Visibility Distance Estimation based on Structure from Motion

Clément Boussard
IMARA
INRIA Paris-Rocquencourt
Rocquencourt, France
clement.boussard@inria.fr

Nicolas Hautière
Universit Paris Est, LEPSiS
LCPC
Paris,France
nicolas.hautiere@lcpc.fr

Brigitte d'Andréa-Novel
CAOR
Mines ParisTech
Paris,France
brigitte.dandrea-novel@ensmp.fr

Abstract—It is obvious to say that perception is necessary to drive. Futhermore we can say that a good visibility is a guaranty for passengers security. The driver will adapt the vehicle speed to the offered visibility. Strong visibility reductions (dense fog for instance) are conditions of risk of accident. We therefore propose here a method to perform an onboard estimation of the visibility distance. Once this estimate is obtained, assistance could be offered to the driver (eg if it runs at a speed not adapted to the current visibility) or to the road infrastructure management so that it can inform other users of risk on his road network. This method uses images acquired by an onboard camera filming the scene and the estimation of vehicle motion. Thus, from this information we will explain how we can achieve a spatial partial structure reconstruction to estimate the visibility distance.

Index Terms—Visibility, Estimation, Structure from Motion.

I. INTRODUCTION

The visibility distance estimation is an important issue for environment perception. The challenge of the work done here is to obtain an on-board estimation of the visibility distance in order to inform the driver if its speed is not adapted to the prevailing visibility condition. Another possibility could be to inform all the vehicles in the area about the risk due to bad visibility via a vehicle to vehicle communication system.

The problem we try to solve is to get information with an onboard camera when the visibility is poor (only few objects can be seen in the image). Among the methods wich estimate the visibility distance, some are based on previous detection. In [1], Mori *et al.* determine a contrast attenuation coefficient from the detection of the preceding vehicle with a radar. This contrast attenuation allows them finding a visibility distance. In [2], Pomerleau estimates the visibility by measuring the contrast attenuation of road markings at various distances in front of the vehicle.

The others do a straightforward calculation of the estimate. In [3] a visibility distance estimation can be obtained by daytime fog, using a monocular method [3] adapted to fog based on Koschmieder's model [4] (model giving the apparent brightness of an object depending on the atmospheric extinction coefficient). In [5] the distance to the furthest element of the scene belonging to the road plane having a contrast greater than 5% (in accordance with the definition given by the CIE [6]) gives the visibility distance. This method is based on stereovision [7] and is not limited to fog. The distance given by this method will vary depending on the presence or not of object in the field of vision (to measure contrast). For this purpose we will talk about mobilized visibility, the maximum achievable mobilized visibility being the mobilizable visibility [8].

The method we aim is as generic as the one based on stereovision but uses a single camera. Thereby, in the same way as this method, we estimate the distance to the furthest object on the plane of the road with a contrast greater than or equal to 5%. Therefore we develop a new method based on Vehicle Dynamics Estimation and Structure from Motion. A complete study of the Vehicle Dynamics Estimation used in the method can be found in [9]. In this paper, we will explain how the Structure from Motion (SFM) takes part in the method and how we develop a new approach based on concepts developped for classic stereovision methods. Furthermore, we implement this method on a car and obtained some results presented at the end of the article.

We will recall our monocular generic method in the next section. Then we will explain in more detail the SFM. Finally we will present some results of on-board visibility distance calculation.

II. MONOCULAR GENERIC METHOD

This method takes into account the definition of visibility given by the International Commission on Illumination [6] and is divided into three steps. The first one is to map the depths of the vehicle environment with Structure From Motion (SFM). The second one is to compute the elements of the image having a contrast greater than 5%. Finally, the visibility distance is obtained from the combination of depth and contrast maps.

Thus we explain how we do SFM with homographic registration, then how we calculate the contrast to, then, estimate a visibility distance.

As we said above, the method we aim is based on the concepts of the method based on stereovision but using a single camera. We will therefore carry out, through the homographic registration, a temporal stereovision process.

The method using the conventional stereovision process (traditional spatial stereovision with at least 2 cameras) uses an extraction of the road surface within the scene, whereas our method of temporal stereovision can, at best, discern the flat

world from vertical objects. Our method will therefore have lower accuracy but gain in ease of use (no calibration of the stereoscopic sensor).

A. Homographic registration

In monovision, it is impossible to go directly to the depth in the images. The standard *Flat World* assumption allows one to associate the image line with a distance, this being true for objects belonging to the road plane.

The challenge is to discern the road objects to others. The most generic way of achieving this is to make successive images registration. Objects belonging to the plane of the road are registered from an image to another. In contrast, the vertical objects are distorted. This method allows, in theory, distinguishing the points that belong to the road plane from the others.

In general, the successive images registration is done using conventional techniques of image processing [10]. These methods match the contours between both images and estimate the transformation between both images. An application for calculating the speed of the vehicle is described in [11].

In our context of low visibility, this approach is not suitable because the contrasts are highly degraded. The originality of our approach is to do homographic registration using the knowledge of vehicle motion, observed or measured by proprioceptive sensors (odometer, Inertial Measurement Unit).

The construction of the depth maps with one camera goes through the understanding of several steps that have been developed in [9]. They begin with the camera modelization, following the application of the Flat World homography, and then, using the vehicle motion to realize the image registration.

We can see in Fig. 1 the result obtained with synthesized images from a simulator [12] after the vehicle has traveled distances of 1 m (top) and 2m (below) before making the homographic registration.
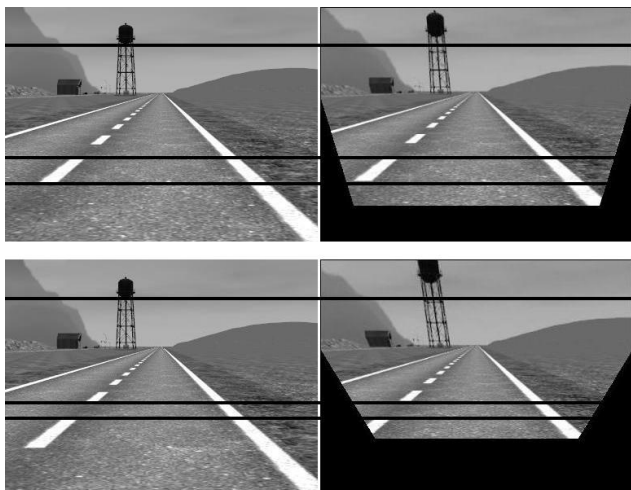

Fig. 1.   left: current image and right: registered image - top: 1m traveled before registering / down: 2m traveled before registering.

Big black lines in the images highlight the fact that objects belonging to the plane of the road are not distorted (road markings, base of the tower, ...) as opposed to objects not belonging to it. Indeed we can notice that the top of the tower is distorted and the more the distance before registering is large.

B. Structure from Motion

Now that we have made the image registration, we will compare them to find connections between two local regions of each image. Initially, we determined areas of interest in the image by retrieving the contours of objects in the scene. Then we calculated the correspondence between these regions of interest of the two images using a correlation method (comparison between different techniques of correlation can be found in [13]). After the correlation calculation performed on all regions of interest of the two images, a *matching distance* d or disparity is calculated. The detailed calculation can be found in [9]. From this disparity calculation we can identify in the image pixel belonging to the road plane from others. The detailed explanation of this process is given in the next section.

C. Estimation of the visibility distance

To estimate the visibility distance, we combine a measure of contrast greater than 5% with the map of pixels belonging to the road plane obtained from the Structure from Motion process. To this end, we compute the local contrast of the image points belonging to the road plane by scanning the image from top to bottom starting from the horizon [14]. Once we found a point with a contrast greater than or equal to 5%, the process stops and the distance to this point is the visibility distance. The distance of a pixel belongings to road plane can be calculated from the projective model (1):

$$Dist = \begin{cases} \frac{\lambda}{v-v_h} & \text{if } v > v_h \\ \infty & \text{if } v \leq v_h \end{cases} \quad \text{where } \lambda = \frac{H\alpha}{cos(\theta)} \quad (1)$$

where H is the mounting height of the camera, α is the ratio between the focal length of the camera and the size of a pixel, θ the pitch angle of the camera and v_h the position of the horizon line in the image.

III. STRUCTURE FROM MOTION BASED ON DISPARITY COMPUTATION

In this section, we detail our SFM approach. We tried to exploit the work already undertaken in the context of spatial stereovision (with at least two cameras). The first approach is based on the notion of V-disparity, the second exploits the geometry of the scene considered.

A. First approach: V-disparity

1) V-disparity in spatial stereovision: Labayrade *et al.* developed the concept of V-disparity [7]: from a pair of rectified stereoscopic images (left and right images), we may for the detected objects, calculate the difference in position between the two images. This allows us to calculate a disparity distance as shown in Fig. 2.
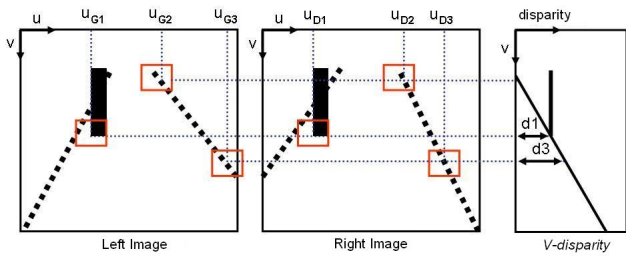
Fig. 2. V-disparity in the case of spatial stereovison. The dotted lines belong to the road, the black stick is a vertical element of the scene.

If we project this distance depending on the height in the image (height pixel v), we get the right image of Fig. 2. We can see that the elements of the road (dotted lines) have a strong disparity in the bottom of the image and a low disparity at the top. Conversely, the vertical elements of the scene (the black stick), will have for all points the same disparity as a function of height in the image.

The disparity of the vertical element is the same for all points belonging to it. We can therefore, from this, discern the road plane (the oblique line in the V-disparity space representation) from the vertical object (the vertical line in the V-disparity space representation).

2) V-disparity for temporal stereovision: The V-disparity concept can be applied to our case of temporal stereovision. This was tested by Alix *et al.* in [15].

In this case, the comparison is done between a normal image and registered image. The objects belonging to the road plane will not be distorted, in contrast to vertical objects (see eg Fig.1). We have shown in Fig. 3, a road (dotted lines) and a vertical element (black stick) in the image and the registered image.
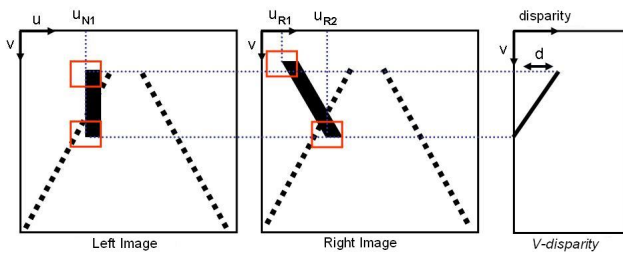

Fig. 3. V-disparity representation in the case of temporal stereovision

Here, the objects belonging to the road plane are the same in both images, unlike the vertical elements. The matching distance or disparity will be zero for elements of the road. For vertical elements it can be approximated, for the sake of simplicity by:

$$d = |u_{N1} - u_{R1}|$$

The projection of this matching distance in the V-disparity space representation gives us the right part of Fig. 3. For vertical elements, the distance will be increased depending on the height position of the points considered. We can

therefore, in theory, from this representation, discern vertical objects (oblique lines) of elements belonging to the road plane (matching distance close to 0).

This approach is interesting but to get good results the registered image must be computed after a sufficiently large displacement of the vehicle in hopes of obtaining a large enough difference to be taken into account. The larger the windows used to calculate the correlation, the larger the disparity between both images will be. But large areas of research involves greater computational power. In our case, to preserve the real-time aspect in our estimation, we turned to another method of spatial stereovision that we applied to our case of temporal stereovision.

B. Second approach: Compensation of perspective projection

The second approach developed by Williamson *et al.* in [16] uses the geometry of the scene. Indeed, if one assumes that we seek in the image elements belonging to the plane of the road and vertical elements, one can change the region of interest for disparity calculation. We will initially explain the approach used by Williamson *et al.* in the case of spatial stereovision and show how we can apply this concept to our case.

1) Compensation for spatial stereovision: Willamson *et al.* operate in [16] the fact that the geometry of elements belonging to the road plane are not the same in the pair of stereoscopic images. Indeed, as shown in Fig. 4, the comparison of regions belonging to the road plane is not effective, unlike regions of elements belonging to vertical objects.
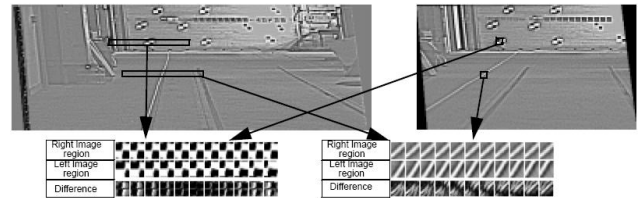

Fig. 4. Calculation of the disparity in the case of stereoscopic images [16]. The elements belonging to vertical objects are the same in both images (picture left). The elements belonging to the road plane (eg the white line), are not the same in both images (right panel).

If you change one of the two images of the pair of stereoscopic images using a flat world hypothesis, we get the image in Fig. 5. As we can see, in this case, the comparison of areas belonging to the road plane gives much better results.
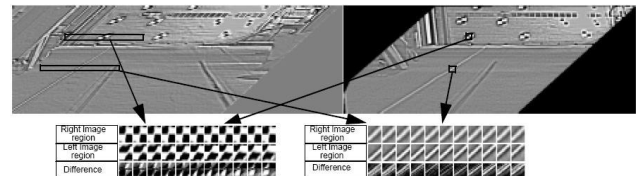

Fig. 5. Calculation of the disparity in the case of stereoscopic images where the image on the right has been adjusted using the perspective projection [16]. The elements belonging to the road plane are, in this case, the same in both images (right panel).

The method proposed by Williamson et al. is based on the difference in correlation score in the case where an image is changed after the flat world hypothesis in the classical case. The different results can discern the elements belonging to the road plane from vertical objects.

Note that the perspective projection may vary during movement of the vehicle, for example if it rolls or pitches. This method should, strictly speaking, take into account the dynamics of the vehicle.

2) Compensation in temporal stereovision: In the case of temporal stereovision objects which not belong to the road plane are distorted upwards and in the edges direction. We have therefore defined searching windows following this deformation. We can account for this distortion in Fig. 1. When the pixel is to the right of the image, that window is distorted upwards and right, when the pixel is left, the window is distorted upwards and the left. This is shown schematically in Fig. 6.
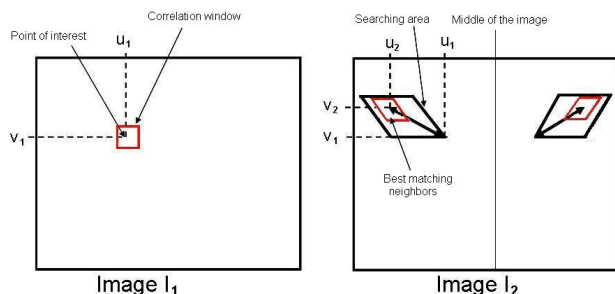

Fig. 6.   Correlation with an oblique windows for the Non-Road hypothesis

In the same way as in the case of spatial stereovision, the deformation of this window will allow us to get better correlation scores for objects not belonging to the road. The idea of our method lies in the fact that for each pixel, we calculate a disparity with normal correlation windows (undeformed) and another with distorted windows. The objects belonging to the road will have a smaller disparity distance with normal windows. Conversely, objects which do not belong to the road will have a disparity distance smaller with oblique windows. This is the approach that we used in our study.

IV. ON-BOARD VISIBILITY DISTANCE CALCULATION

We have seen that knowing of the vehicle dynamics is necessary for our method to calculate the registered image needed for the SFM. In fact, knowing the six degrees of freedom which are the three rotations (roll, pitch and yaw) and three translations (longitudinal T_x, lateral T_y and vertical T_z), we can achieve our homographic registration.

To estimate the vehicle dynamics that we need for our application, we will use an instrumented vehicle. The sensors on this vehicle during our tests are a camera, an inertial measurement unit (IMU) and an odometer:

- the odometer provides information on the vehicle speed.
- the IMU provides the inertial angular velocities of the three axes of rotation of the vehicle (roll, pitch and yaw)

and the accelerations along the three axes (X, Y, Z) of the vehicle.

A. Presentation of the test

The test was conducted on the test track in Versailles-Satory (see Fig. 7) in fog. The three sensors on the vehicle were an IMU, an odometer and a camera.
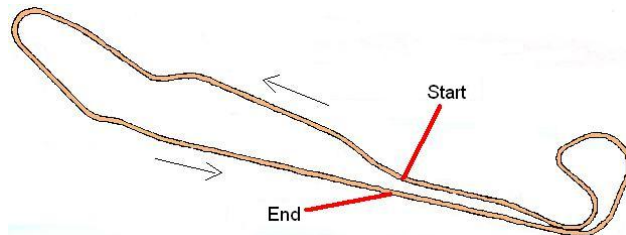

Fig. 7.   Test track in Versailles-Satory

The vehicle starts with a zero speed and then accelerated to a speed of 50 km/h and finally produced a peak at 70 km/h, as shown in Fig. 8.
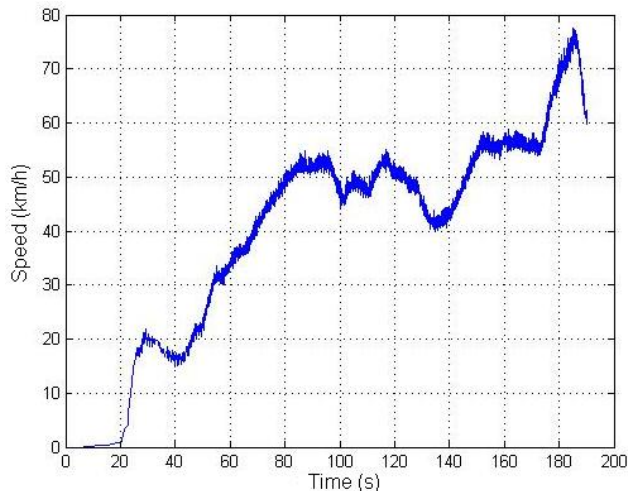

Fig. 8.   Vehicle speed during test

We can see in Fig. 9 on an actual fog image that the combination of the depth map of the elements belonging to the road plane (right) and the map of contrast greater than 5% (left) helps to determine the visibility distance. The distance to the furthest point having a contrast greater than 5% will be our range of vision and it is represented by the horizontal line on the pictures. The complete method and initial results are in [17], [9].

B. Results

We calculated the visibility distance in various parts of the test track. Estimates of visibility distance were made on straight parts or slightly bend. Indeed, since measurements from the IMU are very bad, we were unable to readjust the image correctly at high cornering. You can see these results in Fig. 10.
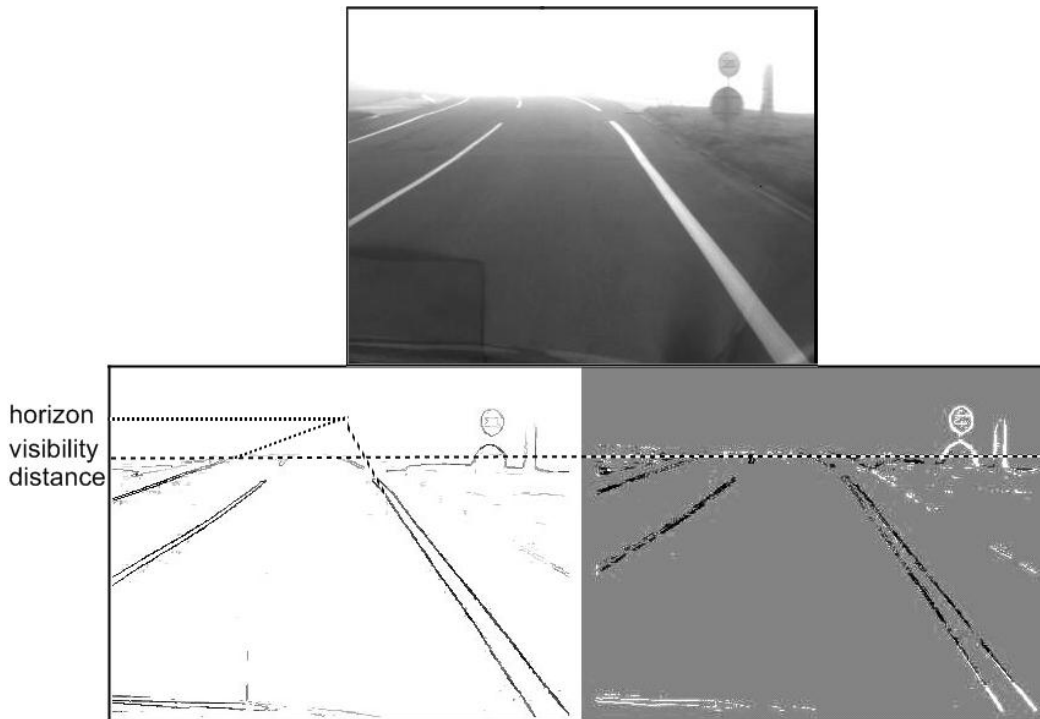
Fig. 9. Visibility distance obtained from a contrast and a depth maps. Up: Image from the camera - Left: contrast map (a contrast higher than 5%) - Right: in white elements not belonging to the road plane, in black elements belonging thereto

The first sequence (left) was acquired in an area with a guardrail and a barrier on the side of the road. Our estimation method provides visibility scores at around 40 m except for the beginning and the end of the sequence where it reached 80 m. These errors are due to vertical objects found on the roadside which were not properly matched in the registered image.

The second sequence (middle) was acquired in an area with trees located a few meters right of the road. Our method considers the visibility distance at a value of around 40 m.

The last sequence (right) was obtained in an area where there are no vertical elements on the roadside. The estimated visibility distance is 40 meters.

In comparison to these results, it was shown in Fig. 11 the results obtained with the method exploiting the effect of haze created by the fog [3]. This method considers the mobilizable visibility distance while our method considers the mobilized visibility distance.

It may be noted that the mobilizable visibility distance (in blue) is estimated to approximately 60 m, while the mobilized visibility distance is 40 m. Overall, if we don't consider the uncertainties due to our methods, the method exploiting the effect of haze considers a range of vision higher than that provided by our method. This result is obvious since the former provides a mobilizable visibility distance while the second provides a range of vision mobilized [8].

V. CONCLUSION AND PERSPECTIVES

We have presented in this paper a method to estimate the visibility distance with a camera. This method is based
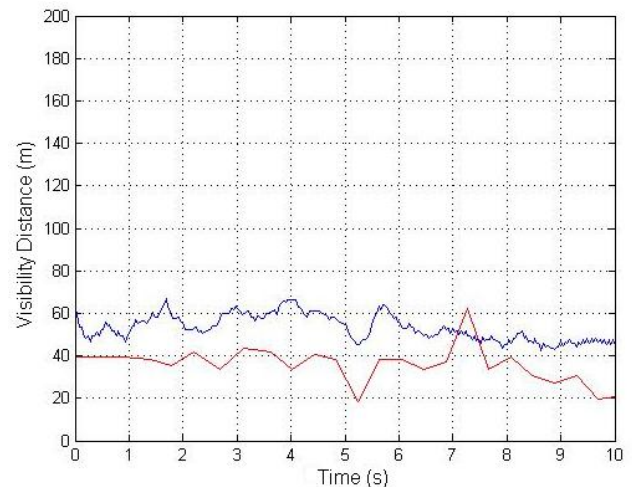

Fig. 11. Comparison of results of visibility distance calculation. Our method in red, in blue the method based on the effect of haze[3].

on SFM and therefore we developed a new SFM method for a monocular sensor. Results have been obtained and a comparison with a method of estimation with mobilizable visibility has been made.

We see two perspectives to this study:

- We have presented in the state of the art a method based on the same type of consideration (coupling between a depth map and a contrast map) where the depth map is determined by stereovision [5]. It would be interesting perspective to this study to evaluate and compare the two
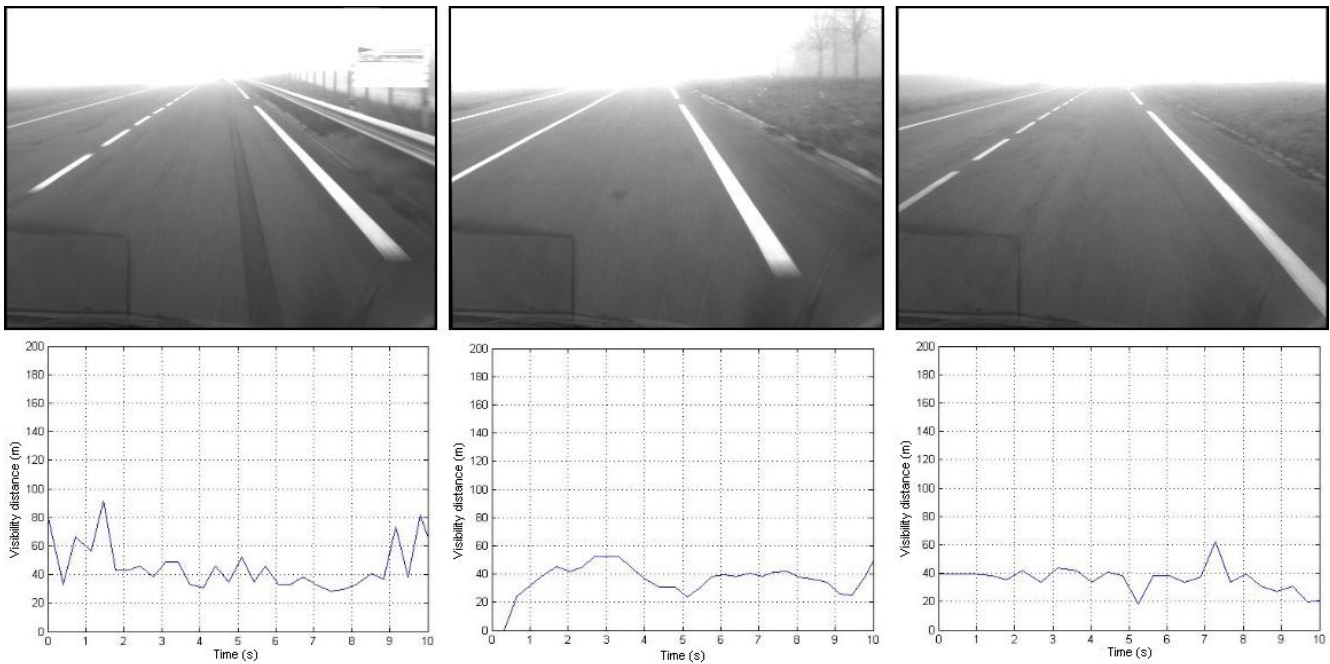
Fig. 10. Visibility distance estimated with our method on three different sequences of the test.

methods.

- Any method propose and tested here is based on the flat world homographic shift. We therefore considered all our transformation from elements found on a plane at an altitude in the reference vehicle frame. One idea might be to consider not only homographic shift transformation but for several different heights. This would validate various hypotheses for the elements of the scene. If an item is considered outside the road plane, we might as well check if it does not belong to a plane lying at a different altitude. We could thus determine the vertical elements and their height.

REFERENCES

[1] K. Mori, T. Kato, T.Takahashi, I. Ide, and H. Murase. Visibility estimation in foggy conditions by in-vehicle camera and radar. In *Internatinal Conference on Innovative Computing, Infomation and Control (ICICIC06)*, volume 2, August 2006.

[2] D. Pomerleau. Visibility estimation from a moving vehicle using the RALPH vision system. In *IEEE Conference on Intelligent Transportation Systems*, pages 906–911, November 1997.

[3] N. Hautière, J.-P. Tarel, J. Lavenant, and D. Aubert. Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine vision and applications journal*, 17(1):8–20, April 2006.

[4] W. Middleton. *Vision through the atmosphere*. University of Toronto Press, 1952.

[5] N. Hautière, R. Labayrade, and D. Aubert. Real-time disparity contrast combination for onboard estimation of the visibility distance. *IEEE Transactions on Intelligent Transportation Systems*, 7(2):201–212, June 2006.

[6] Commission Internationale de l'Eclairage, editor. *International lighting vocabulary*, volume 17 of *4*. 1987.

[7] R. Labayrade and D. Aubert. In-vehicle characterization of obstacles by stereovision. In *1st International Workshop on in-Vehicle Cognitive Computer Vision Systems, Graz, Austriche*, 2003.

[8] N. Hautière, D. Aubert, and E. Dumont. Mobilized and mobilizable visibility distances for road visibility in fog. In 26th *session of the CIE*, Beijing, China, July 2007.

[9] C. Boussard, N. Hautière, and B. d'Andréa Novel. Vehicle dynamics estimation for camera-based visibility distance estimation. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems IROS 2008*, pages 600–605, September 22–26, 2008.

[10] A. Shashua. Projective structure from uncalibrated images: Structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 16(8):778–790, 1994.

[11] G. Stein and A. Shashua. A robust method for computing vehicle ego-motion. In *IEEE Intelligent Vehicle Symposium*, Dearborn, MI, USA, October 2000.

[12] D. Gruyer, C. Royère, and S. Glaser. Sivic, une plate forme de prototypage d'environnement routier et de capteurs virtuels pour la conception et l'évaluation de systèmes d'aide à la conduide. In *Journée Automatique Automobile*, 8-9 November 2005.

[13] M. Perrollaz. Construction d'une carte de disparité et application à la détection d'obstacles routiers. Technical report, INRETS/LCPC, 2006. Technical Report.

[14] N. Hautière, D. Aubert, and M. Jourlin. Mesure du contraste local dans les images, application à la mesure de distance de visibilité par caméra embarquée. *Traitement du Signal*, 23(2):145–158, Septembre 2006.

[15] R. Alix, F. Le Coat, and D. Aubert. Flat world homography for non-flat world on-road obstacle detection. In *IV*, 2004.

[16] T. Williamson and C. Thorpe. Detection of small obstacles at long range using multibaseline stereo. In *Proceedings of the 1998 IEEE International Conference on Intelligent Vehicles*, 1998.

[17] C. Boussard, N. Hautière, and B. d'Andréa Novel. Vision guided by vehicle dynamics for onboard estimation of the visibility range. In *IAV 2007*, Toulouse, France, Septembre 2007.