Sensor Data Fusion for Road Obstacle Detection: A Validation Framework

Raphaël Labayrade, Mathias Perrollaz, Gruyer Dominique, Didier Aubert

▶ **To cite this version:**

Raphaël Labayrade, Mathias Perrollaz, Gruyer Dominique, Didier Aubert. Sensor Data Fusion for Road Obstacle Detection: A Validation Framework. Ciza Thomas. Sensor Fusion and Its Applications, in-tech / Sciyo, 2010, 978-953-307-101-5. hal-00683751

HAL Id: hal-00683751

https://inria.hal.science/hal-00683751

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.

16

Sensor Data Fusion for Road Obstacle Detection: A Validation Framework

Raphaël Labayrade[1], Mathias Perrollaz[2],
Dominique Gruyer[2] and Didier Aubert[2]
[1]*ENTPE (University of Lyon)*
France
[2]*LIVIC (INRETS-LCPC)*
France

1. Introduction

Obstacle detection is an essential task for autonomous robots. In particular, in the context of Intelligent Transportation Systems (ITS), vehicles (cars, trucks, buses, etc.) can be considered as robots; the development of Advance Driving Assistance Systems (ADAS), such as collision mitigation, collision avoidance, pre-crash or Automatic Cruise Control, requires that reliable road obstacle detection systems are available. To perform obstacle detection, various approaches have been proposed, depending on the sensor involved: telemeters like radar (Skutek et al., 2003) or laser scanner (Labayrade et al., 2005; Mendes et al., 2004), cooperative detection systems (Griffiths et al., 2001; Von Arnim et al., 2007), or vision systems. In this particular field, monocular vision generally exploits the detection of specific features like edges, symmetry (Bertozzi et al., 2000), color (Betke & Nguyen, 1998) (Yamaguchi et al., 2006) or even saliency maps (Michalke et al., 2007). Anyway, most monocular approaches suppose recognition of specific objects, like vehicles or pedestrians, and are therefore not generic. Stereovision is particularly suitable for obstacle detection (Bertozzi & Broggi, 1998; Labayrade et al., 2002; Nedevschi et al., 2004; Williamson, 1998), because it provides a tri-dimensional representation of the road scene. A critical point about obstacle detection for the aimed automotive applications is reliability: the detection rate must be high, while the false detection rate must remain extremely low. So far, experiments and assessments of already developed systems show that using a single sensor is not enough to meet these requirements: due to the high complexity of road scenes, no single sensor system can currently reach the expected 100% detection rate with no false positives. Thus, multi-sensor approaches and fusion of data from various sensors must be considered, in order to improve the performances. Various fusion strategies can be imagined, such as merging heterogeneous data from various sensors (Steux et al., 2002). More specifically, many authors proposed cooperation between an active sensor and a vision system, for instance a radar with mono-vision (Sugimoto et al., 2004), a laser scanner with a camera (Kaempchen et al., 2005), a stereovision rig (Labayrade et al., 2005), etc. Cooperation between mono and stereovision has also been investigated (Toulminet et al., 2006).

Our experiments in the automotive context showed that using specifically a sensor to validate the detections provided by another sensor is an efficient scheme that can lead to a very low false detection rate, while maintaining a high detection rate. The principle consists to tune the first sensor in order to provide overabundant detections (and not to miss any plausible obstacles), and to perform a post-process using the second sensor to confirm the existence of the previously detected obstacles. In this chapter, such a validation-based sensor data fusion strategy is proposed, illustrated and assessed.
The chapter is organized as follows: the validation framework is presented in Section 2. The next sections show how this framework can be implemented in the case of two specific sensors, i.e. a laser scanner aimed at providing hypothesis of detections, and a stereovision rig aimed at validating these detections. Section 3 deals with the laser scanner raw data processing: 1) clustering of lasers points into targets; and 2) tracking algorithm to estimate the dynamic state of the objects and to monitor their appearance and disappearance. Section 4 is dedicated to the presentation of the stereovision sensor and of the validation criteria. An experimental evaluation of the system is given. Eventually, section 5 shows how this framework can be implemented with other kinds of sensors; experimental results are also presented. Section 6 concludes.

2. Overview of the validation framework

Multi-sensor combination can be an efficient way to perform robust obstacle detection. The strategy proposed in this chapter is a collaborative approach illustrated in Fig. 1. A first sensor is supposed to provide hypotheses of detection, denoted 'targets' in the reminder of the chapter. The sensor is tuned to perform overabundant detection and to avoid missing plausible obstacles. Then a post process, based on a second sensor, is performed to confirm the existence of these targets. This second step is aimed at ensuring the reliability of the system by discarding false alarms, through a strict validation paradigm.

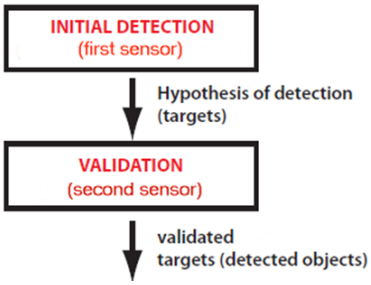

Fig. 1. Overview of the validation framework: a first sensor outputs hypothesis of detection. A second sensor validates those hypothesis.

The successive steps of the validation framework are as follows. First, a volume of interest (VOI) surrounding the targets is built in the 3D space in front of the equipped vehicle, for each target provided by the first sensor. Then, the second sensor focuses on each VOI, and evaluates criteria to validate the existence of the targets. The only requirement for the first

sensor is to provide localized targets with respect to the second sensor, so that VOI can be computed.
In the next two sections, we will show how this framework can be implemented for two specific sensors, i.e. a laser scanner, and a stereovision rig; section 5 will study the case of an optical identification sensor as first sensor, along with a stereovision rig as second sensor. It is convenient to assume that all the sensors involved in the fusion scheme are rigidly linked to the vehicle frame, so that, after calibration, they can all refer to a common coordinate system. For instance, Fig. 2 presents the various sensors taken into account in this chapter, referring to the same coordinate system.

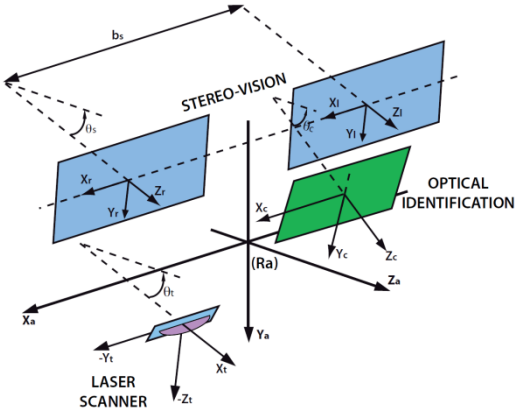

Fig. 2. The different sensors used located in the same coordinate system R_a.

3. Hypotheses of detection obtained from the first sensor: case of a 2D laser scanner

The laser scanner taken into account in this chapter is supposed to be mounted at the front of the equipped vehicle so that it can detect obstacles on its trajectory. This laser scanner provides a set of laser points 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. Fig. 4. shows a (X, -Y) projection of the laser points into the coordinate system linked to the laser scanner and illustrated in Fig. 2.

3.1 Dynamic clustering

From the raw data captured with the laser scanner, a set of clusters must be built, each cluster corresponding to an object in the observed scene (a so-called 'target'). Initially, the first laser point defines the first cluster. For all other laser points, the goal is to know whether they are a member of the existent cluster or whether they belong to a new cluster. In the literature, a large set of distance functions can be found for this purpose.

The chosen distance $D_{i,\mu}$ must comply with the following criteria (Gruyer et al., 2003):
- Firstly, this function $D_{i,\mu}$ must give a result scaled between 0 and 1 if the measurement has an intersection with the cluster μ. The value 0 indicates that the measurement i is the same object than the cluster μ with a complete confidence.
- Secondly, the result must be above 1 if the measurement i is out of the cluster μ,
- Finally, this distance must have the properties of distance functions.

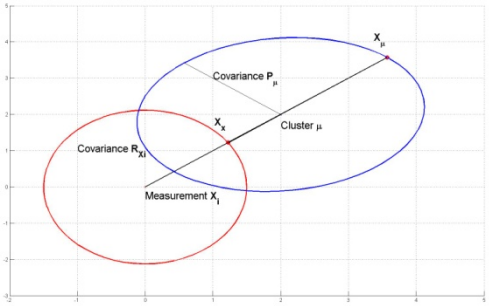

Fig. 3. Clustering of a measurement.

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

$$D_{i,j} = \frac{\sqrt{(X_i - \mu)(X_i - \mu)^t}}{\sqrt{(X_\mu - \mu)} + \sqrt{(X_X - X_i)}} \tag{1}$$

In the normalizing part, the point X_μ represents the border point of a cluster μ (centre μ). This point is located on the straight line between the cluster μ (centre μ) and the measurement i (centre X_i). The same border measurement is used with the measurement. 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 & Popoli, 1999) with ρ_0 the distance and θ_0 the angle:

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

where, using a first order expansion:

$$\begin{cases} \sigma_{x_0}^2 = \sigma_{\rho_0}^2 \cos^2\theta_0 + \sigma_{\theta_0}^2 \rho_0{}^2 \sin^2\theta_0 \\ \sigma_{y_0}^2 = \sigma_{\rho_0}^2 \sin^2\theta_0 + \sigma_{\theta_0}^2 \rho_0{}^2 \cos^2\theta_0 \\ \sigma_{x_0 y_0}^2 = \frac{1}{2}\sin 2\theta 0 \left[\sigma_{\rho_0}^2 - \sigma_{\theta_0}^2 \rho_0{}^2 \right] \end{cases} \quad (3)$$

$\sigma_{\rho_0}^2$ and $\sigma_{\theta_0}^2$ 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 ellipsoid cluster, measurement modeling and the line between the cluster centre μ and the laser measurement X is then deduced:

$$\begin{cases} x = V_{11}\sqrt{\sigma_1{}^2}\cos\Psi + V_{12}\sqrt{\sigma_2{}^2}\sin\Psi \\ y = V_{21}\sqrt{\sigma_1{}^2}\cos\Psi + V_{22}\sqrt{\sigma_2{}^2}\sin\Psi \\ y = ax + b \end{cases} \quad (4)$$

x and y give the position of a point on the ellipse and the position of a point in a line. If x and y are the same in the three equations then an intersection between the ellipse and the line exists. The solution of the set of equations (4) gives:

$$\Psi = \arctan\left(\frac{-\sqrt{\sigma_1{}^2}\left[V_{2,1} - aV_{1,1} \right]}{\sqrt{\sigma_2{}^2}\left[V_{2,2} - aV_{1,2} \right]} \right) \text{ with } \Psi \in \left[\frac{-\pi}{2}, \frac{\pi}{2} \right] \quad (5)$$

From (5), two solutions are possible:

$$X_\mu = P_\mu \sqrt{\sigma^2} \begin{bmatrix} \cos\Psi \\ \sin\Psi \end{bmatrix} \text{ and } X_\mu = P_\mu \sqrt{\sigma^2} \begin{bmatrix} \cos\Psi + \pi \\ \sin\Psi + \pi \end{bmatrix} \quad (6)$$

Then equation (1) is used with X_μ to know if a laser point belongs to a cluster. Fig. 3 gives a visual interpretation of the used distance for the clustering process. Fig. 4 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 (standard deviations).

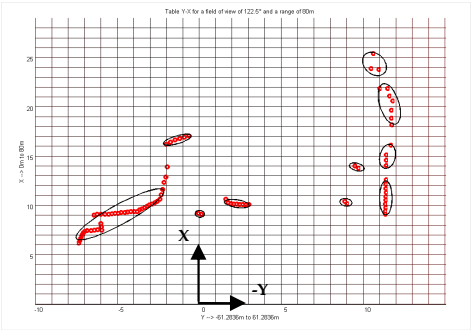

Fig. 4. Example of a result of autonomous clustering (a laser point is symbolized by a little circle, and a cluster is symbolized by a black ellipse).

3.2 Tracking algorithm

Once objects have been generated from laser scanner data, a multi-objects association algorithm is needed to estimate the dynamic state of the targets and to monitor appearances and disappearances 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).

In a general framework, the problem consists to identify 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 to associate perceived objects X_i to known objects Y_j. Belief theory allows assessing the veracity of P_i propositions representing the matching of the different objects.

A basic belief allowing the characterization of a proposition must be defined. This basic belief (mass $m_{\Theta}(\ )$) is defined in a [0,1] interval. This mass is very close to the one used in probabilistic approach, except that it is distributed on all the propositions of the referential of definition 2^{Ω}= { $A/A\subseteq\Omega$} = {$\varnothing$, {Y_1}, {Y_2 },..., {Y_n}, {Y_1,Y_2},…, {Ω}}. This referential is the power set of $\Omega = \{Y_1, Y_2, \cdots, Y_n\}$ which includes all the admissible hypotheses. These hypotheses must also be exclusive $\left(Y_i \cap Y_j = \varnothing, \forall i \neq j\right)$. The masses thus defined are called "*basic belief assignment*" and denoted "*bba*" and verify:

$$\sum_{A\subseteq\Omega} m^{\Omega}(A) = 1 \quad A \in 2^{\Omega}, A \neq \varnothing \tag{7}$$

The sum of these masses is equal to 1 and the mass corresponding to the impossible case $m^{\Omega}_{1..n}\{X_i\}(\varnothing)$ must be equal to 0.

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, the frame of discernment is:

$$\Omega = \{Y_1, Y_2, Y_3, Y_*\}$$

where Y_i means that "X and Y_i are supposed to be the same object"

In order to be sure that the frame of discernment is really exhaustive, a last hypothesis noted "Y_*" is added (Royere et al., 2000). This one can be interpreted as "a target has no association with any of the tracks". In fact each Y_j represents a local view of the world and the "Y_*" represents the rest of the world. In this context, "Y_*" means that "*an object is associated with nothing in the local knowledge set*".

In our case, the definition of the *bba* is directly in relation with the data association applications. The mass distribution is a local view around a target X_i and of a track Y_j. The *bba* on the association between X_i and Y_j will be noted $m_j^{\Omega}\{X_i\}(\ )$. It is defined on the frame of discernment $\Omega = \{Y_1, Y_2, \ldots Y_n, Y_*\}$ and more precisely on focal elements $\{Y, \overline{Y}, \Omega\}$ were $\overline{Y}$ means not Y.

Each one will respect the following meaning:

$m_j^{\Omega}\{X_i\}(Y_j)$: Degree of belief on the proposition « X_i is associated with Y_j »;

$m_j^{\Omega}\{X_i\}(\overline{Y}_j)$: Degree of belief on the proposition « X_i is not associated with Y_j »;

$m_j^{\Omega}\{X_i\}(\Omega)$: Degree on « the ignorance on the association between X_i and Y_j »;

$m_j^{\Omega}\{X_i\}(Y_*)$: mass representing the reject: X_i is in relation with nothing.

In fact, the complete notation of a belief function is: $m_{S,t}^{\Omega}\{X\}[BC_{S,t}](A) \quad A \in 2^{\Omega}$

With S the information source, t the time of the event, Ω the frame of discernment, X a parameter which takes value in Ω and BC the *evidential corpus* or *knowledge base*. This formulation represents the degree of belief allocated by the source S at the time t to the hypothesis that X belong to A (Denoeux & Smets, 2006).

In order to simplify this notation, we will use the following basic belief function notation $m_j^{\Omega}\{X\}(A)$. The t argument is removed because we process the current time without any links with the previous temporal data.

In this mass distribution, X denotes the processed perceived objects and the index j the known objects (track). If the index is replaced by a set of indices, then the mass is applied to all targets.

Moreover, if an iterative combination is used, the mass $m_j^{\Omega}\{X_i\}(Y_*)$ is not part of the initial mass set and appears only after the first combination. It replaces the conjunction of the combined masses $m_j^{\Omega}\{X_i\}(\overline{Y}_j)$. 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 s recurrent stage. For the construction of these combination rules, the work and a first formalism given in (Rombaut, 1998) is used. The use of a *basic belief assignment* generator using the strong hypothesis: "*an object cannot be in the same time associated and not associated to another object*" allows obtaining 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 (Gruyer & Berge-Cherfaoui 1999a; Gruyer & Berge-Cherfaoui 1999b). The rules become:

$$m_{1..n}^{\Omega}\{X_i\}(Y_j) = m_j^{\Omega}\{X_i\}(Y_j)\prod_{\substack{a=1\\a\neq j}}^{n}\left(1-m_a^{\Omega}\{X_i\}(Y_a)\right) \tag{8}$$

$$m_{1..n}^{\Omega}\{X_i\}\left(\{Y_j, Y_*\}\right) = m_j^{\Omega}\{X_i\}(\Omega)\prod_{\substack{a=1\\a\neq j}}^{n} m_a^{\Omega}\{X_i\}(\overline{Y}_a) \tag{9}$$

$$m_{1..n}^{\Omega}\{X_i\}\left(\{Y_j, Y_k, Y_*\}\right) = m_j^{\Omega}\{X_i\}(\Omega).m_k^{\Omega}\{X_i\}(\Omega)\prod_{\substack{a=1\\a\neq j\\a\neq k}}^{n} m_a^{\Omega}\{X_i\}(\overline{Y}_a) \tag{10}$$

$$m_{1..n}^{\Omega}\{X_i\}\left(\{Y_j, Y_k, ..., Y_l, Y_*\}\right) = m_j^{\Omega}\{X_i\}(\Omega).m_k^{\Omega}\{X_i\}(\Omega)....m_l^{\Omega}\{X_i\}(\Omega)\prod_{\substack{a=1\\a\neq j\\a\neq k\\.....\\a\neq l}}^{n} m_a^{\Omega}\{X_i\}(\overline{Y}_a) \tag{11}$$

$$m_{1..n}^{\Omega}\{X_i\}\left(\overline{Y}_j\right) = m_j^{\Omega}\{X_i\}(\overline{Y}_j)\prod_{\substack{a=1\\a\neq j}}^{n} m_a^{\Omega}\{X_i\}(\Omega) \tag{12}$$

$$m_{1..n}^{\Omega}\{X_i\}(\Omega) = \prod_{a=1}^{n} m_a^{\Omega}\{X_i\}(\Omega) \tag{13}$$

$$m_{1..n}^{\Omega}\{X_i\}(Y_*) = \prod_{a=1}^{n} m_a^{\Omega}\{X_i\}(\overline{Y}_a) \tag{14}$$

$$m_{1..n}^{\Omega}\{X_i\}(\varnothing) = 1-\left[\prod_{a=1}^{n}\left(1-m_a^{\Omega}\{X_i\}(Y_a)\right)+\sum_{a=1}^{n} m_a^{\Omega}\{X_i\}(Y_a)\prod_{\substack{b=1\\b\neq a}}^{n}\left(1-m_b^{\Omega}\{X_i\}(Y_b)\right)\right] \tag{15}$$

$m\{X_i\}(Y_*)$ is the result of the combination of all non association belief masses for X_i. Indeed, new target(s) apparition or loss of track(s) because of field of view limitation or objects occultation, leads to consider with attention the Y_* hypothesis which models these phenomena.
In fact, a specialized *bba* can be defined given a local view of X with Y association. In order to obtain a global view, it is necessary to combine the specialized *bbas*. The combination is possible when *bbas* are defined on the same frame of discernment and for the same parameter X.

In a first step, a combination of $m_j^{\Omega}\{X_i\}(\ )$ with $j \in [1..n]$ is done using equations (8) to (15). The result of the combination gives a mass $m_{1..n}^{\Omega}\{X_i\}(\ )$ defined on 2^{Ω}. We can repeat these operations for each X_i and to obtain a set of p *bbas*: $m_{1..n}^{\Omega}\{X_1\}(\ )$, $m_{1..n}^{\Omega}\{X_2\}(\ )$,.. $m_{1..n}^{\Omega}\{X_p\}(\ )$

p is the number of targets and Ω the frame including the n tracks corresponding to the n hypotheses for target-to-track association.

In order to get a decision, a pignistic transformation is applied for each $m_{1..n}^{\Omega_i}\{X_i\}(\ )$ with $i \in [1..p]$. The pignistic probabilities $BetP^{\Omega_i}\{X_i\}(Y_j)$ of each Y_j hypothesis are summarized in a matrix corresponding to the target point of view.
However, this first matrix gives the pignistic probabilities for each target without taking into consideration the other targets. Each column is independent from the others. A dual approach is proposed in order to consider the possible association of a track with the targets in order to have the tracks point of view.
The dual approach consists in using the same *bba* but combined for each track Y.

From the track point of view, the frame of discernment becomes $\Theta = \{X_1,...,X_m,X_*\}$

The X_* hypothesis models the capability to manage either track disappearance or occultation. For one track Y_j, the *bbas* are then:

$m_i^{\Theta}\{Y_j\}(X_i) = m_j^{\Omega}\{X_i\}(Y_j)$: Degree of belief on the proposition « Y_j is associated with X_i »;

$m_i^{\Theta}\{Y_j\}(\overline{X}_i) = m_j^{\Omega}\{X_i\}(\overline{Y}_j)$: Degree of belief on the proposition « Y_j is not associated with X_i »;

$m_i^{\Theta}\{Y_j\}(\Theta) = m_j^{\Omega}\{X_i\}(\Omega)$: Degree of « the ignorance on the association between Y_j *and* X_i ».

The same combination -equations (8) to (15)- is applied and gives $m_{1..p}^{\Theta}\{Y_i\}(\ )$.

These operations can be repeated for each Y_j to obtain a set of n *bbas*:

$$m_{1..p}^{\Theta_1}\{Y_1\}(\ ),\ m_{1..p}^{\Theta_2}\{Y_2\}(\ ),..\ m_{1..p}^{\Theta_n}\{Y_n\}(\ )$$

n is the number of tracks and Θ_j is the frame based on association hypothesis for Y_j parameter. The index j in Θ_j is now useful in order to distinguish the frames based on association for one specific track Y_j for $j \in [1..n]$.
A second matrix is obtained involving the pignistic probabilities $BetP^{\Omega_i}\{Y_i\}(X_j)$ about the tracks.

The last stage of this algorithm consists to establish the best decision from the previously computed associations using the both pignistic probabilities matrices ($BetP^{\Omega_i}\{X_i\}(Y_j)$ and $BetP^{\Omega_i}\{Y_i\}(X_j)$). The decision stage is done with the maximum pignistic probability rule. This rule is applied on each column of both pignistic probabilities matrices.
With the first matrix, this rule answers to the question "*which track Y_j is associated with target X_i?*":

$$X_i = d(Y_j) = \underset{i}{Max}\left[BetP^{\Omega_i}_{\{X_i\}(Y_j)} \right] \tag{16}$$

With the second matrix, this rule answers to the question "*which target X_i is associated to the track Y_j?*":

$$Y_j = d(X_i) = \underset{j}{Max}\left[BetP^{\Theta_j}_{\{Y_j\}(X_i)} \right] \tag{17}$$

Unfortunately, a problem appears when the decision obtained from a pignistic matrix is ambiguous (this ambiguity quantifies the duality and the uncertainty of a relation) or when the decisions between the two pignistic 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 et al., 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 defective or poor sensors perceiving it, then the sensor 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 target is updated using the new position of the target, and so the dynamic state of each target is estimated, i.e. both speed and angular speed.

4. Validation of the hypotheses of detection: case of a stereovision-based validation

In order to validate the existence of the targets detected by the laser scanner and tracked over time as describe above, a stereovision rig is used. The geometrical configuration of the stereoscopic sensor is presented in Fig. 5. The upcoming steps are as follows: building Volumes Of Interest (VOI) from laser scanner targets, validation criteria evaluation from 'obstacle measurement points'.

4.1 Stereovision sensor modeling

The epipolar geometry is rectified through calibration, so that the epipolar lines are parallel. Cameras are described by a *pinhole* model and characterized by (u_0, v_0) the position of the optical center in the image plane, and a = *focal length / pixel size* (pixels are supposed to be square). The extrinsic parameters of the stereoscopic sensor are (0, Y_s^0, Z_s^0) the position of the central point of the stereoscopic baseline, θ_s the pitch of the cameras and b_s the length of

stereoscopic baseline. Given a point P (X_a, Y_a, Z_a) in the common coordinate system R_a, its position (u_r, Δ_s, v) and (u_l, Δ_s, v) in the stereoscopic images can be calculated as:

$$u_r = u_0 + \alpha \frac{X_a - b_s/2}{(Y_a - Y_S^0)\sin\theta_s + (Z_a - Z_S^0)\cos\theta_s} \tag{18}$$

$$u_l = u_0 + \alpha \frac{X_a + b_s/2}{(Y_a - Y_S^0)\sin\theta_s + (Z_a - Z_S^0)\cos\theta_s} \tag{19}$$

$$v = v_0 + \alpha \frac{(Y_a - Y_S^0)\cos\theta_s - (Z_a - Z_S^0)\sin\theta_s}{(Y_a - Y_S^0)\sin\theta_s + (Z_a - Z_S^0)\cos\theta_s} \tag{20}$$

$$\Delta_s = \alpha \frac{b_s}{(Y_a - Y_S^0)\sin\theta_s + (Z_a - Z_S^0)\cos\theta_s} \tag{21}$$

where $\Delta_s = u_l - u_r$ is the disparity value of a given pixel, $v = v_l = v_r$ its y-coordinate.

This transform is invertible, so the coordinates in R_a can be retrieved from images coordinates through:

$$X_a = b_s / 2 + \frac{b_s(u_r - u_0)}{\Delta_s} \tag{22}$$

$$Y_a = Y_S^0 + \frac{b_s((v - v_0)\cos\theta_s + \alpha\sin\theta_s)}{\Delta_s} \tag{23}$$

$$Z_a = Z_S^0 + \frac{b_s(\alpha\cos\theta_s - (v - v_0)\sin\theta_s)}{\Delta_s} \tag{24}$$

The coordinate system $R_\Delta = (\Omega, \boldsymbol{u_r}, \boldsymbol{v}, \boldsymbol{\Delta_s})$ defines a 3D space E_Δ, denoted disparity space.

4.2 Building Volumes Of Interest (VOI) from laser scanner

The first processing step of the validation algorithm is the conversion of targets obtained from laser scanner into VOI (Volumes Of Interest). The idea is to find where the system should focalize its upcoming processing stages. A VOI is defined as a rectangular parallelepiped in the disparity space, frontal to the image planes.

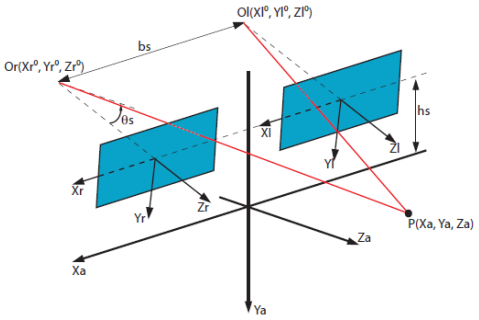

Fig. 5. Geometrical configuration of the stereoscopic sensor in R_a.

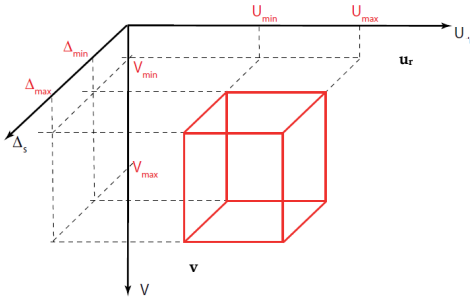

Fig. 6. Definition of the volume of interest (VOI).

Fig. 6 illustrates this definition. This is equivalent to a region of interest in the right image of the stereoscopic pair, associated to a disparity range. This definition is useful to distinguish objects that are connected in the images, but located at different longitudinal positions.
To build volumes of interest in the stereoscopic images, a bounding box V_o is constructed in R_a from the laser scanner targets as described in Fig. 7 (a). Z_{near} , X_{left} and X_{right} are computed from the ellipse parameters featuring the laser target. Z_{far} and Y_{high} are then constructed from an arbitrary knowledge of the size of the obstacles. Fig. 7 (b) shows how the VOI is projected in the right image of the stereoscopic pair. Equations (18-20) are used to this purpose.

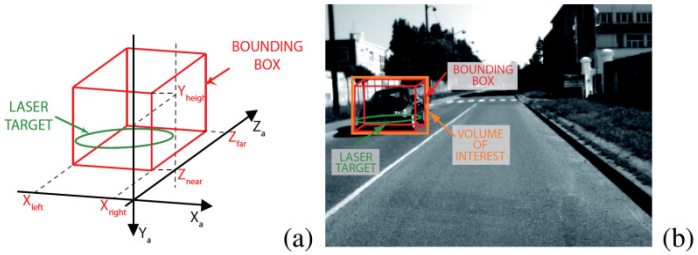

Fig. 7. (a): Conversion of a laser target into bounding box. (b): Projection of the bounding box (i.e. VOI) into the right image of the stereoscopic pair.

4.3 Computation of 'obstacle measurement points'

In each VOI, measurement points are computed, which will be used for the further validation stage of the data fusion strategy. This is performed through a local disparity map computation.

1) Local disparity map computation: The local disparity map for each VOI is computed using a classical Winner Take All (WTA) approach (Scharstein & Szeliski, 2002) based on Zero Sum of Square Difference (ZSSD) criterion. Use of a sparse disparity map is chosen to keep a low computation time. Thus, only high gradient pixels are considered in the process.

2) Filtering: Using directly raw data from the local disparity map could lead to a certain number of errors. Indeed, such maps could contain pixels belonging to the road surface, to targets located at higher distances or some noise due to matching errors. Several filtering operations are implemented to reduce such sources of errors: the cross-validation step helps to efficiently reject errors located in half-occluded areas (Egnal & Wildes, 2002), the double correlation method, using both rectangular and sheared correlation window provides instant classification of the pixels corresponding to obstacles or road surface (Perrolaz et al., 2007). Therefore only obstacle pixels are kept; it is required to take in consideration the disparity range of the VOI in order to reject pixels located further or closer than the processed volume; a median filter rejects impulse noise created by isolated matching errors.

3) Obstacle pixels: Once the local disparity map has been computed and filtered, the VOI contains an 'obstacle disparity map', corresponding to a set of measurement points. For better clarity, we will call *obstacle pixels* the measurement points present in the 'obstacle disparity map'.

We propose to exploit the *obstacle pixels* to reject false detections. It is necessary to highlight the major features of what we call 'obstacles', before defining the validation strategy. These features must be as little restrictive as possible, to ensure that the process of validation remains generic against the type of obstacles.

4.4 Stereovision-based validation criteria

In order to define validation criterion, hypotheses must be made to consider a target as an actual obstacle: 1) its size shall be significant; 2) it shall be almost vertical; 3) its bottom shall be close to the road surface.
The derived criteria assessed for a target are as follows:

1) The observed surface, which must be large enough;
2) The orientation, which must be almost vertical;
3) The bottom height, which must be small enough.

Starting from these three hypotheses, let us define three different validation criteria.

1) Number of obstacle pixels: To validate a target according to the first feature, the most natural method consists in checking that the volume of interest associated to the target actually contains *obstacle pixels*. Therefore, our validation criterion consists in counting the number of *obstacle pixels* in the volume, and comparing it to a threshold.

2) Prevailing alignment criterion: One can also exploit the almost verticality of obstacles, while the road is almost horizontal. We offer therefore to measure in which direction the *obstacle pixels* of the target are aligned. For this purpose, the local disparity map of the target is projected over the v-disparity plane (Labayrade & al., 2002). A linear regression is then computed to find the global orientation of the set of *obstacle pixels*. The parameters of the extracted straight line are used to confirm the detection.

3) Bottom height criterion: A specific type of false detections by stereovision appears in scenes with many repetitive structures. Highly correlated false matches can then appear as objects closer to the vehicle than their actual location. These false matches are very disturbing, because the validation criteria outlined above assume that matching errors are mainly uncorrelated. These criteria are irrelevant with respect to such false detections. Among these errors, the most problematic ones occur when the values of disparities are over-evaluated. In the case of an under-evaluation, the hypothesis of detection is located further than the actual object, and is therefore a case of detection failure. When the disparity is significantly over-evaluated, the height of the bottom of an obstacle can be high and may give the feeling that the target flies without ground support. So the validation test consists to measure the altitude of the lowest *obstacle pixel* in the VOI, and check that this altitude is low enough.

4.5 Detailed architecture for a laser scanner and a stereovision rig

The detailed architecture of the framework, showing how the above mentioned criteria are used, is presented in Fig. 8. As the first sensor of the architecture, the laser scanner produces fast and accurate targets, but with a large amount of false positives. Indeed, in case of strong vehicle pitch or non-plane road geometry, the intersection of the scanning plane with the road surface produces errors that can hardly be discriminated from actual obstacles. Thus, as the second sensor of the architecture, the stereovision rig is aimed at discarding false positives through the application of the above-mentioned confirmation criteria.

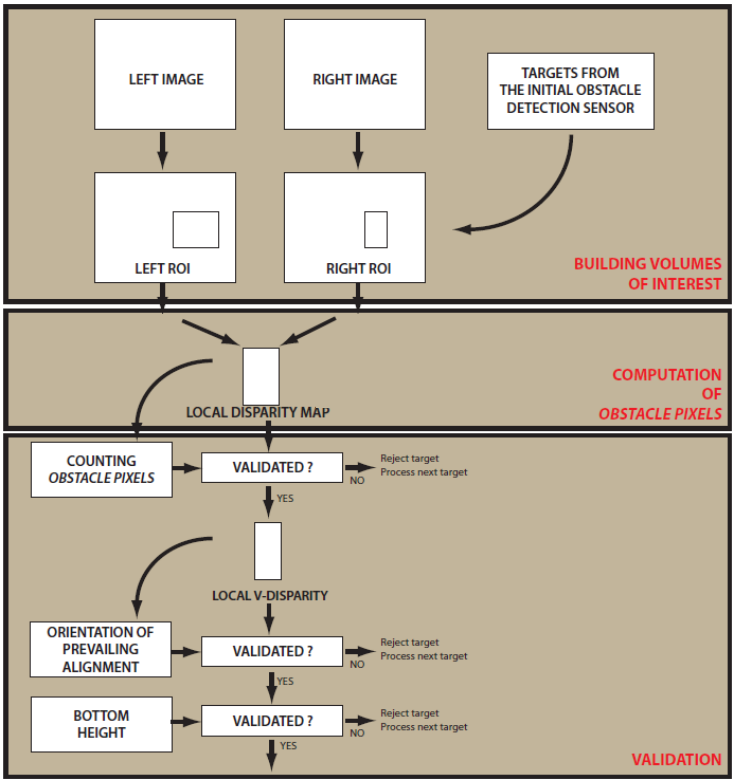

Fig. 8. Detailed architecture of the framework, using a laser scanner as the first sensor, and stereovision as validation sensor.

4.6 Experimental setup and results

The stereoscopic sensor is composed of two *SMaL*™ CMOS cameras, with 6 *mm* focal length. VGA 10 bits grayscale images are grabbed every 30 *ms*. The stereoscopic baseline is 30 *cm*. The height is 1,4 *m* and the pitch $\theta_s = 5^o$.

The telemetric sensor is a *Sick*™ laser scanner which measures 201 points every 26 *ms*, with a scanning angular field of view of 100º. It is positioned horizontally 40 *cm* over the road surface. Fig. 9 shows the laser points projected in the right image of the stereoscopic sensor, as well as bounding boxes around obstacles, generated from the laser point clustering stage.

Fig. 10 presents examples of results obtained in real driving conditions. False positives are generated by the laser scanner, and are successfully discarded after the validation process. A quantitative evaluation was also performed. 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. 7032 images have been processed. The number of false alarms drops from 781 (without the validation step) to 3 (with the validation step). On its part, the detection rate is decreased by 2.6% showing that the validation step hardly affects it.

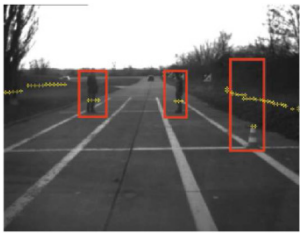

Fig. 9. Right image from stereoscopic pair with laser points projected (cross), and resulting targets (rectangles).

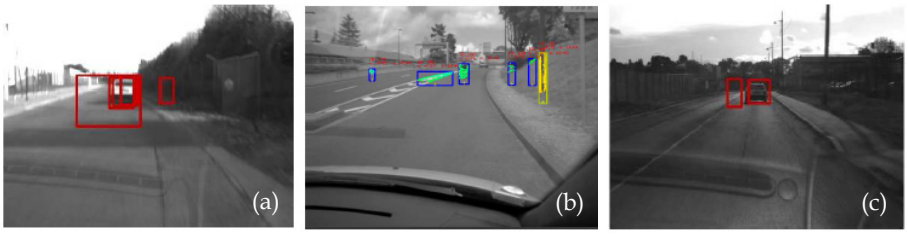

Fig. 10. Common sources of errors in detection using a laser scanner. (a): laser scanning plane intersects road surface. (b): non planar road is seen as an obstacle. (c): laser temporal tracking failed. All of these errors are correctly discarded by the stereovision-based validation step.

5. Implementation with other sensors

The validation framework presented in Fig. 1 is generic and can be used along with arbitrary sensors. Good results are likely to be obtained if the two sensors present complementary features, for instance distance assessment accuracy / obstacle 3D data. For instance, the first sensor providing hypotheses of detection can be a radar or optical identification. In this section we focalize on the latter sensor as the first sensor of the architecture, and keep using the stereovision rig as the second sensor.

5.1 Optical identification sensor

Optical identification is an example of cooperative detection, which is a recently explored way of research in the field of obstacle detection. With this approach, the different vehicles in the scene cooperate to enhance the global detection performance.

The cooperative sensor in this implementation is originally designed for cooperation between obstacle detection and vehicle to vehicle (V2V) telecommunications. It can as well be used for robust high range obstacle detection. The process is divided in two parts: an emitting near IR lamp on the back of an object, emitting binary messages (an unique ID code), and a high speed camera with a band pass filter centered around near IR, associated to an image processing algorithm to detect the sources, track them and decode the messages. This sensor is described more in details in (Von Arnim et al., 2007).

5.2 Building Volumes Of Interest (VOI) from optical identification

VOIs are built in a way similar to the method used for laser scanner. A bounding box around the target, with arbitrary dimensions, is projected into the disparity space. However, ID lamps are localized in decoding-camera's image plane, with only two coordinates. So, to obtain fully exploitable data, it is necessary to retrieve a tri-dimensional localization of the detection in *Ra*. Therefore, it has been decided to fix a parameter: the lamp height is considered as known. This constraint is not excessively restrictive because the lamp is fixed once and for all on the object to identify.

5.3 Experimental results with optical identification sensor

Fig. 11 (a) presents optical identification in action: a vehicle located about 100 m ahead is detected and identified. Fig. 11 (b) presents a common source of error of optical identification, due to the reflection of the IR lamp on the road separating wall. This error is correctly discarded by the stereovision-based validation process. In this implementation, the stereoscopic processing gives the opportunity to validate the existence of an actual obstacle, when a coherent IR source is observed. This is useful to reject false positives due to IR artifacts; another example is the reflection of an ID lamp onto a specular surface (another vehicle for instance).

Fig. 11. (a): Detection from optical identification system projected in the right image. (b): Error in detection: ID lamp reflected on the road separating wall. This error is correctly discarded by the stereovision-based validation step.

6. Conclusion

For the application of obstacle detection in the automotive domain, reliability is a major consideration. In this chapter, a sensor data fusion validation framework was proposed: an initial sensor provides hypothesis of detections that are validated by a second sensor. Experiments demonstrate the efficiency of this strategy, when using a stereovision rig as the validation sensor, which provide rich 3D information about the scene. The framework can be implemented for any initial devices providing hypothesis of detection (either single sensor or detection system), in order to drastically decrease the false alarm rate while having few influence on the detection rate.

One major improvement of this framework would be the addition of a multi-sensor combination stage, to obtain an efficient multi-sensor collaboration framework. The choice to insert this before or after validation is still open, and may have significant influence on performances.

7. References

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

Bertozzi M. & Broggi, A. (1998). Gold: A parallel real-time stereo vision system for generic obstacle and lane detection, *IEEE Transactions on Image Processing*, 7(1), January 1998.

Bertozzi, M., Broggi, A., Fascioli, A. & Nichele, S. (2000). Stereo vision based vehicle detection, In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Detroit, USA, October 2000.

Betke, M. & Nguyen, M. (1998). Highway scene analysis form a moving vehicle under reduced visibility conditions, *Proceedings of the IEEE International Conference on Intelligent Vehicles*, Stuttgart, Germany, October 1998.

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

Denoeux, T. & Smets, P. (2006). Classification using Belief Functions: the Relationship between the Case-based and Model-based Approaches, *IEEE Transactions on Systems, Man and Cybernetics B*, Vol. 36, Issue 6, pp 1395-1406, 2006.

Egnal, G. & Wildes, R. P. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133, 2002.

Griffiths, P., Langer, D., Misener, J. A., Siegel, M., Thorpe. C. (2001). Sensorfriendly vehicle and roadway systems, *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hongrie, 2001.

Gruyer, D. & Berge-Cherfaoui V. (1999a). Matching and decision for Vehicle tracking in road situation, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Koera, 1999.

Gruyer, D., & Berge-Cherfaoui V. (1999b). Multi-objects association in perception of dynamical situation, *Fifteenth Conference on Uncertainty in Artificial Intelligence*, *UAI'99*, Stockholm, Sweden, 1999.

Gruyer, D. Royere, C., Labayrade, R., Aubert, D. (2003). Credibilistic multi-sensor fusion for real time application. Application to obstacle detection and tracking, *ICAR 2003*, Coimbra, Portugal, 2003.

Kaempchen, N.; Buehler, M. & Dietmayer, K. (2005). Feature-level fusion for free-form object tracking using laserscanner and video, *Proceedings of the IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, June 2005.

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

Labayrade, R.; Aubert, D. & Tarel, J.P. (2002). Real time obstacle detection on non flat road geometry through 'v-disparity' representation, *Proceedings of the IEEE Intelligent Vehicles Symposium*, Versailles, France, June 2002.

Labayrade R.; Royere, C. & Aubert, D. (2005). A collision mitigation system using laser scanner and stereovision fusion and its assessment, *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp 440– 446, Las Vegas, USA, June 2005.

Labayrade R.; Royere C.; Gruyer D. & Aubert D. (2005). Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner", *Autonomous Robots, special issue on Robotics Technologies for Intelligent Vehicles*, Vol. 19, N°2, September 2005, pp. 117 - 140.

Mendes, A.; Conde Bento, L. & Nunes U. (2004). Multi-target detection and tracking with a laserscanner, *Proceedings of the IEEE Intelligent Vehicles Symposium*, University of Parma, Italy, June 2004.

Michalke, T.; Gepperth, A.; Schneider, M.; Fritsch, J. & Goerick, C. (2007). Towards a human-like vision system for resource-constrained intelligent Cars, *Proceedings of the 5th International Conference on Computer Vision Systems*, 2007.

Nedevschi, S.; Danescu, R.; Frentiu, D.; Marita, T.; Oniga, F.; Pocol, C.; Graf, T. & Schmidt R. (2004). High accuracy stereovision approach for obstacle detection on non planar roads, *Proceedings of the IEEE Intelligent Engineering Systems*, Cluj Napoca, Romania, September 2004.

Perrollaz, M., Labayrade, R., Gallen, R. & Aubert, D. (2007). A three resolution framework for reliable road obstacle detection using stereovision, *Proceedings of the IAPR International Conference on Machine Vision and Applications*, Tokyo, Japan, 2007.

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

Royere, C., Gruyer, D., Cherfaoui V. (2000). Data association with believe theory, *FUSION'2000*, Paris, France, 2000.

Scharstein, D. & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of Computer Vision*, 47(1-3):7–42, 2002.

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

Skutek, M.; Mekhaiel, M. & Wanielik, M. (2003). Precrash system based on radar for automotive applications, *Proceedings of the IEEE Intelligent Vehicles Symposium*, Columbus, USA, June 2003.

Steux, B.; Laurgeau, C.; Salesse, L. & Wautier, D. (2002). Fade: A vehicle detection and tracking system featuring monocular color vision and radar data fusion, *Proceedings of the IEEE Intelligent Vehicles Symposium*, Versailles, France, June 2002.

Sugimoto, S.; Tateda, H.; Takahashi, H. & Okutomi M. (2004). Obstacle detection using millimeter-wave radar and its visualization on image sequence, *Proceedings of the IAPR International Conference on Pattern Recognition*, Cambridge, UK, 2004.

Toulminet, G.; Bertozzi, V; Mousset, S.; Bensrhair, A. & Broggi. (2006). Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis, *IEEE Transactions on Image Processing*, 15(8):2364–2375, August 2006.

Von Arnim, A.; Perrollaz, M.; Bertrand, A. & Ehrlich, J. (2007). Vehicle identification using infrared vision and applications to cooperative perception, *Proceedings of the IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 2007.

Williamson, T. (1998). *A High-Performance Stereo Vision System for Obstacle Detection*. PhD thesis, Carnegie Mellon University, 1998.

Yamaguchi, K.; Kato, T. & Ninomiya, Y. (2006). Moving obstacle detection using monocular vision, *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006.