VISION GUIDED BY VEHICLE DYNAMICS FOR ONBOARD ESTIMATION OF THE VISIBILITY RANGE

Clément Boussard * **Nicolas Hautière** **
Brigitte d'Andréa-Novel *

* *Ecole Nationale Supérieure des Mines de Paris, ENSMP*
** *Laboratoire Central des Ponts et Chaussées, LCPC*

Abstract: The presence of an area with low visibility conditions is a relevant information to communicate to drivers before they reach this area. In this aim, we develop a generic sensor of visibility using an onboard camera in a vehicle equipped with vehicle-to-infrastructure (V2I) communications. Our approach consists in estimating the range to the most distant object belonging to the plane of the road having at least 5% of contrast. The originality of this approach lies in the fact that the depth map of the vehicle environment is obtained by aligning the road plane in the successive images. This algorithm exploits the dynamics of the vehicle which is given or observed by proprioceptive sensors. In this paper, we present the principle of our approach in terms of image processing and explain how the vehicle dynamics takes part in it with a sensitivity study.

Keywords: Vision, Vehicle dynamics, Visibility range, Image alignment.

1. INTRODUCTION AND OBJECTIVES

Within the framework of the European project REACT (REACT 2005), we are developing a visibility sensor. One goal of the project is to make progress in road safety with vehicle-to-infrastructure (V2I) communications. The vehicle is seen like a sensor, inserted into the traffic, which communicates measurements to a traffic management regional center. In this paper, our objective is to be able to locally measure the visibility range with an onboard camera, in the aim of pursuing low visibility conditions owed by climatic factors.

Different studies about visibility distance measurement exist, among which we can find:

- A method using detection of lane markings: Pomerleau (Pomerleau 1997) estimates the visibility distance by measuring the contrast attenuation of lane markings at different distances in front of the vehicle.

- A mono-camera method adapted to fog using Koschmieder's model (Middleton 1952). We obtain, under daytime foggy weather, an estimation of the meteorological visibility distance (Hautière *et al.* 2006*a*).

- A method using stereo-vision: this method is generic and not limited to fog. Thanks to stereo-vision, a good quality depth map is computed (Labayrade and Aubert 2003). The distance to the farthest point of the road surface with a contrast greater than 5% gives the visibility distance (Hautière *et al.* 2006*b*).

The method using stereo-vision does not make any difference between the geometric and atmospheric visibility distance. Indeed, if there is a curve or a uphills where the visibility is reduced due to physical reasons, the visible road surface will be limited by the road geometry (Brun *et al.* 2006). In these cases, the visibility distance calculated will be the geometric one.

The method we design is as generic as the one based on stereo-vision but uses only one camera. In this aim, we estimate the distance of the farthest object which is part of the road plane with a contrast higher or equal to 5%. This method takes into account the definition of the meteorological visibility distance given by the CIE (Commission Internationale de l'Eclairage 1987) and is decomposed in three parts:

(1) Creation of a pseudo-depth map of the vehicle environment by aligning the road plane in the successive images.
(2) Creation of a contrast map.
(3) The visibility distance is obtained by taking the farthest point (depth map) with a contrast greater than 5% (contrast map).

2. IMAGE PROCESSING

Through use of a single camera, it is impossible to get directly the depth in images. But we can calculate with perspective projection the distance of points belonging to the road. The most generic way to determine the road plane is to use successive images. Objects belonging to the road plane are at the same place from one image to another, at the opposite, verticals objects are deformed. In general, successive images alignment is made with classical image processing techniques, e.g. (Shashua 2004). These methods consist in matching objects in the two images. In our degraded visibility context, this approach is not well adapted because local contrasts are strongly deteriorated. The originality of our approach is to align images with the knowledge of the motion of the camera, which is observed or measured with proprioceptive sensors.

2.1 Image acquisition

In the coordinate system of the camera frame, the position of a pixel in the image plane is given by its coordinates (u, v). The image optical center is denoted (u_0, v_0) in the image frame and considered as the image center.
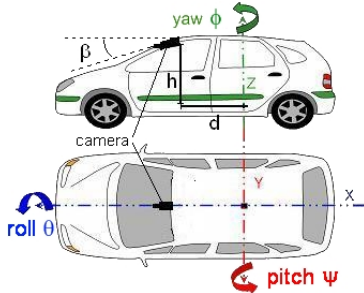

Fig. 1. Position of the camera and vehicle dynamics

The transformation between the vehicle frame (with origin at the center of gravity of the vehicle) and the camera frame, is represented by a vectorial translation $\overrightarrow{t} = d\overrightarrow{X} + h\overrightarrow{Z}$ (Fig. 1) and a rotation around the axes Y of angle β. We denote T the translation matrix and R the rotation matrix. The coordinate change between the image frame and the camera frame can be expressed using a projective matrix M_{proj} (Horaud and Monga 1995):

$$M_{proj} = \begin{pmatrix} u_0 & 0 & -\alpha & 0 \\ v_0 & -\alpha & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad (1)$$

where α represents the ratio between the camera focal length and the size of one pixel. At last, we obtain the transformation matrix T_r from the vehicle frame to the image frame:

$$T_r = M_{proj} R T \qquad (2)$$

If P is a point with homogeneous coordinates $(X, Y, Z, 1)$ in the vehicle frame, its homogeneous coordinates in the image frame become:

$$p = T_r P = (x, y, z)^T \qquad (3)$$

We can now compute the coordinates (u, v) of the projection of P in the image frame:

$$\begin{cases} u = \dfrac{x}{z} = u_0 + \alpha \dfrac{\cos\beta(Z+h) - \sin\beta(X+d)}{\cos\beta(X+d) + \sin\beta(Z+h)} \\ v = \dfrac{y}{z} = v_0 - \alpha \dfrac{Y}{\cos\beta(X+d) + \sin\beta(Z+h)} \end{cases} \qquad (4)$$

2.2 Creation of a transformed image

Flat World Assumption:

If we consider I_1 and I_2 images taken at time t_1 and t_2, the knowledge of the vehicle dynamics allows us, thanks to (4), to obtain an estimation of the image I_2 from the image I_1. Let $\tilde{I}_{12}$ be this estimated image and P a point whose projection in the image frame belongs to it. Let us assume that this point belongs to the road plane, meaning that if (X_2, Y_2, Z_2) are the coordinates of this point in the vehicle frame, then $Z_2 = 0$. So the expression of X_2 and Y_2 is deduced from (4):

$$\begin{cases} X_2 = \dfrac{\cos\beta[dU + \alpha h] + \sin\beta[hU - \alpha d]}{\alpha \sin\beta - \cos\beta U} \\ Y_2 = \dfrac{-hV}{\alpha \sin\beta - \cos\beta U} \end{cases} \qquad (5)$$

where $U = u - u_0$ and $V = v - v_0$

Vehicle motion:

If we know the vehicle motion, we can calculate its movement between time t_1 and t_2. As soon as
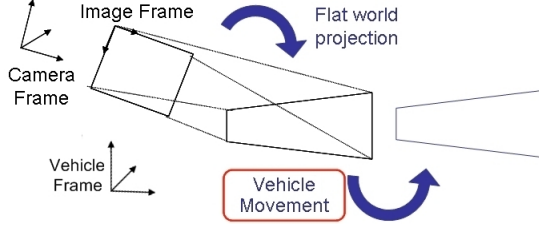
Fig. 2. Movement of the vehicle

we have points in the vehicle frame (5) we can get new points following this movement (see Fig.2).

From the knowledge of the coordinates of a point P and the vehicle dynamics, we can express the coordinates of the point P in the camera frame at time t_1:

$$(x_{12}, y_{12}, z_{12})^T = T_r M (X_2, Y_2, 0)^T \qquad (6)$$

where M is the vehicle rotation/translation matrix between two instants. We obtain the coordinates (u_{12}, v_{12}) of P in the image frame of I_1:

$$u_{12} = \frac{x_{12}}{z_{12}} \text{ and } v_{12} = \frac{y_{12}}{z_{12}} \qquad (7)$$

Example of transformed image:

An example of transformed image $\tilde{I}_{12}$ obtained from an image I_1 is given in Fig. 3. The comparison between image I_1 and the estimated image $\tilde{I}_{12}$ allows to obtain a depth map in the same way as what is done in stereo-vision. As far as we have made the hypothesis that all the points in the image I_2 belong to the road plane, a short distance allows us to validate the flat-world assumption.
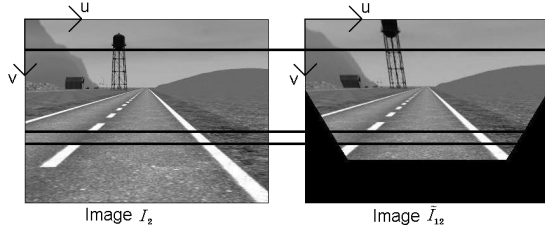

Fig. 3. left: current image / right: transformed image after a 2m displacement. Result obtained with synthetic images

2.3 Pseudo-depth map construction

As far as we can say that a pixel of coordinates (u, v) belongs to the road plane (see Fig.4), we can express the distance d of this pixel (8) :

$$d = \begin{cases} \dfrac{\lambda}{v - v_h} & \text{if } v > v_h \\ \infty & \text{if } v \leq v_h \end{cases} \text{ where } \lambda = \frac{H\alpha}{cos^2 \beta_0} \qquad (8)$$

where H denotes the mounting height of the camera, α the ratio between the camera focal length and the size of one pixel, v_h the position of the horizon in the image and β_0 the camera pitch angle.
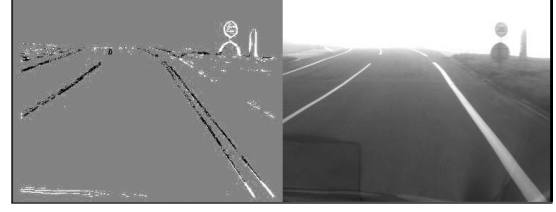

Fig. 4. Image taken by the onboard camera. In black: points belonging to the road plane. In white: points not belonging to the road plane

2.4 Structure of the Image via Correlation Metrics

2.4.1. Pseudo-disparity Computation between two Images We have to match both images. It means that we have to find local correspondences between two neighborhoods from each image. These correspondences are computed via the ZNCC correlation metrics (a comparison of different existing metrics is carried out in (Perrollaz 2006)). To realize this operation, we have to select a pixel $p_1 = (u_1, v_1)$ in the image I_1 and another pixel $p_2 = (u_2, v_2)$ in the transformed image $\tilde{I}_{12}$. Then, we define a centered neighborhood $V(p_1)$ around the pixel p_1 and $V(p_2)$ around the pixel p_2 in which we are computing the ZNCC correlation metric:

$$ZNCC(p_1, p_2) = \frac{\displaystyle\sum_{V(p_1), V(p_2)} \tilde{I}_1 \tilde{I}_2}{\sqrt{\displaystyle\sum_{V(p_1), V(p_2)} \tilde{I}_1^2 \displaystyle\sum_{V(p_1), V(p_2)} \tilde{I}_2^2}} \qquad (9)$$

where

$$\tilde{I}_1 = \big(I_1(u_1 + i, v_1 + j) - \bar{I}_1 \big)$$
$$\tilde{I}_2 = \big(I_2(u_1 + i, v_1 + j) - \bar{I}_2 \big)$$

The more the correlation metrics is close to 1, the more we can consider these two neighborhoods as identical. Working on a single pair of (p_1 and p_2) limits our study. Indeed, some matching errors can occur and a pixel belonging to the road can be incorrectly matched in the image $\tilde{I}_{12}$. That's why we have to extend our study zone. To do it, we have defined a search window. The correlation neighborhood in image I_1 is centered on a point of interest. The correlation neighborhood in the image $\tilde{I}_{12}$ is centered successively around a pixel varying in a search frame (this search frame is centered on the pixel p_1 of image I_1). This principle is schematized in Fig. 5.

As soon as the sweeping of the search window is done, we keep the position (u_2, v_2) of the pixel
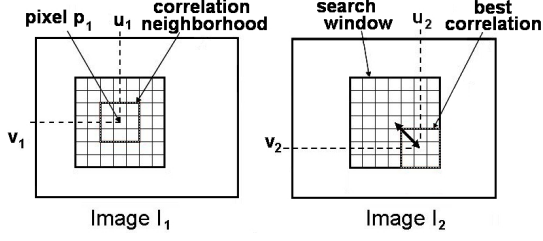
Fig. 5. Correlation neighborhood and search window

with the best correlation score. With these two positions, we calculate a *pairing distance*:

$$d = \sqrt{|u_1 - u_2|^2 + |v_1 - v_2|^2}$$

After that, we only kept points with a small pairing distance, in considering them as points belonging to the road plane.

2.4.2. Road or Non-Road Hypothesis To get a good non-road hypothesis, one can notice that objects not belonging to the road plane are deformed towards the top and the borders of the image. We have defined a search frame on the basis of these deformations. We can have an idea of this deformation in Fig. 3. When the pixel are on the right side of the image, the search frame is deformed towards the top and the right. When the pixel is on the left side, the search frame is deformed towards the top and the left side of the image. This idea comes from what is done in stereovision in (Williamson 1998). This is schematized in Fig. 6. This deformed window gives us some better correlation scores for objects not belonging to the road plane. Finally, the idea of
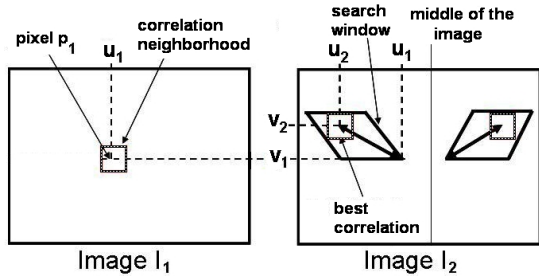

Fig. 6. Correlation with deformed window for the Non-Road hypothesis

our method is based on the fact that for each pixel, we are going to compute a pairing distance with a normal and a deformed searching window. Objects belonging to the road plane have a shorter pairing distance with a normal window. On the contrary, objects not belonging to the road plane have a shorter pairing distance with a deformed window. An example of result is given in Fig.4 using actual images of fog. The majority of pixels belonging to the road plane is successfully recognized, contrary to the pixels belonging to the vertical sign.

3. VEHICLE DYNAMICS

In the previous section, we have seen that the vehicle dynamics is a need for our visibility estimation method. Indeed, knowing the six degrees of freedom, the three rotations (roll, pitch, yaw) and the three translations (longitudinal T_x, lateral T_y and vertical T_z), let us realize the successive images transformations. Sensors that give vehicle dynamics estimation are an odometer and an Inertial Measurement Unit (IMU):

- The odometer gives informations on the numbers of turns done by the wheel.
- The IMU gives angular speed of the three rotations axis of the vehicle (roll, pitch, yaw) and accelerations of the three axis of the vehicle (X, Y, Z).

At first sight, the odometer and the IMU should give us the knowledge of the six degrees of freedom that we need. Indeed, if we consider that the wheel radius is constant, we can have an estimation of the distance covered by this wheel, meaning by the vehicle. Moreover, the numerical integration of the angular speed given by the IMU gives an estimation of the relative angular variations between two instants. The first question we wonder was not to know what type of estimator or observers we should implement to estimate the six degrees of freedom. We wanted to know if the knowledge of any degrees of freedom was really a need for our successive images alignment. This was done in the aim of eliminating some of the degrees of freedom in our process. To do it, we used the notion of *sensitivity* (Arriola and Hyman 2003). If we just look at the nature of the degrees of freedom, we have angles expressed in radians and distances expressed in meters. So we can not directly compare them. The sensitivity allows us to compare the different contributions of the degrees of freedom using of simulated scenarii.

3.1 Sensitivity

Let $M_{vehicule}(T_x, T_y, T_z, \theta, \psi, \phi)$ be the motion function of the vehicle between time t_1 and t_2. From (5) we can compute the coordinates (X_2, Y_2, Z_2) of the point in the world frame at time t_2. This brings us to:

$$\begin{cases} u_2 = \dfrac{x_2}{z_2} = f_u(T_x, T_y, T_z, \theta, \psi, \phi, u, v) \\ v_2 = \dfrac{y_2}{z_2} = f_v(T_x, T_y, T_z, \theta, \psi, \phi, u, v) \end{cases} \quad (10)$$

The detail of (10) is explained in section (2.2) through (5) and (4). (10) allows us to say that we have an algebraic relation between (u_2, v_2) et (u, v). A sensitivity study is done with respect to a criteria (or a cost function). We have to define a

criteria that helps us in knowing which degree of freedom influences the most the successive image transformations.

Sensitivity criteria:
The criteria that seems the most important is the pixel displacement. So we define the following criteria:

$$J(u,v) = \sqrt{(u_2 - u)^2 + (v_2 - v)^2} \qquad (11)$$

With the help of this criteria we are able to quantify the influence of the different degrees of freedom on the displacement of a pixel (u,v) through the transformation (10).

Sensitivity computation:
The parametric sensitivity computation is defined as being the cost function derivative with respect to the studied parameter (12):

$$S_p(u,v) = \frac{\partial J}{\partial p} \times \frac{p}{J} \qquad (12)$$

For example, for the x translation:

$$S_{T_x}(u,v) = \frac{\partial J}{\partial T_x} \times \frac{T_x}{J} \qquad (13)$$

Now, we have to compare results obtained in generalizing this kind of computation for all the degrees of freedom. The vehicle dynamics is time varying. The value of the degrees of freedom is not always the same. We have to define some situations in which we compute sensitivity. This process allows us to say that in specific situations (braking, turning, ...), which are the degrees of freedom that are the most dominating in the image transformations.

3.2 Results on Different Simulated Scenarii

We have designed a prototyping platform (Boussard *et al.* 2006) with which we can simulate the behavior of a vehicle and its onboard sensors, get exactly their motions and see the results of the successive images transformations. We have defined different scenarii to stimulate all the degrees of freedom and to reproduce some of the classic vehicle behaviors. The initial speed was all the time $30km/h$ and the different scenarii are the followings:

- Acceleration and braking in a straight line: the acceleration was between $1.5m/s^2$ and $-1.5m/s^2$.
- Right and left oscillation at constant speed: on a two lane road, we move the vehicle from lane to the other.
- Straight line at a constant speed.

- Long right turn: at a constant speed, we turn the wheel to turn along a circle.

Table 1 shows the maximum value of the sensitivity we obtain for all the degrees of freedom we consider.

	Acceleration Braking	Right-Left oscillation	Constant speed	Right turn
T_x	20	15	20	20
T_y	0	8	0	20
T_z	3	2	0	1
pitch	10	4	0	4
roll	0	3	0	1
yaw	0	50	0	100

Table 1. Pixels displacements obtained with the different scenarii

We can see that the pixel displacement $J(u,v)$ (11) is less sensitive to the translation T_z and the pitch and roll angle, except in the first scenario (acceleration and braking). We can say that, as soon as we are driving at a constant speed, doing a turn or changing lane, the three most important degrees of freedom are the two translations T_x and T_y and the yaw angle. The others can be neglected.

Vehicle dynamic estimation is done with two sensors: an odometer and an inertial measurement unit (IMU). The yaw angle is given directly by the IMU and we have to estimate T_x and T_y. The odometer gives the distance l covered between time t_1 and t_2. When the yaw angle ϕ is not zero, we can consider that the vehicle is moving along an arc of circle with radius R and then use the well known trigonometric equation:

$$\begin{cases} T_x = R\sin(\phi) \\ T_y = R\cos(\phi) + R \end{cases} \text{ with } R = \frac{l}{\phi}$$

4. VISIBILITY

4.1 Contrast estimation

We adapted Köhler's binarization technique (Köhler 1981) in order to measure the local contrasts of images. A pair of pixels (x,x_1) is said to be separated by s if two conditions are met. First, $x_1 \in N_4(x)$. Second, the condition $\min(I(x), I(x_1)) \leq s < \max(I(x), I(x_1))$ is respected. Let $F(s)$ be the set of all couples (x, x_1) separated by s. With these definitions, for every value of s belonging to [0,255], $F(s)$ is built. For every couple belonging to $F(s)$, the mean logarithmic contrast associated to $F(s)$ is then:

$$C(s) = \frac{1}{\#F(s)} \sum_{(x,x_1)\in F(s)} \min\left(\frac{|s - I(x)|}{\max(s, I(x))}, \frac{|s - I(x_1)|}{\max(s, I(x_1))}\right)$$

The best threshold s_0 verifies the following condition:

$$s_0 = \underset{s \in [0,255]}{\mathrm{argmax}}\, C(s) \qquad (14)$$

It is the threshold which has the best mean contrast along the associated border $F(s_0)$. Instead of using this method to binarize images, we use it to measure the contrast locally. The evaluated contrast equals $2C(s_0)$ along the associated border $F(s_0)$.

4.2 Visibility estimation

To estimate the visibility distance, we combine the measurement of contrasts higher than 5% with the map of the pixels belonging to the road plane. In this aim, we locally process the contrast of image points belonging to the road plane by scanning it from top to bottom starting from the horizon line. As soon as we find a point with a contrast greater or equal to 5%, the process stops and the visibility distance is the distance of this point given by (8). We can see in Fig. 7 on an actual fog image an example of a 5% contrast map and the result of previous images alignment. The visibility distance is represented by the horizontal line on the pictures.
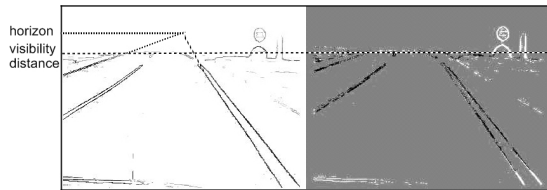

Fig. 7. 5% contrast map (left) and Road/Non-road image (right)

5. CONCLUSION

In this paper, a generic method estimating the atmospheric visibility distance is presented. It detects the farthest picture element belonging to the road plane having a contrast greater then 5% using a single camera. To discern points belonging to the road plane from the others, the road plane is aligned in successive images by exploiting the relative motion of the vehicle between two instants. Contrarily to classical image processing approaches, this relative motion is obtained thanks to the proprioceptive sensors of the vehicle.

To distinguish the dominating degrees of freedom in the image transformations, a sensitivity study is carried out using typical driving scenarii. We found that three degrees of freedom (lateral and longitudinal displacements and yaw angle) are enough in our context. Using this assumption, sample results of visibility estimation are given using actual images of fog.

REFERENCES

Arriola, Leon M and J. M. Hyman (2003). Forward and adjoint sensitivity analysis : with apllication in dynamical system, linear algebra and optimization. In: *Technical report, Los Alamos National Laboratory*.

Boussard, C., N. Hautière and D. Gruyer (2006). Prototypage d'un capteur monoculaire générique de visibilité pour véhicule traceur. In: *MajecSTIC 2006*.

Brun, X., P. Charbonnier and F. Goulette (2006). Modélisation 3d de routes par télémétrie laser embarquée pour la mesure de distance de visibilité.. In: *Journées des Sciences de l'Ingénieur*.

Hautière, N., J.-P. Tarel, J. Lavenant and D. Aubert (2006a). Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine Vision and Applications Journal* **17**(1), 8–20.

Hautière, N., R. Labayrade and D. Aubert (2006b). Real-time disparity contrast combination for onboard estimation of the visibility distance. In: *IEEE Transactions on Intelligent Transportation Systems*. Vol. 7 of *2*.

Horaud, R. and O. Monga (1995). *Vision par ordinateur, outils fondamentaux*. Editions Hermès.

Köhler, R. (1981). A segmentation system based on thresholding. *Graphical Models and Image Processing* **15**, 319–338.

Labayrade, R. and D. Aubert (2003). in-vehicle characterization of obstacles by stereovision. In: *1st International Workshop on In-Vehicle Cognitive Computer Vision Systems*.

Commission Internationale de l'Eclairage, Ed. (1987). *International lighting vocabulary*. Vol. 17.4.

Middleton, W. (1952). *Vision through the atmosphere*. University of Toronto Press.

Perrollaz, M. (2006). Construction d'une carte de disparité et application à la detection d'obstacles routiers. Technical Report.

Pomerleau, D. (1997). Visibility estimation from a moving vehicle using the RALPH vision system. In: *IEEE Conference on Intelligent Transportation Systems*. pp. 906–911.

REACT (2005). *Realizing Enhanced Safety and Efficiency in European Road Transport (http://www.react-project.org)*.

Shashua, A. (2004). Projective structure from uncalibrated images: Structure from motion and recognition. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. Vol. 16 of *8*. pp. 778–790.

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