Vision-based Traffic Data Collection for Floating Car Data Enhancement

D. F. Llorca, M. A. Sotelo, S. Sánchez, C. Fernández, J. J. Vinagre-Díaz and J. Ramiro-Bargueño

***Abstract*— This paper presents a complete vision-based vehicle detection system for Floating Car Data (FCD) enhancement in the context of Vehicular Ad hoc NETworks (VANETs). Three cameras (side, forward and rear looking cameras) are installed onboard a vehicle in a fleet of public buses. Thus, a more representative local description of the traffic conditions (extended FCD) can be obtained. Specifically, the vision modules detect the number of vehicles contained in the local area of the host vehicle (traffic load) and their relative velocities. Absolute velocities (average road speed) and global positioning are obtained after combining the outputs provided by the vision modules with the data supplied by the CAN Bus and the GPS sensor. This information is transmitted by means of a GPRS/UMTS data connection to a central unit which merges the extended FCD in order to maintain an updated map of the traffic conditions (traffic load and average road speed). The presented experiments are promising in terms of detection performance and computational costs. However, significant effort is further necessary before deploying a system for large-scale real applications.**

I. INTRODUCTION

Floating Car Data (FCD) refers to technology that collects traffic state information from a set of individual vehicles which float in the current traffic. Each vehicle, which can be seen as a moving sensor that operates in a distributed network, is equipped with positioning (GPS) and communication (GSM, GPRS, UMTS, etc.) systems, transmitting its location, speed and direction to a central control unit that integrates the information provided by each one of the vehicles.

FCD systems are being increasingly used in a variety of important applications since they overcome the limitations of fixed traffic monitoring technologies (installation and maintenance costs, lack of flexibility, static nature of the information, etc.). We refer to [1] for general background concerning the most representative FCD activities in Japan, Europe and the United States.

FCD can be used by the public sector to collect road traffic statistics and to carry out real-time road traffic control. The information provided by FCD systems can be supplied to individual drivers via dynamic message signs, PDA devices, satellite navigation systems or mobile phones, including dynamic re-routing information. Thus, drivers would be able to make more informed choices, spending less time in congested traffic. In addition, the knowledge of the current traffic situation can be also used to estimate time of arrival of a fleet of public transport vehicles and, furthermore, to plan and coordinate the movements of the fleet (fleet management) so that driving assignments can be carried out more efficiently. Besides previous applications, the use of FCD entails environmental benefits since it can be used to reduce fuel consumption and emissions.

The basic data provided by FCD systems (vehicle location, speed and direction) can be enriched using new onboard sensors (ambient temperature, humidity and light, windshield wiper status, fog light status, fuel consumption, emissions, tire pressure, suspension, emergency brake, etc.) which are centralized by means of the controller-area-network (CAN) bus. Such data can be exploited to extend the information horizon including traffic, weather, road management and safety applications [1]. In addition, computer vision systems can be included in order to improve the automatic detection of potentially interesting events and to document them by sending extended data [2].

In order to provide ubiquitous coverage of the entire road network, a minimum representation of the total passenger car fleet has to be used, since each moving sensor (each vehicle) only supplies information about its status. The fact that everyday road users have to be asked to share information regarding their movements and speeds arises privacy issues that have to be addressed. Another approach consists of using the information supplied by a specific fleet of vehicles. Taxis or public transport buses can be used due to the extended periods of time they spend on the urban road network.

This paper presents a complete vision-based vehicle detection system onboard a fleet of public transport buses with the aim of improving the data collected in FCD applications. The proposed system has been developed in the framework of the GUIADE project. Three cameras covering the local environment of the vehicle are used: forward, rear and side looking cameras. The system obtains the number of vehicles in the local range of the bus as well as their relative position and velocity. This information is combined with the data provided by regular FCD systems (global location, speed and direction), obtaining a more detailed description of the local traffic load and the average speed. The communication system between the vehicles and the central control unit is based on wireless technology via GPRS/UMTS cellular protocols. Finally, the central unit integrates the data collected by the fleet in order to generate updated traffic status maps.

The remainder of this paper is organized as follows: the system description is summarized in Section II. Section III describes the vision-based vehicle detection system. Section

D. F. Llorca, M. A. Sotelo, S. Sánchez and C. Fernández are with the Department of Electronics, University of Alcalá, Madrid, Spain. llorca@depeca.uah.es

J. J. Vinagre-Díaz and J. Ramiro-Bargueño are with the Department of Signal Theory and Communications, Rey Juan Carlos University, Madrid, Spain. juanjose.vinagre@urjc.es

IV presents the spatial and temporal integration of the data collected by the vision system. Experimental results are presented in Section V. Finally, conclusions and future works are discussed in Section VI.

II. SYSTEM DESCRIPTION

The proposed FCD architecture can be seen in Figure 1. Floating car data is supplied by a fleet of public transport buses which corresponds to an inner-city bus line. Each vehicle is equipped with a Global Positioning System (GPS), wireless communication interfaces (GPRS/UMTS and WLAN IEEE 802.11) and a complete vision-based vehicle detection system.

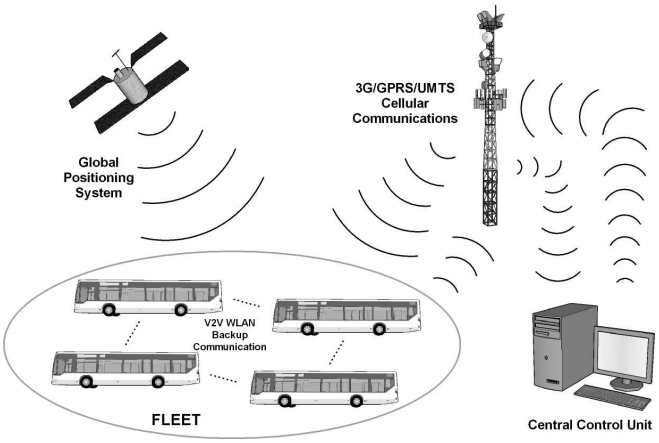

Fig. 1. Overview of the proposed FCD architecture.

The Vehicle-to-Infrastructure (V2I) communication system is based on the geographic coverage provided by cellular networks. General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS) are used to connect each vehicle with the central control unit. Each vehicle provides information that can be divided in three main groups:

1) Standard FCD information: vehicle identifier, timestamp, GPS position, speed and direction.
2) Vehicle status information: ambient temperature, humidity, light, windshield wiper status, fog light status, fuel consumption and emissions.
3) Extended FCD information: globally referenced average traffic load and average road speed for a measured segment travel time.

The central control unit integrates the information provided by each one of the vehicles in order to compute updated traffic and weather maps which will be used for fleet management tasks as well as to estimate the time of arrival. The Vehicle-to-Vehicle (V2V) communication system is defined as a backup communication system based on a wireless-fidelity (WiFi) IEEE 802.11a/b/g interface. In situations where the cellular network is not working, in-range vehicles will exchange the most updated information available.

Each individual vehicle is equipped with three cameras (forward, rear and side looking cameras) that cover the local environment of the bus (see Figure 2). A common hardware trigger synchronizes the image acquisition of the three cameras and an onboard PC houses the computer vision software. Each vehicle detection system provides information about the number of detected vehicles and both their relative position and speed. These results are combined with the GPS measurements and the data provided by the CAN bus in order to provide globally referenced traffic information. This scheme is described in Figure 3. The layers of the proposed architecture of the three vision modules are conceptually the same: lane detection, vehicle candidates selection, vehicle recognition and tracking.

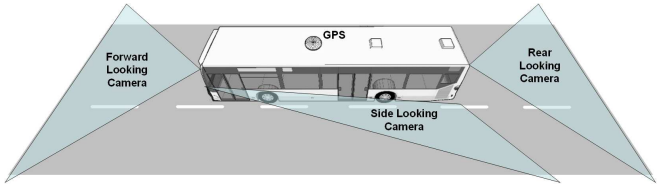

Fig. 2. Main vehicle sensors: three cameras (forward, rear and side looking cameras) and a Global Positioning System.

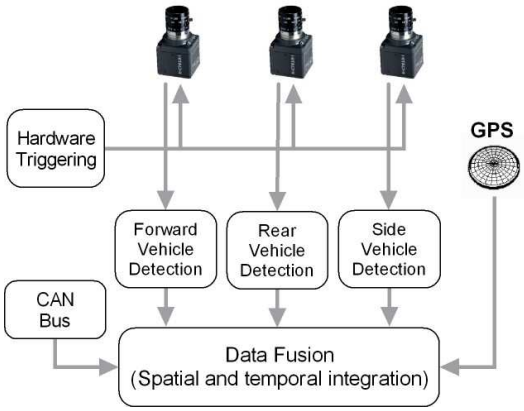

Fig. 3. Stages of the vision-based traffic detection system.

III. VISION-BASED TRAFFIC DETECTION SYSTEM

A. Lane Detection

Lane markings are detected using gradient information in combination with a local thresholding method which is adapted to the width of the projected lane markings. Then, clothoid curves are fitted to the detected markings. The algorithm scans up to 25 lines in the candidates searching area, from 2 meters in front of the camera position to the maximum range in order to find the lane marking measurements. The proposed method implements a non-uniform spacing search that reduces certain instabilities in the fitted curve. The final state vector is composed of 6 variables [3] for each lane on the road. Figure 4 provides some examples of lane markings detection in real outdoor scenarios. Detected lanes determine the vehicle searching area and help reduce false positive detections. In case no

lane markings are detected by the system, fixed lanes are assumed instead.

Fig. 4. Vehicle searching area as a result of the lane markings analysis for forward, rear and side modules.

B. Vehicle detection

Side, forward and rear looking vehicle detection systems share the same algorithmic core. Side vehicle detection includes optical flow cues as described in [4]. The attention mechanism sequentially scans each road lane from the bottom to the maximum range looking for a set of features that might represent a potential vehicle. Firstly, the vehicle contact point is searched by means of the top-hat transformation. This operator allows the detection of contrasted objects on non-uniform backgrounds [5]. There are two different types of top-hat transformations: white hat and black hat. The white hat transformation is defined as the residue between the original image and its opening. The black hat transformation is defined as the residue between the closing and the original image. In our case we use the white hat operator since it enhances the boundary between the vehicles and the road [6]. Horizontal contact points are pre-selected if the number of white top-hat features is greater than a configurable threshold. Then, candidates are pre-selected if the entropy of Canny points is high enough for a region defined by means of perspective constraints and prior knowledge of target objects (see Figure 5)

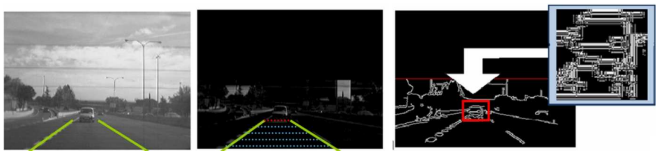

Fig. 5. From left to right: original image; contact point detection on white top-hat image; candidate pre-selected with high entropy of Canny points.

In a second step, vertical edges (S_v), horizontal edges (S_h) and grey level (S_g) symmetries are obtained, so that, candidates will only pass to the next stage if their symmetries values are greater than a threshold. The vertical and horizontal edges symmetries are computed as listed in Algorithm 1. The grey level symmetry computation procedure is shown in Algorithm 2. Some examples of the three types of symmetries are depicted in Figure 6.

Symmetry axes are linearly combined to obtain the final position of the candidate. Finally, a weighted variable is defined as a function of the entropy of Canny points, the three symmetry values and the distance to the host vehicle. We use this variable to apply a non-maximum suppression

Algorithm 1 Vertical/Horizontal edges symmetries

```
Initialize Acc_{0,...,ROI_WIDTH} = 0
for i = 0, ..., ROI_HEIGHT do
    for each pair of vertical/horizontal edges pixels (x_1, i), (x_1, i) do
        Acc_{(x_1+x_2)/2} + +
    end for
end for
S_{v,h} = arg max_i(Acc_i / S_{v,h,MAX})
```

Algorithm 2 Gray level symmetries

```
for each possible symmetry axis x_i do
    Initialize S_i = 0
    for j = 0...ROI_HEIGHT do
        for k = 0...ROI_WIDTH/2 do
            if abs(I(j, x_i + k) − I(j, x_i − k)) < δ then
                S_i + +
            end if
        end for
    end for
end for
S_g = arg max_i(S_i / (area_ROI/2))
```

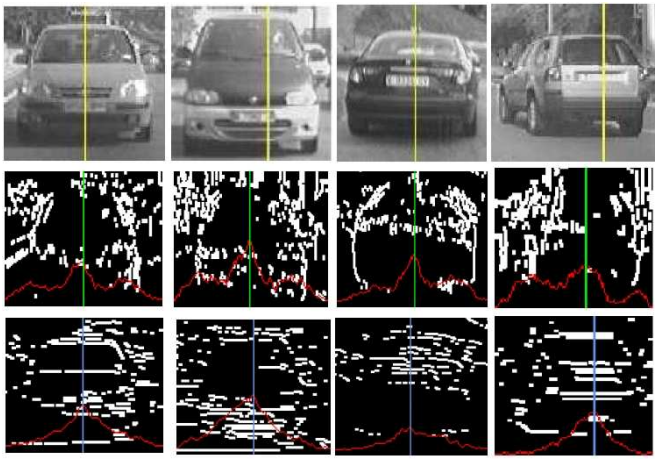

Fig. 6. Upper row: gray level symmetry; Middle row: vertical edges symmetry; Lower row: horizontal edges symmetry.

Fig. 7. overlapped candidates. Right: non-maximum suppression results.

process per lane which removes overlapped candidates. An example of this process is shown in Figure 7.

The selected candidates are classified by means of a linear SVM classifier[7], in combination with Histograms of Oriented Gradients features[8]. We have developed and tested two different classifiers depending on the module (forward and rear/side classifiers). All candidates are resized to a fixed size of 64x64 pixels to facilitate the features extraction process. The rear-SVM classifier is trained with 2000 samples and tested with 1000 samples (1/1 positive/negative ratio) whereas the forward-SVM classifier is trained with 3000 samples and tested with 2000 samples (1/1 positive/negative ratio). Figures 8 and 9 depict some positive and negative samples of the forward and rear/side training and test data sets respectively.

Fig. 8. Forward data set. Upper row: positive samples (vehicles). Lower row: negative samples.

Fig. 9. Rear data set. Upper row: positive samples (vehicles). Lower row: negative samples.

After detecting consecutively an object classified as vehicle a predefined number of times (empirically set to 3 in this work), data association and tracking stages are triggered. The data association problem is addressed by using feature matching techniques. Harris features are detected and matched between two consecutive frames as depicted in Figure 10.

Tracking is implemented using Kalman filtering techniques. For this purpose, a dynamic state model and a measurement model must be defined. The proposed dynamic state model is simple. The state vectoris defined as $x_n = \{u, v, w, h, \dot{u}, \dot{v}, \dot{w}, \dot{h}\}^T$, where u and v are the respective horizontal and vertical image coordinates for the top left corner of every object, and w and h are the respective width and height in the image plane. The measurement vector is defined as $z_n = \{u, v, w, h\}^T$.

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 v 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.

Fig. 10. Data association by features matching. Upper row: Harris features on image t. Lower row: matched Harris features on image $t+1$.

IV. FCD INTEGRATION

FCD integration or Data Fusion module uses three sources of data: the measurements provided by the GPS, the data supplied by the CAN bus and the output obtained from the three vision-based vehicle detection modules. Whereas the GPS and the CAN bus sample frequency is 1Hz, the vision-based system operates in real-time at 25 frames per second (25Hz). The proposed data fusion scheme provides information at the lowest sample frequency (1Hz) covering two consecutive GPS measurements, the vehicle speed v_i^h (via CAN bus) and the outputs of the vision module.

The outputs of the side, forward and rear vehicle detection systems at frame i are the number of detected vehicles N and their corresponding distances to the host vehicle $d_i^{(k)}$. These outputs are combined to cover the whole local environment of the vehicle. The traffic load at frame i is given by next expression:

$$L_i = (N_i + 1)/N_{MAX} \tag{1}$$

where N_{MAX} is the maximum number of vehicles in range that can be detected by the three systems (in our case N_{MAX} is defined as 9 or 13 for two lanes and three lanes roads respectively). The average road speed at frame i is computed as follows:

$$v_i = \frac{1}{N_i+1}\left(\sum_{k=0}^{N_i-1}\left(\frac{\left(d_i^{(k)} - d_{i-1}^{(k)}\right)}{\Delta t} + v_i^h\right) + v_i^h\right) = \\ = \frac{1}{(N_i+1)\Delta t}\sum_{k=0}^{N_i-1}\left(d_i^{(k)} - d_{i-1}^{(k)}\right) + v_i^h \tag{2}$$

where $d_i^{(k)}$ and $d_{i-1}^{(k)}$ represent the distance between the host vehicle and vehicle k at frames i and $i-1$ respectively, Δt corresponds to the sample time, v_i^h is the host vehicle speed provided by the CAN bus, and N_i is the number of detected vehicles. Note that the distance values correspond to filtered measurements since they are obtained from the first two elements of the Kalman filter state vector (u and v) using known camera geometry and ground-plane constraints.

Two consecutive GPS measurements define both a spatial and a temporal segment. The temporal segment corresponds to the GPS sample time (1 second), and the spatial segment will be defined as the globally referenced trajectory between the two GPS measurements. In order to obtain the extended FCD information (i.e., the road traffic load and the road speed) for this spatio/temporal segment we integrate the values supplied by the vision modules during 25 consecutive frames. With this approach a dense coverage of the road traffic load and the road speed can be assured for host vehicle speeds up to 180km/h since the total range of the vision module covers more than 50m (25 meters for both the rear and the forward looking modules; the side range covers up to two third parts of the bus length in the adjacent lane). Obviously this maximum speed will never be exceeded by a public bus. This approach facilitates further map-matching tasks since the extended FCD information between two consecutive points will always be globally referenced.

V. EXPERIMENTAL RESULTS

The system was implemented on a PC Core 2 Duo at 3.0 GHz and tested in real traffic conditions using CMOS cameras with low resolution images (320×240). After training and test, a trade-off point has been chosen at Detection Rate (DR) of 95% and False Positive Rate (FPR) of 5% for the rear-SVM classifier and at DR of 90% and FPR of 6% for the forward-SVM classifier. We have to note that these numbers are obtained in an off-line single-frame fashion, so that, they will be improved in subsequently stages. In addition, the lane detection system reduces the searching area and the number of false candidates passed to further stages.

In order to validate the proposed vision-based vehicle detection system as an extended source for FCD applications we have recorded several video sequences in real traffic conditions and we have manually labeled the number of vehicles in range at every frame (a total of 800 frames). The speed of the host vehicle was around 90km/h so the length of the traveled route was 1km approximately. Both the traffic load L_i and the average road speed v_i are computed at every frame using equations 1 and 2.

Figure 11 shows the ground truth and the number of vehicles detected in range. Most of the errors take places in cases where the host vehicle is passing beneath a bridge due to strong illumination changes (see Figure 12) and in curves or cases where there are strong changes in the vehicle pitch, roll or camera height.

The traffic load L_i, the ground truth and the corresponding absolute error are depicted in Figure 13. The overall RMSE in the traffic load computed by the proposed approach is 0.07

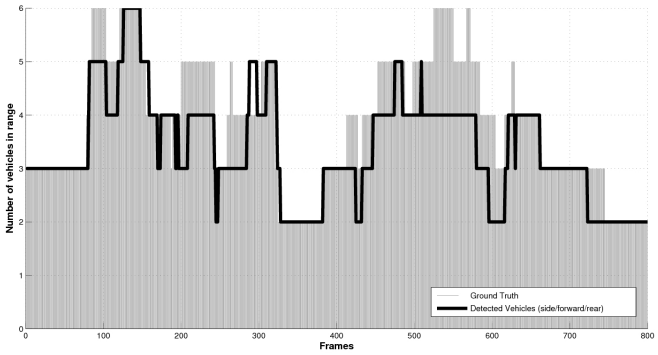

Fig. 11. Number of vehicles detected by the three vision modules compared with the manually labeled ground truth in a real sequence.

Fig. 12. Examples with strong illumination changes after passing beneath a bridge.

(7%). The average road speed v_i at every frame is depicted in Figure 14.

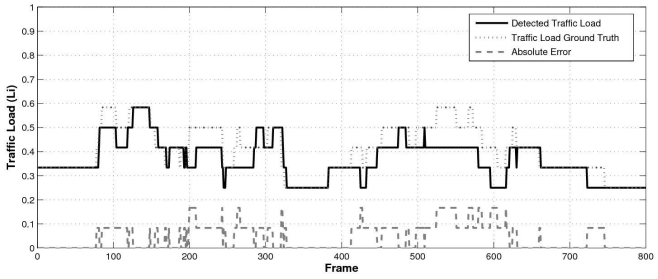

Fig. 13. Traffic load L_i at every frame in a real sequence.

Both the traffic load and the average road speed are provided by the vision modules at a frequency of 25Hz. As the extended FCD information is supplied to the central unit at a frequency of 1Hz the traffic load and the average road speed are finally integrated during 25 consecutive frames. These results are shown in Figures 15 and 16. We use a colour code to describe the level of traffic load and the road speed: green indicates that there is good flow/high speeds, yellow indicates that there is semi-dense traffic/medium speeds, and red shows dense traffic/slow speeds (traffic jams).

After combining the results with the GPS measurements we can obtain the traffic load in Universal Transverse Mercator (UTM) coordinates, as depicted in Figure 17 (note that map-matching is not carried out; the aerial image has been obtained from Google Earth).

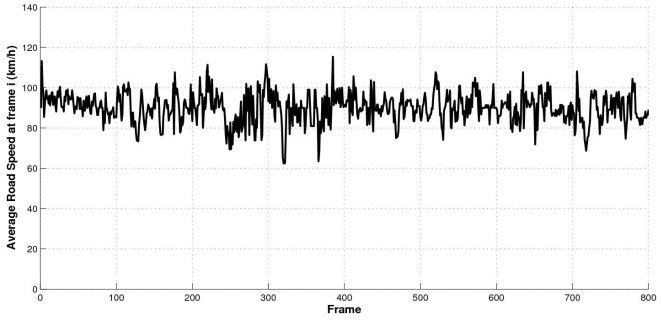

Fig. 14. Average road speed v_i at every frame in a real sequence.

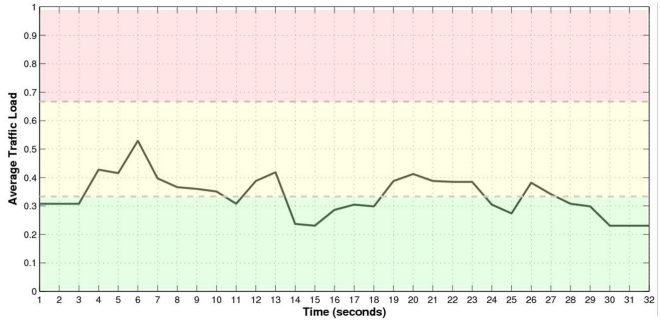

Fig. 15. Average Traffic Load at every second in a real sequence.

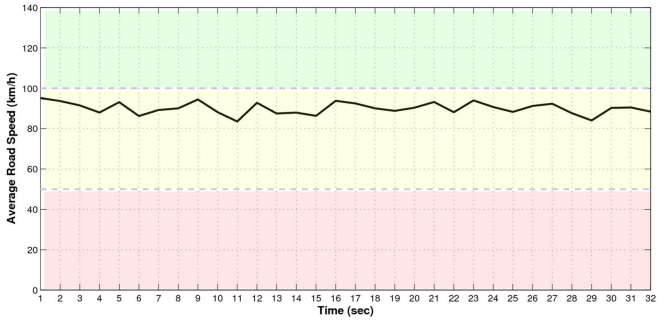

Fig. 16. Average Road Speed at every second in a real sequence.

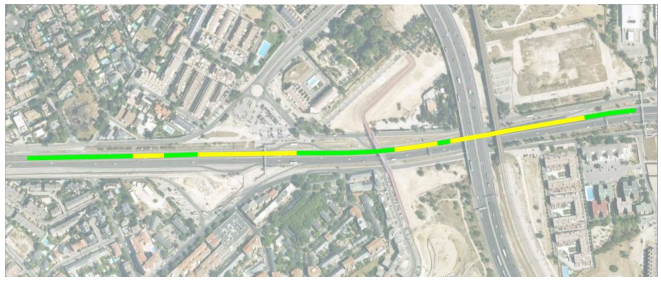

Fig. 17. GPS trajectory and the corresponding traffic load computed at the central unit (the aerial image has been obtained from Google Earth).

VI. CONCLUSIONS AND FUTURE WORKS

This paper presented a complete vision-based vehicle detection system that enhances the data supplied by FCD systems in the context of vehicular ad hoc networks. The system is composed of three vision sub-systems (side, forward and rear sub-systems) that detect the traffic load and the relative velocities of the vehicles contained in the local area of the host vehicle. Absolute velocities and global positioning are obtained after combining the outputs provided by the vision modules with the outputs supplied by the CAN Bus and the GPS sensor. Standard FCD systems provide the vehicle position, speed and direction. The proposed approach extends this information by including more representative measurements corresponding to the traffic load and the average road speed.

In order to cover the entire road network, the proposed vision-based system is defined for being installed onboard a fleet of public buses where privacy is a minor issue. The extended packets collected by each moving vehicle are transmitted to the central unit by means of a GPRS/UMTS data connection. The central unit merges the extended FCD in order to maintain an updated map of the traffic conditions (traffic load and average road speed).

The presented experiments are promising in terms of detection performance and computational costs. However, significant effort is further necessary before deploying a system for large-scale real applications. For this purpose, new experiments will be carried out merging the data collected by more than one vehicle, including map-matching techniques and further analysis on V2V and V2I communications.

VII. ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Development by means of Research Grant GUIADE P9/08.

REFERENCES

[1] R. Bishop, Intelligent Vehicle Technologies and Trends. Artech House. 2005.

[2] S. Messelodi, C. M. Modena, M. Zanin, F.G.B. De Natale, F. Granelli, E. Betterle and A. Guarise. Intelligent extended floating car data collection. *Expert Systems with Applications.* vol. 36, no. 3, part 1, pp 4213-4227, 2009.

[3] M. A. Sotelo, J. Nuevo, L. M. Bergasa, M. Ocaña, I. Parra, D. Fernández. Road Vehicle Recognition in Monocular Images. In. *Proceedings of IEEE ISIE*, pp. 1471-1476, 2005.

[4] M. A. Sotelo and J. Barriga. Blind Spot Detection using Vision for Automotive Applications. *Journal of Zhejiang University SCIENCE A.* 9(10): 1369-1372, 2008.

[5] Y. Derong, Z. Yuanyuan, and L. Dongguo. Fast Computation of Multi-scale Morphological Operations for Local Contrast Enhancement. In. *Proceedings of Engineering in Medicine and Biology 27th Annual Conference.* 2005.

[6] D. Balcones, D. F. Llorca, M. A. Sotelo, M. Gavilán, S. Álvarez, I. Parra and M. Ocaña. Real-time Vision-Based Vehicle Detection for Rear-End Collision Mitigation Systems. *EUROCAST 2009, LNCS 5717,* pp. 320-325, 2009.

[7] J.C. Christopher Burges. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2,121-167. Kluwer Academic Publishers.1. 1998.

[8] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In. *Proceedings of IEEE CVPR,* pp. 886-893, 2005.