Vehicle Identification Using Near Infrared Vision and Applications to Cooperative Perception

Axel von Arnim, Mathias Perrollaz, Arnaud Bertrand, Jacques Ehrlich

Abstract— **Vehicles will be in the next future equipped with V2V telecommunication means to exchange data, such as the presence of an obstacle on the road, or an emergency braking notification. Vehicles are also more and more equipped with perception systems (cameras, laser scanners, radars) that enable them to explore the immediate environment, including other vehicles. We propose in this paper an on-board optical vehicle identification system to enable telecom and perception systems to cooperate. The optical identification identifies which vehicle, in the scene captured by the perception system, is sending information via telecom.**

I. INTRODUCTION

Using vehicle to vehicle telecoms to extend the perception of local on-board systems like vision [5], laser scanners [4] or radars [8] is an attractive idea. For example (see Fig. 1), a collision mitigation application using stereovision techniques can take advantage of information about the braking power of the followed vehicle or an emergency brake event notification via telecoms [6]. But to be relevant in a complex road scene, where several vehicles are involved, the information has to be localized, otherwise we have the knowledge of a vehicle's behaviour (speed, braking action) but are unable to determine which vehicle of the scene is doing so. If the followed and the following vehicles are accurately and reliably localized, it is possible for the following vehicle to match, in the scene captured by perception sensors, which vehicle is sending relevant data. Unfortunately, GPS based localization is not accurate enough to discriminate between vehicles on different driving lanes, and if it were (differential GPS), it would not be reliable since the GPS signal is masked by tunnels, dense trees or urban canyons. So the need for an optical identification system, that delivers the localized identification of vehicles directly in the perception sensors captured scene is strong.

For now, the main systems developed use automatic license plate recognition [1], or spatial pattern based techniques [3]. These systems are limited by the accuracy of vision sensors (which induces a range limitation). Other systems use infrared light to extract the ID information from the view. But they use simple, often binary ID symbols to classify types of vehicles, rather than to individually identify them [10]. A more suitable technology has been tested, that combines two IR beacons with a common signaling protocol by means of radio-communication to identify vehicles [9]. But this system is impeded by both the facts that it is

Axel von Arnim (vonarnim@lcpc.fr), Mathias Perrollaz (mathias.perrollaz@inrets.fr) and Jacques Ehrlich are with LIVIC (INRETS / LCPC), Bât. 824, 14, route de la Minière 78000 Versailles - France.

vulnerable to partial occultation, and that it needs a radio support signal to work. A system using two IR beacons at the rear of vehicles to measure inter-vehicle distances uses an optical identifier protocol (spatial and frequential), but the purpose is to match the two corresponding beacons on each vehicle [2]. So, a robust, vision-only, and long-range system does not, to our knowledge, exist.
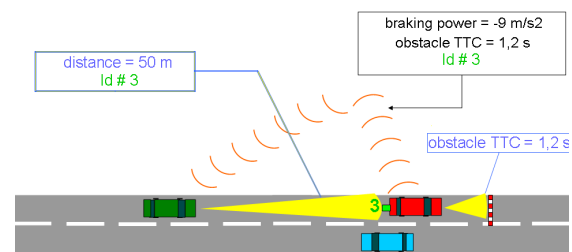

Fig. 1.   Combining perception and telecoms.

In this paper we present an optical identification sensor that enables middle to long-range identification, does not rely on high resolution cameras and is not vulnerable to partial occultation. In part II we give a description of our sensor, from hardware to detailed software. Part III gives some experimental material. Finally, in part IV we present applications of the sensor to three cooperative perception systems.

II. OPTICAL VEHICLE IDENTIFICATION

A. General Overview of the Sensor

1) Working Principles: The goal is to identify vehicles, or more generally objects, in the road scene and localize them in the reference system of the receiver-equipped vehicle. The need to identify vehicles in a camera captured scene, without having to detect spatial patterns, leads to a near IR solution. The proposed system is a cooperative sensor, composed of two parts: the emitter and the receiver.

Let us consider a car following another car and wanting to identify it. Both cars have telecom means.

On the leading car, we attach a led-based infrared emitter, which codes, using time-frequency digital coding (*i.e.* blinking) the unique identification number of the car, for example, its IP address. Time-frequency signals are more robustly and farther perceptible by a vision sensor than spatially coding signals, who are vulnerable to both partial occultation and vision sensors resolution.
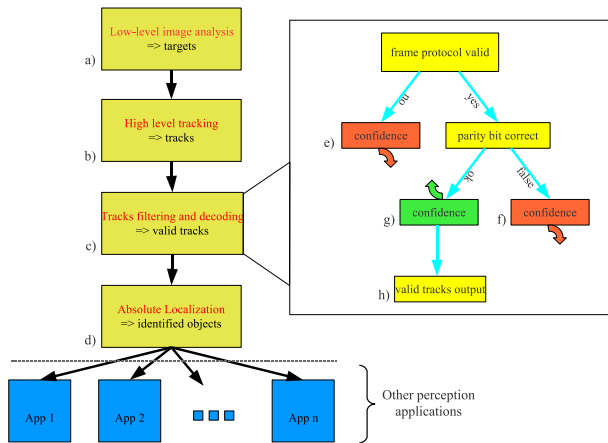
Fig. 2.   Overview of the identification algorithm.

On the following vehicle, we use a low-resolution, high-frequency, infrared band pass filtered digital camera, which detects and decodes the signal, and localizes it in the same reference system as the perception systems. Thus, each detected vehicle can be identified and the information from radio-communication and optical sources can be merged. An overview of the complete process, from image acquisition to delivery of the identified objects to other perception applications is presented on Fig. 2.

2) The emitter: The emitter (Fig. 3) is composed of a panel (named the lamp) of infrared emitting leds (SFH 484-2 emitting at peak $880nm$ in our prototype), an electronic command controller that generates the coding signal, and a power controller delivering the necessary electrical power. The leds could easily be integrated in the rear lights.

Fig. 3.   Prototype of the emitter lamp.

The digital coding signal conforms to a protocol (Fig. 4). A cycling frame is emitted permanently, containing a start sequence (4 bits), the usable code (identifier, as many bits as necessary), a 0 and a parity bit.

...	Start Code	Id Code	'0'	parity	...

Fig. 4.   Emitter signal protocol.

3) The receiver: The receiver is a high-frequency digital camera which captures the image of the road scene, and a processing algorithm which decodes and localizes the signal from the emitter. The camera is equipped with a bandpass near infrared filter.

B. Low-level Detection Algorithm

A low-level algorithm (Fig. 2-a) processes the input image to extract the spots corresponding to emitters. Most false detections (other infrared sources like sky, vegetation) are eliminated by shape analysis. The order of operations is:

1) IR-filtered image is acquired. Emitters appear like bright round spots (Fig. 5-b).
2) Image is binarized using a threshold, so that only the brightest pixels are kept (the most IR emitting sources) (Fig. 5-c).
3) Regions are made more consistant, less fragmented, using morphological dilatation.
4) White pixels are clustered: each connected region is a labelled cluster (Fig. 5-d).
5) Regions are accepted or rejected depending on their size and roundness. Large regions, corresponding to the sky, and regions which don't look like a spot, for example pieces of vegetation, are rejected. The algorithm used to determine the roundness of a region computes the rate of pixels of the region inside its circumcircle.

Accepted spots are named targets. Some objects, like solar reflections or vehicle or traffic lights appear like spots and are then accepted. We will see in the next paragraph that they are eliminated by the high-level tracks filtering algorithm.
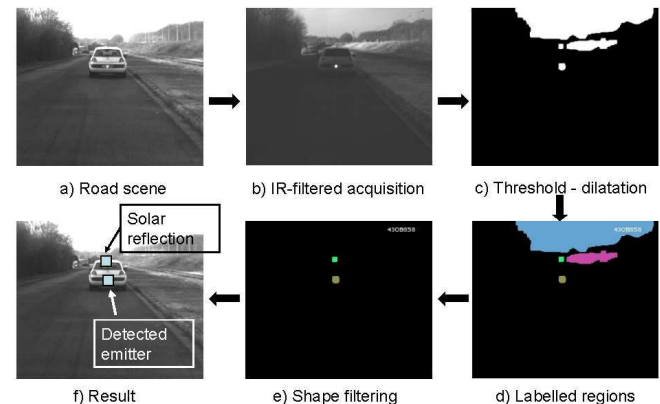

Fig. 5.   Low-level Image Analysis.

C. High-Level Filtering, Tracking and Decoding

A high-level algorithm (Fig. 2-b) tracks detected targets over time (in successive images) to follow their movement and analyze their logical state (0=extinct, 1=lit). In the following, a tracked target is named a track.

The tracking is based on a very simple neighborhood prediction. The neighborhood in which a track is expected in the next image is a region around the last position where it was seen. As long as the track is extinct, the neighborhood is growing, faster horizontally than vertically, because the horizontal movement of the followed vehicle in the image is obviously greater than its vertical movement. When the track appears again, the neighborhood is reduced to the center of

this new observation. A Kalman approach should improve the results and will be explored in the future.

The observed states of each track are translated into bits, which are recorded into a FIFO. This resulting bitstream is continuously decoded as a message. Besides, a confidence is attached to each track. A new track starts with maximum confidence, and during the tracking, this confidence is managed as follows (Fig. 2-c):

- If the message does not conform to the communication protocol (Fig. 2-e), or if the parity bit of the message is incorrect (Fig. 2-f), the track either is a false detection or has a temporary communication error and is marked as incorrect. It will continue to be tracked. Its confidence keeps decreasing as long as the problem remains.
- If the track has been too long incorrect, *i.e.* the confidence has reached zero, it is deleted and ignored in the next images.
- If the message conforms to the communication protocol and the parity bit is ok (Fig. 2-g), its code (identifier) and localization in the image are attached to it. If the track was incorrect before, its confidence increases, but it remains incorrect as long as the confidence has not reached its maximum value. The confidence increase is quicker than the decrease, to avoid accumulating too much delay when a track is temporarily incorrect due to short masking for example. Of course, The track will continue to be tracked in successive images.

False tracks, created by solar reflections or vehicle and traffic lights that passed the low-level image analysis, are here eliminated because they don't blink, or they blink without conforming to the communication protocol (random blinking).

Finally, valid tracks are output (Fig. 2-h) to the absolute localization algorithm (Fig. 2-d) with their position in the image coordinate system and their identification number.

Figure 6 shows three tracks. One, in thick white, is a valid track and two, in thin grey, are invalid tracks. They correspond to the front lights of a vehicle coming on the opposite lane and are rightly rejected.

Fig. 6.   Two bad tracks (thin) and one valid (thick).

D. Absolute Localization

Once the emitter-equipped objects have been detected, identified and localized in the identification image, they have to be localized in the coordinate systems of the other sensors of the vehicle, in order to be shared (Fig. 2-d). For that

purpose, an absolute coordinate system R_a is defined. As all sensors are rigidly linked to the car frame, they can all refer to R_a. Therefore, whichever sensor is used, the absolute localization problem finally consists in an inverse mapping of the identified tracks into R_a. The position of the tracks can then be transferred into any sensor's coordinate system.

The only 2 camera coordinates (U_{id}, V_{id}) are not sufficient to solve this 3 variable problem. This issue adds a new constraint on the emitter: its height Y_L, relative to the road surface, is supposed to be known. That constraint is not so strong, since the height of the emitter can be coded into the identifying code itself.

1) Geometrical Description: The geometry of our receiver is presented on figure 7. The sensor is a video camera represented by a *pin-hole* model whose behaviors are: α_{id}^u and α_{id}^v, respectively the focal length measured in pixels horizontally and vertically (pixels are rectangular), and (U_{id}^0, V_{id}^0) the projection of the optical center among the image plane. The camera is placed in R_a on the $(X_{id}^0, h_{id}, Z_{id}^0)$ position. Yaw and roll angles are considered null and constant. The pitch angle is θ_{id}.
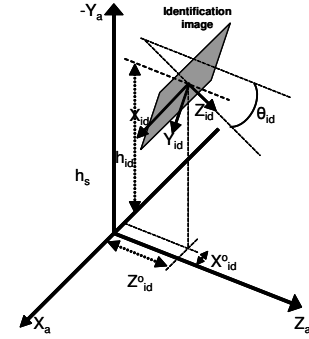

Fig. 7.   Geometry of the identification sensor in the absolute coordinate system.

2) Coordinate Transform: The easiest approach to localization relies on a plane world assumption. Indeed, considering that the road surface is the $(X_a, 0, Z_a)$ plane, and according to the projection equations of the *pin-hole* model, the coordinates $(X_a, Y_a = Y_l, Z_a)$ of the emitter in R_a can be easily computed using equation system 1.

$$\begin{cases} Z_a = Z_{id}^0 + \dfrac{(Y_L + h_{id})(\alpha_{id}^v \cos\theta_{id} - (V_{id} - V_{id}^0)\sin\theta_{id})}{(V_{id} - V_{id}^0)\cos\theta_{id} + \alpha_{id}^v \sin\theta_{id}} \\[3mm] X_a = X_{id}^0 + \dfrac{(U_{id} - U_{id}^0)((Y_L + h_{id})\sin\theta_{id} + (Z_a - Z_{id}^0)\cos\theta_{id})}{\alpha_{id}^u} \end{cases}$$

(1)

In operation, h_{id} and θ_{id} parameters can change significantly, due to vehicle pitch. Thus, a pitch sensor is required for a correct coordinate transform. For this purpose, any inertial sensor could be efficient. Stereovision is also well designed for that application. Indeed, as described in [5], the height and pitch of the stereoscopic sensor can be retrieved using the *v-disparity* approach, by analyzing the near road

surface. h_{id} and θ_{id} are then immediately deduced, because the stereoscopic sensor is rigidly linked to the identification sensor.

So, even if it is not mandatory, the use of a separate stereoscopic sensor for the estimation of these parameters is an interesting solution. Moreover, it may be used in parallel for obstacle detection.

III. EXPERIMENTAL RESULTS

A. Scenarii

Here are the parameters which have an effect on the experiments:

- speed of receiver-equipped vehicle
- camera iris aperture of the receiver. In all the following test scenarii, we use an auto-iris, to have optimal iris aperture for our application
- meteorological (light) conditions. A cloudy day has best light conditions, and a sunny day, the worst ones
- camera framerate of the receiver (related to exposure time), expressed in fps (frames per second)
- camera resolution of the receiver
- emitter frequency, expressed in Hz, equivalent to $bits/s$. The ratio between the camera and the emitter frequencies gives the number of acquired images per bit emitted. This ratio, because of Shannon's theorem, must be greater than 2

Here are the features of the system that we wanted to evaluate. They are:

- static identification range: maximum range where a correct identification of the object is possible. Measured on static emitter-equipped object, with static receiver. We placed an emitter-equipped vehicle at the end of a long straight portion of road, and we placed a receiver at different longitudinal positions on the road until we found the limit detection range.
- identification time: time between first detection of the emitter and correct identification. Measured on static emitter-equipped object, with static receiver. We chose an optimal position for the receiver.
- false detection rate: number of false detections per minute or km. Moving receiver (vehicle). We drove a receiver-equipped vehicle on public roads with normal to heavy traffic, during a reasonable time.
- non detection rate: number of emitter-equipped objects not detected. Moving receiver. Measured on the here-after explained dynamic scenario.
- dynamic identification range: same feature as the static one, but measured on the dynamic scenario hereafter explained. Static emitter (road sign), moving receiver.
- localization accuracy: measured by placing an emitter on a vehicle moving away on a straight road. The results of the absolute localization (among Z_a axis) are compared to a reference, measured using a lidar, with centimetric precision.

Dynamic scenario on test track: we placed emitter-equipped road signs on the roadside, driving a receiver-equipped vehicle at various known speeds. Despite the system is mainly a vehicle identification sensor, we designed the scenario with static emitters, in order to control the previously listed parameters. And how could static emitters be better employed than being attached to road signs? But we also tested the system while following a car.
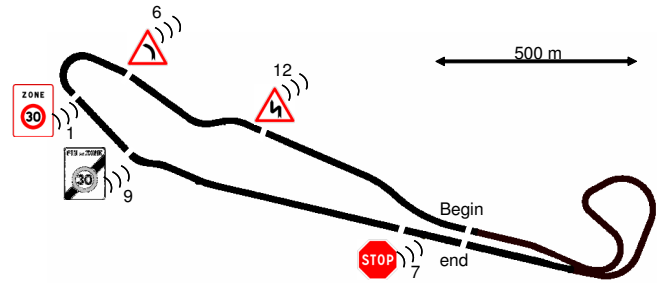

Fig. 8.   Experimental Scenario.

Figure 8 shows the test track, and the positions of the emitter-equipped road signs. Each road sign was emitting an identifying code (number) which is shown on the figure. The various positions of the emitters ensured that the receiver detected them with various light expositions.

B. Results

In all experiments, we used a Pulnix greyscale high speed CCD camera, in binning mode (320x120) to increase the acquisition framerate and lower the CPU consumption. This low resolution is sufficient because we do not detect spatial patterns, but bright spots. The focal length was 12mm. The computer ran a Pentium D 940 with 1Gb RAM, on Windows XP. The emitters used 16 bit frames, including start code and parity bit.

1) Static experiments: The main parameters here are a null speed of the receiver and the emitter (static), a variable camera framerate, thus a variable emitter frequency (according to Shannon's theorem). We tested the features with two values of the meteorological conditions parameters: cloudy and sunny day. The images per bit value is here a measured value (during a calibration procedure), and not the theoretical ratio between camera and emitter frequencies. The difference is due to discretization artefacts and threshold effects.

TABLE I

STATIC EVALUATION OF THE PERFORMANCES.

Camera Frame rate	410	514	595	650
Camera exposure time (μs)	1000	500	250	100
Emitter frequency (Hz)	175	210	250	275
Images per bit	2.29	2.36	2.32	2.17
Id. range - Best case (m)	**360**	**400**	**378**	**83**
Id. range - Worst case (m)	**65**	**110**	**107**	**62**
Theoretical min Id. time (ms)	90	73	62	53
Id. time (ms)	**105**	**100**	**76**	**65**

We can see on table I that the receiver has an optimal identification range with an exposure time of $500\mu s$, which gives the best contrast between emitters and environmental IR sources.

The theoretical identification time is given by the formula: $t_{id}^{th} = n * ib/f_c$, where n is the total number of bits of a message, including parity bit and start code sequence, ib is the number of images per bit acquired, and f_c is the camera framerate. The experimental identification time is always slightly greater than the theoretical one because the emitter and the receiver are not synchronized.

The measured identification time for our optimal configuration is $100ms$, which is a value that enables to identify objects moving very fast, or to identify static objects while driving very fast by.

So for the next presented experiments, we used following combination of parameters: $camera\, framerate = 514fps$, $emitter\, frequency = 210Hz$. The receiver algorithm always runs in real-time, consuming an average 20% CPU.

2) Dynamic experiments: Table II shows the false detection rate for a scenario on public road without emitter equipped cars, by sunny weather (bad condition). We had a few false detections due to solar reflections, with very little duration.

TABLE II

FALSE DETECTIONS ON PUBLIC ROADS (WORST CASE).

Total scenario duration	16 min
Total scenario length	21 km
False detection occurrences	5
Total false detections duration	138 ms
Average false detection duration	28 ms
False detection rate (nb/minute)	0.31
False detection rate (nb/km)	0.24

For other dynamic features, we used the formerly presented scenario with emitter-equipped road signs on our test track. The main parameters were: variable speed of receiver (from 50 to $110km/h$), cloudy day (good condition), and emitters' positions as shown on figure 8.

On table III, you can see the list of identified road signs: #12, 6, 1, 9, 7 and the average identification range when we drove along at different speeds. The standard deviation from the average value is very low, so we only report here the average values. This means that the tested features have similar values at all different speeds.

TABLE III

RESULTS OF THE DYNAMIC EVALUATION.

road sign ID	12	6	1	9	7
avg. id. range (m)	282	109	42	170	236
non detections	0				
false detections	0				

The ranges depend on the location of the road signs. Road signs #12 and 7 are seen from far away (Fig. 8) because they

are located on straight road portions. #1 appears on the image just after the tight turn, so the range is short.

Figure 9-top shows what we call a chronogram of the scenario and represents the identification code detected versus time. So you can see the history of the scenario: the time a road sign is detected, the duration of the correct detection (*i.e.* its identification range, knowing the speed of the vehicle). It shows how the id ranges and times are related to the positions of the road signs on the track.
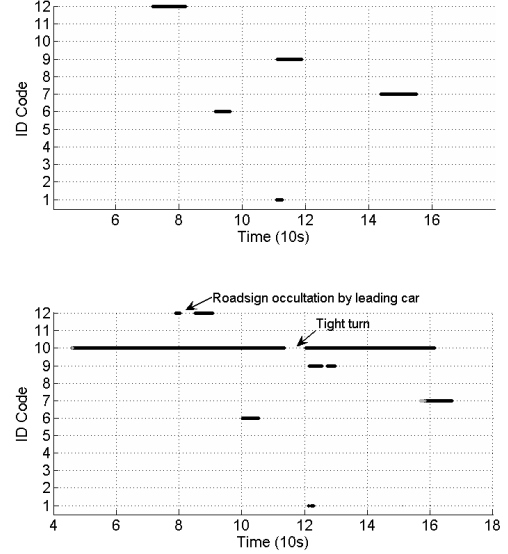

Fig. 9. Chronograms of the scenario at speed = 90 km/h (top), and following a vehicle (bottom).

Figure 9-bottom is another chronogram of the same scenario, but this time we followed an emitter-equipped vehicle. The interdistance varied between $5m$ and $100m$. You can see that the vehicle is continuously detected and identified, apart from a short period of time in the tight turn, near $Time = 12$ (due to camera field of view). A short break in the identification of road sign #12 appears after $Time = 8$, because the road sign is momentarily masked by the leading vehicle.

In a multi lane scenario, multiple leading cars were correctly identified and tracked.

Other tests with this scenario were performed with less optimal meteorological conditions (sunny weather). The id ranges for road signs located on straight lines were reduced, but remained compatible with the scenario and let us drive the receiver at the same speeds, up to $110km/h$.

Finally, we have similar results at different speeds of the receiver-equipped vehicle, and the ranges are relatively high, allowing interesting applications, as explained in part IV.

3) Localization Accuracy: The performances of the localization algorithm have been assessed using stereovision as a height/pitch sensor (two $1/4$ PAL video cameras, with $8.5mm$ lens, baseline $1m$). Figure 10 gives the results of the process among the Z_a axis.

The localization algorithm gives satisfying results on short ranges (until $25m$). For longer ranges, the accuracy decreases
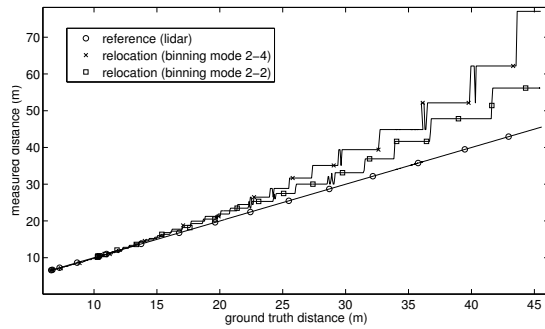
Fig. 10. Results of the absolute localization step in Z_a, compared to a reference measured with a lidar.

because of the low vertical accuracy of each pixel. As expected, the accuracy is better when the identification camera is configured in binning mode 2-2 (320*240 pixels) than in binning mode 2-4 (320*120 pixels), but with no spectacular improvement.

IV. APPLICATIONS TO COOPERATIVE PERCEPTION

A. Telecom-Enriched Cooperative Detection

Let us consider a vehicle equipped with an obstacle detection system that provides relative position and speed. V2V telecom can provide high level information such as event notifications, masked obstacle alerts or characteristics of vehicles. As presented in figure 11, thanks to our identification system, the localized identifiers can be associated with the corresponding detected objects. Then, the telecom messages can be matched with their detected senders.

Fig. 11. Telecom enriched detection: a) a car is detected thanks to the lidar and a message is received from vehicle 8. b) an object identified as 8 is perceived and localized in the scene. c) both detections are associated: the car in front is sending an emergency brake message.

Besides, irrelevant messages can also be discarded by filtering the identifiers outside the driving lane, using lane marking detection or any path prediction algorithm.

The enriched perception system is adapted to design a collision anticipation application. By extension, when applying this system to a pool of vehicles in a driving lane, it is possible to dramatically reduce the global time of reaction on a road event [6].

B. Cooperative Road Sign

If an emitter is placed on a road sign, you can detect and recognize it with the onboard receiver in vehicles. This is an alternative to other sign recognition systems, based on vision, GPS and maps or WiFi tags.

C. Long Range Cooperative Detection

As our optical identification sensor is able to detect equipped vehicles up to $300m$, it can be used for long range obstacles detection. A basic approach consists in just localizing the identified tracks in R_a. This kind of detection is not precise enough for driving assistance purposes, but provides seeds for more accurate vision systems. For example, when stereovision is also in use in the vehicle, the identified tracks, localized in R_a can be used to define a region of interest in the stereoscopic images. Then, the distance to the obstacle is more precisely measured by stereovision. Moreover, using a confirmation strategy [7] can lead to a very robust detection, extending or replacing lidar capabilities.

V. CONCLUSION

This paper has presented a vision-only identification system, which is long-range and computation cost effective. It has proved to be efficient in many road scenarii. It enables to combine V2V telecoms with shorter range perception systems to make possible applications like extended or enriched perception. Besides, other applications like smart road signs have been successfully tested. Future works will deal with the improvement of the localization algorithm in case of non plane roads. Afterwards a complete evaluation of the performances of the sensor will be driven.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] A. Albiol, J.M. Mossi, A. Albiol, and V. Naranjo. Automatic license plate reading using mathematical morphology. In *Proc. of VIIP Conf.*, Marbella, Spain, 2004.

[2] M.P. Blmal and W.H. Taichi for Rockwell International Corporation. Inter-vehicle distance measuring system, 1993. European Patent - EP 0 561 353 A1.

[3] J. Bruce and M. Veloso. Fast and accurate vision-based pattern detection and identification. In *Proc. of IEEE ICRA*, Taipei, Taiwan, 2003.

[4] N. Kaempchen, M. Buehler, and K. Dietmayer. Feature-level fusion for free-form object tracking using laserscanner and video. In *Proc. of IEEE IV Symp.*, Las Vegas, USA, 2005.

[5] R. Labayrade, D. Aubert, and J.P. Tarel. Real time obstacle detection on non flat road geometry through 'v-disparity' representation. In *Proc. of IEEE IV Symp.*, Versailles, France, 2002.

[6] B. Mourllion, D. Gruyer, and A. Lambert. A study on the safety-capacity tradeoff improvement by warning communications. In *Proc. of IEEE ITS Conf.*, Toronto, Canada, 2006.

[7] M. Perrollaz, R. Labayrade, C. Royère, N. Hautière, and D. Aubert. Long range obstacle detection using laser scanner and stereovision. In *Proc. of IEEE IV Symp.*, Tokyo, Japan, 2006.

[8] H. Takahashi S. Sugimoto, H. Tateda and M. Okutomi. Obstacle detection using millimeter-wave radar and its visualization on image sequence. In *Proc. of ICPR*, Cambridge, UK, 2004.

[9] K. Takada, H. Fujii, and O. Hayashi. Multiple vehicle identification in a longitudinal ranging system. In *Proc. of ITS World Congress*, Berlin, Germany, 1997.

[10] M. Usami, K. Aoki, and Y. Suzuki. Development of cooperative vehicle following system for ahs in mixed traffic. In *Proc. of ITS World Congress*, Berlin, Germany, 1997.