Extended Floating Car Data System - Experimental Study -

R. Quintero, A. Llamazares, D. F. Llorca, M. A. Sotelo, L. E. Bellot, O. Marcos, I. G. Daza, C. Fernández

Abstract—This paper presents the results of a set of extensive experiments carried out in daytime and nighttime conditions in real traffic using an enhanced or extended Floating Car Data system (xFCD) that includes a stereo vision sensor for detecting the local traffic ahead. The detection component implies the use of previously monocular approaches developed by our group in combination with new stereo vision algorithms that add robustness to the detection and increase the accuracy of the measurements corresponding to relative distance and speed. Besides the stereo pair of cameras, the vehicle is equipped with a low-cost GPS and an electronic device for CAN Bus interfacing. The xFCD system has been tested in a 198-minutes sequence recorded in real traffic scenarios with different weather and illumination conditions, which represents the main contribution of this paper. The results are promising and demonstrate that the system is ready for being used as a source of traffic state information.

I. INTRODUCTION

The so-called 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 global positioning and communication systems, transmitting its global 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 [1]. If this system achieves a sufficient penetration rate (1.5% as described in [2]), the service quality in urban traffic would be sufficient. The most representative FCD projects in Japan, Europe and the United States until 2005 are described in [3].

The basic data provided by FCD systems can be extended (xFCD) using new specific devices and sensors endowed in modern vehicles [4], [5]. Such data can be exploited to extend the information horizon including traffic, weather, road management, and safety applications [3]. A second generation of xFCD has been recently proposed by including vision-based sensors in order to estimate the local traffic conditions. For example in [6] a stereo vision-based detection module is used as a vehicle ahead checker (excluding lateral road lanes) in order to validate the traffic alarms generated by the traffic level analyser.

Based on our previous works related with vision-based vehicle detection in the context of intelligent vehicles applications [7], [8], [9], we have proposed to use the information provided by the cameras to supply xFCD systems with a more representative measurements of the traffic conditions [10], [11]. A set of monocular vision-based modules (forward-rear and side-looking) were used to have a nearly 360 degree field of view. Instead of excluding lateral lanes [6] the relative distance and speeds of the vehicles detected in adjacent lanes were incorporated to the xFCD structure providing variables such as the local traffic load (number of vehicles, road capacity, etc.) and the average road speed (not only the floating vehicle speed) of the specific local area. Thus, a more detailed description (less discrete) of the traffic state can be obtained (for example, in cases where the floating vehicle is stopped in a congested lane but adjacent lanes have different levels of congestion).

In this paper we describe an improved approach of the vision-based vehicle detection systems described in [10] and [11], by means of the use of stereo vision. Stereo information improves both the detection performance and the accuracy of the measurements (host-to-vehicle -H2V- relative distance and speed). Previous results [11] were obtained in sequences of a few seconds (800 frames). Here we provide extensive results in sequences of a total duration of 198 minutes (428400 frames, 36fps) recorded in real traffic scenarios with different weather (rainy/cloudy) and lighting conditions (nighttime/daytime), and different levels of congestion.

The remainder of this paper is organised as follows: Section II provides a global overview of the system. The description of the stereo vision-based vehicle detection system is presented in Section III. The experimental results are listed in Section IV, including sensor accuracy analysis. Finally, conclusions and future works are discussed in Section V.

II. SYSTEM OVERVIEW

The experimental vehicle used in this work is a car-like robot (a modified Citroën C4) which can be seen in Figure 1. It has an onboard computer housing the image processing system, a RTK-DGPS which is connected via RS232 serial port, a pair of synchronized low cost digital cameras connected via FireWire port, a specific electronic device for CAN Bus interfacing and a cellular communication system. The stereo vision sensor uses 320 × 240 pixel greyscale images with a baseline of approximately 300mm and a focal length of 8.5mm.

The global architecture of the system can be seen in Figure 2. The results obtained by the stereo vision module are combined with the GPS measurements and the data provided...
by the CAN bus in order to have globally referenced traffic state information. Among the variables that are directly (or indirectly, after some processing) available via CAN Bus we remark: speed, acceleration, number of stop&go (based on the speed), outside temperature and humidity, windshield wiper state, different lights state (fog, emergency, high beams, indicators, etc.), number of gear changes, number of lane changes (based on the indicator lights), etc.

III. VISION-BASED VEHICLE DETECTION

The global scheme of the proposed vision-based vehicle detection system is depicted in Figure 3. The first step consists of reducing the searching space in the image plane by detecting the road lane markings. The detected lanes are used as the guidelines that drive the vehicle searching process. 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 [11]. Then several region of interests (ROIs) generation modules are triggered in parallel.

Monocular ROIs are selected by combining white top-hat and canny features with different types of symmetries (grey level, vertical edges and horizontal edges symmetries) as described in [11]. Stereo processing results in a dense disparity map which allows the 3D position and the relative speed of the vehicles ahead to be estimated accurately. The camera pitch angle is estimated dynamically by means of the so-called virtual disparity map from which regions corresponding to the ground-plane can easily be removed

Nighttime ROIs selection mechanism is firstly based on adaptive thresholding. Candidates are then obtained by searching pair of clusters (which are usually rear lights) with similar vertical location, using again the flat-world assumption, perspective constraints and prior knowledge of target objects. Although this is a simple method, it is very effective in practice. It provides a low number of false positives since the adaptive thresholding mostly highlights the rear lights of the vehicles (see Figure 5).

The three ROIs selection modules provide different candidates, some of them overlapped as they can refer to the
same vehicle. All these candidates are classified by means of
linear SVM classifiers [13] in combination with histograms
of oriented gradients (HOG) features [14]. Two specialized
classifiers have been trained depending on the lighting condi-
tions. Candidates selected by monocular and stereo modules
are classified using a daytime SVM classifier. On the other
hand, a nighttime SVM classifier is used for nighttime ROIs
(see Figure 3). Figure 6 shows some of the positive and
negative training samples. The number of samples used
for developing the training test differs depending on the
classifier. Daytime classifier was trained with 19031 nega-
tive and 9248 positive samples. Nighttime classifier was obtained
with 2486 negative and 1847 positive samples. The number
of samples used in nighttime scenarios is lower because
the intra-class variability is much lower than in daytime
conditions.

![Fig. 6. Upper row: positive samples. Lower row: negative samples. Left: daytime samples. Right: nighttime samples.](image)

The three detection modules may generate candidates
related with the same vehicle. Accordingly a kind of non-
maximum suppression technique is used to group multiple
overlapped candidates, trying to generate only one can-
didate per vehicle. In practice, we have observed that good
candidates (candidates that are well fitted to the actual
vehicle contour) usually provide SVM results farther to the
hyperplane that separates the vehicle and non-vehicle classes,
i.e., their classification result has higher confidence. Thus,
we use the distance to the hyperplane as the main variable
to select the best candidate of a set of overlapped candidates
(overlapping has to be larger than 70% of the area of the
smaller candidate).

Finally, the vehicles detected in this single-frame way
are passed to the multi-frame validation and tracking module
(see Figure 3). A predefined number (empirically set to 3) of
consecutive identifications of an object classified as a vehicle
triggers the data association and tracking stages. The data
association problem is addressed by using feature matching
techniques. Harris features are detected and matched between
two consecutive frames, as in [11]. Tracking is implemented
using a Kalman filter with a constant velocity model [11].

IV. EXPERIMENTAL RESULTS

A. System validation

In a first experiment, we evaluated the different parameters
provided by the vision sensor, i.e., H2V relative distance
and speed, together with their corresponding errors. A set
of sequences was recorded in which a driver was requested
to perform an overtaking manoeuvre in a two-lane road at
approximately 25km/h. A vehicle was parked in the right-
hand (non-overtaking) lane, so that the host vehicle overtakes
it at a relative speed of around 25km/h. Besides the stereo
vision sensor, two DGPSs are used, one placed at the lead
vehicle’s position and the other on board the host vehicle (see
Figure 7). The measurements supplied by these DGPSs (after
linear interpolation due to their low sampling frequency -
5Hz) are taken to be the ground truth.

![Fig. 7. Scheme of the manual driving experiment.](image)

Figure 8 plots the stereo and the DGPS distance estimates,
including the absolute error and the stereo depth estimation
uncertainties [15]. The DGPS H2V, which is taken to be the
ground truth, lies mostly within the limits of the stereo depth
estimate including their corresponding uncertainties. As can
be observed, the greater the depth the greater the absolute
error. Depth estimates may not be reliable at long distances
(absolute errors). However, the absolute error decreases as
the H2V distance decreases.

![Fig. 8. Stereo and DGPS H2V relative distance including absolute and depth estimation errors in the experiment in which the car is overtaken.](image)

Although the depth estimation errors are sufficiently small
for the precision required in the present application, those of
the discrete estimate of the relative speed from the Kalman
state variable are not. As demonstrated in [16] discrete difference
between two noisy depth estimation values introduces a
considerable error in the relative speed computation for non-
infinitiesimal Δt, and this clearly limits the vision system’s
precision. Those authors propose an optimal value for Δt
which is directly proportional to depth. In the present case,
we define a practical value of Δt = 1seg and compute the
average speed of the last 25 frames. This approach is
very effective in practice. Figure 9 depicts the relative speed
obtained from the CAN bus (recall that the leading vehicle is
parked in this experiment), that computed by means of DGPS
H2V distance values, the discrete relative speed provided by
the Kalman filter, and the relative speed computed at each
second. One observes that the discrete values of the stereo
relative speed are not at all useful. However, the proposed
approach described above provides relative speed estimates
that are accurate enough for the application’s requirements -
the root-mean-squared-error (RMSE) is around 3km/h.
B. SVM Classifier

As described in Section III, the candidates are classified by means of two linear SVM classifiers (daytime and nighttime). In order to define the SVM decision thresholds we use the ROC curves defining the work points in terms of the relationship between the Detection Rate (DR) and the False Positive Rate (FPR). In these experiments 2/3 of the samples were used for training and 1/3 for test. The ROC curves are showed in Figure 10. On the one hand, the decision threshold for the daytime classifier is fixed to 0.03 with a DR of 96.1% and a FPR of 6.4%. On the other hand the decision threshold for the nighttime classifier is fixed to 0.07 with a DR of 91.4% and FPR of 7.8%. We have to consider that these results are single-frame results, so they will be improved in multi-frame validation and tracking stages. The final classifier is trained with all the available samples.

C. xFCD results

The proposed system was tested on data from real traffic conditions. A set of video sequences were recorded on the same route through the Madrid (Spain) M-30 highway (see Figure 11). The route distance is approximately 15 kilometres and it was driven up to 4 times, going and returning from the A and B points highlighted in Figure 11. Accordingly, a total of 150 kilometres were driven approximately. The video sequences were recorded during the same day, from 7:20 AM to 10:38 AM, including different lighting (nighttime/daytime) and weather (rainy/cloudy) conditions, as well as different levels of congestion.

Global results are presented in Table I. We have labelled the total amount of vehicles in range in all the sequences (ground truth). Note that we distinguish between true positives and vehicles that are detected from a previously detected vehicle (identifier exchanged). In both cases we consider the vehicle as detected, assuming that some errors will appear in the relative distance computation when a new vehicle is associated with a previously tracked one. Accordingly, we obtain a detection rate of 88.44% and 90.99% in nighttime and daytime conditions respectively. 0.84 false positives per minute are obtained in nighttime conditions whereas 0.69 false positives per minute are detected in daytime conditions. In addition, the 6.67% and the 5.82% of the vehicles are counted more than one time in nighttime and daytime conditions respectively (i.e. a new identifier is generated).

Figures 12 and 13 depict some of the variables provided by the system in two sequences (daytime and nighttime respectively) with different levels of traffic congestion. The following variables are drawn: host speed (via CAN bus interface), the average road speed, the distance between the host vehicle and the vehicle ahead, and the number of vehicles detected per each frame. As can be observed, there is a strong relationship between the road speed and the H2V distance. In fact we consider this variable as one of the most relevant in terms of traffic congestion. The maximum number of vehicles detected per frame is 6 (note that the minimum number is 1 since we always count the host vehicle). The number of vehicles detected does not seem to affect to the average road speed computation which implies that the relative distances of the vehicles detected in adjacent lanes and the proposed method for computing the relative speed are coherent up to a point (extensive ground truth data will be needed to assure the actual goodness of the relative distance and speed measures).

The average difference between the host speed and the average road speed is around 0.4km/h. This can be explained by the fact that the host vehicle is driving in the middle
### TABLE I

**GLOBAL XFCD RESULTS**

<table>
<thead>
<tr>
<th># frames</th>
<th># vehicles in range (ground truth)</th>
<th>True Positives</th>
<th>Identifier exchanged</th>
<th>Missed vehicles</th>
<th>False Positives</th>
<th>Multiple counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nighttime (from 7:20am to 8:35am)</td>
<td>161907</td>
<td>727</td>
<td>625</td>
<td>18</td>
<td>84</td>
<td>63</td>
</tr>
<tr>
<td>Daytime (from 8:35am to 10:38am)</td>
<td>266493</td>
<td>1133</td>
<td>1007</td>
<td>24</td>
<td>102</td>
<td>86</td>
</tr>
<tr>
<td>Total (198 minutes)</td>
<td>428400</td>
<td>1860</td>
<td>1632</td>
<td>42</td>
<td>186</td>
<td>149</td>
</tr>
</tbody>
</table>

---

**Fig. 12.** Results in daytime, cloudy and rainy conditions with different levels of traffic congestion.

**Fig. 13.** Results in nighttime, cloudy and rainy conditions with different levels of traffic congestion.
lane of a three lane highway most of the time, so the higher speed of the vehicles located on the left lane is compensated somehow with the lower speed of vehicles located on the right lane. However we have detected some cases where the host vehicle is stopped (host speed equal to 0 km/h) whereas the vehicles located on the left lane are moving at speeds up to 10km/h. This effect can be observed in Figure 14. In some cases the difference between the host speed and the average road speed achieves more than 3km/h.

Although the results are promising, new advances have to be developed to improve the detection rate and reduce the number of false positives. A more sophisticated ensemble classifier will be introduced using different classifiers depending on the lane, and including a new class for trucks since most of the missed vehicles are trucks.

Finally, a set of experiments have to be addressed to clearly define the limits of the proposed system. We will compare the xFCD results with data provided by fixed loops that are installed throughout the same highway used in the experiments. New vehicles (fleet) will be used, and xFCD will be combined with data from fixed loops to assess applications in the context of traffic management and control.

VI. CONCLUSIONS AND FUTURE WORKS

This paper presented the results of an xFCD system equipped with a stereo vision-based local traffic detector in a set of extensive experiments (198 minutes) carried out in real traffic conditions with different lighting and weather conditions as well as different levels of traffic congestion. The vision-based vehicle detection system combines different approaches (monocular, stereo and nighttime) to supply generic obstacles that are classified by means of two linear SVM classifiers (daytime/nighttime) and HOG features. The vehicles are then validated in a multi-frame fashion and tracked using a Kalman filter. The proposed approach provides data, not only from the host vehicle, but from the vehicles located in the field of view of the host, including the vehicle ahead and the vehicles located in adjacent lanes (when available). Thus, standard variables such as the host vehicle speed (via CAN bus) can be enriched and supported with new variables such as the distance to the vehicle ahead, the average road speed, the number of vehicles in range, etc. Although the use of stereo vision implies to manage some errors when estimating the relative distance (errors that are squared proportional to the depth), it allows us to have an accurate estimation of the relative speed by integrating relative distance values during 1 second. The global results show that the proposed approach produces good results in different traffic conditions, succeeding when detecting the traffic congestion.

Fig. 14. Specific situation: host vehicle is stopped whereas the vehicles located on the left lane are moving.

V. CONCLUSIONS AND FUTURE WORKS

This paper presented the results of an xFCD system equipped with a stereo vision-based local traffic detector in a set of extensive experiments (198 minutes) carried out in real traffic conditions with different lighting and weather conditions as well as different levels of traffic congestion. The vision-based vehicle detection system combines different approaches (monocular, stereo and nighttime) to supply generic obstacles that are classified by means of two linear SVM classifiers (daytime/nighttime) and HOG features. The vehicles are then validated in a multi-frame fashion and tracked using a Kalman filter. The proposed approach provides data, not only from the host vehicle, but from the vehicles located in the field of view of the host, including the vehicle ahead and the vehicles located in adjacent lanes (when available). Thus, standard variables such as the host vehicle speed (via CAN bus) can be enriched and supported with new variables such as the distance to the vehicle ahead, the average road speed, the number of vehicles in range, etc. Although the use of stereo vision implies to manage some errors when estimating the relative distance (errors that are squared proportional to the depth), it allows us to have an accurate estimation of the relative speed by integrating relative distance values during 1 second. The global results show that the proposed approach produces good results in different traffic conditions, succeeding when detecting the traffic congestion.