HAL
open science

La stéréovision appliquée à l'analyse de scènes routières

Pierre Charbonnier, Valérie Muzet, Philippe Nicolle, Nicolas Hautiere, Jean Philippe Tarel, Didier Aubert

To cite this version:

Pierre Charbonnier, Valérie Muzet, Philippe Nicolle, Nicolas Hautiere, Jean Philippe Tarel, et al.. La stéréovision appliquée à l'analyse de scènes routières. Bulletin des Laboratoires des Ponts et Chaussées, 2008, 272, pp.57-73. hal-00362952

HAL Id: hal-00362952

https://hal.science/hal-00362952v1

Submitted on 19 Feb 2009

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.

Stereovision applied to road scene analysis

Pierre CHARBONNIER*
Valérie MUZET
LRPC de Strasbourg, France
Philippe NICOLLE
LCPC, Centre de Nantes, France
Nicolas HAUTIERE
Jean-Philippe TAREL
LCPC, Centre de Paris, France
Didier AUBERT
LIVIC, INRETS-LCPC, France

ABSTRACT

3D reconstruction techniques through use of stereovision have experienced considerable development over the past two decades. Their field of application extends from non-destructive control to augmented reality and includes remote sensing. Following a review of the basic principles behind these methods, this paper will present a summary of the most significant developments of stereovision in the field of road scene analysis, within the network of *Ponts et Chaussées* laboratories (LPC) over the last few years. The targeted applications include driving assistance and road inspection tools, as well as safety studies. The research described in this paper is thus intended over the short term to yield a stereo version of the range of LPC products dedicated to road imaging, i.e. IRCAN (road imaging using digital camera) and IREVE (road imaging calibration, visualization and utilization). Stereovision has led to devising operational obstacle detectors in addition to being used for determining atmospheric visibility distances and estimating longitudinal road profiles.

La stéréovision appliquée à l'analyse de scènes routières

RÉSUMÉ

Les techniques de reconstruction 3D par stéréovision ont connu un développement considérable au cours des deux dernières décennies. Leur champ d'application s'étend du contrôle non destructif à la réalité augmentée, en passant par la télédétection. Après un rappel des principes de base de ces méthodes, on propose ici une présentation synthétique des principales mises en œuvre de la stéréovision au sein du réseau des LPC (laboratoires des ponts et chaussées) ces dernières années dans le domaine de l'analyse de scènes routières. Les applications visées concernent, d'une part, les aides à la conduite et, d'autre part, la connaissance du patrimoine routier et les études de sécurité routière. Ainsi, les travaux décrits devraient déboucher à court terme sur une version « stéréo » des produits LPC dédiés à l'imagerie routière (IRCAN, imagerie routière par caméra numérique, et IREVE, imagerie routière étalonnages visualisations exploitations). La stéréovision a permis la mise au point de détecteurs opérationnels d'obstacles sur chaussée. Elle est également utilisée pour la détermination des distances de visibilité atmosphérique et pour l'estimation du profil en long de la chaussée.

*Corresponding author:
Pierre CHARBONNIER
Pierre.Charbonnier@developpement-durable.gouv.fr

INTRODUCTION

Road scenes are images of the physical road and its local environment, with these images being acquired by a traveling vehicle. Within the scope of road network management or road safety studies, such images are taken at regular distance intervals along travel itineraries in order to create image databanks, which have become increasingly requested by facility managers. Beyond simple visual interpretation, an offline analysis of these images makes it possible to extract quantitative information

that enhance knowledge of road infrastructure and safety issues. More specifically, a strong demand is felt with respect to detection, localization and dimensional measurements of target objects. In the area of driving assistance systems, road scenes are acquired at video pace. Their interpretation, in real time, is aimed at providing information of benefit, for example, to vehicle guidance assistance systems or road obstacle detection systems.

Though these two applications feature different objectives and constraints, they also show some obvious common ground; in particular, an identification of the three-dimensional structure of scenes as information gets degraded by their projection onto the camera image plane is indeed mandatory. Techniques that make use of a single view often rely upon a reference plane, which restricts their domain of application to just the road plane. Moreover, this so-called *flat world* hypothesis is limited in practice by variations caused by the dynamics of the camera-mounted vehicle and the actual road geometry. For this reason, several teams from the Transport Ministry's scientific and technical network have shown keen interest in stereovision methods, i.e. to enable overlapping three-dimensional information based on several views, in taking advantage of the considerable body of research published in this field. Once the basic principles of stereovision have been recalled, this article will propose a summary description of developments introduced by the RST research teams, all in a spirit of cooperation.

PRINCIPLES OF STEREOVISION

Stereovision has generated a vast body of literature over the past twenty years. This section of the paper will offer a very brief state-of-the-art on existing passive vision methods, which refer to techniques that do not rely upon radiation projection but instead based solely on image content. For a more detailed presentation, the interested reader is referred to a number of summary articles such as **[1]** or to essential works: **[2]** for an overview, **[3]** for a more detailed presentation, and **[4]** for an introduction to more sophisticated methods, like stereovision in *N* views and self-calibration techniques. The similar *Structure From Motion* (SFM) approach, based on several successive views of a single camera in motion, has not been considered in this research effort due to the relative lack of a structured framework for observed scenes, the geometric sensor configuration and an excessive distance between successive images. Readers more interested in these aspects can still refer to Chapter 8 of publication **[2]** for a good introduction to the subject and to **[5]**, as an example, for application to an urban transport context.

The goal of stereovision is to compute the spatial position of points on the basis of their image coordinates in two different views, in the aim of either performing measurements or reconstructing the three-dimensional scene structure. Since the cameras are modeled by a central projection, all points positioned on a light ray are projected onto the same pixel p of the left-hand image **(Figure 1a)**. The image of this ray in the right-hand image is called the epipolar line, where *p'*, the correspondence pixel of *p*, needs to be identified in order to compute the position of *P* in space by means of triangulation, knowing the relative position and orientation of the cameras. The keys to stereovision thus revolve around 4 points: geometric sensor configuration, estimation of system geometry or calibration, matching, and reconstruction.

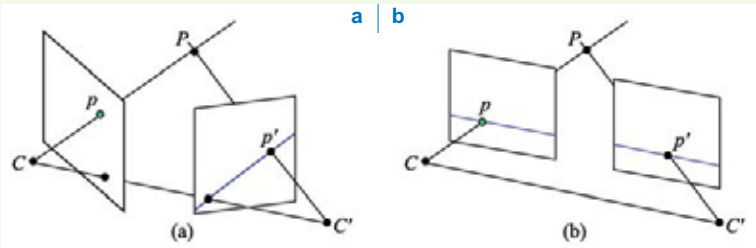


Figure 1
Epipolar geometry:
a: ordinary
b: rectified

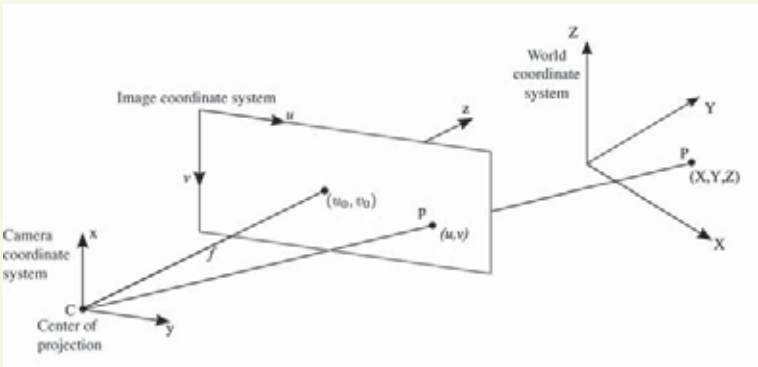
Geometric sensor configuration

It is first necessary to generate images from the same scene; this is classically achieved by using a set-up comprising 2 or, in some cases, 3 cameras (the use of 3 cameras serves to simplify the matching phase **[3]**). The simultaneity of photographs taken is an absolute necessity should the cameras or scene components be in motion. Special attention must therefore be paid to such issues in the overall design of onboard instrumentation.

Geometry: Calibration and rectification

Calibration consists of estimating the values of geometric model parameters **(Figure 2)**. Such parameters are of two types: extrinsic and intrinsic. Extrinsic or external parameters are geometric indications of camera orientation and position within a coordinate system associated with the scene, i.e. the so-called world coordinate system.

Figure 2
Geometric model (according to **[3]***)*



In the following discussion, the triplet *(X,Y,Z)* will designate the set of coordinates within this system and *(u,v)* will denote in a generic manner the column and row coordinates within the image. In the case of stereovision, index *l* or *r* will be used to specify whether the coordinates belong to the left or right view, respectively.

Intrinsic parameters, on the other hand, are internal to the cameras and enable describing their optical properties and the CCD matrix characteristics. The following mathematical notations have been employed:

- *(u0, v0)*projection of the optical center in the image;
- *f* focal length;
- t_u, t_v pixel size along u and v;
- $\alpha_u = f/t_u$ and $\alpha_v = f/t_v$; given the cameras used, $\alpha_u = \alpha_v = \alpha$ will be assumed.

These parameters are correlated with a geometric camera model and should not be expected to correspond exactly with values provided by the manufacturer. For this reason, they must be estimated, just like with the extrinsic parameters. According to the classical approach, calibration is based on a calibration pattern, containing points whose position is known both in the world coordinate system and within the image. The estimation is then conducted by means of nonlinear optimization techniques that have already been well identified. This calibration step offers access to the epipolar geometry of the sensor. It thereby proves possible to return to a simpler, so-called rectified, geometry, in which the epipolar lines are parallel and horizontal **(Figure 1b)**. It is now possible to compute, thanks to interpolation techniques, two new images, stemming from *virtual* cameras, with the same optical centers yet oriented in parallel. This situation can even be used to eliminate the radial optical distortion and equalize the intrinsic parameters of both cameras. Such a rectification step is also well identified; it features the advantage of simplifying the two subsequent steps of the process: matching and reconstruction.

Matching and reconstruction

In order to reconstruct the scene, it first proves necessary to identify in the two views the pixels corresponding to the same point in space. In rectified geometry and without radial distortion, the search for correspondence is confined to a 1D exploration on the horizontal lines of the image. Many matching methods exist, and we refer the reader to **[1]** for an introductory approach to the subject. Let's simply point out that the procedure may be dense (i.e. with the homologue of each image point actually being sought) or sparse (concentrating on key points such as corners, contours). The simplest and most widespread techniques are based on a local search for similarity between small regions, typically by optimizing a correlation function. The images can also be segmented and the resulting regions be matched. Local techniques are sensitive to occlusions (when a point is only visible on one of the two views), as well as to minimally-textured regions and repetitive patterns. Other more robust techniques, which tend to involve heavier computations, rely upon more global constraints than just resemblance, like order constraints or disparity gradient constraints. The most widespread techniques are based on dynamic programming optimization. Some recent approaches employ an explicit model of the scene; technically speaking, this stage no longer entails matching, but rather searching for a model position that yields an enhanced *explanation* of observed views by optimizing a criterion. This inverse problem can then be solved by means of variational methods.

Once the points have been matched, their horizontal position deviation can be measured. Conversely proportional to the depth of the point in the scene, this deviation is called disparity. Disparity maps, as well as the space where they dip play a key role in scene structure analysis. The properties of this disparity space **[6]**, also referred to as (u, v, disparity), are quite numerous and will be directly utilized depending on the application, as described in the following section. Moreover, since both disparity and the stereo sensor parameters are known, the last step consists of solving a system of 3 unknowns (position of the point in space) with 4 equations (relating to the coordinates of both points). This step becomes one of computing the intersection of light rays originating from the two corresponding pixels.

APPLICATIONS WITHIN THE TRANSPORT MINISTRY'S SCIENTIFIC AND TECHNICAL NETWORK

This section will offer a description of four applications using the stereovision technique; they involve: an analysis of road scenes to improve road network knowledge and safety studies, road obstacle detection, visibility distance estimation under adverse meteorological conditions, and reconstruction of the longitudinal road profile.

Moving towards a stereo version of the set of LPC road scene analysis tools

These stereovision techniques are applied for the purpose of conducting measurements beyond the pavement plane and thereby stretching the limitations of single-camera systems currently used for studies on road facilities and safety. The developments presented herewith predate the *stereovision* version produced by MLPC IRCAN hardware and associated operating software IREVE.

› Presentation of the experimental 3D vision sensor

ERA 27 from the Strasbourg regional Ponts et Chaussées Laboratory is equipped with an experimental stereo photography device for analyzing road scenes. Their set-up includes a passenger vehicle, instrumented by the CECP Prototype Design and Construction Center in Angers (western France). The vehicle comprises a topometer for measuring the vehicle's forward motion, an electrical installation, an industrial PC and a suspension bracket for attaching camera supports.

This bracket serves to easily adjust camera position and orientation. The spacing between the two cameras can be modulated from 0.50 and 1.50 m and fixed by means of stop blocks **(Figure 3)**. A dedicated software program has been developed by LCPC's Metrology and Instrumentation Division

Figure 3
The ERA 27 experimental vehicle designed by the Strasbourg regional Ponts et Chaussées laboratory (images courtesy of CECP)

for the purpose of assisting operators in performing preliminary camera alignment within a geometry as close as possible to the rectified epipolar geometry. The software also oversees the acquisition of stereo pairs. Let's point out that the two cameras are identical to those from the MLPC IRCAN hardware and that development of the experimental software relies upon the same libraries as those called up by the software associated with IRCAN. This situation should allow for a faster transfer towards an operational application. As previously mentioned, synchronization proves critical in the case of onboard equipment and for that reasons has been physically introduced into this set-up. The electrical pulse output at regular distance intervals by the topometer is duplicated and sent to the cameras, which are then able to freeze their images. The two views will then be returned onto the PC via an IEEE 1394 interface. The equipped vehicle has enabled demonstrating system feasibility: 1 280 × 1 024-pixel stereo pairs can be recorded at 5-m intervals for vehicle speeds reaching 110 km·h^{-1}. On this basis, the development of a stereo version of the MLPC IRCAN package is underway. In conjunction with this effort, adaptation of the system to cameras with a resolution of 1 920 × 1 024 pixels and extension to shooting 3 or 4 simultaneous views will also be pursued.

› Experimental operating software designed for stereovision

Prior to being run, the stereo pairs acquired using the equipment described above must first undergo processing. As a preliminary, each camera is calibrated according to classical techniques, accessible from the MLPC IREVE software. Note that an automatic technique described in **[7]** has been used herein to simultaneously detect the ground marking lines that make up the calibration patterns.

Upon completion of this calibration step, the epipolar geometry is known; it then becomes possible to perform a software-based epipolar rectification, since the mechanically-generated alignment was naturally imperfect. This transformation is carried out on each acquired stereo pair. At the same time, the images obtained are corrected for optical distortion and, if necessary, compressed **(Figure 4)**. Here once again, compatibility with operational tools has been directly incorporated into the software development, so as to ensure quick transfer to IREVE.

A demonstrator program, called STEREO, was also launched by ERA 27; this program enables the use of stereo pairs for performing road network analysis. Its main functionalities include: naviga-

Figure 4
Rectified stereo pair, following distortion correction. The green line in the right-hand image traces the epipolar line associated with the point marked by a cross in the left-hand view.

tion within the image sequence, matching and dimensional measurement, as well as depth map computation.

The point-matching technique employed in this application is based on the zero mean normalized cross-correlation (ZNCC) between windows of variable dimensions. Furthermore, this technique applies the uniqueness constraint **[3]**, i.e. only the correspondences detected in both the left-right and right-left directions are to be validated. Shape matching has also been introduced for geometric objects such as columns, panels and miscellaneous supports (**Figures 5** and **6**, left view).

The computation of disparity maps, whether dense or sparse, is also based on correlation techniques. An analysis of these maps, according to the u/v-disparity technique, as proposed by the Laboratory on Vehicle-Infrastructure-Driver Interactions (LIVIC) and discussed in the following section, highlights vertical objects present in the scene (**Figure 6**, right view). The depth map measurement and computation functionalities described in this section are aimed at being integrated into an upcoming version of the IREVE operating software, as part of the stereo IRCAN development effort.

Figure 5
Graphic interface of the STEREO demonstrator software. Measurement of the dimensions of a J6 beacon (left); point-specific matching and diagram of (dis)similarity (right).

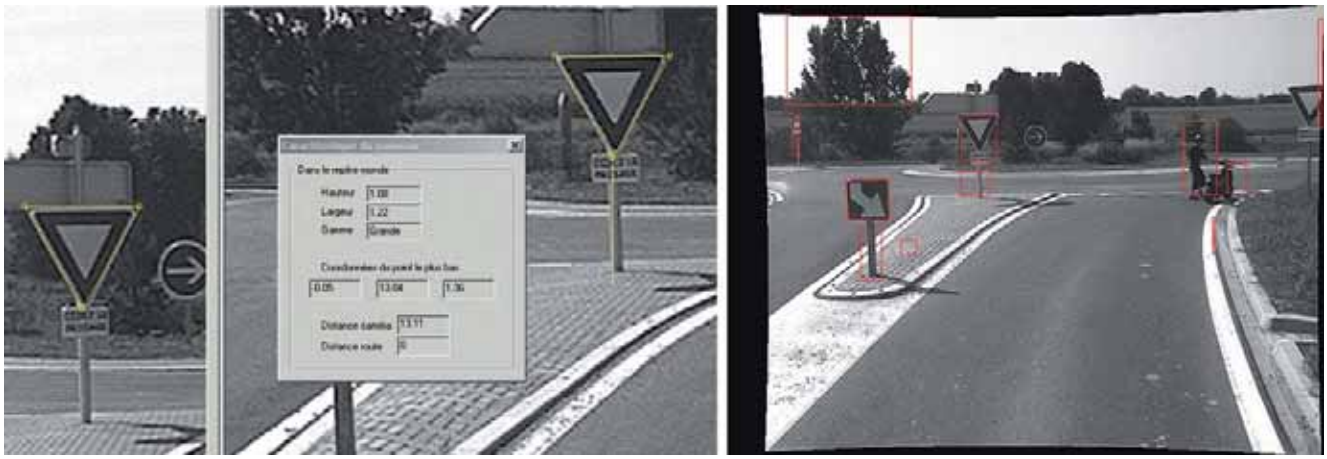


Figure 6
Matching of a triangular shape and measurement of the dimensions of an AB3a panel (left); detection of vertical objects by means of disparity map analysis (right)

Obstacle detection on pavements by means of u/v-disparity projection

The automatic detection of objects located on the pavement in front of the vehicle, is a key task in furthering the development of driving assistance systems; as an example, this capability serves to set up emergency braking or avoidance systems. In this aim, the LIVIC has devised an original approach, called v-disparity **[8, 9]**, which will be described below. Its guiding principle consists of extracting the longitudinal road profile and then searching for the shapes located above this profile.

› Principle implemented

According to the proposed approach, it is considered that the road comprises a succession of portions of both horizontal and oblique planes with respect to the stereoscopic sensor plane. As such, an object positioned on the pavement (vehicle, pedestrian, tree, etc.) is characterized by a vertical or quasi-vertical portion of plane in the scene **(Figure 7)**. The objective then entails seeking such planes based on a stereoscopic view of the scene. The originality and robustness of the method lies in the procedure by which this particular step is executed.

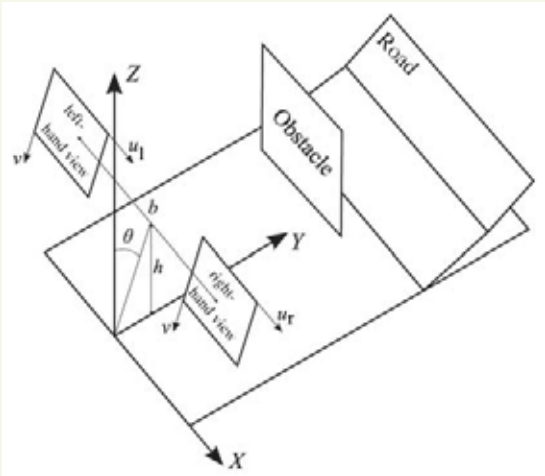


Figure 7
Geometric configuration of the stereovision sensor and representation of the scene by use of horizontal and oblique planes for the road, and vertical planes for the obstacles

The images are assumed to be corrected of any radial distortion and the method begins within a rectified geometry, where the intrinsic camera parameters are considered to be identical. In this configuration, the epipolar lines are parallel and overlap with the image scanning lines. The projection of a point P of the scene thus lies on the same line in both the left and right-hand views, i.e. $v_r = v_l = v$, where:

$$v = \frac{(v_0 \cos\theta + \alpha \sin\theta)Y + (v_0 \sin\theta - \alpha \cos\theta)(Z - h)}{Y \cos\theta + (Z - h)\sin\theta} \quad (1)$$

in denoting the following, for the extrinsic sensor parameters **(Figure 7)**:

– θ the angle between the direction of the camera optical axis and the horizontal (pitch);
– h the camera height with respect to the ground;
– b the spacing between cameras (also called the stereoscopic *base*).

The disparity Δ for point P, i.e. the column position difference in the P projection between left and right views, is given by the following equation:

$$\Delta = u_l - u_r = \frac{\alpha b}{Y \cos\theta + (Z - h)\sin\theta} \quad (2)$$

Let's assume a plane described by the equation $Y = aZ + d$ within the scene (according to the geometric configuration shown in **Figure 7**) and any point M on this plane, with coordinates $(X, aZ + d, Z)^T$. By expressing Z as a function of v, derived in Equation **(1)**, and by substituting this value into Equation **(2)**, it is shown that the disparity of M equals:

$$\Delta_M = \frac{b}{ah + d}(v - v_0)(\sin\theta + a\cos\theta) + \frac{b}{ah + d}\alpha(\cos\theta - a\sin\theta) \quad (3)$$

Accordingly, all points on the plane of equation $Y = aZ + d$ are projected onto the straight line given by Equation **(3)** within a space (v, Δ). Let's now note that if $a = 0$ in **(3)**, the equation obtained

corresponds to the projection of a vertical plane located at distance $Y = d$ (e.g. a plane defining an obstacle), and if a tends to infinity, the equation obtained corresponds to the projection of a horizontal plane Z = 0 (e.g. plane defining the road).

At this point, let's consider an image with a coordinate system positioned on the upper left-hand side, with an x-axis indicating disparity and a y-axis, oriented downwards, representing the image line number. This image, denoted IvD, will be called the v-disparity projection of disparity ID. This search for scene planes that contain the x-axis consists of finding straight lines in the IvD projection image. In the current LIVIC sensor configuration, i.e. with $b = 1$ m, $h = 1.40$ m and $\theta = 5°$, the straight line corresponding to an obstacle is just about vertical (i.e. slope of less than 0.1), while the straight line corresponding to the road displays a much different slope (0.7). In this manner, we are able to make the transition from a three-dimensional representation of planes to a two-dimensional one, which turns out to use much less computing time to analyze and moreover allows for much more robust extractions, thanks to the accumulation effect provided by this latter representation.

› Road slopes

Since the search for road straight lines takes place within the v-disparity space, the initial step must be to perform matching between information stemming from the stereo pair in order to estimate the disparity. Given the very stringent constraints in terms of computation time, a sparse disparity map (solely along the contours) is built thanks to the simplest possible realigned similarity measurements.

The contour extraction phase in both views (based on a gray-level gradient computation) yields images IG_r and IG_l shown in **Figure 8**. A comparison of windows along these contours then enables associating a disparity value, resulting in map ID, presented in the lower left-hand view on **Figure 8** (the color chart employed serves to highlight decreasing disparities, over the spectrum from blue to red).

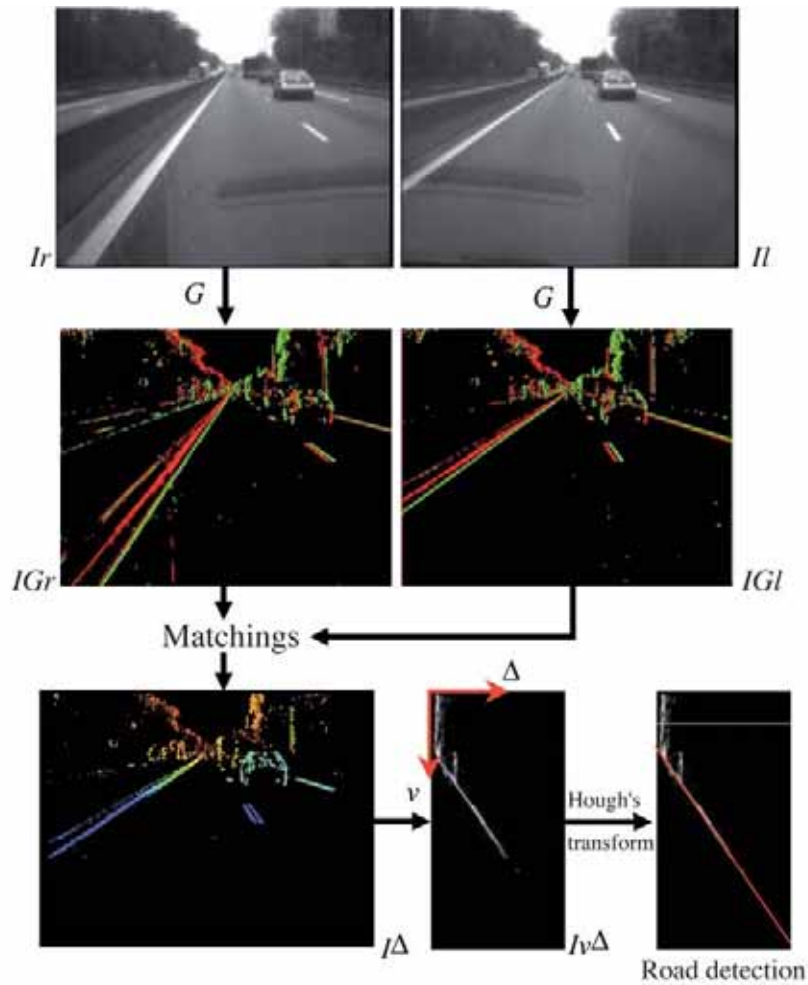


Figure 8
Flowchart of the detection method employed. A gradient extraction (G function) highlights the contours of perceived shapes (red: black-white transitions, green: white-black transitions). The matching of contours between images IGr and IGl yields the ID disparity map. Colors range from blue to red when making the transition from maximum to minimum disparity. The disparity map accumulation technique along the image lines provides the v-disparity projection, IvD. Lastly, a line segment extraction technique (Hough's transform in this case) reveals the road line in the v-disparity.

For each image line of $I\Delta$, points with the same disparity are then summed, which provides the v-disparity projection image IvD. For image line j therefore, the x-coordinate u_M of point M in IvD corresponds to disparity Δ_M and its gray level I_M corresponds to the number of points featuring the same disparity Δ_M on line j:

$$I_M = \sum_{Pixels} \delta_{v_M, j} \delta_{\Delta_M, \Delta_{u,v}} \quad (4)$$

where $\delta_{i,j}$ denotes Kronecker's symbol ($\delta_{i,j} = 1$ if $i = j$; $\delta_{i,j} = 0$ otherwise).

As pointed out above, the planes composing the pavement are projected in straight lines in IvD, whose equations are correlated with plane parameters (see the IvD image in **Figure 8**). As a result of the IvD construction, the complex 3D plane detection problem can be reduced to a much simpler alignment detection problem within a 2D space. Straight line detection in IvD may be performed by means of robust methods, like Hough's transform or an M-estimator regression (see the oblique straight line representing the road in **Figure 8**).

When the road is flat, it appears as an oblique straight line in IvD, which makes the longitudinal road profile a straight line. In the presence of an ascent or a descent, the road profile becomes a simple curve (a cylindrical surface) that may be approximated by either a polynomial or a piecewise linear curve, with the surface thus being modeled by a succession of plane portions. Instead of searching for a peak in Hough's transform, the method records the m maximum values lying in the vicinity of a peak. These m selected points correspond to m straight line segments in IvD.

The target piecewise linear curve is either the upper envelope (when approaching an ascending slope) or lower envelope (when approaching a descent) of the family of the m generated lines. The choice between these two is made by identifying the envelope that superimposes the best with the pixels contained in IvD. This process involves progressing along each of the two curves and summing the value of the . This process involves progressing along each of the two curves and summing the value of the IuD pixels encountered, with the curve ultimately selected being the one that yields the highest cumulative value **(Figure 9)**.

Figure 9
Extraction of the longitudinal road profile. From left to right: an image of the stereo pair corresponding to an upcoming descent, the corresponding IvD image, the associated Hough's transform (the white rectangle encompasses the search zone for the k highest values in Hough's space), the bundle of m straight lines generated, the two calculated envelopes, and the longitudinal profile of the downward-sloping road.

› Road obstacle detection

The intended objects correspond to vertical straight lines located above the curve showing the pavement within the v-disparity space. The base of these straight lines corresponds to the point of intersection between the considered plane and the pavement surface (see the intersection of the two red lines in the v-disparity image on **Figure 10**). Moreover, segment height provides the object height. In analyzing the segments representing vertical planes of the scene, both the base and summit of a box bounding the targeted object are obtained.

In order to identify the sides of this bounding box, it is necessary to detect the edge corresponding to the summit of the considered vertical plane. To proceed, the IuΔ projection, called u-disparity,

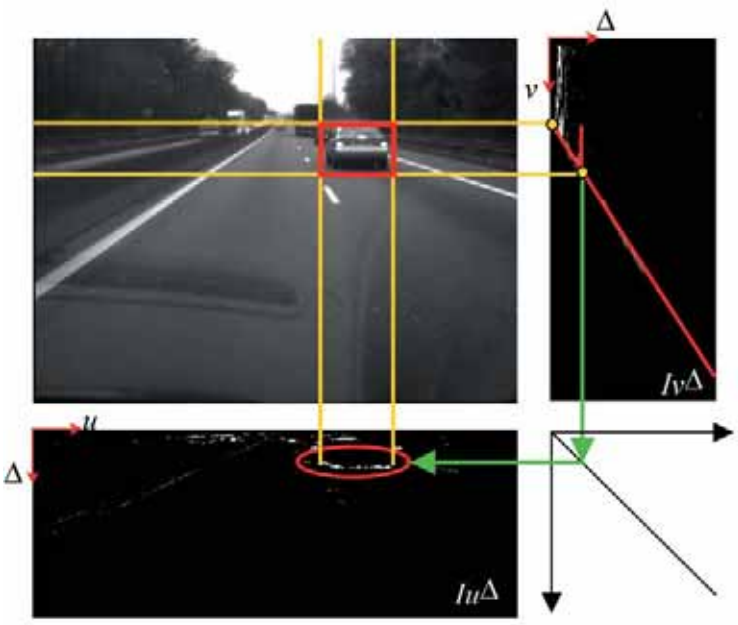


Figure 10
Localization of each object on the road. The straight vertical segment corresponding to the plane of the targeted object (in back of the vehicle in this image) detected in the v-disparity image IvD offers the base and height of the given object. This same plane seen from above produces a straight horizontal segment in the u-disparity image IuD (visible in the red ellipse), which provides an indication on the object extension widthwise. The object detected on the road is then framed by a bounding box.

is introduced. Like for v-disparity, u-disparity is built by summing points with the same disparity, yet in this case in each column *u* of the disparity image (see the lower left-hand image shown in **Figure 10**).

This cumulative projection will serve to draw straight line segments within the u-disparity space. The ends of each segment are in fact the sides of the boxes bounding the targeted objects. Within the scope of defining driving assistance systems, this technique has been successfully implemented for the onboard detection of vehicles or pedestrians **(Figure 11)**.

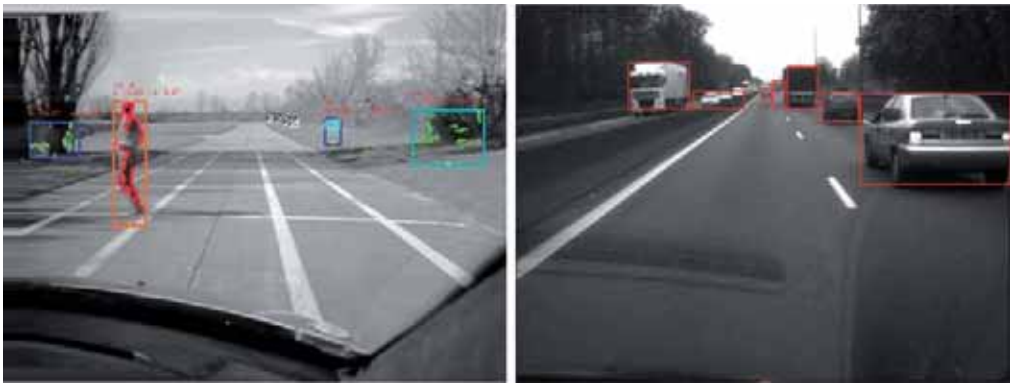


Figure 11
Detection of obstacles on the road and vertical elements contained in the scene (as indicated by the bounding boxes)

Contextual visibility and stereovision

In the presence of poor meteorological conditions, fog in particular, onboard camera operations are far less efficient. Similarly, driver perception becomes altered. Being able to detect deteriorated sensor operations or inappropriate driver behavior, given the meteorological conditions, proves to be a critical feature. The tools presented herewith, as developed at the LIVIC in collaboration with LCPC-ESE (LCPC's Operations, Signaling and Lighting Division), describe such a contextual visibility diagnostic method with the help of a stereoscopic sensor.

› Mobilized and mobilizable visibility distances

Under foggy weather conditions, visual information gets modified. A drastic drop is found in contrast vs. observation distance. This effect may be described by the meteorological visibility V_{met}, defined as the greatest distance at which a suitably-sized black object can be recognized during the

daytime against the background sky. A derived and patented method estimates this meteorological visibility using just a single camera **[10, 11]**.

This visibility distance however is merely the reflection of an atmospheric parameter and does not incorporate the type of road scene. As such, two new visibility distances have been defined: the mobilized visibility distance V_{mob}, which indicates the distance to the object lying furthest on the pavement yet still visible; and the mobilizable visibility distance V_{max}, i.e. the maximum distance at which a potential object on the pavement could be seen. For a visibility threshold of 5% contrast, an order-of-magnitude relationship among the three aforementioned distances was established, i.e. $V_{mob} < V_{max} \leq V_{met}$ **[12]**.

› Estimation of mobilized visibility

In order to estimate the mobilized visibility distance, a three-stage process was engaged:

❶ Construction of a precise road surface depth map

❷ Detection of the image elements whose contrast exceeds 5%

❸ Combination of information from the first two stages.

Computation of a precise road scene depth map by means of stereovision

In two iterations, the v-disparity approach yields a high-quality disparity map of the road surface **(Figure 12)**. To proceed, once the profiles have been extracted from the v-disparity image, each couple of previously-matched pixels is verified as belonging or not to one of the profiles. In this manner, the number of false matchings can be tremendously reduced **(Figure 12d)** **[13]**. This technique was utilized for the purpose of measuring the mobilized visibility distance.

Nonetheless, this kind of disparity map is quite sparse, meaning that disparity is only known on the vertical contour points. Under adverse visibility conditions, this would cause problems, in particular at the level of the summit of vertical objects, such as vehicles, where many false matchings could arise.

To remedy this shortcoming, a complementary approach has been developed; it introduces a disparity propagation technique, in a way analogous to regional growth, like in **[14]**, where germs composed of matchings assumed to be correct subsequent to an initial algorithm iteration are

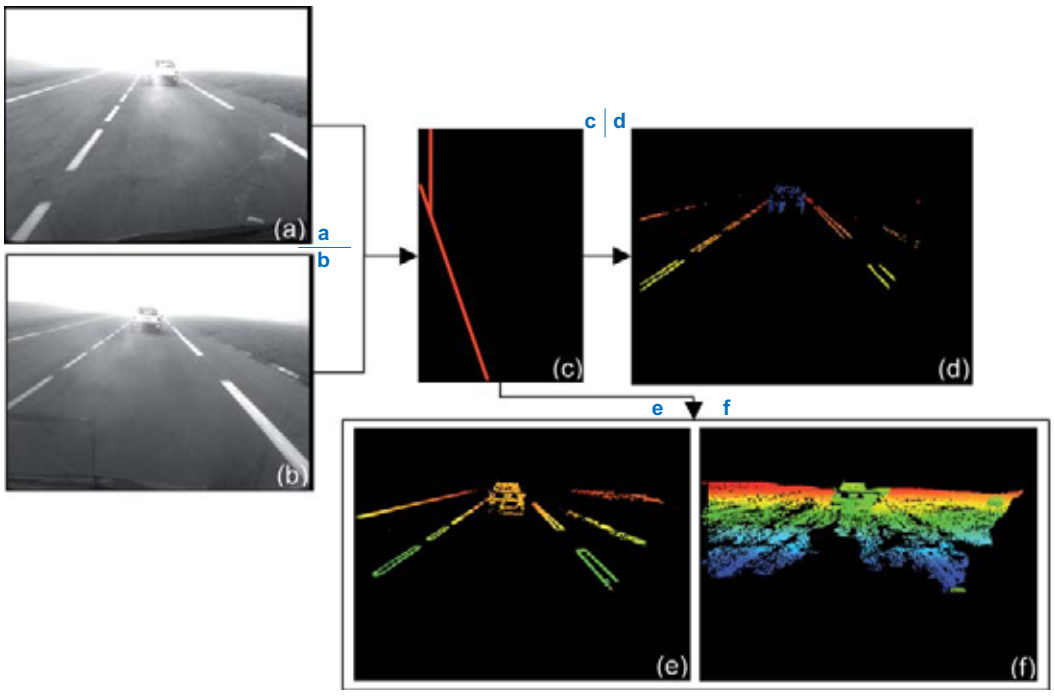


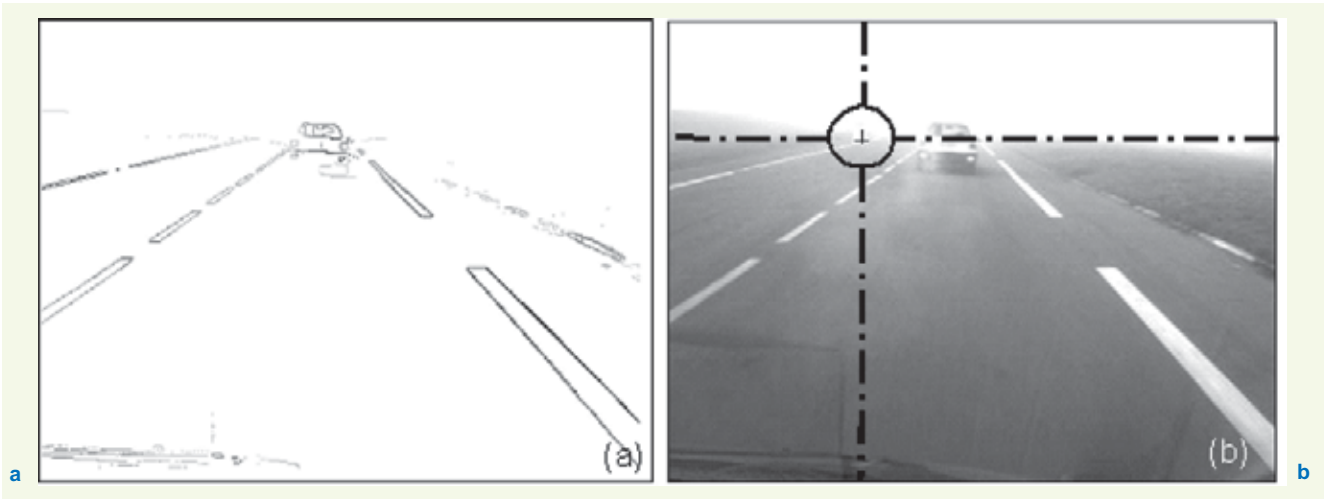
Figure 12
Overview of the v-disparity approach in two passes on a pair of stereo images acquired during foggy weather:
a: left-hand image
b: right-hand image
c: v-disparity profile calculated based on images (a) and (b)
d: improved sparse disparity map
e and f: examples of quasi-dense disparity maps obtained using two passes (see text for description)

propagated. This technique is also inspired from the second algorithm iteration described in **[13]**, where matching is validated both by using the lines extracted from the v-disparity and by modeling the road profile. More specifically, the proposed method propagates matchings after having verified that they belong to one of the v-disparity image profiles. Thanks to this method, the disparity map obtained becomes quasi-dense **(Figures 12e** and **12f)**, especially on the horizontal contours. In comparison with the direct computation of dense disparity maps, this method offers the advantage of being less costly in terms of computation time; moreover, it allows for more precise positioning, by means of the u-disparity approach, of the bounding boxes around road objects **[15]**.

Figure 13
Result from application of the method for estimating mobilized visibility distance:
a: local contrasts exceeding 5%
b: estimation of the mobilized visibility distance; the furthest window displaying a contrast of at least 5% and over which depth is known has been painted white; the point is marked by a black cross on this same window.

Estimation of contrast exceeding 5%

The local contrast computation in an image has not been addressed to any considerable extent in the literature. By relying upon a classical image segmentation method, an original approach has been developed; it consists of scanning the image through use of small windows cut in two along a boundary so as to maximize contrast, from a logarithmic standpoint, between the two cut parts **(Figure 13a)**. This method has been compared with relatively obscure existing methods and proves to be just as precise while committing fewer false detections **[16]**.



Quick combination of the two pieces of information

The objects encountered move closer to the instrumented vehicle the further down they appear in the disparity map. For this reason, the algorithm entails scanning the disparity map from top to bottom, beginning at the horizon line and then computing the contrast for each pixel of known disparity. Once a pixel with a contrast exceeding 5% has been found, the computation routine stops. By knowing the disparity of this point, its depth can be derived, and this constitutes the mobilized visibility distance **(Figure 13b)**. This method has been detailed in **[12]** and patented **[17]**.

It was compared with the first method to enable estimating meteorological visibility through use of reference data. The two methods show nearly equal precision for meteorological visibility distances of between 50 and 200 m. These reference data were obtained on a specific site, designed and built at the Satory track, which contains large-sized targets for estimating the meteorological visibility distance given the attenuation of their contrast vs. distance **(Figure 14)**.

For these specific tests, the cameras were adjusted in a realistic manner as regards use conditions, i.e. minimum exposure time in order to reduce motion blur and reliance upon an auto-iris lens to avoid both over- and underexposures. Details of these validation experiments are provided in **[18, 19]**.

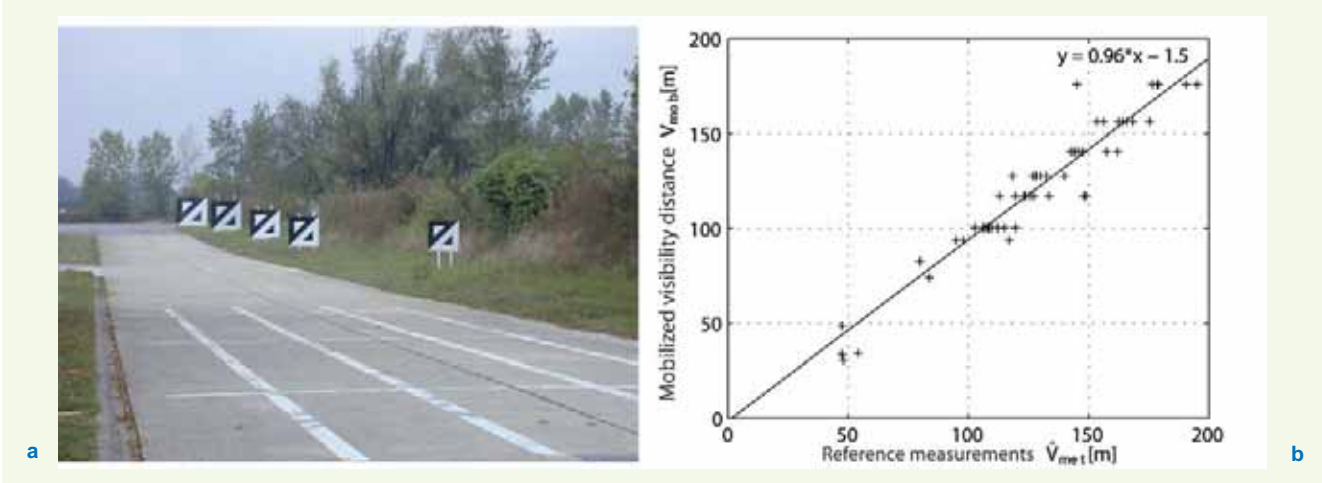


Figure 14
Systematic evaluation of results:
a: photograph of the Satory experimental validation site under good weather conditions
b: points: estimation of the mobilized visibility distance vs. the reference visibility distance obtained using large-sized targets. Right: Least-squares adjustment line, whose equation is listed on the graph. The coefficient of linear correlation equals approximately 97%.

Robust matching for road profile reconstruction

This discussion will refocus on the problem of accurately determining the road profile, which offers multiple applications from driving assistance features to an evaluation of geometric visibility distances. According to the v-disparity approach described above, the road is modeled by a succession of oblique planes (*see* **Figure 7**), which serve to define straight lines within the v-disparity space. After matching contour points, the method proceeds in two stages, i.e.: construction of a scene representation in the v-disparity space; then, adjustment of a bundle of straight lines within this space (**Figure 9**).

According to the approach initially proposed in **[20]**, the piecewise plane model is replaced by a polynomial model. In employing the notations from **Figure 7**, the following can be expressed:

$$u_r = u_l + a_0 + a_1 v + \cdots + a_n v^n = u_l + A^T V(v) \quad (5)$$

It is to be noted that this approach, which extends the linear model **(3)**, implicitly uses the v-disparity representation, since **(5)** is equivalent to $\Lambda_M = A^T V(v)$. Yet, rather than building the cumulative v-disparity image, the method simultaneously performs point matching and road profile reconstruction *via* the robust estimation of polynomial coefficients A. This problem set-up enables introducing the notion of confidence into model adjustment through the covariance matrix associated with the estimation. Such information may prove valuable, e.g. when integrating profiles obtained on successive image pairs in order to reconstitute road relief over a certain distance.

Lastly, it should be remarked that the model shown in **Figure 7** only includes angle θ. One advantage of the representation in **(5)** is its ease of extension for better representing actual pavements. As an illustration, the following is chosen as a model in the space (u,v,disparity):

$$u_r = u_l + A^T V(v) + a_u u_l \quad (6)$$

where a_u is an additional coefficient to be estimated. This model turns out to be well adapted for incorporating a roll angle (rotation around the y-axis): it corresponds to a first-order approximation of the projection of a polynomial pavement model with an incline. Similarly, by multiplying the additional term $a_u u_l$ by a polynomial in v, it becomes possible to include an incline variable with distance, without changing the nature of the algorithm.

› Formulation of the optimization problem

Equation **(5)** relates the coordinates of point pairs that correspond between the left- and right-hand images, via a polynomial whose coefficients A must be estimated. The correspondence between points is obviously unknown upfront and must also be determined. The classical approach adopted to solve this type of problem consists of minimizing the quadratic error between the left image after application of the target transform and the right image, which is set as the reference. Since the information contained inside regions of constant intensity within the images is not discriminating, all image points are not considered, but only those contour points extracted from images, which substantially limits problem complexity.

Let's denote *(i,j)* and *(k,j)* as the point coordinates potentially in correspondence on each line *j* of the left and right contour images. The method proposed herein consists of minimizing a non-quadratic error function, written as follows:

$$e(A) = \sum_{(i,j),k} r\big((i,j),(k,j)\big)\phi\left(\frac{1}{2s^2}\left(k + A^T V(j) - i\right)^2\right) \qquad (7)$$

where $r((i,j),(k,j))$ is a factor introduced to account for local similarities (correlation or ZNCC in gray levels, a decreasing function of the color gradient difference) between pixels in correspondence *(i,j)* and *(k,j)*; moreover, $\phi(n^2) = \varphi(n)$ is a robust potential function, for which n is a scale parameter.

The role of function ϕ is to minimize the influence of erroneous correspondences on the estimation of parameters *A*. Such functions are known in the field of robust statistics **[21]** under the name M-estimators; they also get used for semi-quadratic regularization in image reconstruction **[22,23]** and called in this case φ-function models. A few typical examples of ϕ functions are shown in **Figure 15**.

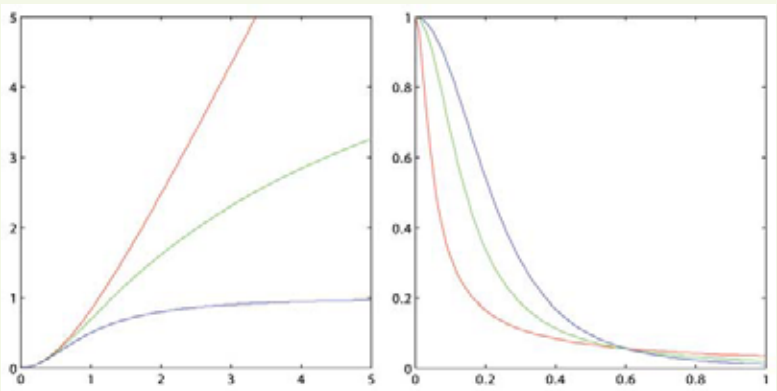


Figure 15
Examples of functions $\phi(n^2) = \varphi(n)$ (left) and associated weighting functions $\phi'(n^2) = \varphi'(n)/2n$ (right, with an arbitrary weighting coefficient on the x-axis u)

› Estimation algorithm

In order to minimize **(7)** with respect to *A*, inspiration is derived from semi-quadratic theory **[22, 23]** to linearize the problem and propose a simple iterative algorithm. The Lagrangian formalism proposed in **[24]** and detailed in **[25]** is to be applied; this step requires that the ϕ function satisfies hypotheses imposed by semi-quadratic theory, i.e. the function must be defined and continuous over $[0,+\infty[$, as are both its first and second derivatives, and be increasing and concave. For further details on the formalism introduced for algorithm derivation (which ensures algorithm convergence at a local criterion minimum), the interested reader is referred to **[7, 24, 25]**. The iterated, weighted least squares algorithm generated can be written as follows:

❶ Initialize A_0 and set $t = 1$

❷ For all indices *(i,j,k)*, compute the auxiliary variables:

$$w_{i,j,k} = \frac{1}{2s^2}\left(k + A_t^T{}_1 V(j) - i\right)^2$$

which represent the quadratic calibration error on each pair of points.

❸ For all indices *(i,j,k)*, compute the weightings (Lagrangian coefficients):

$$\lambda_{i,j,k} = r\left((i,j),(k,j)\right)\phi'\left(w_{i,j,k}\right)$$

❹ Solve the following linear system in order to determine A_t :

$$\sum_{i,j,k} \lambda_{i,j,k} V(j) V(j)^T A_t = \sum_{i,j,k} \lambda_{i,j,k} (k-i) V(j)$$

❺ If $| A_t - A_{t-1} | > \varepsilon$, increment *t* and return to Step 2. Otherwise, set $A = A_t$

Steps 2 and 3 introduce a *soft* matching between the points of indices *i* and *k* on line *j*, via the *score* provided by $\lambda_{i,j,k}$. This approach incites correspondences both in sync with the model, i.e. associated with a small quadratic error (which means that weighting ϕ' lies close to 1, *see* **Figure 15**), and photometrically consistent (the value of *r((i,j),(k,j))* then also lies close to 1). Next, Step 4 updates the estimation of the road model coefficients *A*, by ascribing greater weight to the *highly-scored* matchings.

It is to be noted that the parameter s enables a multi-scale search, within a continuation approach called gradual non-convexity (GNC) **[26]**: an initial estimation is performed at a rather coarse scale, with this result then serving as an initialization on a more refined scale, and so forth. This heuristic enables significantly improving algorithm convergence. Furthermore, let's indicate that in practice, in order to accelerate the algorithm, indices *(i,k)* are not all considered for each line: *r((i,j),(k,j))* gets cancelled when its value is small (the corresponding weighting value $\lambda_{i,j,k}$ amounts to nearly zero, which brings about practically no change at all).

› Experimental results

Figure 16 illustrates algorithm robustness in the presence of many perturbations caused by vertical elements such as trees, houses or vehicles. The proposed algorithm is a robust estimator on the matchings between points. The robust function ϕ may be chosen so as to accept up to 50% of erroneous matchings **[25]**. Let's note that a large share of these erroneous data may be eliminated *a priori* by using the pavement segmentation technique described in **[27]**. **Figure 16** also reveals the benefit of applying a polynomial model: the left-hand transform image contours overlap much better with those from the right-hand image than when employing a plane model. In this experiment, the model in **(6)** has been implemented in order to incorporate the roll angle, which enhances results even further.

Lastly, a recalibration experiment was performed between successive local profiles **(Figure 17)**. Once the adjustment has been carried out, it is indeed possible to deduce the spatial point positions, particularly their altitude. Since the same point had been visible in several stereo pairs, it also proves possible to recalibrate the various profiles between themselves by means of an iterative procedure, which generates a global profile over the studied itinerary. It is obvious that this procedure is subject to alteration, and the profile derived is quantitatively quite removed from that capable of being extracted from the itinerary map. Yet, from a qualitative perspective, the profile corresponds to the relief actually observed. These preliminary results are thus promising from the standpoint of integration into a more complete system yielding absolute information, e.g. *via* a GPS.

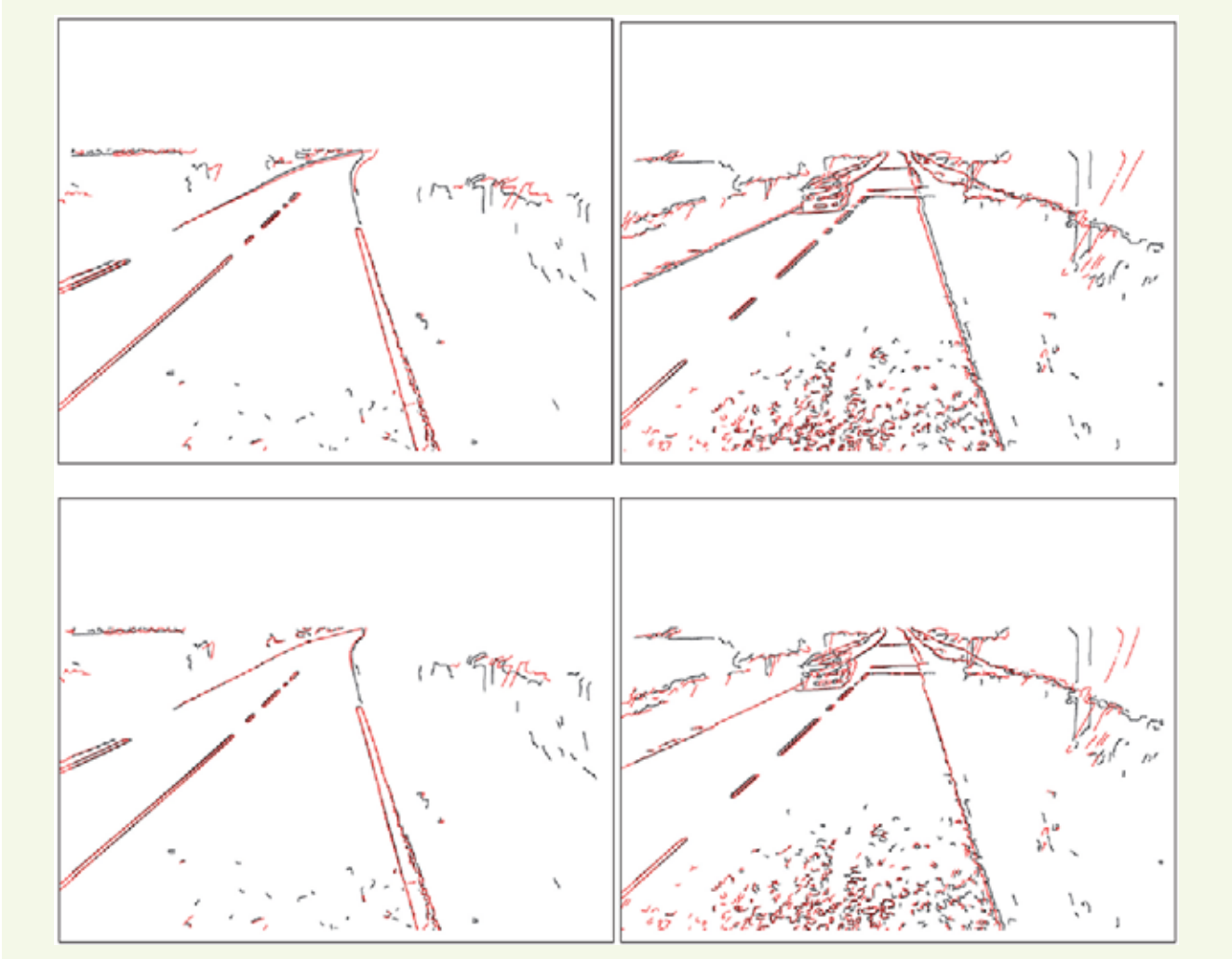


Figure 16
Comparison of the readjusted contour map (shown in red) with respect to the reference map (black) for two pairs of stereo images. Top: a plane model is used; bottom: a 6-order polynomial model is applied. The roll angle has also been included.

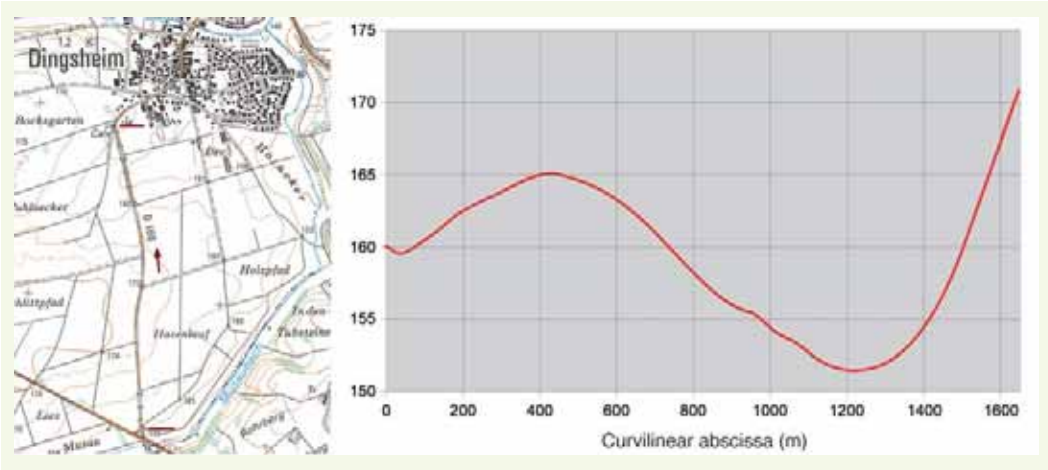


Figure 17
Map of the studied itinerary: the local RD 166 road (France's Bas-Rhin department) traveled in the direction of the arrow, between the intersection with RD 41 (lower horizontal red line) and the Dingsheim Calvary monument (upper horizontal line, IGN. Right: elevation profile obtained by means of recalibrating local profiles. The absence of absolute recalibration causes a shift with respect to the profile potentially extracted from this map.

CONCLUSION

In this article, we have proposed a summary of the various research projects devoted to applying stereovision techniques to an analysis of road scenes, as recently conducted by several teams belonging to the Transport Ministry's RST network. These efforts have enabled developing a unified methodological framework; moreover, they have led to operational or nearly operational solutions, whether specific to facility studies, through the development of a new generation of IRCAN and IREVE tools from the Ponts et Chaussées Laboratories, or to driving assistance systems, through the devel-

opment of patented visibility distance estimation tools or onboard road obstacle detection systems. Stereovision techniques offer a general interest, which extends beyond the domain of road scene analysis, and many dimensional measurement and shape analysis applications could be enhanced within the field of civil engineering.

REFERENCES

1 **Brown M.**, **Burschka D.**, **Hager D.**, Advances in computational stereo, *IEEE Transactions on Pattern analysis and machine intelligence*, **2003**, volume **25**, **8**, pp. 993-1008.

2 **Trucco E.**, **Verri A.**, *Introductory techniques for 3-D computer vision*, Prentice Hall, Upple Saddle River, USA, **1998**, 343 pages.

3 **Horaud R.**, **Monga O.**, *Vision par ordinateur – Outils fondamentaux – 2e édition revue et augmentée*, Traité des Nouvelles Technologies, Série informatique, Hermès, Paris, **1995**, 425 pages.

4 **Hartley R.**, **Zisserman A.**, *Multiple view geometry in computer vision*, Cambridge University Press, Cambridge, Angleterre, **2000**, 607 pages.

5 **Leibe B.**, **Cornelis K.**, **Van Gool L.**, Dynamic 3D Scene Analysis from a Moving Vehicle, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, Minneapolis, USA, juin **2007**.

6 **Tarel J.-P.**, Global 3D Planar Reconstruction with Uncalibrated Cameras and Rectified Stereo Geometry, *Machine Graphics & Vision*, **1997**, volume **6**, **4**, pp. 393-418.

7 **Tarel J.-P.**, **Charbonnier P.**, **IengS.S.**, Simultaneous Robust Fitting of Multiple Curves, *International Conference on Computer Vision Theory and Applications (Visapp 2007)*, Barcelone, Espagne, 8-11 mars **2007**, pp. 175-182.

8 **Aubert D.**, **Labayrade R.**, Détection d'obstacles routiers par stéréovision : l'approche v-disparité, numéro spécial « systèmes de transports intelligents », *Annales des Télécommunications*, **2005**, **3-4**, volume **60**, pp. 299-325.

9 **Labayrade R.**, **Aubert D.**, Robust and fast stereovision based obstacles detection for driving safety assistance, *IEICE Transactions on Information and Systems, Special Section on Machine Vision Applications*, **2004**, volume **E87-D**, **1**, pp. 80-88.

10 **Hautière N.**, **Tarel J.-P.**, **Lavenant J.**, **Aubert D.**, Automatic Fog Detection and Estimation of Visibility Distance through use of an Onboard Camera, *Machine Vision and Applications Journal*, **2006**, volume **17**, **1**, pp. 8-20.

11 **Lavenant J.**, **Tarel J.-P.**, **Aubert D.**, Procédé de détermination de la distance de visibilité et procédé de détermination de la présence d'un brouillard, *Brevet français #0201822 soumis par LCPC / INRETS*, **2002**.

12 **Hautière N.**, **Labayrade R.**, **Aubert D.**, Real-Time Disparity Contrast Combination for Onboard Estimation of the Visibility Distance, *IEEE Transactions on Intelligent Transportation Systems*, **2006**, volume **7**, **2**, pp. 201-212.

13 **Labayrade R.**, **Aubert D.**, In-vehicle obstacles detection and characterization by stereovision, *International Workshop on In-Vehicle Cognitive Computer Vision Systems*, Graz Autriche, **2003**.

14 **Lhuillier M.**, **Quan L.**, Match propagation for image-based modeling and rendering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2002**, volume **24**, **8**, pp. 1140-1146.

15 **Hautière N.**, **Labayrade R.**, **Perrollaz M.**, **Aubert D.**, Road scene analysis by stereovision : a robust and quasi-dense approach, *IEEE International Conference on Automation, Robotics, Control and Vision*, Singapore, **2006**.

16 **Hautière N.**, **Aubert D.**, **Jourlin M.**, Mesure du contraste local dans les images, application à la mesure de distance de visibilité par caméra embarquée, *Traitement du Signal*, **2006**, volume **23**, **2**, pp. 145-158.

17 **Hautière N.**, **Labayrade R.**, **Aubert D.**, Dispositif de mesure de distance de visibilité, Distance de visibilité par stéréovision, *Brevet français #0411061 soumis par LCPC/INRETS*, **2004**.

18 **Hautière N.**, **Aubert D.**, **Dumont E.**, **Tarel J.-P.**, Validation expérimentale de méthodes dédiées à l'estimation embarquée de la visibilité atmosphérique, *Actes des journées scientifiques du LCPC*, 4es Journées des Sciences de l'Ingénieur, Marne-la-Vallée, **2006**.

19 **Hautière N.**, **Aubert D.**, **Dumont E.**, **Tarel J.-P.**, Experimental Validation of Dedicated Methods to In-Vehicle Estimation of Atmospheric Visibility Distance, *IEEE Transactions on Instrumentation and Measurement*, **2008**, volume **57**, **10**, pp. 2218-2225.

20 **Tarel J.-P.**, **Ieng S.S.**, **Charbonnier P.**, Accurate and robust image alignment for road profile reconstruction, *International Conference on Image Processing (ICIP'07)*, San Antonio, USA, 16 au 16 septembre, **2007**, volume **V**, pp. 365-368.

21 **Huber P.-J.**, *Robust statistics*, John Wiley and Sons, New York, USA, **1981**, 328 pages.

22 **Geman D.**, **Reynolds G.**, Constrained restoration and the recovery of discontinuities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1992**, volume **14**, **3**, pp. 367-383.

23 **Charbonnier P.**, **Blanc-Féraud L.**, **Aubert G.**, **Barlaud M.**, Deterministic edge-preserving regularization in computed imaging, *IEEE Transactions on Image Processing*, **1997**, volume **6**, **2**, pp. 298-311.

24 **Tarel J.-P.**, **Ieng S.S.**, **Charbonnier P.**, Using robust estimation algorithms for tracking explicit curves, *European Conference on Computer Vision (ECCV'02)*, Copenhague, Danemark, **2002**, volume **1**, pp. 492-507.

25 **Tarel J.-P.**, **Ieng S.S.**, **Charbonnier P.**, *Robust Lane Marking Detection by the Half Quadratic Approach*, **2007**, ERLPC, CR 49, 74 pages.

26 **Blake A.**, **Zisserman A.**, *Visual Reconstruction*, MIT Press, Cambridge, MA, USA, **1987**, 240 pages.

27 **Bigorgne E.**, **Tarel J.-P.**, Backward segmentation and region fitting for geometrical visibility range estimation, *Asian Conference on Computer Vision (ACCV'07)*, Tokyo, Japon, 18 au 22 novembre **2007**, volume **III**, pp. 817-826.