An in-vehicle vision system for dangerous situation detection

Chiara Corridori, Dimitri Giordani, Paolo Lombardi, Stefano Messelodi, Carla Maria Modena, Michele Zanin
{corridori, dimi, lombardi, messelod, modena, mizanin}@itc.it

Istituto Trentino di Cultura - Centro per la Ricerca Scientifica e Tecnologica (ITC-irst)
Via Sommarive 18, 38050 Povo, Trento, Italy

Abstract. This paper presents the on-board computer vision system under development at ITC-irst inside the automotive project DIPLODOC. The main task of the system is to detect hazard situations in order to alert the driver. The paper describes the status of the obstacle detection and the road recognition modules.

1 Introduction

DIPLODOC (DIstributed Processing of LOcal Data for On-line Car services) [1] is a three years project partially funded by the *Provincia Autonoma di Trento* that started on April, 2002. Three research partners are involved: ITC-irst, CRF (*Centro Ricerche FIAT*), and the University of Trento.

The goal of the project is to design and develop a system based on a distributed architecture where intelligent vehicles communicate with a remote traffic control center. Each vehicle integrates different technologies to provide more comfort and driver safety. Speech recognition and synthesis techniques are used to interact with the user. Computer vision and image understanding are applied to the extraction of traffic parameters and to accident avoidance by detection and recognition of obstacles on the road. Wireless telecommunication is used to send and receive traffic data and route planning information to/from the control center. The practical result of the project will be a simulation of the traffic control center and a demonstrative vehicle (Figure 1) equipped with vision, speech and telecommunication devices.

The DIPLODOC architecture has been designed in order to satisfy four services, defined as follows:

– xFCD (eXtended Floating Car Data): the vehicles act as mobile probes for the collection of remote data. Each vehicle transmits its position provided by the satellite positioning system and data coming from the engine control unit and from the on-board vision system. Moreover, still images are sent to a specialized vision module hosted in the center for a deeper analysis. The system at the traffic control center exploits and integrates the information coming from the vehicle fleet in order to deduce the local traffic conditions.

Fig. 1. DIPLODOC demonstrative vehicle

– xEC (eXtended Emergency Call): the user can activate manually or by voice an emergency call to the control center. The emergency call allows the operator in the traffic control center to be timely informed about the on-board situation. Extended information, like video sequences and audio recording, are sent in addition to the position data of standard emergency call.
– DRP (Dynamic Route Planning): the driver can request by voice a route plan for the travel from the traffic center: how to arrive at a desired destination starting from the current vehicle position, considering also information about traffic levels on the possible paths.
– FOR (Front Obstacle Recognition): this service aims at warning the driver when pedestrians, vehicles or obstacles are in close proximity to the driver's intended path, using information coming from the on-board vision sensors, the vehicle sensor data, and possibly from environmental conditions or driver activity to modulate the alarm level.

The Technologies of Vision team (TeV) at ITC-irst is mainly involved in the realization of the xFCD and FOR services. For this project the research focuses on designing innovative solutions, on comparing the performances of novel and known approaches, and on the development of intelligent strategies for information fusion. We also develop mathematical models for improved tracking of high-curvature roads for applications in mountain areas. The following sections discuss the FOR computer vision architecture, recent advances, and some open issues.

2 Front Obstacle Recognition architecture

The FOR system has to detect threat of collision combining information coming from the on-board vision system and the vehicle data, actually the speed. This system works on-board without exchanges with the control center and one of its main requirements is operating in hard real time.

The vehicle is endowed with two image acquisition devices: a color stereo camera head and a monocular high resolution camera, both mounted in the front of the vehicle and looking ahead. The obstacle detection system receives its input from the stereo device. For research, we currently employ a couple of IEEE 1394 digital cameras distributed by Videre Design together with the provided software for the camera calibration and the disparity map computation: Small Vision System (SVS) by SRI International [2, 3].

The system architecture is outlined in Figure 2. Input images are pre-processed with noise reduction, edge detection and disparity map computation. Such information is delivered to a set of expert modules. Individual modules are summarized in Table 1.

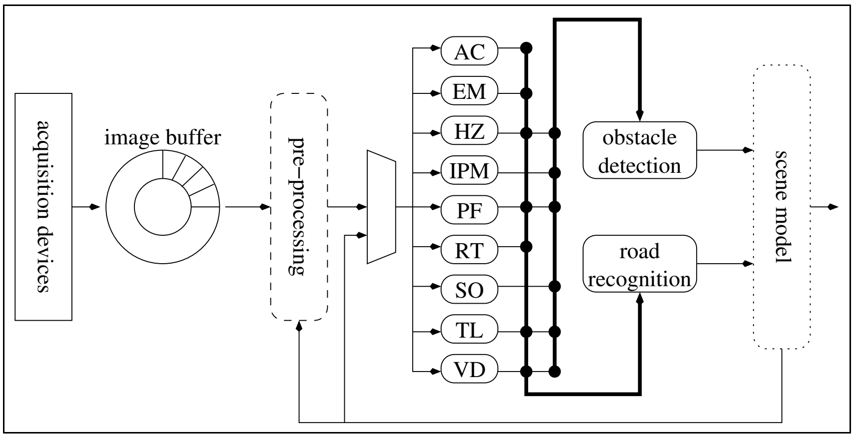


Fig. 2. General architecture of DIPLODOC on-board vision system. Active Contours (AC), Ego Motion (EM), HoriZon line estimation (HZ), Inverse Perspective Mapping (IPM), Plane Fitting (PF), Road Tracking (RT), Stereo obstacle Outliner (SO), TextureLess region detection (TL) and V-Disparity mapping (VD) are the expert modules.

In the following the Road Recognition and the Obstacle Detection modules are described. They gather the contributions from the experts, perform information fusion based on a priori knowledge of scene geometry and expected appearance, and finally update the scene description. We are considering classical fusion methods like voting, cumulative voting, and time-adaptive weighted average, both at pixel-wise and region-wise level.

Feedback is provided for the next frame after a time-updating step.

3 Road recognition

3.1 Investigated approaches

We can distinguish the approaches to road recognition from two different points of view: *static* vs. *dynamic* and *real world* vs. *image plane*. The former takes

Table 1. DIPLODOC Vision modules

name	*description*
AC	Active B-splines follow the roadway center by integrating edge information from both roadway sides [4, 5].
EM	It estimates the vehicle motion from optical flow and focus of expansion [6] by extracting and tracking salient features.
HZ	It filters temporally the position of the apparent horizon based on stereo analysis.
IPM	The anti-perspective transformations, *i.e.* the Inverse Perspective Map of both the left and right images are computed. By analyzing their difference [7, 8] this module estimates the free frontal space and the candidate obstacles position.
PF	It fits a plane on the disparity map by means of linear regression of disparity values. It works under the assumption of a locally flat road.
RT	It implements Kalman filtering with an innovative high-curvature road model specifically designed for high-curvature roads [9].
SO	It segments the regions exhibiting a constant disparity in the stereo disparity map. Thus, it generates candidates for obstacles.
TL	It extracts regions having certain radiometric characteristics in the color space and analyses their statistics signature. Textureless regions are typical of asphalt roads and large obstacles, like walls.
VD	It computes the v-disparity map, *i.e.* the histogram of disparity values for each image line. It allows the estimation of road geometry and obstacle distance [10, 11].

into account if images are analyzed one at a time (*static*) or using the temporal information, modeling the dynamic behavior of the system (*dynamic*). The latter indicates if the road model is described in real world coordinates or exclusively on the image plane.

Our Road Tracking module (RT) falls into the *dynamic - real world* approaches, while AC, PF, TL and VD modules fall into the *static - image plane* ones. The experts rely on different techniques, providing independent road shape descriptions. The outputs of these modules will be fused into a consistent representation by the Road Recognition higher-level module.

A preliminary version of the Road Tracking module is described in [9] and gives a 3D-world road description. It is related to these previous works: Dickmanns *et al.* [12], that introduces a clothoid[1] based road model; Khosla's [14] that presents a two clothoids road model. In [12], the road ahead of the vehicle is represented by a single clothoid segment whose parameters are estimated together with a physical vehicle model. This approach is designed for a limited horizontal curvature, and it cannot model the abrupt changes of some clothoid parameters when more than one clothoid is present in the camera's field of view.

[1] A clothoid is a spiral curve whose radius of curvature changes linearly with the distance along the curve. Italian roads are designed by civil engineers as sequences of clothoid segments [13].

In [14], the road in front of the vehicle is modeled as the conjunction of two different clothoids. The transition point between the two clothoids is assumed to be at a constant distance from the observer. Our RT module considers a road model that covers the possibility of having in the field of view either one or two clothoids, switching between these two situations dynamically. The RT module can compute clothoid parameters also on high curvature roads, that is necessary for road recognition in non trivial situations. Furthermore, our algorithm is able to compute the transition point coordinates at each instant. RT module needs an accurate estimation of vehicle motion (provided by the EM module) and the coordinates of some points on the roadway center (from AC, HZ, PF, TL and VD modules).

The Active Contours module approximates the roadway center and the roadway boundaries in the image plane with B-splines curves, implementing a variant of the algorithm proposed in [5]. The AC module exploits the information about the position of the apparent horizon, that is computed by the HZ module.

The PF, TL and VD modules output image regions that are classified as road surface. PF and VD are based on stereo information, while TL uses image raw data directly.

3.2 Results

Up to now, results from RT module have been studied mainly in a synthetic environment, under the assumption of ideal performances of the other involved modules and the hypothesis of locally flat roads. In these conditions the 3D road model and the relative position (offset and yaw with respect to the road) of the vehicle are extracted with good accuracy [9]. Figure 3 shows an example of high curvature road and Figure 4 the RT estimation error along the curve. Experiments on real images are undergoing.

Figure 5 presents an example of the road image regions detected by the PF, TL and VD modules. When considered separately, these modules do not guarantee a reliable road shape estimation. Their fusion with the other road recognition modules leads to a more robust and complete road description.

4 Obstacle detection

4.1 Investigated approaches

Driver assistance for collision avoidance is a key issue to help decreasing the frightening number of injures and deaths annually caused by accidents. The FOR system currently detects obstacles on the road by integrating the functionalities of the VD, PF, IPM, TL, SO and HZ experts.

Individual contributions of experts are as follows. Both VD and IPM independently provide estimations of obstacle positions. They both use stereo images but in different ways: VD scans the disparity map, whereas IPM projects one image onto the other. Thus their outputs are expected to be poorly correlated,

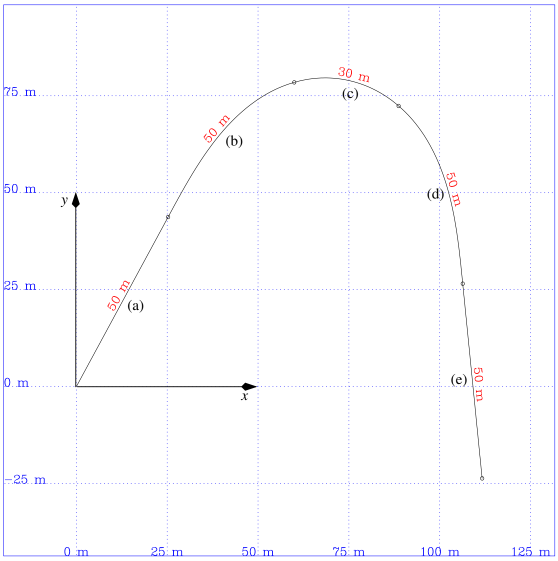


Fig. 3. An example of road curve. (a) and (e) are straight line segments, (c) is an arc of a circle with radius $R = 31.8$m, (b) and (d) are clothoid segments.

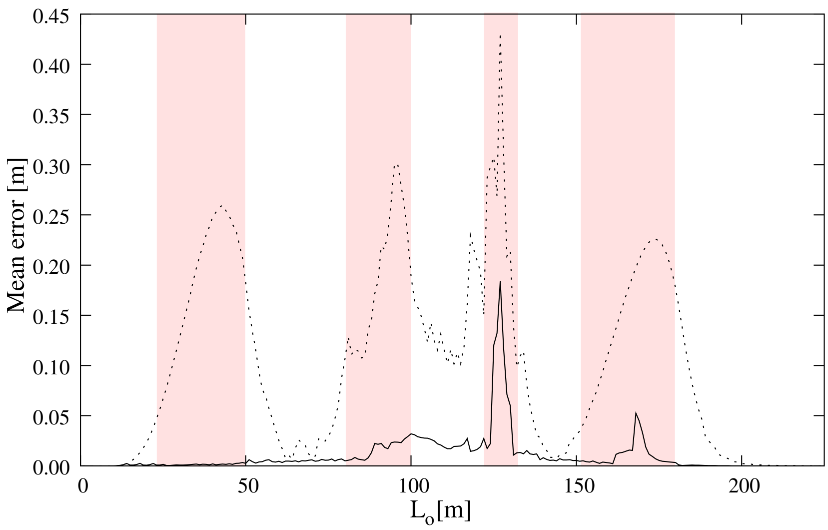
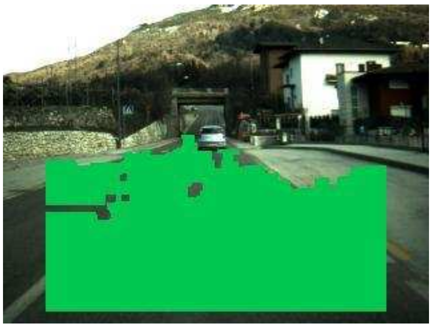
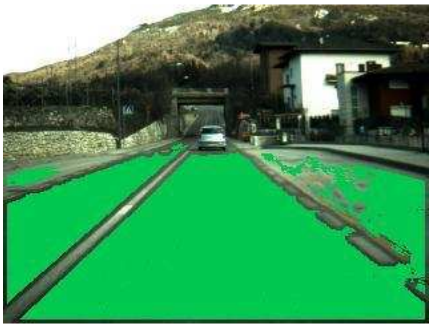


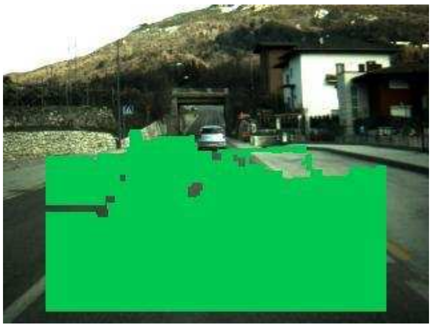
Fig. 4. Mean error on estimated point positions for an observer moving along the road of Figure 3. The dotted line shows the single clothoid model performances, while the solid one shows our hybrid road model performances.



(a) original image

(b) Plane Fitting





(c) TextureLess region detection

(d) V-Disparity mapping

Fig. 5. Road recognition results from the involved modules.

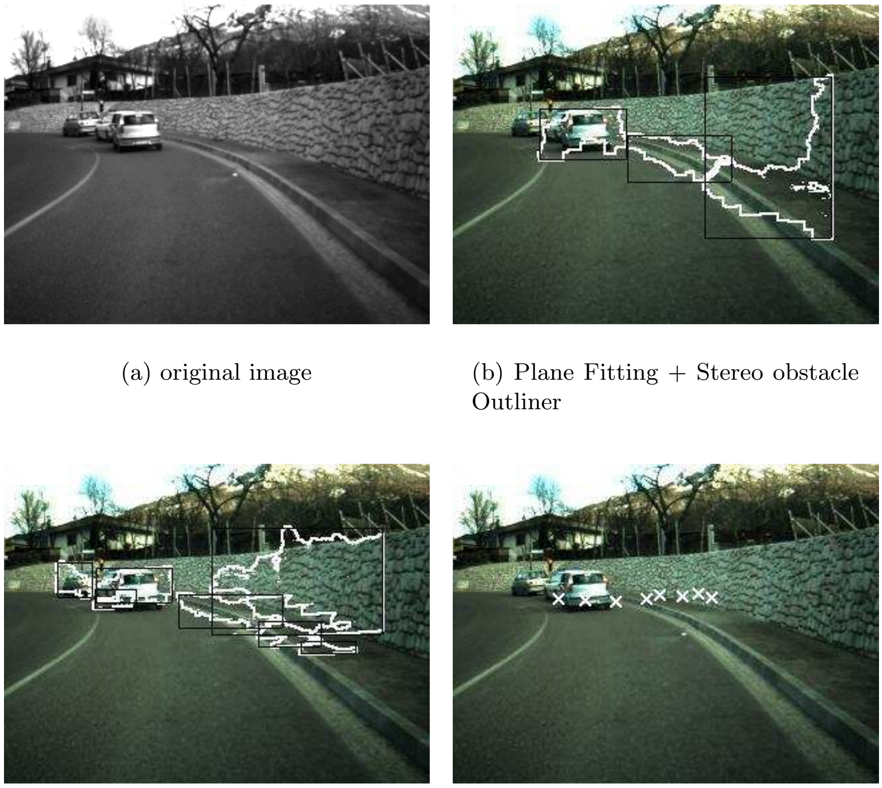
apart from the natural correlation coming from the same acquisition device. VD detects obstacles if their disparity falls in one, or at most two bins [10]. This results in a peak over the road line on the v-disparity map. VD is less efficient when obstacles cover various disparity steps (*e.g.* walls along the road, an overtaking vehicle) or if the input image is cluttered. IPM estimates well the obstacle/road contact point thanks to the inverse perspective effect (one pixel in a distance becomes many when transformed on the birdview perspective) [7]. However, various parts of complex obstacles are difficult to merge.

SO outlines connected regions displaying a constant disparity value and performs split and merge based on appearance heuristics of vehicles and other street participants. However, detected regions sometimes wrongly belong either to the road plane or to objects well above the road. TL, PF and HZ help disambiguating these situations.

4.2 Results

Development of the Obstacle Detection is still on-going. We are currently focusing on refining results of individual modules, leaving their integration for a later stage.

Hereafter we show some qualitative results by individual modules (Fig. 6). PF feeds a road estimation to the SO module, which then grows disparity regions starting from the road border (Fig. 6(b)). VD detects obstacle candidates on the v-disparity map. Such hypotheses are then refined by SO with splitting and merging on the disparity image (Fig. 6(c)). Finally, IPM provides the image contact point between road plane and obstacles. Its performance is most reliable when the obstacle is frontal. However, IPM does not provide useful information on the total image area occupied (Fig. 6(d)).



(a) original image

(b) Plane Fitting + Stereo obstacle Outliner

(c) V-Disparity + Stereo obstacle Outliner

(d) Inverse Perspective Mapping

Fig. 6. Obstacle detection results.

5 DIPLODOC demo description

We have set up a demo of the individual modules and the fusion of some of them. Modules work on data previously acquired with the experimental DIPLODOC vehicle moving on two-lane rural roads. Data types include stereo images, vehicle status (*e.g.* brakes on/off), and other measurements (*e.g.* on-board speedometer, GPS).

6 Future work

Current development of RT module tackles introducing vertical curvature estimation in order to deal with hilly roads and performing the estimation on real world image sequences.

For what concerns fusion, we plan to compare the performance of current methods to Bayesian fusers [15] and possibly others (*e.g.* neural networks, Dempster-Shafer, etc).

Presently, no optimization has yet been considered for real-time operation. However, we are investigating new solutions in the field of projective fusers, *i.e.* fusers selecting the output of only one sensory modality at a time [16]. These strategies may help improving robustness while approaching the required real-time constraints for real driver assistance.

Finally, algorithms have to be developed for tracking and classification of obstacles, in order to rank threat of collision situations.

References

1. (2002) DIPLODOC project website. [Online]. Available: http://www.dit.unitn.it/~diplodoc/
2. K. Konolige, "Small vision systems: Hardware and implementation," in *Eighth International Symposium on Robotics Research*, Hayama, Japan, October 1997, pp. 209–214.
3. (1997) SRI Stereo Engine. [Online]. Available: http://www.ai.sri.com/~konolige/svs/
4. Y. Wang, D. Shen, and E. K. Teoh, "Lane detection using spline model," *Pattern Recognition Letters*, vol. 21, no. 8, pp. 677–689, July 2000.
5. Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using B-Snake," *Image and Vision Computing*, vol. 4, no. 22, pp. 269–280, Apr. 2004.
6. A. Giachetti, M. Campani, and V. Torre, "The use of optical flow for road navigation," *IEEE J. Robot. Automat.*, vol. 14, no. 1, pp. 34–48, 1998.
7. M. Bertozzi and A. Broggi, "GOLD: A Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection," *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 62–81, Jan. 1998.
8. M. Bertozzi, A. Broggi, and A. Fascioli, "Stereo Inverse Perspective Mapping: Theory and Applications," *International Journal of Computer Vision*, vol. 16, pp. 585–590, 1998.

9. C. Corridori and M. Zanin, "High Curvature Two-Clothoid Road Model Estimation," in *Proceedings of IEEE Conference on Intelligent Transportation Systems (ITSC 2004)*, Washington DC, USA, Oct. 3-6 2004.
10. R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation," in *Proceedings of IEEE Intelligent Vehicle Symposium*, Versailles, France, June 2002.
11. R. Labayrade and D. Aubert, "A Single Framework for Vehicle Roll, Pitch, Yaw Estimation and Obstacles Detection by Stereovision," in *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2003)*, Colombus, OH, USA, June 9-11 2003, pp. 31–36.
12. E. D. Dickmanns and B. D. Mysliwets, "Recursive 3-D Road and Relative Ego-State Recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 199–213, Feb. 1992.
13. Colombo, *Manuale dell'ingegnere.* Hoepli, 1997.
14. D. Khosla, "Accurate estimation of forward path geometry using two-clothoid road model," in *Proceedings of IEEE Intelligent Vehicle Symposium*, Versailles, France, June 2002.
15. P. Lombardi and B. Zavidovique, "Architectural Design Issues for Bayesian Contextual Vision," in *Proceedings of 17th International Conference on Pattern Recognition (ICPR 2004)*, Cambridge, UK, August 23-26 2004.
16. N. S. V. Rao, *Multisensor Fusion.* Kluwer Academic Pub., 2002, ch. Multisensor fusion under unknown distributions: Finite sample performance guarantees, pp. 1–39.