

# Vision-based Target Detection in Road Environments

S. ÁLVAREZ, M. A. SOTELO, M. OCAÑA, D. FERNÁNDEZ, I. PARRA

Department of Electronics

University of Alcalá

Edificio Politécnico, Campus Universitario s/n. 28871, Alcalá de Henares (Madrid)

SPAIN

sergio.alvarez, sotelo, mocana, llorca, [parra@depeca.uah.es](mailto:parra@depeca.uah.es)

<http://www.depeca.uah.es/personal/michael/irv>

*Abstract:* - This paper describes a target detection system on road environments based on Support Vector Machine (SVM) and monocular vision. The final goal is to provide car-to-car time gap. The challenge is to use a single camera as input, in order to achieve a low cost final system that meets the requirements needed to undertake serial production in automotive industry. The basic feature of the detected objects are first located in the image using vision and then combined with a SVM-based classifier. An intelligent learning approach is proposed in order to better deal with objects variability, illumination conditions, partial occlusions and rotations. A large database containing thousands of object examples extracted from real road images has been created for learning purposes. The classifier is trained using SVM in order to be able to classify cars and trucks. In addition, the vehicle detection system described in this paper provides early detection of passing cars and assigns lane to target vehicles. In the paper, we present and discuss the results achieved up to date in real traffic conditions.

*Key-Words:* - SVM (Support Vector Machine), Vision, Target detection and tracking.

## 1 Introduction

Insufficient distance keeping is a major source of rear-end accidents as many drivers find it difficult to keep adequate headway distance because it requires taking into account both the distance to the vehicle ahead and the travelling speed of their vehicle. The importance of keeping sufficient headway for reduction of accidents is recognized by traffic authorities worldwide and is being enforced in an increasing number of countries. Monocular vision can be used for vehicle detection and range measurement, and also to apply lane analysis in order to measure road geometry and curvature to determine the Closest In-Path Vehicle. The vision system can also detect and classify targets ahead of the host vehicle and send range and range rate (relative velocity) information to the ACC (Adaptive Cruise Control) controller to maintain a constant time gap between the host and followed vehicles. In such a case, the ACC controller automatically adjusts the speed of the host vehicle to maintain the desired headway by using throttle control and braking, and resumes to the set speed when the lane ahead is clear. All these vision-based functions can be globally considered as part of an Advanced Warning and Adaptive Cruise Control System. In this paper, we propose a monocular vision system for vehicle detection mainly intended for Adaptive Cruise Control functionality, although the results of this system can also be used for Headway Monitoring Warning and Emergency Braking.

Some previous developments use available sensing methods such as radar [9], stereo vision or a combination of stereo-vision and laser [6]. In [1] the authors propose the fusion of stereo-vision and radar for creating a hybrid velocity adaptive control system called HACC. Only a few works deal with the problem of monocular vehicle detection using symmetry and colour features [2] or pattern recognition techniques [8], including Support Vector Machines (SVM) [7]. In [2] the authors propose the use of horizontal edges and vertical symmetry together with a shape-dependent process for removing objects that are too small or too big in the image plane. In [5] the authors propose the use of a geometrical model for vehicle characterization using evolutionary algorithms, assigning different geometrical models depending on the vehicle lane. In [3] the authors develop an algorithm that provides night time vehicle detection by combination with Lane Departure Warning (LDW) in one-way roads for reducing false positive detections. Let us remark that the pattern recognition techniques used by all these systems for vehicle recognition can also be used for other eSafety applications such as Pedestrian Detection because of their generalization capability

In the current work, the searching space in the image plane is reduced in an intelligent manner in order to increase the performance of the vehicle detection module. Accordingly, road lane markings are detected and used as the guidelines that drive the

vehicle searching process. The area contained by the limits of the lanes is scanned in order to find vehicle candidates that are passed on to the vehicle recognition module. This helps reduce the rate of false positive detections. In case that no lane markings are detected, a basic *area of interest* is used instead covering the front part ahead of the ego-vehicle. The presence of collections of horizontal edges together with vertical symmetries triggers the attention mechanism.

## 2 Candidates Selection

An attention mechanism is necessary in order to filter out inappropriate candidate windows based on the lack of distinctive features, such as horizontal edges and vertical symmetrical structures, which are essential characteristics of road vehicles. This has the positive effect of decreasing both the total computation time and the rate of false positive detections. Each road lane is sequentially scanned, from the bottom to the horizon line of the image looking for collections of horizontal edges that might represent a potential vehicle. The scanned lines are associated in groups of three. For each group, a horizontality coefficient is computed as the ratio of connected horizontal edge points normalized by the size of the area being analysed. The resulting coefficient is used together with a symmetry analysis in order to trigger the attention mechanism. Apart from the detected road lanes provided by a Lane Departure Warning System (LDWS) developed by the authors in previous works [7], additional virtual lanes have been considered so as to cope with situations in which a vehicle is located between two lanes (for example, if it is performing a change lane manoeuvre). Virtual lanes provide the necessary overlap between lanes, avoiding both misdetections and double detections caused by the two halves of a vehicle being separately detected as two potential vehicles. A virtual lane is located to provide overlap between two adjoining lanes.

An adaptive thresholding process is implemented in order to obtain robust edges from the road images. This adaptive process is based on an iterative algorithm that gradually increases the contrast of the image, and compares the number of edges obtained in the contrast increased image with the number of edges obtained in the actual image. If the number of edges in the actual image is higher than in the contrast increased image the algorithm stops. Otherwise, the contrast is gradually increased and the process resumed. After thresholding, horizontal edges in the scanned regions given by the Lane Departure Warning (LDW) system are examined to detect the

rear part of potential vehicles. In order to decide if the collection of horizontal lines represents a possible vehicle candidate, its width is compared to that of an ideal car. The ideal car width is obtained for each vertical coordinate using the camera pinhole model. To calculate the  $Z$  distance, the pinhole model is used again. The origin of the vehicle coordinate system is located at the central point of the camera lens. The  $X_V$  and  $Y_V$  coordinates of the vehicle coordinate system are parallel to the image plane and the  $Z_V$  axis is perpendicular to the plane formed by the  $X_V$  and  $Y_V$  axes. A vehicle at a look-ahead distance  $Z$  from the camera will be projected into the image plane at a vertical and horizontal coordinates  $(u, v)$  respectively. The vertical road mapping geometry following this nomenclature is depicted in figure 1. The vertical model considers the flat terrain assumption and uses the following parameters:

**I:** Image plane

**Z:** look-ahead distance for planar ground (m)

**$h_{CAM}$ :** elevation of the camera above the ground (m)

**Pitch:** relative camera pitch angle (rad)

**$\theta_Z$ :** incident angle of the preceding vehicle's contact-to-asphalt point relative to vehicle pitch axis (rad)

**$v$ :** vertical image coordinate (pixels)

**HEIGHT:** vertical size of the CCD (pixels)

**$F_v$ :** vertical focal length (pixels)

**$F_u$ :** horizontal focal length (pixels)

**$k_v$ :** vertical scaling factor for the camera (pixels/mm)

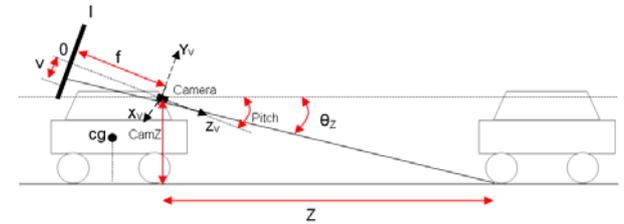


Figure 1. Vertical road mapping geometry.

According to figure 1, the vertical mapping geometry is mainly determined by the camera elevation  $h_{CAM}$  above the local ground plane as well as the pitch angle. The longitudinal axis of the vehicle is assumed to be always tangential to the road at the vehicle centre of gravity (cg). For each image scan line at  $v$ , there corresponds a pitch angle relative to the local tangential plane given by (1):

$$\theta_Z = Pitch + a \tan\left(\frac{v}{f K_v}\right) \quad (1)$$

Based on this, the planar look-ahead distance corresponding to  $v$ , is obtained as:

$$Z = \frac{h_{CAM}}{\tan(\theta_z)} \quad (2)$$

Applying a coordinate change due to the fact that the image origin in our case is on the top of the image instead of in the centre, the new vertical coordinate  $v(top)$  is given by:

$$v(top) = 2v(centre) - HEIGHT \quad (3)$$

In (4), the vertical scaling factor of the camera is introduced in the distance length parameter:

$$F_v = f K_v \quad (4)$$

The equation for computing the look-ahead distance  $Z$  becomes:

$$Z = \frac{h_{CAM}}{\tan\left(Pitch + atan\left(\frac{2 \cdot v - HEIGHT}{F_v}\right)\right)} \quad (5)$$

Once the car width is computed at the current frame it is compared to the collection of horizontal lines found after the thresholding analysis. If they are similar to some extent defined by an empirical value, a square area above the collection of horizontal lines, denoted as candidate ROI, is considered for further analysis. The aim is to compute the entropy of the candidate ROI and its vertical symmetry. Only those regions containing enough entropy and symmetry are identified as potential vehicles. Figure 2 depicts two examples of the detection step, while figure 3 shows a detailed block diagram of the detection procedure.

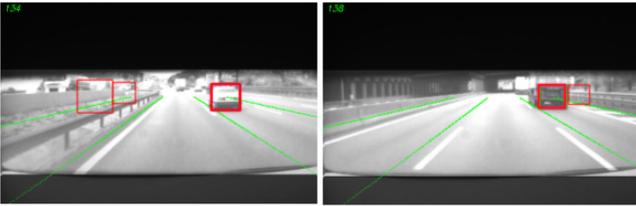


Figure 2. Examples of potential candidates detected by the attention mechanism (vehicles and non-vehicles).

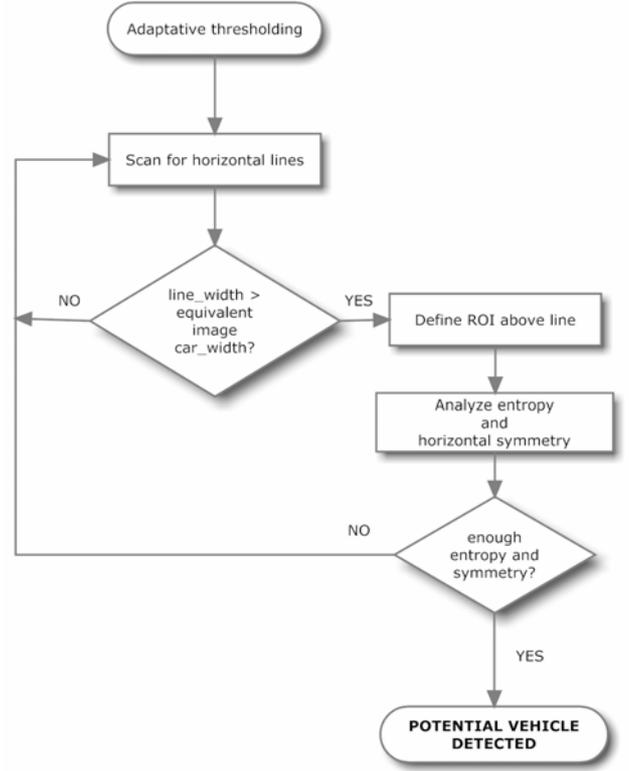


Figure 3. Block diagram of the vehicle detection mechanism.

Accurate detection of the wheel-to-road contact point of the preceding vehicle is essential for assuring maximum precision of the host-to-vehicle estimated distance. Thus, the error committed in estimating the host-to-vehicle distance  $Z_{err}$  due to a vehicle detection error of  $n$  pixels in the image plane is given by:

$$Z_{err} = Z_n - Z = \frac{F_v h_{CAM}}{v + n} - Z = \frac{-nZ^2}{F_v h_{CAM} + nZ} \quad (6)$$

where  $v$  is the vertical coordinate of the wheel-to-road contact point in the image plane,  $Z$  is the estimated host-to-vehicle distance, and  $h_{CAM}$  represents the camera height (as previously defined). Considering an error of one pixel  $n=1$  and  $F_v h_{CAM} \gg nZ$ ,  $Z_{err}$  becomes:

$$Z_{err} \approx \frac{nZ^2}{F_v h_{CAM}} \quad (7)$$

For example, for a 640x480 image, a focal length of 740 pixels, and a camera height  $h_{CAM}=1.2m$ , an error of 1 pixel ( $n=1$ ) becomes a relative 5% error at a distance:

$$Z = \frac{Z_{err}}{Z} F_v h_{CAM} = 0.05 * 740 * 1.2 = 44m \quad (8)$$

On the other hand, the error at 90m is 10%. These values are more than enough for the ACC function. What is really important is the measurement of relative host-to-vehicle velocity. Relative velocity  $Rv$  is computed using the following equation:

$$Rv = \frac{\Delta Z}{\Delta t} \quad (9)$$

Based on the scale change  $s$  of detected objects in the image plane, the optimal value of  $\Delta t$  that minimizes the estimation noise can be calculated. Let  $W$  denote the width (in meters) of the preceding vehicle,  $w$  and  $w'$  the width of the preceding vehicle in the image plane when it is located at distances  $Z$  and  $Z'$ , respectively, with regard to the host vehicle. The scale change  $s$  can be defined as:

$$s = \frac{w - w'}{w'} \quad (10)$$

Finally, the estimated relative velocity can be computed as follows:

$$Rv = \frac{\Delta Z}{\Delta t} = \frac{Z \frac{w - w'}{w'}}{\Delta t} = \frac{Zs}{\Delta t} \quad (11)$$

### 3 Vehicle Recognition and Tracking

Detected candidates are classified as vehicles or non-vehicles depending on features obtained from the vehicle ROI using Support Vector Machines (SVM). Support vector machines (SVMs) are a set of related supervised learning methods used for classification and pattern recognition. One special property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin. Hence they are also known as maximum margin classifiers. Support vector machines map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed, by maximizing the distance between both classes. An assumption is made that the larger the margin or distance between these parallel hyperplanes the better the generalization error of the classifier will be. The output of the SVM,  $D$ , is simply the signed distance of the test instance from the separating hyperplane. This output indicates whether the analyzed object corresponds to a vehicle (+1, in theory) or not (-1, in theory) and can be used as a threshold for separating them. Two aspects are essential in the deployment of SVM classifiers: the training strategy and the classifier structure.

#### 3.1 Training Strategy

The first step in the design of the training strategy is to create representative databases for learning and testing. The following considerations must be taken into account when creating the training and test sets.

- The ratio between positive (vehicles) and negative (others) samples has to be set to an appropriate value. A very large number of positive samples in the training set may lead to a high percentage of false positive detections during on-line classification. On the contrary, a very large number of negative samples produce misslearning.
- The size of the database is a crucial factor to take care of. As long as the training data represent the problem well, the larger the size of the training set the better for generalization purposes.
- A sufficiently representative test set must be created for verification. The content of the test set has similar characteristics to those of the training sets in terms of variability and ratio of positive/negative samples.

#### 3.2 Classifier Structure

An input vector for the classifier was defined. This vector is composed of different parameters which are computed for all candidates and define the state vector for the SVM. Those parameters are local histograms of oriented gradients (HOG) [4]. The aim of this method is to describe an image by a set of local histograms which count occurrences of gradient orientation in a local part of the image (the selected candidate ROI). As a general overview, the algorithm is composed of the following steps:

- Parameters of the detected objects (HOG) are computed and used as inputs (SVM Feature Vector) to the SVM classifier.
- Once the parameters vector is computed, the SVM process analyzes this vector and returns a value which is simply the signed distance of the test instance from the separating hyperplane.

#### 3.3 Vehicle Tracking

After detecting consecutively an object a predefined number of times (empirically set to 3 in this work), tracking is implemented using Kalman filtering techniques. The purpose of the Kalman filtering is to obtain a more stable position of the detected vehicles. Besides, oscillations in vehicles position due to the unevenness of the road makes  $y$  coordinate of the detected vehicles change several pixels up or down. This effect makes the distance detection unstable, so a Kalman filter is necessary for minimizing these

kinds of oscillations. As a future idea, even though an image correction and filtering can be done, it would be much more efficient to go through this problem by introducing an oscillation sensor in the car.

#### 4 Implementation and Results

The system was implemented on a PC Pentium IV at 2.4 GHz onboard a Citroën C4 and tested in real traffic conditions using a 640x320 CMOS camera. The training database contains 10.000 representative samples while the test set has 3.000 samples. In both cases, a positive/negative ratio of 1:2 has been observed. The size of the database (10.000 samples) represents a crucial factor to take care of. To obtain a sufficiently representative set we have taken cars and trucks as positive samples, and crash barriers, median strip, pieces of road, etc, like negative samples. The samples have been taken in different weather conditions (with and without rain, shadows, etc). The content of the test set has similar characteristics to those of the training set in terms of variability and ratio of positive/negative samples. The size of the test set (3.000 samples) is appropriate for verification of the overall system. To create the samples sets, we have developed a tool called “ACC Database”. This tool represents an extended option of the main software used for vehicle detection. The tool allows entering the candidates extracted by the car detection system as positive or negative samples in the database. Using the ACC Database tool an intensive training stage was accomplished. Table 1 shows the number of samples obtained for training and testing.

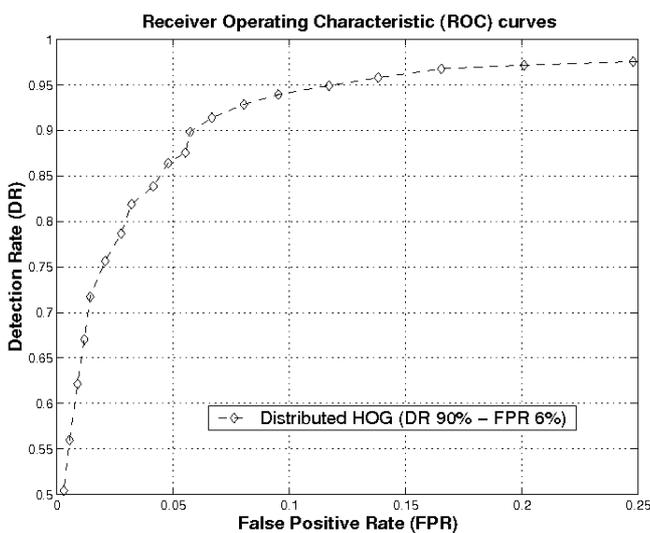


Figure 4. Receiver Operating Characteristic Curve (ROC Curve).

Figure 4 depicts the Receiver Operating Characteristic (ROC) Curve. Table 2 provides a summary of statistics concerning global system performance. The table shows the results achieved using the previously described database containing 13.000 samples. The average processing time per frame (Tpf) is given in ms, as well as the number of detected vehicles, missing vehicles and number of false alarms. As can be observed from table 2, not only the detection rate and false alarm rate are provided, but also the reasons that cause it. The system yields a global Detection Rate of 90.32% with 1 False Alarm. Miss detections mainly occur with motorcycles and trucks under heavy rain. Motorcycles miss detections can be solved by incorporating a sufficient number of motorcycle images in the database. For this purpose, the candidate selection mechanism should be modified in order to raise candidates with the shape and aspect of motorcycles, not only cars, trucks and buses. Indeed, if the effect of miss-detected motorcycles is neglected, a Detection Rate of 96.77% is achieved for cars, trucks and buses. In order to diminish the number of false alarms due to road artefacts, such as road fences, these types of elements should be included in the database as negative samples. Although these elements are already included in the current database, it should be further enlarged and enriched until proper generalization will be achieved.

#### 5 Conclusions

We have developed and implemented a vehicle detection system based on Support Vector Machine (SVM) and monocular vision with the objective of providing car-to-car time gap measurement for Adaptive Cruise Control (ACC) applications in the framework of Intelligent Transportation Systems (ITS). Vehicle candidates are raised using an attention mechanism based on horizontal edges, vertical symmetry and entropy. The detected objects are passed on to a SVM-based classifier. After classification, detected vehicles are tracked using Kalman filtering. A large database containing thousands of vehicle examples extracted from real road images has been created for learning purposes. The classifier is trained using SVM in order to be able to classify cars and trucks. In addition, the vehicle detection system described in this paper provides early detection of passing cars and assigns lane to target vehicles based on the use of a Lane Departure Warning System (LDWS). After assessment of the practical results achieved in our experiments, the following general conclusions can be summarized:

Table 1. Number of samples obtained from video sequences for training and testing.

Video	Frames	Positive	Negative	Training/Test
Sequence_1 (rain)	4962	721	2183	Training
Sequence_2 (dry, bridges)	2214	954	3071	Training
Sequence_3 (dry, glare)	2215	502	765	Training
Sequence_4 (rain)	2213	297	1218	Test
Sequence_5 (cloudy)	2664	1327	423	Training
Sequence_6 (cloudy)	1338	725	808	Test

Table 2. Global performance of the ACC system.

Video	#Frames	Tpf (ms)	Detected	Missed	FA	Cause
Sequence_7 (cloudy)	2.456	62	1	0	0	--
Sequence_8 (sunny)	3.765	33	9	1	0	Motorcycle
Sequence_9 (rain)	1.987	30	1	1	0	Truck (rain)
Sequence_10 (cloudy)	2.367	31	8	0	1	Road fence
Sequence_11 (rain)	2.678	53	9	1	0	Motorcycle

- The global performance of the monocular daytime ACC developed and described in this paper yields a Detection Rate above 90% for a False Alarm Rate around 1%.
- The performance of ACC is significantly increased by building on the output provided by the LDWS function.
- The presence of large shadows on the asphalt due to vehicles circulating along the road produces negative effects on the candidate selection mechanism, yielding to inaccuracy in measuring the distance to the vehicles.

#### Acknowledgements

This work has been supported by means of Research Grant DPI2005-07980-C03-02.

#### References:

- [1] L. Bombini, P. Cerri, P. Medici, and G. Alessandretti. "Radar-vision fusion for vehicle detection". In *Procs. International Workshop on Intelligent Transportation*, pp. 65-70, Hamburg, Germany, March 2006.
- [2] A. Broggi, P. Cerri, and P. C. Antonello. "Multi-Resolution Vehicle Detection using Artificial Vision". *IEEE Intelligent Vehicles Symposium*. Parma, Italy. 2004.
- [3] M. Y. Chem and P. C. Hou. "The Lane Recognition and Vehicle Detection at Night for a Camera-Assisted Car on Highway". *IEEE International Conference on Robotics & Automation*, 2003.
- [4] N. Dalal and B. Triggs. "Histogram of oriented gradients for human detection". In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886-893. 2005.
- [5] C. Hilario, J. M. Collado, J. M. Armingol, and A. de la Escalera. "Visual Perception and Tracking of Vehicles for Driver Assistance Systems". *IEEE Intelligent Vehicles Symposium*. 2006.
- [6] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, "Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner". In *Proc. Int. Conf. On Advanced Robotics*, pp. 1538-1543, 2003.
- [7] M. A. Sotelo, J. Nuevo, L. M. Bergasa, and M. Ocaña. "A monocular solution to vision-based ACC in road vehicles". *Lecture Notes in Computer Science*. Vol. 3643, 507-512. 2005.
- [8] G. P. Stein, O. Mano, and A. Sashua, "Vision-based ACC with a single camera: bounds on range and range rate accuracy". In *Proc. Int. Conf. Intelligent Vehicles*, 2002.
- [9] G. R. Widman, W. A. Bauson, and S. W. Alland, "Development of collision avoidance systems at Delphi Automotive Systems". In *Proc. Int. Conf. Intelligent Vehicles*, pp. 353-358, 1998.