grey-level image. Kim (2010) showed that the fused data could provide an improved predictive accuracy in neural networks and be more informative and efficient than single data alone.

In this study, several types of fused data have been explored to determine the best type of input vector in a different learning model for video-based vehicle detection. For the leaning algorithm, Support Vector Machines(SVMs) have been used because they have been popular and shown to provide higher performance than other types of models, such as *k*-Nearest Neighbor and the neural network model (Wang *et al.*, 2000; Wan *et al.*, 2012; Pan *et al.*, 2012).

2. Support Vector Machines Learning Algorithm

Support Vector Machines(SVMs) which were introduced by Vapnik(1995) have been successfully applied to numerous pattern recognition problems, such as handwritten character recognition (Cortes and Vapnik, 1995; Schölkopf *et al.*, 1995; Schölkopf et al, 1996; Dong et al., 2005; John et al., 2012), face recognition (Hotta, 2008), object detection (Juang et al., 2009), pedestrian detection (Cao and Qiao, 2008), and traffic sign detection (Greenhalgh and Mirmehdi, 2012; Park and Kim, 2013). Support Vector Machines (SVMs) are based on statistical learning theory and the basic idea of the algorithm has been widely described in previous researches (Kim, 2004; Kim et al., 2007).

Let the training data $(\mathbf{x}_i, y_i)$, for $i = 1,\ldots,l$, be $y_i \in \{\pm 1\}$ $\mathbf{x}_i \in \mathbf{R}^N$. Then the support vector algorithm simply looks for the optimal hyper-plane with the largest margin. This can be formulated as follows:

$$\min \ \tau(\mathbf{w}, \xi) = \frac{1}{2}\left|\mathbf{w}\right|^2 + C \sum_{i=1}^{l} \zeta_i \tag{1}$$

$$\text{s. t. } \ y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i, \quad i = 1,\ldots, l \tag{2}$$

where $\mathbf{w}$ is a normal to the hyper-plane, b is the bias or offset, and C is the upper bound for the Lagrange multiplier, λ_i, i.e., $0 \leq \lambda_i \leq C$. The primal form of the objective function can be:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\left|\mathbf{w}\right|^2 - \sum_{i=1}^{l} \alpha_i (y_i((\mathbf{x}_i \cdot \mathbf{w}) + b) - 1) \tag{3}$$

The Lagrangian L has to be minimized with respect to the primal variables $\mathbf{w}$ and b and maximized with respect to the dual variables α_i. From the Karush-Kunhn-Tucker conditions, the derivative of L with respect to the primal variables must vanish, subject to the constraints $\alpha_i \geq 0$, i.e.,

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, b, \boldsymbol{\alpha}) = 0, \quad \frac{\partial}{\partial b} L(\mathbf{w}, b, \boldsymbol{\alpha}) = 0 \tag{4}$$

Equation (4) leads to

$$\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i \text{ and } \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{5}$$

Equation (5) is equality constraints in the dual formulation, and the following equation (6) which is the Wolfe dual of the optimization problem is given by substituting them into equation (3).

$$\max W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \tag{6}$$

$$\text{s.t. } \alpha_i \geq 0, \text{ and } \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{7}$$

The hyper-plane decision function can thus be given as

$$f(x) = \operatorname{sgn}\left((\mathbf{x} \cdot \mathbf{w}) + b\right) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b\right) \tag{8}$$

The algorithm explained above is for linearly inseparable classification problem. To solve nonlinear classification problems, SVMs are using kernel functions. The most popular kernel function is the Gaussian kernel of equation (9)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \left|\mathbf{x}_i - \mathbf{x}_j\right|^2\right) \tag{9}$$

where γ determines the width of the kernel function.

3. Data Fusion and Training and Test Data

The significant sources of information in vision-based traffic monitoring system are the edge information and the gray value of each pixel. Most commonly, they have been used as a single input vector in learning systems. However, the combined information could be more efficient than grey values on the pixel or edges alone. More importantly, the combined data of edge information and grey value with some pre-processing to reduce the complexity of input vector would be efficient both in terms of computing time (simplicity of input vectors) and predictive accuracy.

Even though there might be various types of data fusion methods, four types of fused data which are input vectors in learning models will be considered in this study. The four data fusion methods for combining two information sources in traffic scene analysis are as follows; i.e. Type 1 is a combined data with 15 by 30 pixels of 256 gray level and Sobel edge image with 15 by 30 pixels. Type 2 is a combined data with 8 by 15 pixels of 256 gray level and Sobel edge image with 15 by 30 pixels. Type 3 is a combined data with 15 by 30 pixels of 256 gray level and Sobel edge image with 8 by 15 pixels. Type 4 is a combined data with 15 by 30 pixels of 256 gray level and binarized Sobel edge image with 15 by 30 pixels.

With four different types of fused data, this study will investigate how the shrinking image size and the binarization of Sobel image can affect to data fusion and the predictive performance of SVMs. The well-known Sobel edge detector uses a pair of 3-by-3 convolution masks and produces bell-shaped edge information. The binarization of Sobel edge detection image may provide a different source of information as input vectors for learning models. In addition, the shrunken images have often been used in neural networks in order to reduce the number of input units and the complexity of input vectors. The shrunken image which is achieved by selecting every other pixel may conserve the geometric relationships between pixels and take less computation effort.

For the implementation of experiments, traffic scene images which have been used in previous researches (Suh *et. al.*, 2013; Kim, 2010) have also been used in this study. The learning task is to learn and recognize three different patterns, Pattern A, Pattern B, and Pattern C, obtained from traffic scene images (see Figure 1). Figure 1 shows four types of fused data with three different patterns where the image of Pattern A corresponds to the front part of a vehicle, while Pattern B images the top of a vehicle, and Pattern C is a non-vehicle image on the road. Table 1 shows the number of data sets used for training and testing the SVMs.

Table 1. Data Sets used for Training and Testing the SVMs

Pattern data	Number of trainings	Number of test data	Data for each pattern
Pattern A	100	300	400
Pattern B	100	300	400
Pattern C	30	100	130
Total	230	700	930

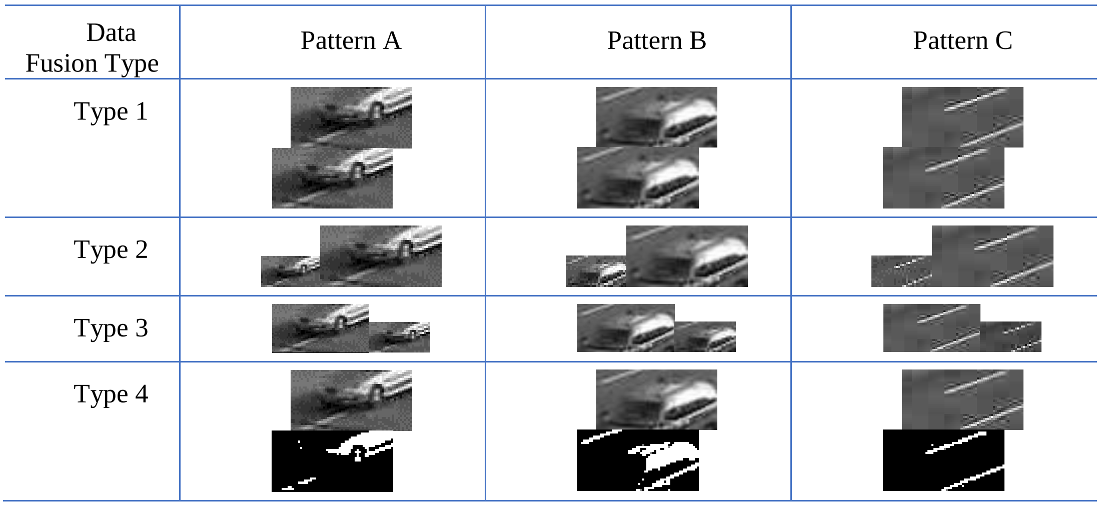

Figure 1. Image Patterns of a Fused Data

4. Experimental Results

For the application of the SVMs model, two parameters, C and γ, should be determined in advance. The parameter C is a positive regularization parameter that controls the tradeoff between the complexity of the machine and the allowed classification error, and γ is the parameter of the Gaussian kernel. For the parameter *C*, all experiments have been carried out in this study with $C = 1.0$, because the value of this parameter does not affect the predictive performance. However, the parameter γ of the Gaussian RBF (Radial Basis Function) kernel may affect the predictive performance, and the experimental results show a different performance according to the parameter value. In this study, the best prediction performance was achieved with Gamma (γ) = 0.04 for Type 1, Gamma (γ) = 0.06 for Type 2, Gamma (γ) = 0.06 for Type 3 and Gamma (γ) = 0.07 for Type 4 (see Table 2).

Table 2 shows the predictive performance of the SVMs on the test data set according to four different forms of fused data. With Type 1 data which is a combined data with 15 by 30 pixels of 256 gray level and Sobel edge image with 15 by 30 pixels, the prediction errors are 12 in the test of a total of 700 data sets. However, the number of prediction errors was increased in Type 2, Type3 and Type 4. With Type 2 which is a combined data with 8 by 15 pixels of 256 gray level and Sobel edge image with 15 by 30 pixels, the prediction errors are 16. With Type 3 which is a combined data with 15 by 30 pixels of 256 gray level and Sobel edge image with 8 by 15 pixels, the average prediction errors are 15 in the test of a total of 700 data sets. On the other hand, with Type 4 which is a combined data with 15 by 30 pixels of 256 gray level and binarized Sobel edge image with 15 by 30 pixels, average prediction errors are 19 in the test of a total of 700 data sets.

This study shows completely different results from previous study (Kim, 2010), in which the Backpropagation model has been adopted as the network learning method to study four different forms of fused data (see Table 3). As shown on the table 3, the best predictive performance was achieved with Type 4 of fused data in the Backpropagation model. However, the best performance in terms of predictive accuracy was achieved with Type 1 of fused data in SVMs.

More importantly, in the comparison of predictive performance of two learning models, Backpropagation and SVMs, the SVMs were superior to the Backpropagation model in any types of fused data. The best performance regarding recognition accuracy was 98.29% on the SVMs with Type 1 of fused data, even though the best predictive accuracy was 92.22% in Backpropagation model with Type 4 of fused data.

Table 2. Performance on the Combined Input Vectors with SVMs - Experimental Results of this Study

	Fusion Type of Input Vectors			
	Type 1	Type 2	Type 3	Type 4
C	1.0	1.0	1.0	1.0
Gamma(γ)	0.04	0.06	0.06	0.07
Mean Squared Error (on Test)	0.0214286	0.0314286	0.03	0.0485714
Prediction errors	12	16	15	19
Prediction accuracy (%)	98.2857	97.7143	97.8571	97.2857

Table 3. Performance on the Combined Input Vectors with Backpropagation

	Fusion Type of Input Vectors			
	Type 1	Type 2	Type 3	Type 4
Epoch	62.33	81.97	75.57	133.73
RMSE	0.001841	0.001845	0.002454	0.003668
Average Prediction Errors	75.57	77.83	92.73	54.43
Variance	588.60	924.21	367.31	392.19
Prediction accuracy (%)	89.20	88.88	86.75	92.22

Source: Kim (2010)

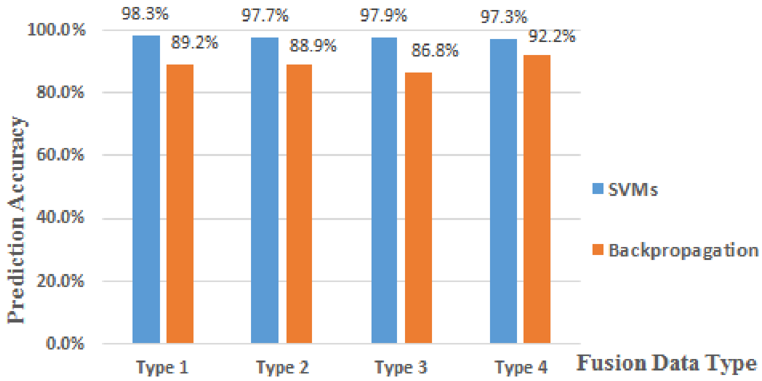

Figure 2. Performance Comparison of Two Models-SVMs and Backpropagation

5. Conclusion

Data fusion which combines two more sources of raw data may create a new source of information in learning models. Moreover, the fused data could be more informative and synthetic than the original inputs. An important issue of this paper is to apply different kinds of fused data in SVMs and compare the predictive performance of each. Four data fusion methods, Type 1, Type 2, Type 3 and Type 4 have been evaluated using SVMs. The experimental result shows that there is no significant performance difference, in terms of predictive accuracy, among the four data fusion methods in SVMs.

Important results obtained in this study can be summarized as follow: First, the prediction performance in SVMs may be decreased by reducing complexity of input vectors. On the other hand, the prediction performance can be improved in Backpropagation model by reducing complexity of input vectors. Even though the fused data may provide much better performance than raw data, the prediction performance can be different according to the learning algorithm. Second, the best fusion data for the application of SVMs is Type 1 which is a combined data with 15 by 30 pixels of 256 gray level and Sobel edge image with 15 by 30 pixels. Third, SVMs could provide much better performance than Backpropagation model.

Acknowledgment

This study was financially supported by Chonnam National University, 2013.

References

[1] X. Cao and H. Qiao, “A Low-Cost Pedestrian-Detection System with a Single Optical Camera”, IEEE Transactions on Intelligent Transportation Systems, vol. 9, no. 1, **(2008)**, pp. 58-67.
[2] F. Castanedo, “A Review of Data Fusion Techniques”, Scientific World Journal, vol. 2013, **(2013)**, pp. 1-19.
[3] C. Cortes and V. Vapnik, “Support vector networks”, Machine Learning, vol. 20, **(1995)**, pp. 273-297.
[4] J. Dong, A. Krzy-zak and C. Y. Suen, “An improved handwritten Chinese character recognition system using support vector machine”, Pattern Recognition Letters, vol. 26, **(2005)**, pp. 1849-1856.
[5] J. Greenhalgh and M. Mirmehdi, “Real-Time Detection and Recognition of Road Traffic Signs”, IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 4, **(2012)**, pp. 1498-1506.
[6] K. Hotta, “Robust face recognition under partial occlusion based on support vector machine with local Gaussian summation kernel”, Image and Vision Computing, vol. 26, **(2008)**, pp. 1490-1498.
[7] J. John, K. V. Pramod and K. Balakrishnan, “Unconstrained Handwritten Malayalam Character Recognition using Wavelet Transform and Support vector Machine Classifier”, Procedia Engineering, vol. 30, **(2012)**, pp. 598-605.
[8] C. Juang, W. Sun and G. Chen, “Object detection by color histogram-based fuzzy classifier with support vector learning”, Neurocomputing, vol. 72, **(2009)**, pp. 2464-2476.

[9] D. Kim, "Prediction Performance of Support Vector Machines on Input Vector Normalization Methods". International Journal of Computer Mathematics, vol. 81, no. 5, **(2004)**, pp. 547-554.
[10] D. Kim, "Pre-processing of inputs to a neural network model for better performance in traffic scene analysis", Civil Engineering and Environmental Systems, vol. 27, no. 1, pp. 23-31, **(2010)**.
[11] D. Kim, S. Lee and S. Cho, "Input Vector Normalization Methods in Support Vector Machines for Automatic Incident Detection", Transportation Planning and Technology, vol. 30, no. 6, **(2007)**, pp. 593-608.
[12] L. A. Klein, "Sensor and data fusion: A tool for information assessment and decision making", SPIE Press. ISBN 0-8194-5435-4, **(2004)**.
[13] Z. Lie and Q. Chi, "Structural Damage Detection Based on Semi-supervised Fuzzy C-means Clustering", The 9th International Conference on Computer Science & Education (ICCSE 2014) August 22-24, 2014. Vancouver, Canada, **(2014)**, pp. 551-556.
[14] R. O'Malley, M. Glavin and E. Jones, "Vision-based detection and tracking of vehicles to the rear with perspective correction in low-light conditions", IET Intelligent Transport Systems, vol. 5, no. 1, **(2011)**, pp. 1-10.
[15] J. Oh, "Vehicle Detection Using Video Image Processing System: Evaluation of PEEK Videotrak", ASCE Journal of Transportation Engineering, vol. 129, no. 4, **(2003)**, pp. 462-465.
[16] S. Pan, S. Iplikci, K. Warwick and T. Z. Aziz, "Parkinson's Disease tremor classification - A comparison between Support Vector Machines and neural networks", Expert Systems with Applications, vol. 39, **(2012)**, pp. 10764-10771.
[17] J. Park and K. Kim, "Design of a visual perception model with edge-adaptive Gabor filter and support vector machine for traffic sign detection", Expert Systems with Applications, vol. 40, **(2013)**, pp. 3679-3687.
[18] B. Schölkopf, B. Burges and V. Vapnik, "Extracting support data for a given task", U. M. Fayyad, R. Uthurusamy (Eds.) Proceedings of 1st International Conference on Knowledge Discovery & Data Mining, AAAI Press, Menlo Park, **(1995)**.
[19] B. Schölkopf, C. Burges and V. Vapnik, "Incorporating invariances in support vector learning machines", In: C. von der Malsburg, W. von Seelen, J. C. Vorbr¨uggen, B. Sendhoff (Eds.) Artificial Neural Networks - ICANN'96. Vol. 1112. Springer Lecture Notes in Computer Science, Berlin, **(1996)**, pp. 47-52.
[20] L. W. Tsai, J. W. Hsieh and K. C. Fan, "Vehicle detection using normalized color and edge map", IEEE Transaction on Image Processing, vol. 16, no. 3, **(2007)**, pp. 850-864.
[21] V. Vapnik, "The Nature of Statistical Learning Theory", Springer Verlag, **(1995)**.
[22] C. H. Wan, L. H. Lee, R. Rajkumar and D. Isa, "A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine", Expert Systems with Applications, vol. 39, **(2012)**, pp. 11880-11888.
[23] F. Wang, L. Vuurpijl and L. Schomaker, "Support vector machines for the classification of western handwritten capitals", Eur. Workshop Software Process Technol, **(2000)**, pp. 115-130.

