HAL
open science

Obstacle Detection Based on Fusion Between Stereovision and 2D Laser Scanner

Raphaël Labayrade, Dominique Gruyer, Cyril Royere, Mathias Perrollaz, Didier Aubert

To cite this version:

Raphaël Labayrade, Dominique Gruyer, Cyril Royere, Mathias Perrollaz, Didier Aubert. Obstacle Detection Based on Fusion Between Stereovision and 2D Laser Scanner. Sascha Kolski. Mobile Robots: Perception & Navigation, Pro Literatur Verlag, 2007, 3-86611-283-1. hal-00683758

HAL Id: hal-00683758

https://inria.hal.science/hal-00683758

Submitted on 29 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

5

Obstacle Detection Based on Fusion Between Stereovision and 2D Laser Scanner

Raphaël Labayrade, Dominique Gruyer, Cyril Royere,
Mathias Perrollaz, Didier Aubert
LIVIC (INRETS-LCPC)
France

1. Introduction

Obstacle detection is an essential task for mobile robots. This subject has been investigated for many years by researchers and a lot of obstacle detection systems have been proposed so far. Yet designing an accurate and totally robust and reliable system remains a challenging task, above all in outdoor environments. The DARPA Grand Challenge (Darpa, 2005) proposed efficient systems based on sensors redundancy, but these systems are expensive since they include a large set of sensors and computers: one can not consider to implement such systems on low cost robots. Thus, a new challenge is to reduce the number of sensors used while maintaining a high level of performances. Then, many applications will become possible, such as Advance Driving Assistance Systems (ADAS) in the context of Intelligent Transportation Systems (ITS).

Thus, the purpose of this chapter is to present new techniques and tools to design an accurate, robust and reliable obstacle detection system in outdoor environments based on a minimal number of sensors. So far, experiments and assessments of already developed systems show that using a single sensor is not enough to meet the requirements: at least two complementary sensors are needed. In this chapter a stereovision sensor and a 2D laser scanner are considered.

In Section 2, the ITS background under which the proposed approaches have been developed is introduced. The remaining of the chapter is dedicated to technical aspects. Section 3 deals with the stereovision framework: it is based on a new technique (the so-called "v-disparity" approach) that efficiently tackles most of the problems usually met when using stereovision-based algorithms for detecting obstacles. This technique makes few assumptions about the environment and allows a generic detection of any kind of obstacles; it is robust against adverse lightning and meteorological conditions and presents a low sensitivity towards false matches. Target generation and characterization are detailed. Section 4 focus on the laser scanner raw data processing performed to generate targets from lasers points and estimate their positions, sizes and orientations. Once targets have been generated, a multi-objects association algorithm is needed to estimate the dynamic state of the objects and to monitor appearance and disappearance of tracks. Section 5 intends to present such an algorithm based on the Dempster-Shaffer belief theory. Section 6 is about fusion between stereovision and laser scanner. Different possible fusion schemes are introduced and discussed. Section 7 is dedicated to experimental results. Eventually, section 8 deals with trends and future research.

Open Access Database www.i-techonline.com

Source: Mobile Robots: Perception & Navigation, Book edited by: Sascha Kolski, ISBN 3-86611-283-1, pp. 704, February 2007, Plv/ARS, Germany

2. Intelligent Transportation Systems Background

In the context of Intelligent Transportation Systems and Advanced Driving Assistance Systems (ADAS), onboard obstacle detection is a critical task. It must be performed in real time, robustly and accurately, without any false alarm and with a very low (ideally nil) detection failure rate. First, obstacles must be detected and positioned in space; additional information such as height, width and depth can be interesting in order to classify obstacles (pedestrian, car, truck, motorbike, etc.) and predict their dynamic evolution. Many applications aimed at improving road safety could be designed on the basis of such a reliable perception system: Adaptative Cruise Control (ACC), Stop'n'Go, Emergency braking, Collision Mitigation. Various operating modes can be introduced for any of these applications, from the *instrumented mode* that only informs the driver of the presence and position of obstacles, to the *regulated mode* that take control of the vehicle through activators (brake, throttle, steering wheel). The *warning mode* is an intermediate interesting mode that warn the driver of an hazard and is intended to alert the driver in advance to start a manoeuver before the accident occurs.

Various sensors can be used to perform obstacle detection. 2D laser scanner (Mendes 2004) provides centimetric positioning but some false alarms can occur because of the dynamic pitching of the vehicle (from time to time, the laser plane collides with the ground surface and then laser points should not be considered to belong to an obstacle). Moreover, width and depth (when the side of the object is visible) of obstacles can be estimated but height cannot. Stereovision can also be used for obstacle detection (Bertozzi, 1998 ; Koller, 1994 ; Franke, 2000 ; Williamson, 1998). Using stereovision, height and width of obstacles can be evaluated. The pitch value can also be estimated. However, positioning and width evaluation are less precise than the ones provided by laser scanner.

Fusion algorithms have been proposed to detect obstacles using various sensors at the same time (Gavrila, 2001 ; Mobus, 2004 ; Steux, 2002). The remaining of the chapter presents tools designed to perform fusion between 2D laser scanner and stereovision that takes into account their complementary features.

3. Stereovision Framework

3.1 The "v-disparity" framework

This section deals with the stereovision framework. Firstly a modeling of the stereo sensor, of the ground and of the obstacles is presented. Secondly details about a possible implementation are given.

Modeling of the stereo sensor: The two image planes of the stereo sensor are supposed to belong to the same plane and are at the same height above the ground (see Fig. 1). This camera geometry means that the epipolar lines are parallel. The parameters shown on Fig. 1 are:

- θ s the angle between the optical axis of the cameras and the horizontal,
- h is the height of the cameras above the ground,
- b is the distance between the cameras (i.e. the stereoscopic base).

(R_a) is the absolute coordinate system, and O_a lies on the ground. In the camera coordinate system (R_{ci}) (i equals l *(left)* or r *(right)*), the position of a point in the image plane is given by its coordinates (u_i,v_i). The image coordinates of the projection of the optical center will be denoted by (u_0,v_0), assumed to be at the center of the image. The intrinsic parameters of the camera are f (the focal length of the lens), t_u and t_v (the size of pixels in u and v). We also use $a_u=f/t_u$ and $a_v=f/t_v$. With the cameras in current use we can make the following approximation: $\alpha_u \approx \alpha_v = \alpha$.

Using the pin-hole camera model, a projection on the image plane of a point P(X,Y,Z) in (R_a) is expressed by:

$$\begin{cases} u = \alpha \dfrac{X}{Z} + u_0 \\ v = \alpha \dfrac{Y}{Z} + v_0 \end{cases} \tag{1}$$

On the basis of Fig. 1, the transformation from the absolute coordinate system to the right camera coordinate system is achieved by the combination of a vector translation ($\vec{t} = -h\vec{Y}$ and $\vec{b} = (b/2)\vec{X}$) and a rotation around $\vec{X}$, by an angle of $-\theta$. The combination of a vector translation ($\vec{t} = -h\vec{Y}$ and $\vec{b} = -(b/2)\vec{X}$) and a rotation around $\vec{X}$, by an angle of $-\theta$ is the transformation from the absolute coordinate system to the left camera coordinate system.

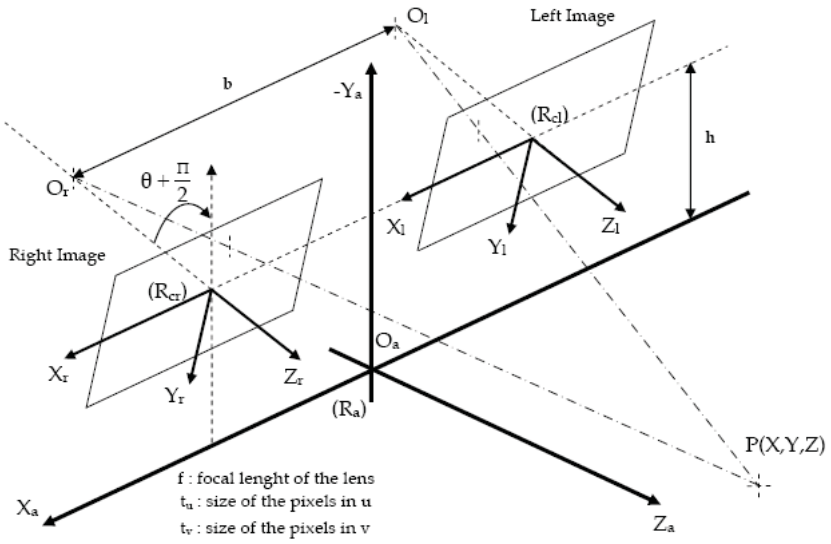

Fig. 1. The stereoscopic sensor and used coordinate systems.

Since the epipolar lines are parallel, the ordinate of the projection of the point P on the left or right image is $v_r = v_l = v$, where:

$$v = \frac{\left[v_0 \sin\theta + \alpha\cos\theta\right](Y+h) + \left[v_0\cos\theta - \alpha\sin\theta\right]Z}{(Y+h)\sin\theta + Z\cos\theta} \tag{2}$$

Moreover, the disparity Δ of the point P is:

$$\Delta = u_l - u_r = \frac{\alpha b}{(Y+h)\sin\theta + Z\cos\theta} \tag{3}$$

Modeling of the ground: In what follows the ground is modeled as a plane with equation: $Z=aY+d$. If the ground is horizontal, the plane to consider is the plane with equation $Y=0$.
Modeling of the obstacles: In what follows any obstacle is characterized by a vertical plane with equation $Z = d$.
Thus, all planes of interest (ground and obstacles) can be characterized by a single equation: $Z = aY+d$.

The image of planes of interest in the "v-disparity" image: From (2) and (3), the plane with the equation $Z = aY+d$ in (R_a) is projected along the straight line of equation (1) in the "v-disparity" image:

$$\Delta_M = \frac{b}{ah-d}(v-v_0)(a\cos\theta+\sin\theta)+\frac{b}{ah-d}\alpha(a\sin\theta+\cos\theta) \tag{4}$$

N.B.: when $a = 0$ in equation (1), the equation for the projection of the vertical plane with the equation $Z = d$ is obtained:

$$\Delta_M = \frac{b}{d}(v_0-v)\sin\theta+\frac{b}{d}\alpha\cos\theta \tag{5}$$

When $a\rightarrow\infty$, the equation of the projection of the horizontal plane with the equation $Y = 0$ is obtained:

$$\Delta_M = \frac{b}{h}(v-v_0)\cos\theta+\frac{b}{h}\alpha\sin\theta \tag{6}$$

Thus, planes of interest are all projected as straight lines in the "v-disparity" image.
The "v-disparity" framework can be generalized to extract planes presenting roll with respect to the stereoscopic sensor. This extension allows to extract any plane in the scene. More details are given in (Labayrade, 2003 a).

3.2 Exemple of implementation

"v-disparity" image construction: A disparity map is supposed to have been computed from the stereo image pair (see Fig. 2 left). This disparity map is computed taking into account the epipolar geometry; for instance the primitives used can be horizontal local maxima of the gradient; matching can be local and based on normalized correlation around the local maxima (in order to obtain additional robustness with respect to global illumination changes).
The "v-disparity" image is line by line the histogram of the occurring disparities (see Fig. 2 right). In what follows it will be denoted as $\boldsymbol{I_{v\Delta}}$.
Case of a flat-earth ground geometry: robust determination of the plane of the ground: Since the obstacles are defined as objects located above the ground surface, the corresponding surface must be estimated before performing obstacle detection.

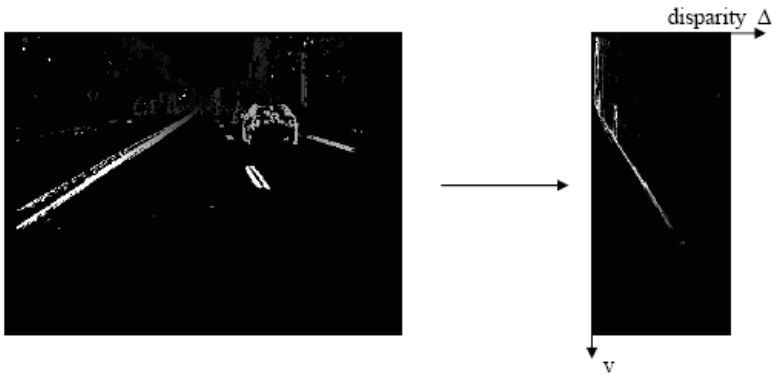

Fig. 2. Construction of the grey level "v-disparity" image from the disparity map. All the pixels from the disparity map are accumulated along scanning lines.

When the ground is planar, with for instance the following mean parameter values of the stereo sensor:

- θ = 8.5°,
- h = 1.4 m,
- b = 1 m,

the plane of the ground is projected in $I_{v\Delta}$ as a straight line with mean slope 0.70. The longitudinal profile of the ground is therefore a straight line in $I_{v\Delta}$. Robust detection of this straight line can be achieved by applying a robust 2D processing to $I_{v\Delta}$. The Hough transform can be used for example.

Case of a non flat-earth ground geometry: The ground is modeled as a succession of parts of planes. As a matter of fact, its projection in $I_{v\Delta}$ is a piecewise linear curve. Computing the longitudinal profile of the ground is then a question of extracting a piecewise linear curve in $I_{v\Delta}$. Any robust 2D processing can be used. For instance it is still possible to use the Hough Transform. The k highest Hough Transform values are retained (k can be taken equal to 5) and correspond to k straight lines in $I_{v\Delta}$. The piecewise linear curve researched is either the upper (when approaching a downhill gradient) or the lower (when approaching a uphill gradient) envelope of the family of the k straight lines generated. To choose between these two envelope, the following process ca be performed. $I_{v\Delta}$ is investigated along both curves extracted and a score is computed for each: for each pixel on the curve, the corresponding grey level in $I_{v\Delta}$ is accumulated. The curve is chosen with respect to the best score obtained. Fig. 3 shows how this curve is extracted. From left to right the following images are presented: an image of the stereo pair corresponding to a non flat ground geometry when approaching an uphill gradient; the corresponding $I_{v\Delta}$ image; the associated Hough Transform image (the white rectangle show the research area of the k highest values); the set of the k straight lines generated; the computed envelopes, and the resulting ground profile extracted.

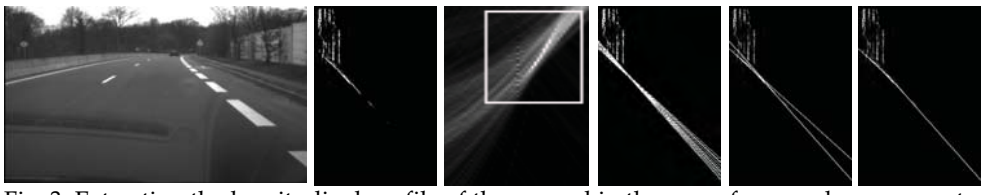

Fig. 3. Extracting the longitudinal profile of the ground in the case of a non planar geometry (see in text for details).

Evaluation of the obstacle position and height: With the mean parameter values of the stereo sensor given above for example, the plane of an obstacle is projected in $I_{v\Delta}$ as a straight line nearly vertical above the previously extracted ground surface. Thus, the extraction of vertical straight lines in $I_{v\Delta}$ is equivalent to the detection of obstacles. In this purpose, an histogram that accumulates all the grey values of the pixels for each column of the $I_{v\Delta}$ image can be built; then maxima in this histogram are looked for. It is then possible to compute the ordinate of the contact point between the obstacle and the ground surface (intersection between the ground profile and the obstacle line in the "v-disparity" image, see Fig. 4). The distance D between the vehicle and the obstacle is then given by:

$$D = \frac{b(\alpha \cos\theta - (v_r - v_0)\sin\theta)}{\Delta} \tag{7}$$

where v_r is the ordinate of the ground-obstacle contact line in the image.
The height of the obstacle is given by the height of the straight line segment in the "v-disparity" image (see Fig. 4). The lateral position and left and right border of the obstacle can be estimated by similar processing in the "u-disparity" image (The "u-disparity" image is column by column the histogram of the occurring disparities). Thus, a target detected by stereovision is characterized by its (X,Z) coordinates, its height and its width.
Moreover, a dynamic estimation of the sensor pitch θ can be obtained from the horizon line, at each fame processed:

$$\theta = \arctan\left(\frac{v_0 - v_{hor}}{\alpha}\right) \tag{8}$$

where v_{hor} is the ordinate of the horizon line. Since the horizon line belongs to the ground surface and is located at infinite distance (which corresponds to nil disparity), v_{hor} is the ordinate of the point located on the ground profile for a nil disparity (see Fig. 4).

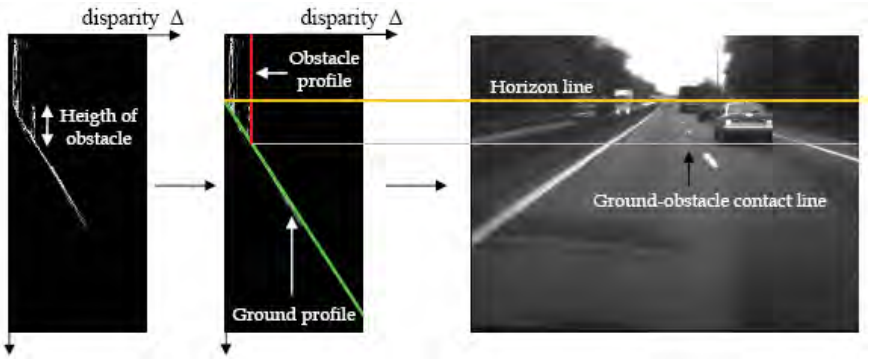

Fig. 4. Extracting obstacles and deducing obstacle-ground contact line and horizon line.

Practical good properties of the algorithm: It should be noticed that the algorithm is able to detect any kind of obstacles. Furthermore, all the information in the disparity map is exploited and the accumulation performed increases the density of the alignments in $I_{v\Delta}$. Any matching error that occur when the disparity map is computed causes few problems as the probability that the points involved will generate coincidental alignments in $I_{v\Delta}$ is low. As a matter of fact, the algorithm is able to perform accurate detection even in the event of a lot of noise or matching errors, and when there is only a few correct matches or a few amount of correct data in the images: in particular in night condition when the majority of the pixels are very dark. Eventually, the algorithm works whatever the process used for computing the disparity map (see (Scharstein, 2001)) or for processing the "v-disparity" image. Eventually, as detailed in (Labayrade, 2003 b), it is possible in a two-stages process to improve the disparity map and remove a lot of false matches.

4. Laser Scanner Raw Data Processing

The 2D laser scanner provides a set of laser impacts on the scanned plane: each laser point is characterized by an incidence angle and a distance which corresponds to the distance of the nearest object in this direction (see Fig. 6). From these data, a set of clusters must be built, each cluster corresponding to an object in the observed scene.

Initially, the first laser impact defines the first cluster. For all other laser points, the goal is to know if they are a membership of the existent cluster or if they belong to a new cluster. In the literature, a great set of distance functions can be found for this purpose. The chosen distance $D_{i,j}$ must comply with the following criteria

Firstly, this function $D_{i,j}$ must give a result scaled between 0 and 1. The value 0 indicates that the measurement i is a member of the cluster j,

Secondly, the result must be above 1 if the measurement is out of the cluster j,

Finally, this distance must have the properties of the distance functions.

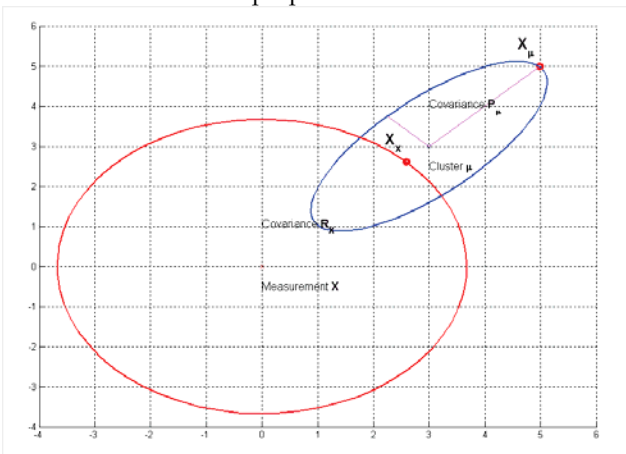

Fig. 5. Clustering of a measurement.

The distance function must also use both cluster and measurement covariance matrices. Basically, the chosen function computes an inner distance with a normalisation part build from the sum of the outer distances of a cluster and a measurement. Only the outer distance uses the covariance matrix:

$$D_{i,j} = \frac{\sqrt{(X-\mu)(X-\mu)^t}}{\sqrt{(X_\mu - \mu)} + \sqrt{(X_X - X)}} \tag{9}$$

In the normalisation part, the point X_μ represent the border point of a cluster i (centre μ). This point is localised on the straight line between the cluster i (centre μ) and the measurement j (centre X). The same border measurement is used with the cluster. The computation of X_μ and X_X is made with the covariance matrices R_x and P_μ. P_μ and R_x are respectively the cluster covariance matrix and the measurement covariance matrix. The measurement covariance matrix is given from its polar covariance representation (Blackman 1999) with ρ_0 the distance and θ_0 the angle:

$$R_x = \begin{bmatrix} \sigma^2_{x_0} & \sigma^2_{x_0 y_0} \\ \sigma^2_{x_0 y_0} & \sigma^2_{y_0} \end{bmatrix} \tag{10}$$

where, using a first order expansion:

$$\begin{aligned} \sigma^2_{x_0} &= \sigma^2_{\rho_0} \cos^2\theta_0 + \sigma^2_{\theta_0} \rho_0{}^2 \sin^2\theta_0 \\ \sigma^2_{y_0} &= \sigma^2_{\rho_0} \sin^2\theta_0 + \sigma^2_{\theta_0} \rho_0{}^2 \cos^2\theta_0 \end{aligned} \tag{11}$$

$$\sigma^2_{x_0 y_0} = \frac{1}{2}\sin 2\theta_0 \left[\sigma^2_{\rho_0} - \sigma^2_{\theta_0}\rho_0{}^2\right]$$

$\sigma^2_{\rho_0}$ and $\sigma^2_{\theta_0}$ are the variances in both distance and angle of each measurement provided by the laser scanner. From this covariance matrix, the eigenvalues σ and the eigenvectors V are extracted. A set of equations for both ellipsoid cluster and measurement modelling and line between the cluster centre μ and the laser measurement X is then deduced:

$$\begin{aligned} x &= V_{11}\sqrt{\sigma_1{}^2}\cos\Psi + V_{12}\sqrt{\sigma_2{}^2}\sin\Psi \\ y &= V_{21}\sqrt{\sigma_1{}^2}\cos\Psi + V_{22}\sqrt{\sigma_2{}^2}\sin\Psi \\ y &= ax + b \end{aligned} \tag{12}$$

The solution of this set of equations gives:

$$\Psi = \arctan\left(\frac{-\sqrt{\sigma_1{}^2}\left[V_{2,1} - aV_{1,1}\right]}{\sqrt{\sigma_2{}^2}\left[V_{2,2} - aV_{1,2}\right]}\right)$$

$$\Psi \in \left[\frac{-\pi}{2}, \frac{\pi}{2}\right] \quad \text{with} \tag{13}$$

From (13), two solutions are possible:

$$X_\mu = P_\mu\sqrt{\sigma^2}\begin{bmatrix}\cos\Psi \\ \sin\Psi\end{bmatrix}$$

$$X_\mu = P_\mu\sqrt{\sigma^2}\begin{bmatrix}\cos\Psi + \pi \\ \sin\Psi + \pi\end{bmatrix} \tag{14}$$

and

Then equation (9) is used with X_μ to know if a laser point belongs to a cluster.

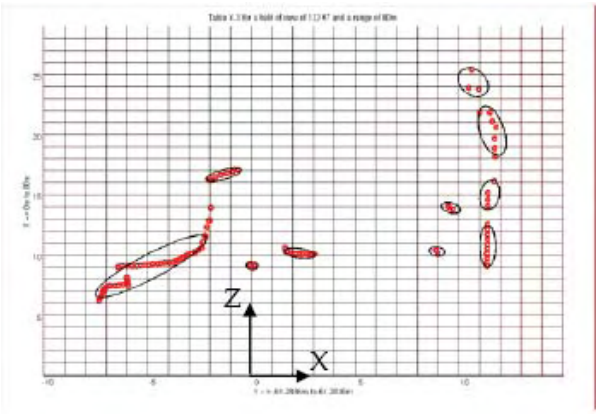

Fig. 6. Example of a result of autonomous clustering (a laser point is symbolized by a little circle, and a cluster is symbolized by a black ellipse).

Fig. 5 gives a visual interpretation of the used distance for the clustering process. Fig. 6 gives an example of a result of autonomous clustering from laser scanner data. Each cluster is characterized by its position, its orientation, and its size along the two axes.

5. Multi-Objects Association

Once targets have been generated from stereovision or from laser scanner, a multi-objects association algorithm is needed to estimate the dynamic state of the obstacles and to monitor appearance and disappearance of tracks. The position of previously perceived objects is predicted at the current time using Kalman Filtering. These predicted objects are already known objects and will be denoted in what follows by Y_j. Perceived objects at the current time will be denoted by X_i. The proposed multi-objects association algorithm is based on the belief theory introduced by Shafer (Shafer, 1976).

5.1 Generalities

In a general framework, the problem consist in identifying an object designated by a generic variable X among a set of hypotheses Yi. One of these hypotheses is supposed to be the solution. The current problem consists in associating perceived objects X_i to known objects Y_j. Belief theory allows to assess the veracity of P_i propositions representing the matching of the different objects.
A magnitude allowing the characterization of a proposition must be defined. This magnitude is the basic belief assignment (mass $m_\Theta(\ )$) defined on [0,1]. This mass is very close to the probabilistic mass with the difference that it is not only shared on single elements but on all elements of the definition referential $2^\Theta = \{ A/A \subseteq \Theta\} = \{\varnothing, Y_1, Y_2, ..., Y_n, Y_1 \cup Y_2, \ldots, \Theta\}$. This referential is built through the frame of discernment $\Theta = \{Y_1, Y_2, \cdots, Y_n\}$, which regroups all admissible hypotheses, that in addition must be exclusive. ($Y_i \cap Y_j = \varnothing$, $\forall$ i $\neq$ j). This distribution is a function of the knowledge about the source to model. The whole mass obtained is called "*basic belief assignment*". The sum of these masses is equal to 1 and the mass corresponding to the impossible case $m(\varnothing)$ must be equal to 0.

5.2. Generalized combination and multi-objects association

In order to succeed in generalizing the Dempster combination rule and thus reducing its combinatorial complexity, the reference frame of definition is limited with the constraint that a perceived object can be connected with one and only one known object.
For example, for a detected object, in order to associate among three known objects, frame of discernment is:

$$\Theta = \{Y_1, Y_2, Y_3, *\}$$
$$\text{where } Y_i \text{ means that "X and } Y_i \text{ are supposed to be the same object"}$$

In order to be sure that the frame of discernment is really exhaustive, a last hypothesis noted "*" is added. This one can be interpreted as an association of a perceived object with any of the known objects. In fact each Y_j represents a local view of the world and the "*" represents the rest of the world. In this context, "*" means well: "*an object is associated with nothing in the local knowledge set*".
The total ignorance is represented by the hypothesis Θ which is the disjunction of all the hypotheses of the frame of discernment. The conflict is given by the hypothesis $\varnothing$

which corresponds to the empty set (since the hypotheses are exclusive, their intersections is empty).

A distribution of masses made up of the masses is obtained:

$m_{i,j}(Y_j)$: mass associated with the proposition « X_i and Y_j. are supposed to be the same object »,

$m_{i,j}(\overline{Y}_j)$: mass associated with the proposition « X_i and Y_j. are not supposed to be the same object »,

$m_{i,j}(\Theta_{i,j})$: mass representing ignorance,

$m_{i,.}(*)$: mass representing the reject: X_i is in relation with nothing.

In this mass distribution, the first index i denotes the processed perceived objects and the second index j the known objects (predictions). If one index is replaced by a dot, then the mass is applied to all perceived or known objects according to the location of this dot.

Moreover, if an iterative combination is used, the mass $m_{i,.}(*)$ is not part of the initial mass set and appears only after the first combination. It replaces the conjunction of the combined masses $m_{i,j}(\overline{Y}_j)$. By observing the behaviour of the iterative combination with n mass sets, a general behaviour can be seen which enables to express the final mass set according to the initial mass sets. This enables to compute directly the final masses without any recurrent stage. For the construction of these combination rules, the work and a first formalism given in (Rombaut, 1998) is used. The use of an initial mass set generator using the strong hypothesis: *"an object can not be in the same time associated and not associated to another object"* allows to obtain new rules. These rules firstly reduce the influence of the conflict (the combination of two identical mass sets will not produce a conflict) and secondly the complexity of the combination. The rules become:

$$m_{i,.}(Y_i) = 0 \quad \text{if } H_1 \tag{15}$$

$$m_{i,.}(Y_j) = K_{i,.} \cdot m_{i,j}(Y_j) \cdot E_1 \quad \text{if } H_2 \tag{16}$$

$$\text{with} \quad \begin{cases} E_1 = \prod_{\substack{k=1\cdots n \\ k\neq j}} \left(1 - m_{i,k}(Y_k)\right) \\ H_1 \Leftrightarrow \exists j, m_{i,j}(Y_j) = 0 \\ H_2 \Leftrightarrow \forall j, m_{i,j}(Y_j) \neq 0 \end{cases} \tag{17}$$

$$m_{i,.}(*) = 0 \quad \text{if } H_1 \tag{18}$$

$$m_{i,.}(*) = K_{i,.} \prod_{j=1\cdots n} m_{i,j}(\overline{Y}_j) \quad \text{if } H_2 \tag{19}$$

$$\text{with} \quad \begin{cases} H_1 \Leftrightarrow \exists j, m_{i,j}(Y_j) \neq 0 \\ H_2 \Leftrightarrow \forall j, m_{i,j}(Y_j) = 0 \end{cases} \tag{20}$$

$$m_{i,.}(\theta) = K_{i,.} \cdot E_1 . E_2 \quad \text{if } H_1 \tag{21}$$

$$m_{i,.}(\theta) = K_{i,.} \cdot E_1 \quad \text{if } H_2 \tag{22}$$

$$m_{i,.}(\theta) = K_{i,.} \cdot (E_3 - E_4) \quad \text{if } H_3 \tag{23}$$

$$\text{with}\quad \begin{cases} E_1 = \prod_{l=1\cdots n} m_{i,l}(\theta_{i,l}) \\ E_2 = \prod_{\substack{j=1\cdots n \\ j\neq l}} \left(m_{i,j}(\theta_{i,j}) - m_{i,j}(\bar{Y}_j)\right) \\ E_3 = \prod_{j=1\cdots n} \left(m_{i,j}(\theta_{i,j}) - m_{i,j}(\bar{Y}_j)\right) \\ E_4 = \prod_{j=1\cdots n} m_{i,j}(\bar{Y}_j) \\ H_1 \Leftrightarrow \exists l \in [1\cdots n], \quad m_{i,l}(Y_l) \neq 0 \\ H_2 \Leftrightarrow \forall j \in [1\cdots n], \quad m_{i,j}(Y_j) \neq 0 \\ H_3 \Leftrightarrow \forall j \in [1\cdots n], \quad m_{i,j}(Y_j) = 0 \end{cases} \tag{24}$$

$$K_{i,.} = \frac{1}{1 - m_{i,.}(\varnothing)} = \frac{1}{E_1 + E_2} \tag{25}$$

$$\text{with}\quad \begin{cases} E_1 = \prod_{j=1\cdots n} \left(1 - m_{i,j}(Y_j)\right) \\ E_2 = \sum_{j=1\cdots n} m_{i,j}(Y_j) \cdot \prod_{\substack{k=1\cdots n \\ k\neq j}} \left(1 - m_{i,k}(Y_k)\right) \end{cases} \tag{26}$$

From each mass set, two matrices $\mathrm{M}^{cr}_{i,.}$ and $\mathrm{M}^{cr}_{.,j}$ are built which give the belief that a perceived object is associated with a known object and conversely. The sum of the elements of each column is equal to 1 because of the re-normalization.
The resulting frames of discernment are:

$$\Theta_{.,j} = \{Y_{1,j}, Y_{2,j}, \cdots, Y_{n,j}, Y_{*,j}\}$$

$$\text{and}\quad \Theta_{i,.} = \{X_{i,1}, X_{i,2}, \cdots, X_{i,m}, X_{i,*}\}$$

The first index represents the perceived object and the second index the known object. The index "*" is the notion of "emptiness" or more explicitly "*nothing*". With this hypothesis, it can be deduced if an object has appeared or disappeared.
The following stage consists in establishing the best decision on association using these two matrices obtained previously. Since a referential of definition built with singleton hypotheses is used, except for Θ and *, the use of credibilistic measure will not add any useful information. This redistribution will simply reinforce the fact that a perceived object is really in relation with a known object. This is why the maximum of belief on each column of the two belief matrices is used as the decision criterion:

$$d(Y_{i,.}) = \underset{j}{Max}[M^{Cr}_{i,j}] \tag{27}$$

This rule answers the question "*which is the known object Y_j in relation with the perceived object X_i*"? The same rule is available for the known objects:

$$d(X_{.,j}) = \underset{i}{Max}[M^{Cr}_{i,j}] \tag{28}$$

Unfortunately, a problem appears when the decision obtained from a matrix is ambiguous (this ambiguity quantifies the duality and the uncertainty of a relation) or when the decisions between the two belief matrices are in conflict (this conflict represents antagonism between two relations resulting each one from a different belief matrix). Both problems of

conflicts and ambiguities are solved by using an assignment algorithm known under the name of the Hungarian algorithm (Kuhn, 1955 ; Ahuja, 1993). This algorithm has the advantage of ensuring that the decision taken is not " *good*" but "*the best*". By the "*best*", we mean that if a known object has some defective or poor sensor to perceive it, then it is unlikely to know what this object corresponds to, and therefore ensuring that the association is good is a difficult task. But among all the available possibilities, we must certify that the decision is the "*best*" of all possible decisions.
Once the multi-objects association has been performed, the Kalman filter associated to each object is updated using the new position of the object, and so the dynamic state of each object is estimated.

6. Fusion

So far, the chapter has described the way in which the two sensors (stereovision and 2D laser scanner) are independently used to perform obstacle detection. Tables 1 and 2 remind the advantages and drawbacks of each sensor.

	Detection range	**Obstacle position accuracy**	**Frequency**	**False alarms occurrence**
Stereovision	Short to medium range (up to 50 m).	Decreases when the obstacle distance increases.	Video frame rate.	When the disparity map is of poor quality.
Laser scanner	Medium to long range (up to 120 m).	Usually a few cm. Independent to the obstacle distance.	Usually higher than the stereovision.	When the laser plane collides with the ground surface.

Table 1. Features of the stereovision and 2D laser scanner sensors.

	Detection failure occurrence	**Ground geometry**	**Width, height, depth, orientation**
Stereovision	Adverse lighting conditions, very low obstacles (<30 cm).	Provide ground geometry, including roll, pitch, longitudinal profile.	Provide width and height
Laser Scanner	When the laser plane passes above obstacle.	Cannot provide ground geometry.	Provide orientation, width and depth (when the side of the obstacle is visible)

Table 2. Features of the stereovision and 2D laser scanner sensors (continued).

From Tables 1 and 2, some remarks can be made. Laser scanner and stereovision are complementary sensors: laser scanner is more accurate but a lot of false alarms can occur when the laser plane collides with the ground (see Fig. 7); stereovision is less accurate but can distinguish the ground from an obstacle, because it can provide a 3D modelling of the scene. The question is then to know how the data provided by stereovision and laser scanner can be combined and/or fused together in order to obtain the best results.

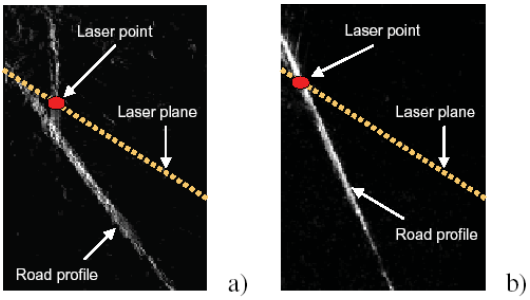

Fig. 7. "v-*disparity*" view of the laser scanner plane. a) An obstacle is detected. b) The ground is viewed as an obstacle, due to sensor pitch.

In this section we discuss several possible cooperative fusion schemes.

6.1 Laser scanner raw data filtering and clustering

The idea is here to use the geometric description of the ground provided by stereovision in order to filter the laser raw data or clustered objects which could be the result of the collision of the laser plane with the ground surface.

Two possibilities are available :

. Strategy 1: firstly, remove laser points that could be the result of the collision of the laser plane with the ground surface from the laser raw data; secondly, cluster laser points from the filtered raw data (see Fig. 8),

. Strategy 2: firstly, cluster impacts from the laser raw data; secondly, remove clustered objects that collide partially or totally the ground surface (see Fig. 9).

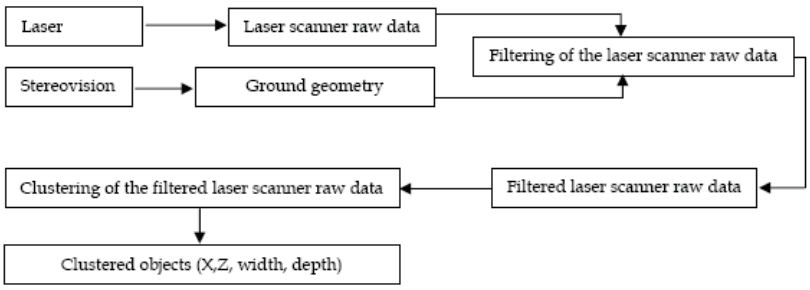

Fig. 8. Laser scanner raw data filtering and clustering. Strategy 1.

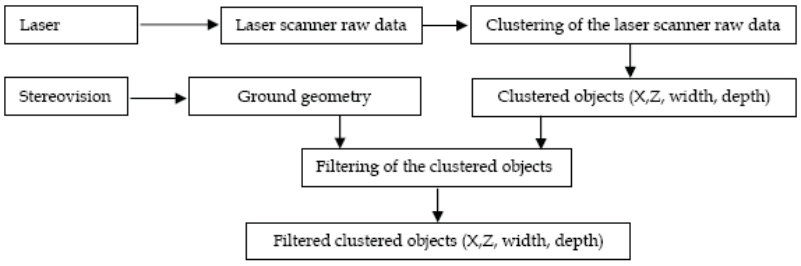

Fig. 9. Laser scanner raw data filtering and clustering. Strategy 2.

6.2 Simple redundant fusion

At this step, filtered objects from laser scanner and stereovision are available. The idea of the first fusion strategy is very simple. It consists in introducing redundancy by matching the set of obstacles detected by stereovision with the set of obstacles detected by laser scanner. If an obstacle detected by laser scanner is located at the same position than an obstacle detected by stereovision, this obstacle is supposed to be real, otherwise it is removed from the set of obstacles (see Fig. 10). However this scheme provide no information about the dynamic states (velocities, etc.) of the obstacles.

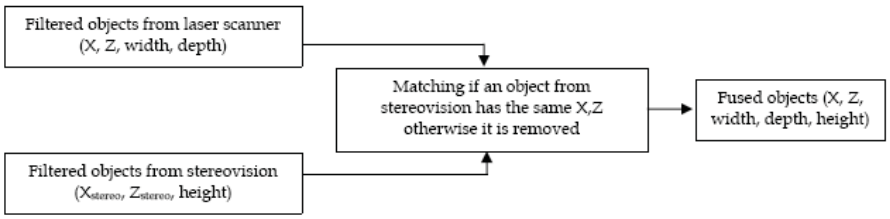

Fig. 10. Simple redundant fusion.

6.3 Fusion with global association

More complex strategies consist in introducing global association, using the algorithm presented in section 5. The idea consists in: a) performing multi-obstacles tracking and association for each sensor in order to obtain multi-tracks for each sensor; b) performing multi-track association between the tracks from the stereovision and the tracks from the laser scanner; c) fusing the tracks together in order to increase their certainty. Fig. 11 presents a fusion scheme including tracking and association for both stereovision and laser scanner sensor, and global fusion.

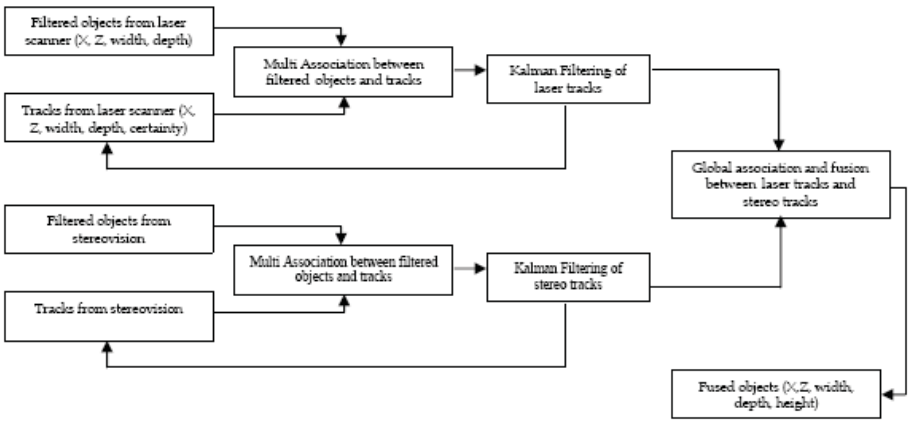

Fig. 11. Fusion with tracking and global association.

From our experiments, it seems that the tracking is difficult to perform for the stereovision tracks when the obstacles are beyond 15 meters, because of the unaccuracy of the positioning provided by the stereovision and resulting in noisy speed used in the linear

Kalman filter. Another strategy is then to perform multi-obstacles tracking and association for the single laser scanner, and then to check whether an obstacle has been detected by the stereovision at the tracked positions. If so, the certainty about the track is increased. Fig. 12 presents the corresponding fusion scheme.

Fig. 13 shows another scheme which consists in using the stereovision only to confirm the existence of an obstacle tracked by the laser scanner: the stereovision detection is performed only at the positions corresponding to objects detected by the laser scanner, in order to save computational time (indeed the stereovision will only be performed in the part of the image corresponding to the position of obstacles detected by laser scanner). Then the existence of an obstacle is confirmed if the stereovision detects an obstacle at the corresponding position. This scheme presents the advantage to work with complex ground geometry since this geometry can be estimated locally around the position of the tracked laser objects.

For each fusion scheme, the resulting positioning of each obstacle is the centimetric positioning provided by laser scanner. The estimated velocity is estimated through a linear Kalman filter applied on laser clustered data. Orientation, width and depth come from laser scanner, and height comes from stereovision.

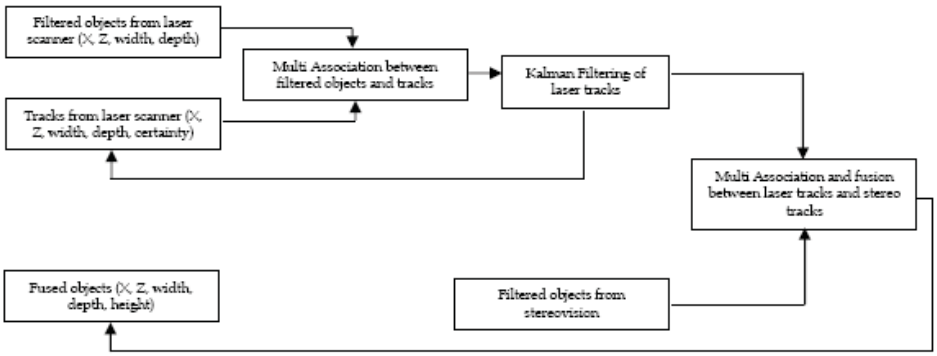

Fig. 12. Fusion with tracking of laser objects.

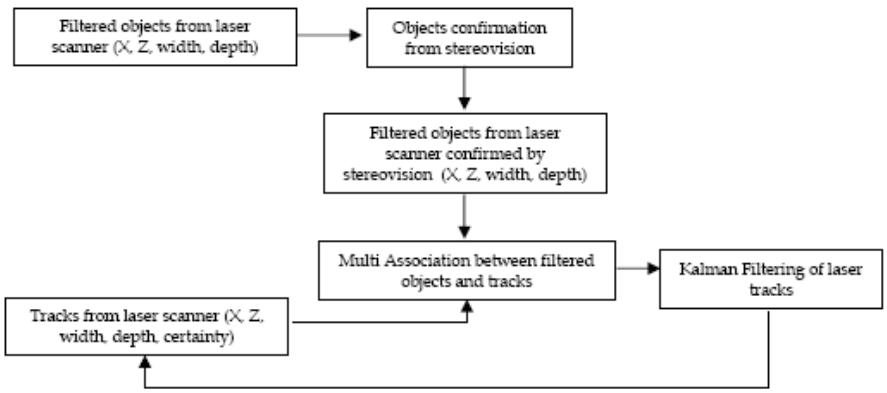

Fig. 13. Fusion with laser scanner tracking and confirmation by stereovision.

6.5 Stereovision-based obstacle confirmation criteria

To confirm the existence of an obstacle in a region of interest given by the projection of a laser-tracked object onto the image, three approaches can be used.

Number of obstacle-pixels: The first approach consists in classifying the pixels of the region of interest. A local ground profile is first extracted using the "v-disparity" iamge. Afterwards, the (u_r,Δ,v) coordinates of each pixel are analyzed to determine whether it belongs to the ground surface. If not, the pixel is classified as an obstacle-pixel. At the end of this process, every pixel in the region of interest has been classified as ground or obstacle. The number of obstacle-pixels gives a confidence on the existence of an object over the ground surface. Therefore, an obstacle is confirmed if the confidence is above a threshold.

The obstacle-pixels criterion has the advantage to avoid any assumption on the obstacles to detect. Moreover, this method gives a confidence, in an intuitive way. However, as it considers each pixel individually, it can be strongly influenced by errors in the disparity map.

Prevailing alignment orientation: Assuming that the obstacles are seen as vertical planes by the stereoscopic sensor, an other confirmation criterion can be defined (Fig. 4 and 7 a). The prevailing alignment of pixels in the local "v-disparity" image is extracted using the Hough transform. The confirmation of the track depends on the orientation of this alignment: a quite vertical alignment corresponds to an obstacle. Other alignments correspond to the ground surface. The Prevailing Alignment criterion relies on a global approach in the region of interest (alignment seeking). This makes it more robust with respect to the errors in the disparity map.

Laser points altitude: Many false detections are due to the intersection of the laser plane with the ground (see Fig. 4). The knowledge of the longitudinal ground geometry allows to deal with such errors. Therefore, the local profile of the ground is estimated through "v-disparity" framework. The altitude of the laser points is then compared to the altitude of the local ground surface. An obstacle is confirmed if this altitude is high enough.

7. Experimental Results

7.1 Experimental protocol

The algorithm has been implemented on one of the experimental vehicle of LIVIC to assess their behaviour in real conditions. The stereoscopic sensor is composed of two Sony™ 8500C cameras featuring Computar™ Auto Iris 8.5 mm focal length. Quarter PAL 8 bits gray-scale images are grabbed every 40 ms. The baseline is b = 1 m, the height h = 1.4 m and the pitch θ = 5°. The laser sensor is a Sick™ scanner which measures 201 points every 26 ms, with a scanning angular field of view of 100 °. It is positioned horizontally 40 cm over the ground surface. The whole algorithm runs at video frame rate on a dual Intel Xeon™ 1.8 GHz personal computer.

7.2 Results

The main objective is to obtain a correct detection rate and almost no false detections. Several aspects must be highlighted: the global performances (rates of non detections and false detections), the robustness of the criteria with respect to errors in the local disparity map, and the ability to work with various types of obstacles.

False detections: To assess the false detection rate, the test vehicle has been driven on a very bumpy and dent parking area to obtain a large number of false detections due to the

intersection of the laser plane with the ground surface. The results are reported in Table 3 (7032 images have been processed).
False detections are globally correctly invalidated using the obstacle-pixels and prevailing alignment criteria. The laser points altitude criterion provides more false

	Laser scanner	Number of obstacle-pixels	Prevailing alignment orientation	Laser points altitude
False Detections	781	3	10	167

Table 3. Number of false detections with the different criteria.

detections than expected, because of its high sensibility to the calibration errors between stereovision and laser scanner. Indeed, a slight error in the positioning of the scanner relative to the cameras can lead to a serious error in laser points projection, especially at long ranges. The other criteria are not dramatically affected by this issue. Most of the remaining false detection occur when the local ground surface is uniform, without any texture allowing to match pixels. So they can be removed using simple heuristics as: no obstacle can be confirmed without enough information in the region of interest. It hardly affects the detection rate, and the false detection rate of obstacle-pixels criterion almost falls to zero.
The main source of errors for the prevailing alignment algorithm comes from cases where the ground surface has non relevant texture, but where the region of interest contains a small part of a nearby object (wall, vehicle, . . .).
Detection failure: The rate of correct laser detections that have been confirmed by the different criteria has been assessed. To check, at the same time, that it can indifferently deal with various kinds of obstacles, this test has been realized with two different obstacles: a vehicle followed by the instrumented vehicle (1268 images processed), and a pedestrian crossing the road at various distances (1780 images processed). The confirmation rate of each criterion (number of obstacles detected by the laser / number of obstacles confirmed) for these two scenarios is reported in Table 4. The three criteria can successfully confirm most of the detections with both kinds of obstacles.

	Number of obstacle-pixels	Prevailing alignment orientation	Laser points altitude
Car	97.4 %	98.5 %	95.2 %
Pedestrian	91.9 %	94.9 %	97.8 %

Table 4. Rate of correct detection successfully confirmed.

Conclusion of the comparison: None of the presented obstacle confirmation criteria really outperforms the others. The obstacle-pixels is based on an intuitive approach and can deal with any types of obstacles. But it is seriously influenced by the quality of the disparity map. The more global feature of the prevailing alignment criterion makes it more robust to this kind of errors.
The laser points altitude is not sufficiently reliable to be exploited alone. Thus an efficient architecture for the application consists in using the laser points altitude to invalidate some false laser targets before the tracking step. Then the tracked obstacles are confirmed using obstacle-pixels criterion.
Performances of the perception system embedded in a collision-mitigation system: a collision mitigation system has been designed on the basis of the fusion scheme described above. This collision mitigation system can be divided into three sub-systems and a decision

unit that interconnects these sub-systems. The first sub-system is the very obstacle detection system, implementing the number of obstacle-pixels criteria for confirmation; the second sub-system is a warning area generation system that predict the path the vehicle will follow and that uses an odometer and an inertial sensor. The decision unit checks whether an obstacle is located in the warning area, and whether its Time To Collision (i.e. distance / relative speed) is under 1 second; if so, a warning message is sent to the third sub-system. The third sub-system is an automatic braking system, based on an additional brake circuit activated when a warning message is received.

The detection rate has been tested on test tracks on the basis of different driving scenarios, including cross roads and suddenly appearing obstacles. The detection rate is 98.9 %.

Then, to assess the false alarm rate, this collision mitigation system has been tested in real driving conditions, on different road types: freeway, highways, rural roads and downtown. All these tests took place on the French road network around Paris. The automatic braking system was turned off and only the warning messages were checked. In normal driving situations, an automatic system should never be launched. Each time an emergency braking would have been launched is thus considered as a false alarm. The tests have been carried out under various meteorological situations: sunny, cloudy, rainy, and under various traffic situations: low traffic to dense traffic.

403 km have been ridden on freeways. The velocity was up to 36 m / s. No false alarm was observed during these tests. Fig. 14 (a) and (b) presents some typical freeway situations under which the system has been tested. 78 km have been ridden on highways and 116 km on rural roads. The velocity was up to 25 m / s. No false alarm was observed during these tests. Fig. 14 (c) (d) presents some typical highway situations, and Fig. 14 (e) (f) some rural road situations under which the system has been tested. The downtown tests are certainly the most challenging tests since the context is the more complex. 140 km have been ridden in downtown and in urban areas. The velocity was up to 14 m/s. A false alarm was observed twice. The first one is due to a matching error during association, and the second one is due to a false target detected by stereovision on a uphill gradient portion. Fig. 15 presents some typical urban situations under which the system has been tested.

For the 737 km ridden, two false alarms were observed. The false alarm rate is thus 2.7 false alarms for 1000 km. No false alarm was observed either on freeways or on highways and rural roads. The two remaining false alarms were observed in downtown. Thus, the false alarm rate in downtown is thus 1.4 false alarm for 100 km. These results are quite promising, even if the false alarm rate must be reduced by a factor of about 1000 before the system can be envisaged to be put in the hands of common driver.

Fig. 14. Typical images of freeway and rural road situations. (a) truck following on a freeway, dense traffic - (b) freeway with low traffic - (c)(d) peri-urban highway - (e)(f) rural road with tight uphill gradient.

8. Trends and Future Research

Experiments of the proposed system give the feeling that an accurate and totally robust and reliable obstacle detection system can be designed on the basis of the techniques described in this chapter. Some tuning of the different modules are required to still improve the performances: for instance, combination of various confirmation criteria should allow to avoid any false alarm. Yet, some issues still need to be tackled, such as the auto-calibration of the set of sensors. Moreover, laser scanner remain a quite.

Fig. 15. Typical images of urban situations. (a) pedestrian crossing - (b) road works - (c) car driving out of parking lot - (d) car and bus traffic - (e) narrow road and tight curve - (f) tight curve, non flat road - (g) dense traffic - (h) road with high roll - (i) narrow paved road, tight curve.

Expensive device. Designing a medium range cheap obstacle detection system featuring high performances is still a challenge for the next years but should be possible. The key could be to use only the stereovision sensor and to implement various competitive stereovision algorithms designed to confirm each other. On a global view, a first algorithm could generate a set of targets that would be tracked along time and confirmed by the other algorithms. The confirmation criteria presented above could be used for this purpose. To reach acceptable accuracy, sub-pixel analysis should be used. Auto-calibration techniques are also required, above all for long baseline stereo sensors. Since stereovision algorithms require massive computations, real-time performance could be achieved only at the cost of a dedicated powerful chipset. Once designed, such a chipset should be not expansive to produce. Thus, a breakthrough in the field of robotics is foreseeable and would result in many applications that can not be considered nowadays because of the dissuasive cost of state-of-the-art obstacle detection systems.

9. References

Ahuja R. K., Magnanti T. L. &, Orlin J. B. (1993), *Network Flows, theory, algorithms, and applications* , Editions Prentice-Hall.

Bertozzi M. & Broggi A. (1998). *GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection*, IEEE Transaction on image processing, Vol. 7, N°1.

Blackman S. and Popoli R. (1999). *Modern Tracking Systems*, Artech.

Franke U. & Joos A. (2000). *Real-time stereovision for urban traffic scene understanding*, IEEE Intelligent Vehicle Symposium, Dearborn, USA.

Gavrila M., Kunert M.& Lages U. (2001), *A Multi-Sensor Approach for the Protection of Vulnerable Traffic Participants: the PROTECTOR Project,* IEEE Instrumentation and Measurement Technology Conference, Budapest, Hungary.

Kuhn H. W. (1955), *The Hungarian method for assignment problem*, Nav. Res. Quart., 2.

Koller D., Luong T. & Malik J. (1994). *Binocular stereopsis and lane marker flow for vehicle navigation: lateral and longitudinal control*, Technical Report UCB/CSD 94-804, University of California at Berkeley, Computer Science Division.

Labayrade R. & Aubert D. (2003 a), *A Single Framework for Vehicle Roll, Pitch, Yaw Estimation and Obstacles Detection by Stereovision*, IEEE Intelligent Vehicles Symposium 2003, Columbus, USA.

Labayrade R. & Aubert D. (2003 b). *In-Vehicle Obstacle Detection and Characterization by Stereovision*, IEEE In-Vehicle Cognitive Computer Vision Systems, Graz, Austria.

Mendes A., Conde Bento L. & Nunes U. (2004), *Multi-target Detection and Tracking with a Laserscanner, IEEE Intelligent Vehicles Symposium 2004*, pp 796-801, Parma, Italy.

Mobus R. & Kolbe U. (2004), *Multi-Target Multi-Object Tracking, Sensor Fusion of Radar and Infrared,* IEEE Intelligent Vehicles Symposium 2004, pp 732-737, Parma, Italy.

Rombaut M. (1998). *Decision in Multi-obstacle Matching Process using Theory of Belief,* AVCS'98, Amiens, France.

Scharstein D., Szeliski R. & Zabih R. (2001). *A taxonomy and evaluation of dense two-frame stereo correspondence algorithm*s, IEEE Workshop on Stereo and Multi-Baseline Vision, Kauai, HI.

Shafer G. (1976). *A mathematical theory of evidence*, Princeton University Press.

Steux B. (2002), *Fade: A Vehicle Detection and Tracking System Featuring Monocular Color Vision and Radar Data Fusion,* IEEE Intelligent Vehicles Symposium 2002, Versailles, France.

Williamson T-A. (1998). *A high-performance stereo vision system for obstacle detection*, PhD, Carnegie Mellon University.