Vehicle Detection by Means of Stereo Vision-Based Obstacles Features Extraction and Monocular Pattern Analysis

Gwenaëlle Toulminet, Massimo Bertozzi, Stéphane Mousset, Abdelaziz Bensrhair, and Alberto Broggi, *Senior Member, IEEE*

***Abstract*—This paper presents a stereo vision system for the detection and distance computation of a preceding vehicle. It is divided in two major steps. Initially, a stereo vision-based algorithm is used to extract relevant three-dimensional (3-D) features in the scene, these features are investigated further in order to select the ones that belong to *vertical* objects only and not to the road or background. These 3-D vertical features are then used as a starting point for preceding vehicle detection; by using a symmetry operator, a match against a simplified model of a rear vehicle's shape is performed using a monocular vision-based approach that allows the identification of a preceding vehicle. In addition, using the 3-D information previously extracted, an accurate distance computation is performed.**

***Index Terms*—Extraction of three-dimensional (3-D) edges of obstacle, platooning, stereo vision, vehicle detection.**

I. Introduction

In THE past few years, extensive research has been carried out in the field of driver assistance systems in order to increase road safety and comfort when driving. For instance, the adaptive cruise control (ACC) and the electronic stability program (ESP) are already installed in many recent vehicles. Prototypes of intelligent vehicles dedicated to *road following* have been designed and tested in real conditions on highways [1]–[3], in urban environment [4], on country roads [5], and off-road [6]. Furthermore, in order to assist drivers in urban traffic environment, some other promising work has been performed on *automatic parking* functionality [7]–[9], on the challenging *Stop&Go* system [10]–[12] and on *traffic sign recognition* [13]–[15].

Manuscript received March 24, 2004; revised June 15, 2005. This work was supported by a GALILEE 2000 program, a French-Italian program between the Institut National des Sciences Appliquées (INSA) of Rouen (France) and the Dipartimento di Ingegneria dell'Informazione of the University of Parma (Italy). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bruno Carpentieri.

G. Toulminet, S. Mousset, and A. Bensrhair are with the Laboratoire Perception Systèmes Information, Université/INSA de Rouen, Mont-Saint-Aignan Cedex, France (e-mail: gwenaelle.toulminet@insa-rouen.fr; stephane.mousset@insa-rouen.fr; abdelaziz.bensrhair@insa-rouen.fr).

M. Bertozzi and A. Broggi are with the Dipartimento di Ingegneria dell'Informazione, Università di Parma, I-43100 Parma, Italy (e-mail: bertozzi@ce.unipr.it; broggi@ce.unipr.it).

Digital Object Identifier 10.1109/TIP.2006.875174

Concerning automatic platooning (the automatic following of the preceding vehicle), processing can be based on:

- the use of radar only (ACC);
- the fusion of an active sensor (laser, radar, lidar) and monocular vision [16];
- monocular vision only.

A widely used approach for monocular vision-based vehicle detection is the search of specific patterns [17], for example: shape [18], [19], motion [20], color [16], [18], symmetry [16], [21], [22], shadow [16]–[22], texture [22], or the use of a specific model [23]. These specific patterns are usually associated to extract regions of interest in the image. These regions of interest correspond to potential vehicles of the observed scene which are identified by using bounding boxes. Usually, this two-dimensional (2-D) detection is not sufficient and other processing methods are often useful to validate the presence of a vehicle or reject false detection. They are for example: classification of the regions of interest, their localization with respect to the road or their tracking. However, a major problem still remains open: the distance to the detected vehicles cannot be accurately computed without the cooperation of other sensors.

This paper is the result of four years of joint research collaboration between the *Institut National des Sciences Appliquées* of Rouen (France) and the *Dipartimento di Ingegneria dell' Informazione* of the University of Parma (Italy). It is derived from a number of activities that date back to the late 1990s. This paper introduces a stereo vision algorithm specifically tailored for Vehicle Detection. Compared to a traditional stereo-vision algorithm, the discussed approach is not aimed at a complete three–dimensional (3-D) world reconstruction but to the mere extraction of 3-D features potentially belonging to a vehicle, namely only 3-D vertical edges. The list of 3-D features is used by a monocular vision system that performs Vehicle Detection by means of a match with a simple vehicle model. As a consequence, the system draws advantages from having additional information on edges' distances from the camera and from working on actual vertical characteristics, i.e., without misinterpretations caused by artefacts or road infrastructures. In addition, besides a more reliable detection, an accurate estimation of vehicle distance is also obtained.

This article is organized as follows. Section II introduces the two main steps of the Vehicle Detection process which are described in Sections III and IV. Some experimental results are presented in Section V. Finally, a Conclusion ends the paper.

II. Vehicle Detection Algorithm

The vision-based algorithm discussed below has been explicitly designed to support automatic platooning based on the localization of the preceding vehicle only.

The Vehicle Detection algorithm is divided in two major steps: *Three-Dimensional features extraction* and *preceding vehicle detection*. In the first step, a stereo vision approach is used to identify 3-D features that belong to potential obstacles on the scene. In the second stage, a symmetry-based algorithm is used to identify elements among all extracted 3-D features that may belong to the preceding vehicle, i.e., to specifically detect that vehicle.

In a previous paper, a monocular approach for feature extraction was used [24]. Anyway, there were cases in which a complex scenario led the algorithm to false detections or to group together objects at different distances [25]. Since the 3-D feature extraction is now able to discriminate 3-D vertical edges and compute edges' distance from the camera, this knowledge is exploited to filter out features that do not belong to potential obstacles (i.e., features belonging to the road) so that the detection of the preceding vehicle can be computed only on edges that actually represent 3-D vertical objects. At the same time, features that do not lie at similar distances will not be used together for detecting the same vehicle. Moreover, a first check on vehicle's size with respect to its distance can be performed at an early stage: candidates of vehicles' left and right vertical edges can be filtered out if they represent an object too narrow or too large with respect to a vehicle lying at the distance estimated by the stereo-based procedure.

The two following sections describe these two main steps.

III. 3-D Features Extraction

Within the framework of road obstacle detection, road features can be classified into two classes: *Nonobstacle* and *Obstacle*. An obstacle is defined as something that obstructs or may obstruct the intelligent vehicle driving path. Vehicles, pedestrians, animals, security guardrails are examples of *Obstacles*. Lane markings, artefacts are examples of *Nonobstacles*.

The vehicle detection functionality (Section IV) is based on a monocular pattern analysis and the extraction of 3-D features of obstacles. It is used to enhance the robustness and reliability of the monocular pattern analysis.

The extraction of 3-D features of obstacles starts with the construction of a 3-D sparse map of a specific area of the image. This area of interest is identified on the basis of road position and perspective constraints. Then, 3-D edge shapes are constructed from the 3-D sparse map. Finally, the 3-D shapes are identified as edges of obstacles or edges of nonobstacles. The 3-D shapes of the class *Obstacle* are extracted.

A. Construction of 3-D Sparse Maps

The algorithm for the construction of 3-D sparse maps is a line by line processing designed for a specific configuration of vision sensor: the optical axes of the two camera-lens units are parallel, and the straight line joining the two optical centers is parallel to each image horizontal line in order to respect an epipolar constraint (cf. Fig. 1).

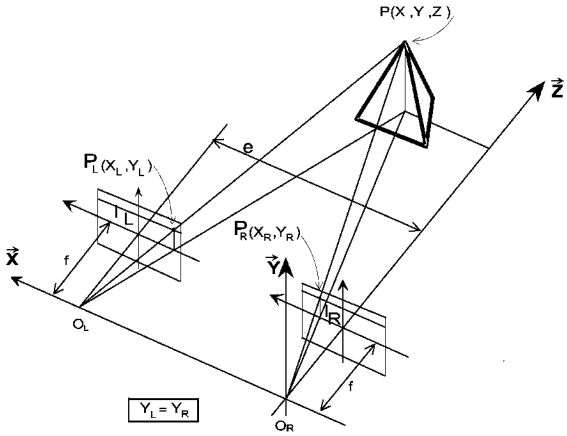

Fig. 1. Configuration of the stereo vision system.

In the first step of the algorithm, the edge points of the right and left images are segmented by a self-adaptive and mono-dimensional operator, called *declivity* (check [26] for the original definition). In a second step, the edge points of the right image are matched with the edge points of the left image, using a dynamic programming method. The matching algorithm provides 3-D information, based on the positions of left and right edge points. The coordinates (X_P, Y_P, Z_P) of a 3-D point P of an object are given by (1)

$$X_P = \frac{x_l \times e}{p_x \times \delta} \quad Y_P = \frac{y_l \times e}{p_x \times \delta} \quad Z_P = \frac{f \times e}{p_x \times \delta} \tag{1}$$

where e is the distance between the two optical centers, p_x is the width of the CCD pixel, f is the focal length of the two lenses, δ is the disparity of P. $\delta = x_l - x_r$, (x_l, y_l), and (x_r, y_l) are the coordinates of the projections of P respectively in the left image and in the right image.

The result of the matching algorithm is a 3-D sparse map. An evaluation of this algorithm applied on the entire image is described in [27] which first introduced the algorithm: 92.6% of the right edge points are associated with a left edge point; among these associations, 98% are correct. For an inside scene, the number of edge points are about 15 000 in each of the two stereoscopic images whose resolution is 512 ×512 × 8 bits. For an outside scene, they are about 21 000. An example of experimental result is presented in Fig. 2.

B. Construction of 3-D Shapes

A 3-D shape is defined as a chain of connected 3-D points belonging to the same object in the scene. Like 3-D points, a 3-D shape is characterized by its projections in the right and in the left image. Then, the projection in the right image of an actual 3-D shape is a chain of connected right edge points; and the projection in the left image of this actual 3-D shape is a chain of left edge points.

The construction of 3-D shapes [28] starts with the construction of their projections in the right image using connecting, depth and uniqueness criteria. They are the projections of actual 3-D shapes in the right image. The result of the matching algorithm provides the estimation of their projections in the left image: a constructed 3-D shape may be partially false as the

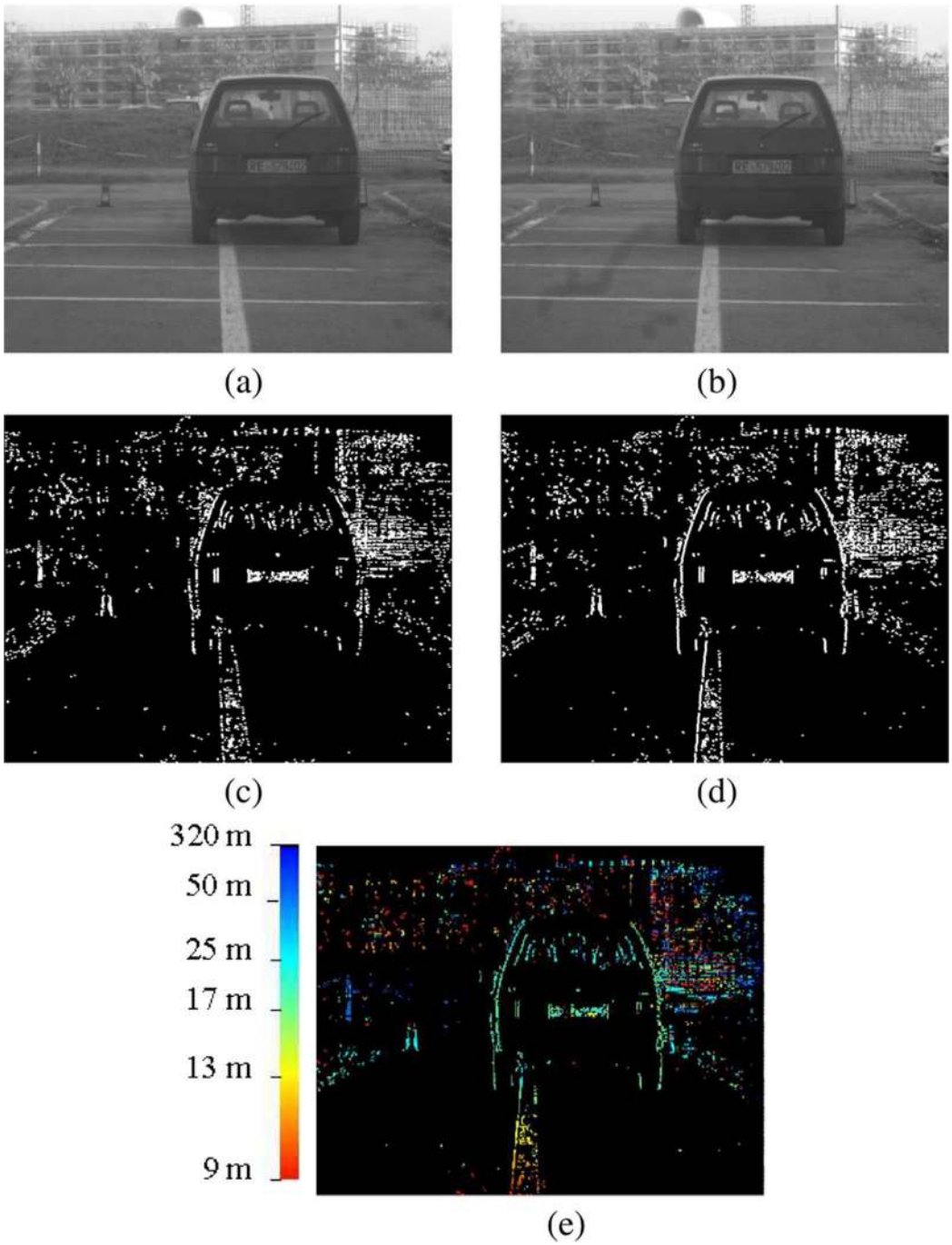

Fig. 2. Experimental results: (a) left image; (b) right image; (c) left edge points; (d) right edge points; and (e) 3-D sparse map coded with grey-level values and its depth scale. (Color version available online at http://ieeexplore.ieee.org.)

depth criterion allows a constructed 3-D shape to contain a certain number of 3-D points resulting from wrong matchings. It may also be incomplete as the construction allows a constructed 3-D shape to contain a certain number of 2–D right points which have not been matched.

C. Detection and Correction of Wrong Matchings

In order to increase the robustness and the reliability of the extraction of 3-D shapes that belong to obstacles, the estimation of the actual 3-D shapes (i.e., the constructed 3-D shapes) must be improved. Because road environment is structured, the edges of road scenes are smooth 3-D shapes. Then, wrong edge points associations are isolated 3-D points. They are detected and corrected taking the constraint of disparity continuity into account (check [29] for the original definition).

In the first step of the algorithm [28], small 3-D shapes are deleted. This operation eliminates most of wrong edge points associations. In a second step, we suppose that most edge point associations of remaining 3-D shapes are correct. And if the coordinates of a 3-D point belonging to a 3-D shape do not validate a criterion of disparity continuity, then this 3-D point is detected

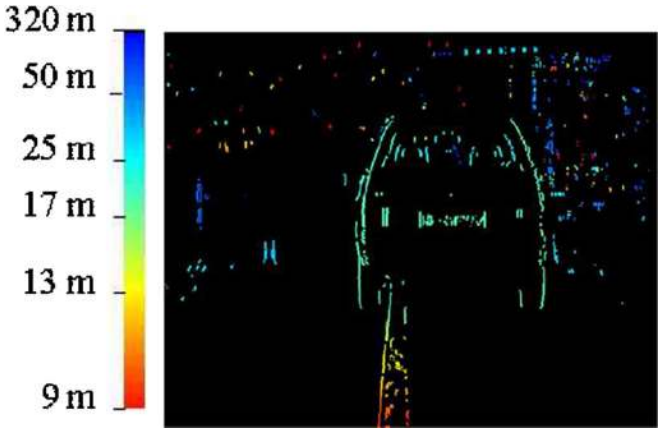

Fig. 3. Result of the detection and correction of wrong matchings of the 3-D map of Fig. 2. (Color version available online at http://ieeexplore.ieee.org.)

as the result of a wrong edge points association. Finally, the correction step is applied on each right edge point that has not been matched or that has been wrongly matched: for each of these points, we look for a left edge point that has not been matched or wrongly matched and that validates a criterion of disparity continuity.

Usually, errors in disparity measurements are modeled by a stochastic component (which usually has a magnitude of 1 pixel) and a wrong matching component (which can have a magnitude of many pixels). We consider that the steps described in this section eliminate the wrong matching component.

The benefit of the detection and correction of wrong matchings of the 3-D map in Fig. 2 is presented in Fig. 3. The pair of stereoscopic images of Fig. 2 has been intentionally chosen because of the grill in the background that creates a lot of false matchings in the 3-D map of Fig. 2. In Fig. 3, most of them have been erased.

D. Identification of 3-D Shapes

The identification of the 3-D shapes as edges of obstacles or not has been conceived as a cooperation of two methods. The first one selects 3-D shapes by thresholding the disparity value of their 3-D points. The threshold values are calculated based on the detection of the road. We chose to model the road by a plane even if this modeling is not always valid, particularly due to dynamic pitching of the vehicle or when approaching uphill or downhill roads. The second method selects 3-D straight segments by thresholding their inclination angle with respect to the road plane. From these two selection results, 3-D shapes that are edges of obstacles are identified. The 3-D shapes that do not belong to the *Obstacle* class, belong to the *Nonobstacle* class.

The following sections describe the three steps of identification of 3-D shapes. The first one presents the method of detection of the road plane. The two other ones present the two methods of selection of 3-D shapes.

1) Detection of the Road Plane: Our method of detection of the road plane derives from the research work of the LIVIC [30]: the authors have introduced an original representation called *V-Disparity* (V means the coordinate of a pixel in the (u, v) right image coordinate system). This method requires information in the image such as lane markings, road edges, shadows,... any sign belonging to the road plane. Let P be a 3-D point whose disparity is δ. Let (u, v) be the coordinates of p_r its projection in the right image. The coordinates v and δ of P are transferred to a grey level map. This operation is repeated for each point P of all 3-D shapes to construct a grey level map (v, δ): an accumulator $H(v, \delta)$ accumulates the points with the same disparity that occurs on a given line v of the image. In a case of a planar road, its representation in the grey level map is a straight line. It is extracted using a *Hough Transform*. Then, we compute the height H of the stereo vision system and the inclination angle α of the road plane with respect to the referee of the stereo vision system (cf. Fig. 4. Note that we do not take the roll angle into account.

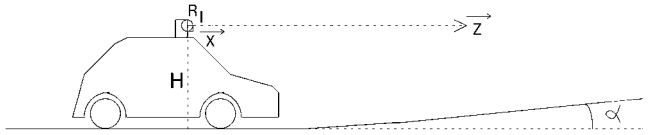

Fig. 4. Inclination angle α of the road plane and height H of the stereo vision system.

2) Selection of 3-D Shapes by Thresholding Disparity Values: Using the principles of the homographic transformation and the specific configuration of our stereo vision system, we associate each line y of the right image to a disparity value $\delta_{\text{road}}(y)$. The function δ_{road} represents the disparity of the road. It is computed from the inclination angle of the road plane α, the height H of the stereo vision system and the calibration of the stereo vision system. It is equal to

$$\delta_{\text{road}} : y \mapsto \begin{cases} \frac{e \times \left(p_y \times \left(y - \frac{h}{2}\right) \times \cos\alpha + f \times \sin\alpha\right)}{p_x \times (H - y \times p_y)}, & \text{if } y \geq y_{\text{horizon}} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

with $y_{\text{horizon}} = ((h/2) - (f/p_y) \times \tan\alpha)$

where e is the distance between the two optical centers, p_x and p_y are the width and height of the CCD pixel, f is the focal length of the two lenses, and $w \times h$ is the resolution in pixels of the cameras. The function δ_{road} is used as a threshold function to select 3-D shapes that are supposed to belong to obstacles: if the disparity of a 3-D point of line y is higher than $\delta_{\text{road}}(y)$, then it is considered to be above the road plane. The 3-D shapes above the road plane are selected and supposed to belong to obstacles.

Fig. 5 shows an experimental result of extraction of 3-D shapes by thresholding the disparity values of their 3-D points. The computed height of the stereo vision sensor is $H = 1.23$. The computed inclination angle of the road plane is $\alpha = 17^o$. The static calibration of the stereo vision sensor gave $H = 1.20$ m and $\alpha = 10^o$

3) Selection of 3-D Segments by Thresholding Their Inclination Angle: As road environment is structured, the constructed 3-D shapes can be approximated by means of one or several 3-D straight segments. By an iterative partition method, 3-D shapes are decomposed into 3-D segments. In order to select 3-D segments that belong to obstacles, we calculate and threshold the inclination angles of 3-D segments with respect to the road plane. The inclination angle β of a 3-D segment with respect to the

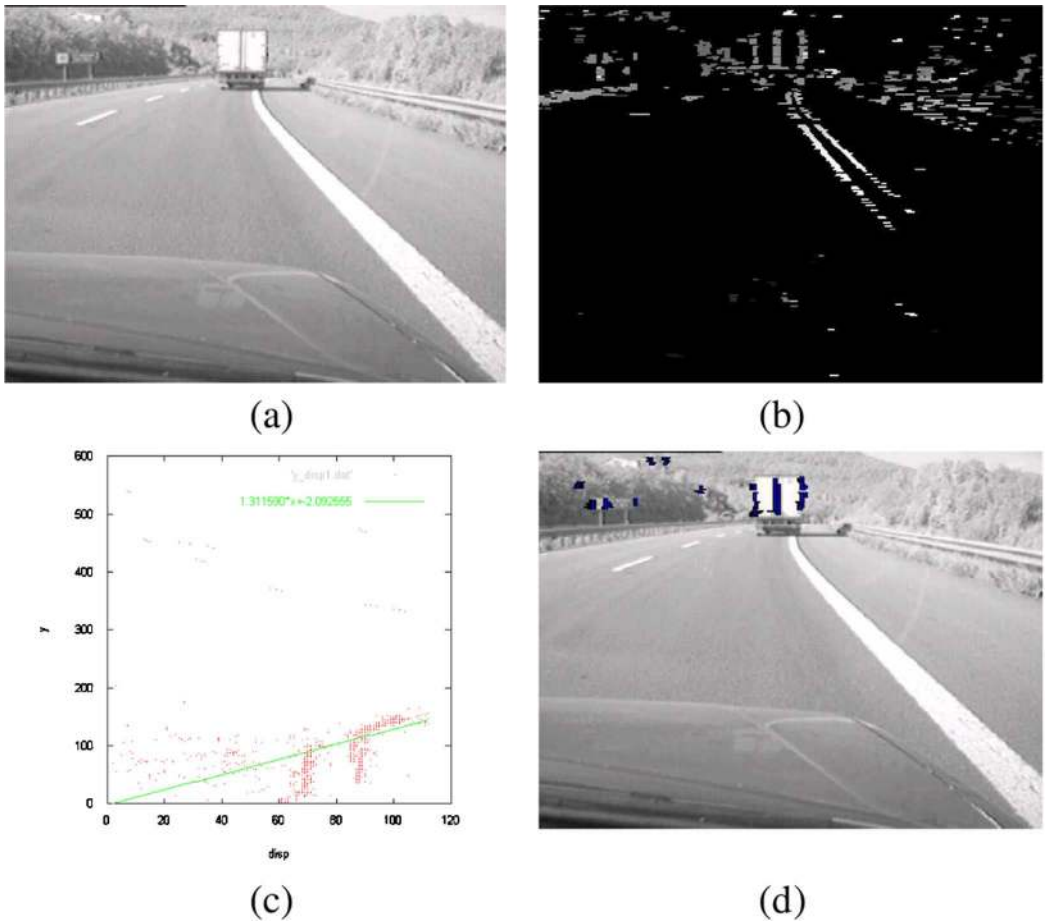

Fig. 5. Experimental results of extraction of 3-D shapes by thresholding disparity values: (a) right image; (b) 3-D sparse map coded with grey level values; (c)(v, δ) representation; and (d) extracted 3-D shapes superimposed in blue on the right image. (Color version available online at http://ieeexplore.ieee.org.)

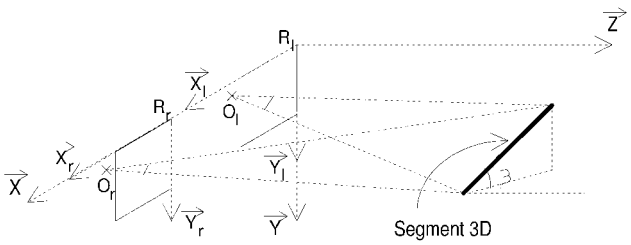

Fig. 6. Inclination angle β of a 3-D segment.

road plane (cf. Fig. 6) is a pattern that can be used to detect obstacles. Indeed, if the inclination angle of a 3-D segment tends towards 0^o, then the 3-D segment usually belongs to the road plane (road markings, road borders, etc.). If this angle tends to 90^o, then it belongs to obstacles.

In a first step, we compute the vector $\overrightarrow{V}$ of the straight line which contains the 3-D segment in the referee of the stereo vision system [cf. Fig. 6 and (3), shown at the bottom of the page].

m_r, m_l, b_r, and b_l are calculated by a least square method and are such that

$$f_r : x_r \rightarrow m_r \times y + b_r \quad f_l : x_l \rightarrow m_l \times y + b_l \tag{4}$$

where f_r and f_l are, respectively, the equations of the projections of the 3-D segment in the right and in the left image.

In a second step, the inclination angle of a 3-D segment is computed with respect to the road plane by using geometry properties.

- Case 1: if $V_z \neq 0$ and $V_x \neq 0$

$$\beta = \arctan\left(-\frac{V_y}{\sqrt{V_x^2 + V_z^2}}\right) - \arctan\left(\frac{V_z \times \tan\alpha}{\sqrt{V_x^2 + V_z^2}}\right). \tag{5}$$

- Case 2: if $V_z = 0$ and $V_x = 0$

$$\beta = \frac{\Pi}{2} - \alpha. \tag{6}$$

$$\overrightarrow{V} = \begin{pmatrix} V_x = p_x \times ((m_r \times b_l - m_l \times b_r) + \frac{w}{2} \times (m_l - m_r)) \\ V_y = p_y \times ((b_l - b_r) + \frac{h}{2} \times (m_l - m_r)) \\ V_z = f \times (m_l - m_r) \end{pmatrix} \tag{3}$$

Fig. 7. Experimental results of the extraction of 3-D segments by thresholding their inclination angle: (a) 3-D sparse map coded with grey level values and (b) extracted 3-D segments superimposed in blue on the right image; the predefined threshold of extraction is $17°$. (Color version available online at http://ieeexplore.ieee.org.)

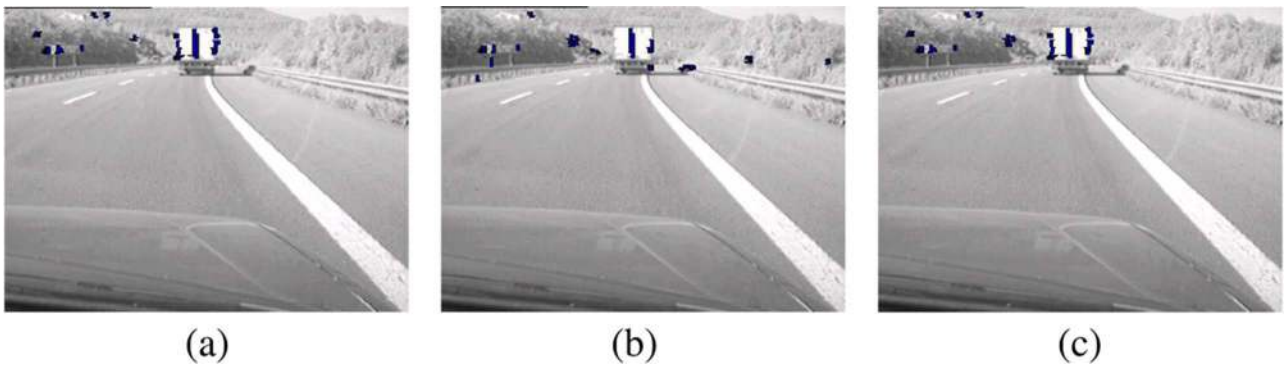

Fig. 8. Contribution of the second step of the cooperation of the two methods to the extraction of the 3-D edges of obstacles: (a) extracted 3-D shapes superimposed in blue on the right image; (b) extracted 3-D segments superimposed in blue on the right image; the predefined threshold of extraction is $30°$. Due to a higher threshold, 3-D edges of obstacles are missing compared to Fig. 7(b). (c) extracted 3-D edges of obstacles. The extraction is robust and reliable despite a poor extraction of the second method on this image. (Color version available online at http://ieeexplore.ieee.org.)

Finally, we extract the 3-D segments whose inclination angle is higher than a predefined threshold. Fig. 7 shows an experimental result of extraction of 3-D segments by thresholding their inclination angle.

4) Extraction of 3-D Edges of Obstacles: In this section, a cooperation of the two previous methods is proposed. The first method (Section III-D.2) has the advantage of being insensitive to noise in disparity measurements, but the robustness and the reliability of this method of extraction depends on the modeling and the method used to detect the road.

This is not the case with the second method (Section III-D.3) which does not suffer from approximate modeling and detection of the road, but some 3-D edges of the obstacles can have a null inclination angle (trailer, trucks with no loading, safety railings along highways, etc.). In addition, this method is sensitive to errors in disparity measurements. As mentioned in Section III-C, the detection and correction of wrong matchings eliminate the wrong matching component. Then, we consider that the second method of extraction is sensitive to the stochastic component only.

These two methods for the extraction of 3-D edges of obstacles do not have the same sensitivity. Considering their different sensitivity, a cooperation of these two methods is proposed as follows. The cooperation is performed is two steps. The first step has been designed to increase the reliability of the extraction of 3-D edges of obstacles (compared with the two methods of extraction used separately). The second step, concerns the robustness of the cooperation.

Let O_1 be the set of 3-D shapes that have been selected by the first method (Section III-D.2). Let O_2 be the set of 3-D segments that have been selected by the second method (Section III-D.3). And, let O be the set of 3-D shapes identified as edges of obstacles. In the first step of the construction of O, O_3 is constructed so that $O_3 = O_1 \cap O_2$. As a result, O_3 contains 3-D segments (that are portions of 3-D shapes) that have been identified as obstacles by the two methods. Then, for each 3-D shape S, if

$$\frac{\text{number of } 3-\text{D points of } S \text{ belonging to } \mathrm{O}_3}{\text{number of } 3-\text{D points of } S} \geq \text{threshold} \tag{7}$$

then S belongs to O. At the end of this fist step, O contains 3-D shapes considered as edges of obstacles.

In a second step, O is completed with 3-D shapes of O_1 or exclusively O_2 with respect to the following considerations. If a 3-D shape of $O_1 \cup O_2$ is:

- an actual 3-D edge of obstacle in the scene;
- at the same height with respect to the road as a 3-D shape of O;
- closer to the stereo vision system than the 3-D shape of O;

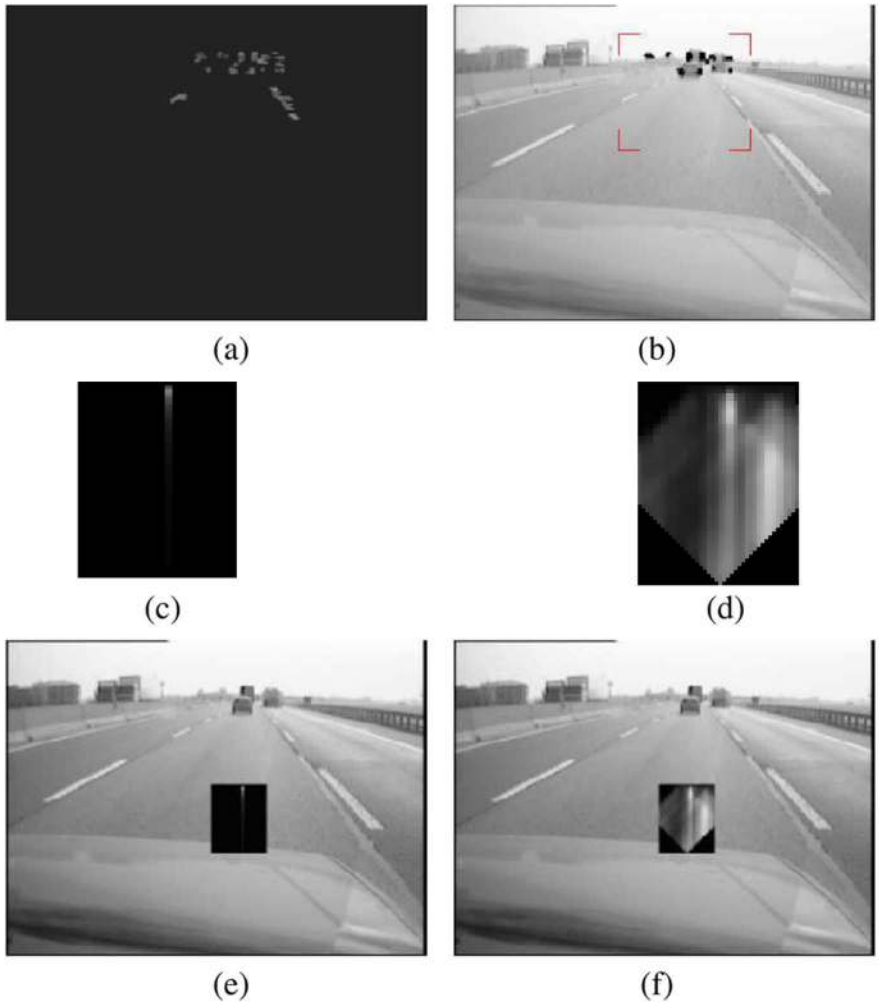

Fig. 9. Steps of the computation of combined symmetry: (a) 3-D shapes computed in the area of interest of the image; (b) 3-D edges of obstacles superimposed in black on the right image; (c) symmetry map computed on 3-D features; (d) combined symmetry map; (e) symmetry map computed on 3-D features superimposed on the right image; and (f) combined symmetry map superimposed on the right image. (Color version available online at http://ieeexplore.ieee.org.)

then it is at the same height in the image as the 3-D shape of O On the contrary, a 3-D shape of $O_1 \cup O_2$ which is:

- an actual 3-D edge of the road in the scene (which has therefore been wrongly extracted by one of the two methods);
- closer to the stereo vision sensor than a 3-D shape of O;

is under the projection of the 3-D shape of O in the image.

The second step concerns the robustness of the cooperation. It has been designed to compensate a poor extraction by one of the two methods (cf. Fig. 8).

The extraction of 3-D edges of obstacles is robust and reliable due to the cooperation of two methods of extraction which do not have the same sensitivity [31].

IV. Preceding Vehicle Detection

The detection of the vehicle ahead is based on the assumptions that a preceding vehicle (framed from the rear) is generally symmetric, characterized by a nearly-rectangular bounding box which satisfies specific aspect ratio constraints, and placed in a specific area of the image. Initially, an area of interest is identified on the basis of road position and perspective constraints. This area is searched for possible vertical symmetries. Once the width and position of the symmetrical area have been detected, a new search begins, aimed at the detection of the two bottom corners of a rectangular bounding box. Finally, the top horizontal limit of the vehicle is searched for, and the preceding vehicle is localized.

The following tracking phase is performed through the maximization of the correlation between the portion of the image contained in the bounding box of the previous frame (partially stretched and reduced to take into account small size variations due to the increment or reduction of the relative distance) and the new frame.

A. Symmetry Detection

To determine the symmetry content inside the area of interest, a *symmetry map* is used, namely an image whose pixels encode the symmetry content.

Each possible symmetry axis within the area of interest is considered. For each axis, different widths for the symmetry area around the axis are in turn examined; the result is a new image (the symmetry map) in which the vertical coordinate of each pixel is related to the horizontal width of the image area

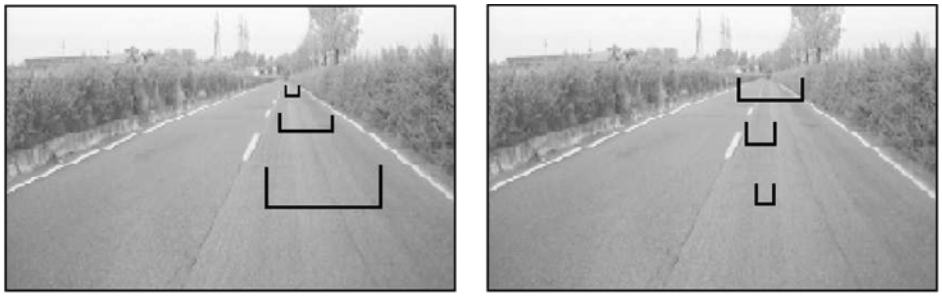

Fig. 10. Detection of the lower part of the bounding box: (a) correct position and size, taking into consideration perspective constraints and knowledge on the acquisition system setup, as well as typical vehicles' size and (b) examples of incorrect bounding boxes if the previous considerations are not taken into account. (Color version available online at http://ieeexplore.ieee.org.)

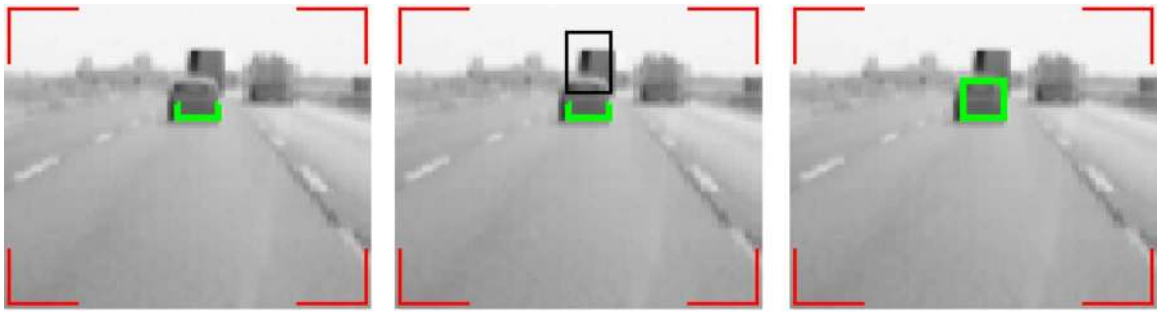

Fig. 11. Bounding box detection: (a) detection of the lower part of the bounding box superimposed on the right image; (b) the search area for the upper part of the bounding box. It is delimited by the black rectangle. It takes into account knowledge about the typical vehicles' aspect ratio, and (c) the resulting bounding box superimposed on the right image. (Color version available online at http://ieeexplore.ieee.org.)

considered for computing the symmetry and the horizontal coordinate refers to the position of the axis within the area of interest. The brighter the pixel, the higher the symmetry content.

A combined symmetry map is built as a weighted sum of the symmetry maps of four monocular images [25] (a symmetry map obtained from the grey level image and the ones obtained by the analysis of horizontal and vertical edges) and the symmetry map of 3-D features.

In the first step of the construction of the symmetry map of 3-D features, 3-D points that belong to 3-D shapes whose disparity variance is lower than 1 pixel are examined. Indeed, a 3-D shape whose disparity variance is low is more likely to belong to an obstacle than a 3-D shape whose disparity variance is high. In the second step, all pairs of examined 3-D points that are symmetrical with respect to the axis and that have a similar distance from the vision system are matched. Candidates of vehicle's left and right 3-D edges are not considered if they represent an object which is considered larger than a vehicle lying at the distance estimated by the stereo vision sensor. Otherwise, a symmetry content is computed. The symmetry content is proportionally increased if one or both candidates are 3-D edges of obstacles.

Fig. 9 illustrates steps of the computation of combined symmetry map.

B. Bounding Box Detection

The symmetry axis and symmetry width corresponding to the maximum value into the combined symmetry map are analyzed to detect the presence of a vehicle shape, namely of a bounding box. Initially, the presence of two corners representing the bottom of the bounding box around the vehicle is checked; a traditional pattern matching technique is used. In addition, size constraints are used to speed up the search: potential bounding boxes too large or too small to be a vehicle at the distance computed thanks to the 3-D features are discarded without checking the presence of the corners. This process is followed by the detection of the top part of the bounding box, which is looked for in a specific region whose location is again determined by perspective and size constraints.

Fig. 10 shows possible and impossible bottom parts of the bounding box, while Fig. 11 presents the results of the lower corner detection and shows the search area.

C. Backtracking

Sometimes it may happen that, when considering the symmetry maximum, no correct bounding boxes exist. Therefore, a backtracking approach is used: the symmetry map is again scanned for the next local maximum and a new search for a bounding box is performed. Fig. 12 shows a situation in which the first symmetry maximum, generated by a building, does not lead to a correct bounding box; on the other hand, the second maximum leads to the correct detection of the vehicle.

D. Tracking

Once a preceding vehicle has been detected, a monocular tracking procedure is used to speed up the computation. Since there will be a high correlation between the vehicle's rears in

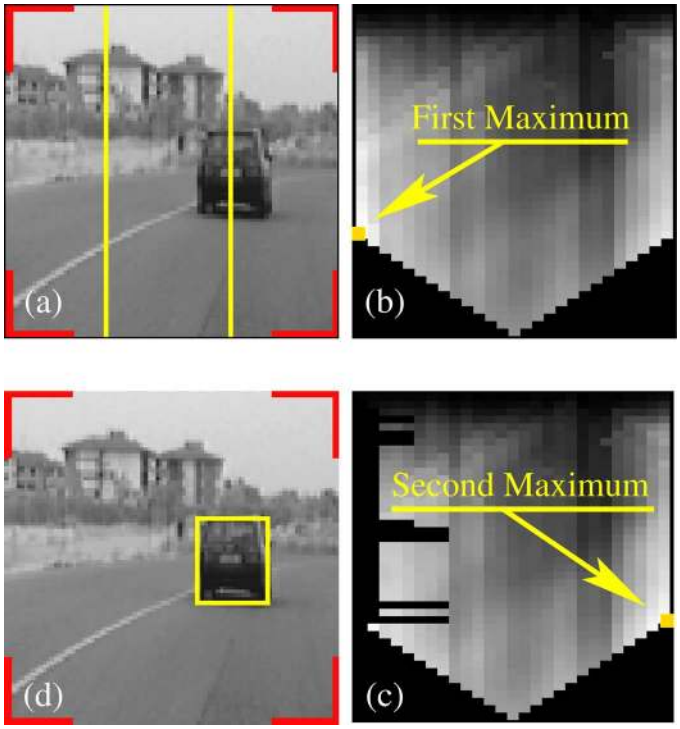

Fig. 12. A case in which the background symmetry is higher than the vehicle symmetry: (a) original image; (b) first symmetry map; (c) second symmetry map after the backtracking process has removed the peak near the maximum; and (d) final bounding box detection. (Color version available online at http://ieeexplore.ieee.org.)

two subsequent frames, a simple correlation is used. Anyway, area-based tracking is sensitive to window size [32]. In order to take into account fluctuations in size of the detected vehicle due to the increment or reduction of relative distance, a number of templates are computed expanding and reducing the portion of image that contains the vehicle (namely the bounding box). Each template is matched against the new frame and the correlation (C_T) is computed using the formula

$$C_T(\triangle x, \triangle y) = \sum_{x=0}^{X_A} \sum_{y=0}^{Y_A} ((L(x,y) - R(x+\triangle x, y+\triangle y))^2$$

where:

- x and y are the pixel relative coordinates within the template;
- X_A and Y_A represent the template size;
- functions R and L return pixel intensity;
- $\triangle x$ and $\triangle y$ are varied by shifting the template on the new frame in a area where the vehicle is expected to be found.

The minimum value of C_T identifies an area into the new frame where the vehicle is looked for.

E. Distance Computation

The distance to the leading vehicle can be computed by examining the 3-D features inside the bounding box. Since, the rear side of the vehicle can be approximated to a vertical plane, they should present similar values for the distance from the vision system.

Unfortunately, it may happen that some noisy 3-D features due to other objects than the vehicle can be found inside the bounding box. For this reason, a simple average is of no use.

A histogram of the distances of the 3-D features is built. For each distance from the vision system (inside an interest interval 0–100 m), the corresponding 3-D features inside the bounding box are counted. Therefore, the histogram shows how many 3-D features at a given distance are present inside the bounding box. A local average operator is used to smooth the histogram. Then the peak of the histogram is detected, it refers to the most frequent distance that can be found in the 3-D features, and thus is assumed as the distance to the preceding vehicle. The absolute accuracy ΔZ of this measured distance is obtained from (1) and is given by (8)

$$\Delta Z = \frac{f \times e}{p_x \times \Delta\delta} \tag{8}$$

where e is the distance between the two optical centers, p_x is the width of the CCD pixel, f is the focal length of the two lenses, $\Delta\delta$ is the absolute accuracy of disparity. The accuracy of *declivity* position is subpixel [26]. Then, $\Delta\delta$ is better than 2 pixels.

V. Experimental Results

The algorithm of vehicle detection has been tested in a number of different situations. Fig. 13 shows some experimental results. The stereo images have been acquired and processed by GOLD, the stereo vision system of the *Dip. di Ingegneria dell' Informazione* of the University of Parma. The two synchronous cameras were installed in a prototype vehicle of the University of Parma, behind the top corners of the windscreen and feature a 6-mm focal length. The stereo images come from daylight highway sequences. They have been acquired in real time (50 Hz) at the format of $768 \times 288 \times 8$ bits. They have been processed at the format of $384 \times 288 \times 8$ bits.

Thanks to stereo-based 3-D feature extraction, which is able to discriminate between features that belong to the road or background and features of vertical objects, the tests demonstrated that the system is reliable and robust with respect to noise caused by shadows, different road textures, varying illumination conditions, or complex scenarios.

The efficiency of the stereo approach used for 3-D feature selection and the exploitation of specific vehicle characteristics in the following detection phase allowed to obtain robust search and tracking. The system has been tested on a AMD Athlon 2.8 GHz with 1-GB memory; the average time required for the whole processing is 95 ms. Table I shows performance results.

As shown on Fig. 13, the vehicle detection system can face different complex situations. The most critical situations, presented in Fig. 14, are due to an incorrect bounding box selection. Other edges than the vehicle's ones [Fig. 14(a)–(c)] as well as internal vehicles edges [Fig. 14(d)–(g)] may cause an incorrect bounding box displacement. Occasionally, 3-D feature extraction can also be confused by a complex background, thus provoking false detections [Fig. 14(h)]. Moreover, since the algorithm has been designed for platooning functionality, it is tailored for finding only one vehicle; unfortunately, the closest ve-

Fig. 13. Results of the preceding vehicle detection in different situations, the detected vehicle is shown using a red bounding box superimposed onto the original image. The four red corners indicate the area of interest when no vehicle is being tracked; anyway, when in tracking mode, vehicles can be detected outside this area. As shown, even vehicles coming toward the vision system are detected as well. The range distance of the vehicle is shown below the red box. If the number of the 3-D points of the detected vehicle is below a predefined threshold, then its distance is not computed. (Color version available online at http://ieeexplore.ieee.org.)

TABLE I
PERFORMANCE RESULTS FOR THE DIFFERENT ALGORITHMS STEPS

Average time (ms)	
3D features extraction	62
Symmetry computation	21
Bounding Box detection	12
Total	95

hicle is not necessarily the one featuring the highest symmetry content [Fig. 14(i)].

VI. CONCLUSION

In this paper, a stereo vision-based algorithm for vehicle detection and distance computation has been presented. In the first stage, 3-D edges of obstacles are extracted. The second stage exploits the vertical symmetry characteristics of a vehicle when it is framed from the rear: a symmetry operator investigates the 3-D features previously computed, the grey level image and the ones obtained by the analysis of horizontal and vertical edges. Then, a match against a simplified model of a vehicle's rear shape allows the detection of a preceding vehicle and the computation of its distance. It is performed in a specific area of the image computed by the symmetry operator and using perspective constraints. Finally, a following tracking phase is performed.

In this research work, techniques of image processing have been highly used. The intensive tests performed proved that the detection is robust and reliable. However, image processing has its own limits: depth accuracy depends on an precise calibration of the stereo vision sensor. In addition, it is inversely proportional to depth, to put it another way, it rapidly decreases with depth. Reduced visibility conditions like fog, rain, or snow limit also the performance of a system based on vision only. In order to improve our system of vehicle detection based on stereo vision, a fusion with another kind of sensor could be investigated. The use of an active sensor, like radar or lidar, could be a promising solution considering the research work already done on the fusion of monocular image and these kinds of active sensor [16], [33].

The vehicle detection system based on stereo vision is the result of a joint research collaboration between the *Institut National des Sciences Appliquées* of Rouen (France) and the *Dip. di Ingegneria dell' Informazione* of the University of Parma (Italy).

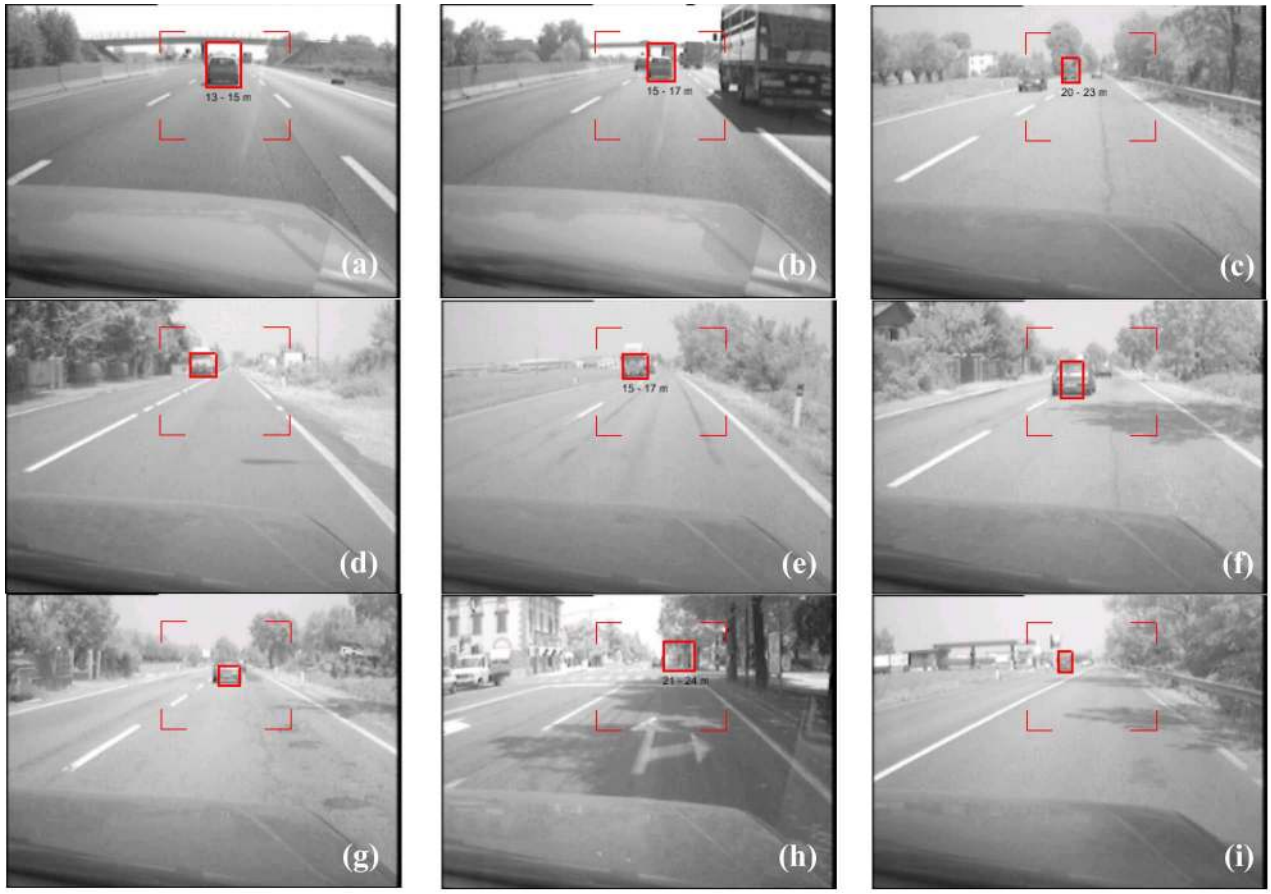

Fig. 14. Critical situations: (a)–(c) the bounding box detection fails in finding the upper part of the vehicle; (d)–(g) only a portion of the vehicle is detected; (h) a complex background fools the algorithm causing a false detection; and (i) a far vehicle features a bigger symmetry content that a closer one.

REFERENCES

[1] Q. Li, N. Zheng, and H. Cheng, "Springrobot: A prototype autonomous vehicle and its algorithms for lane detection," *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 4, pp. 300–308, Dec. 2004.

[2] Y. U. Yim and S. Y. Oh, "Three-feature based automatic lane detection algorithm (TFALDA) for autonomous driving," *IEEE Trans. Intell. Transport. Syst.*, vol. 4, no. 4, pp. 219–225, Dec. 2003.

[3] C. Rotaru, T. Graf, and J. Zhang, "Extracting road features from color images using a cognitive approach," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 298–303.

[4] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 4, pp. 309–318, Dec. 2004.

[5] M. S. von Trzebiatowski, A. Gern, U. Franke, U-P. Kaeppeler, and P. Levi, "Detecting reflection posts - lane recognition on country roads," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 304–309.

[6] R. Aufrère, V. Marion, J. Laneurit, C. Lewandowski, J. Morillon, and R. Chapuis, "Road sides recognition in non-structured environments by vision," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 329–334.

[7] K. Fintzel, R. Bendahan, C. Vestri, S. Bougnoux, and T. Kakinami, "3-D parking assistant system," in *Proc. IEEE Intelligent Vehicles Symp. 2004*, Parma, Italy, 2004, pp. 881–886.

[8] S. Funck, N. Möhler, and W. Oertel, "Determining car-park occupancy from single images," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 325–328.

[9] N. Kaempchen, U. Franke, and R. Ott, "Stereo vision based pose estimation of parking lots using 3-D vehicle models," presented at the IEEE Intelligent Vehicles Symp., Versailles, France, 2002.

[10] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Paetzold, and C. Wohler, "Autonomous driving goes downtown," *IEEE Intell. Syst.*, vol. 13, no. 6, pp. 40–48, Jun. 1999.

[11] Z. Eizad and L. Vlacic, "A control algorithm and vehicle model for stop and go cruise control," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 401–406.

[12] M. M. Minderhoud and F. Zuurbier, "Empirical data on driving behaviour in stop-and-go traffic," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 676–681.

[13] A. de la Escalera, J. M. Armingol, J. M. Pastor, and F. J. Rodriguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 2, pp. 57–68, Jun. 2004.

[14] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 566–571.

[15] F. Lindner, U. Kressel, and Kaelberer, "Robust recognition of traffic signals," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 49–53.

[16] B. Steux, C. Laurgeau, L. Salesse, and D. Wautier, "Fade: A vehicle detection and tracking system featuring monocular color vision and radar data fusion," presented at the IEEE Intelligent Vehicles Symp., Versailles, France, 2002.

[17] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta, "Artificial vision in road vehicles," *Proc. IEEE*, vol. 90, no. 7, pp. 1258–1271, Jul. 2002.

[18] T. Xiong and C. Debrunner, "Stochastic car tracking with line- and color-based features," *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 4, pp. 324–328, Dec. 2004.

[19] G. S. K. Fung, N. H. C. Yung, and G. K. H. Pang, "Vehicle shape approximation from motion for visual traffic surveillance," in *Proc. IEEE Conf. Intelligent Transportation Systems*, Oakland, CA, 2001, pp. 610–615.

[20] C. Demonceaux, A. Potelle, and D. Kachi-Akkouche, "Obstacle detection in a road scene based on motion analysis," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1649–1656, Nov. 2004.

[21] C. Hoffmann, T. Dang, and C. Stiller, "Vehicle detection fusing 2D visual features," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 280–285.

[22] T. K. ten Kate, M. B. van Leewen, S. E. Moro-Ellenberger, B. J. F. Driessen, A. H. G. Versluis, and F. C. A. Groen, "Mid-range and distant vehicle detection with a mobile camera," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, 2004, pp. 72–77.
[23] R. Gregor, M. Lützeler, M. Pellkofer, K. Siedersberger, and E. Dickmanns, "Ems-vision: A perceptual system for autonomous vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 3, no. 1, pp. 48–59, Mar. 2002.
[24] M. Bertozzi, A. Broggi, A. Fascioli, and S. Nichele, "Stereo vision-based vehicle detection," presented at the IEEE Intelligent Vehicles Symp., Dearborn, MI, 2000.
[25] A. Broggi, M. Bertozzi, A. Fascioli, C. Guarino Lo Bianco, and A. Piazzi, "Visual perception of obstacles and vehicles for platooning," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, no. 3, pp. 164–176, Sep. 2000.
[26] P. Miché and R. Debrie, "Fast and self-adaptive image segmentation using extended declivity," *Ann. Telecommun.*, vol. 50, no. 3–4, 1995.
[27] A. Bensrhair, P. Miché, and R. Debrie, "Fast and automatic stereo vision matching algorithm based on dynamic programming method," *Pattern Recognit. Lett.*, vol. 17, pp. 457–466, 1996.
[28] G. Toulminet, S. Mousset, and A. Bensrhair, "Fast and accurate stereo vision-based estimation of 3D position and axial motion of road obstacles," *Int. J. Image Graph. Special Issue 3-D Object Recognition*, vol. 4, no. 1, pp. 99–126, 2004.
[29] T. O. Binford, *Stereo vision: Complexity and Constraints* (in M. Brady and R. Paul, Eds., Robotics Research 1). Cambridge, U.K.: MIT Press, 1984, pp. 475–487.
[30] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation," presented at the IEEE Intelligent Vehicles Symp., Versailles, France, Jun. 2002.
[31] A. Bensrhair, M. Bertozzi, A. Broggi, A. Fascioli, S. Mousset, and G. Toulminet, "Stereo vision-based feature extraction for vehicle detection," presented at the IEEE Intelligent Vehicles Symp., Versailles, France, Jun. 2002.
[32] G. I. M. Salama, "Monocular and Binocular Visual Tracking," Ph.D. dissertation, Dept. Elect. Comput. Eng., Virginia Polytechnic Inst. State Univ., Blacksburg, 1999.
[33] L. Andreone, F. Tango, U. Scheunert, H. Cramer, G. Wanielik, and A. Amditi, "A new driving supporting system, integrating an infrared camera and an anti-collision micro-wave radar: The euclide project," presented at the IEEE Intelligent Vehicles Symp., Versailles, France, Jun. 2002.

Gwenaëlle Toulminet received the M.Eng. degree in electrical engineering from the Superior School, Rouen, France, in 1999, and the Ph.D. degree in physics from the National Institute of Applied Sciences (INSA), Rouen, in 2002. Her Ph.D. work focused on the extraction of three-dimensional edges of obstacles using stereo vision for driving assistance.

She is currently an Assistant Professor with the Department of Information, Systems, and Architecture of the INSA, University of Rouen, and a member of the Vision Systems Division of the Laboratory of Perception, Systems, Information (PSI). Her research focuses on intelligent vehicles, sensors, and stereo vision analysis.

Massimo Bertozzi received the Dr.Eng. degree in electronic engineering and the Ph.D. degree in information technology, both from the Università di Parma, Parma, Italy, in 1994 and 1997, respectively. His Ph.D. dissertation was on the implementation of simulation of Petri nets on the CM-2 massive parallel architecture.

From 1994 to 1997, he Chaired the local IEEE student branch. He is currently a Researcher in the Dipartimento di Ingegneria dell'Informazione, Università di Parma. His research interests focus mainly on the application of image processing to real-time systems and vehicle guidance.

Stéphane Mousset graduated from the École Normale Supérieure of Cachan and received the Ph.D. degree in computer science from the University of Rouen, Rouen, France, in 1997. His Ph.D. dissertation focused on the estimation of axial motion using a stereo vision system.

He is currently an Assistant Professor at the Laboratory of Perception, Systems, Information (PSI), National Institute of Applied Sciences (INSA), University of Rouen. He also teaches at the Technology University Institute of Rouen. His research interests include stereo vision analysis, road application, driving assistance systems, and sensors.

Abdelaziz Bensrhair received the M.S. degree in electrical engineering and the Ph.D. degree in computer science from the University of Rouen, Rouen, France, in in 1989 and 1992, respectively.

From 1992 to 1999, he was an Assistant Professor with the Department of Physic and Instrumentation, University of Rouen. He is currently a Professor in the Department of Information, Systems, and Architecture, National Institute of Applied Sciences, and the Co-Director of the Laboratory Perception System Information (PSI), both at the University of Rouen.

Alberto Broggi (SM'89) received the Dr.Eng. degree in electronic engineering and the Ph.D. degree in information technology from the Università di Parma, Parma, Italy, in 1990 and 1994, respectively.

From 1994 to 1998, he was an Associate Researcher with the Dipartimento di Ingegneria dell Informazione, Università di Parma, and from 1998 to 2001, an Associate Professor of artificial intelligence with the Dipartimento di Informatica e Sistemistics. Since 2001, he has been a Professor of computer science. His research interests include real-time computer vision approaches for the navigation of unmanned vehicles and the development of low-cost computer systems to be used in autonomous agents. He has authored more than 120 refereed publications in international journals, book chapters, and conference proceedings, and delivered invited talks in many international conferences.

Dr. Broggi is the Editor-in-Chief of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and a member of the IEEE ITS Council Executive Committee. He has been an Editor on the subject of ITS for the *IEEE Intelligent Systems Magazine* since 1999.