

Accuracy in obstacle localization using vision

Rudy ALIX^{1,2}, Sebastien GLASER²

¹RENAULT S.A (Research Division)
Technocentre Renault – 1 Av. du Golf – 78288 Guyancourt Cedex

²LIVIC (INRETS/LCPC)
13, route de la Minière, Bt 140, 78000 Versailles Satory, France. E-mail : rudy.alix@renault.fr

Abstract – The goal of our research is to design perception devices dedicated to driving safety improvement such as : Collision Warning and Avoidance systems, Emergency braking... This device is supposed to be a piece of new driving assistance systems aiming at increasing the safety on road. Many safety systems which are emerging nowadays in our vehicles use a “distance to obstacle” information obtained thanks to telemeters such as radars, laser scanners (or lidars), ultrasounds... At least, this kind of systems achieve a great precision in locating objects relative to the sensors, but they are not able to provide their localisation compared to the road or the lane. This is the reason why some systems are involving passive sensors like video integrated inside vehicles. A device founded partially on computer vision allowed to compensate this lack. But the localisation computed by vision needs to be analysed in term of precision. This paper attempts to explore differences in localisation accuracy between systems involving only one camera (monocular vision) and systems involving two cameras (stereovision). A complete study of the errors commit on depth reconstruction is shown.

Keywords: monocular vision, stereovision, active safety, depth and distance reconstruction, obstacle detection, retro-projection.

1 Introduction

In this paper, we propose to estimate and compare accuracies of perception systems based on the use of vision (monocular and binocular).

Our research has important implications for many domains, including automotive, aeronautic, robot guidance and generally speaking, embedded systems. In fact, we try to show, while most systems require two or more video cameras for high precision, how can monocular and binocular vision go to similar results with different accuracies. The main common goal of these vision methods is to detect objects and locate them from the camera in terms of distance, angular position or time to collision.

1.1 The benefits of using a vision system

Vision, whether mono camera or stereo based, is a very dynamic field with plenty of applications in transport [Franke 00]. It can provide information on the road environment, like the lane [Luong *et al* 94], [Tarel 00] and roadside [Dickmanns 92], [Chausse *et al* 00] markings, the distance between vehicles [Labayrade *et al* 02b], or indicators of visibility distances.

Though the provided measures look all right, not much has been written about their accuracy. Yet systems based on RADAR or laser scanners have some serious drawbacks as regards security applications. These are mainly due to the lack of information on the position of the detected object

relatively to the road and objects recognition (for instance many on-board RADAR are unable to make a distinction between a bridge and a car).

A system using a vision sensor in addition to a rangefinder can help to solve those problems. In this paper we are going to study the implementation and capabilities of two standard systems based on mono and stereo vision, which implies the modeling of the vision sensor. Then knowing this, we will underscore the reconstruction of the scene in 3D, through a simple method making it possible to simultaneously define the position of the object and the possible mistakes on this positioning. Assuming that the object points detected are static, and the video sensors’ parameters (intrinsic and extrinsic) are well known, we finally analyze this information on simulated and real data.

1.2 Monocular vision

One of the most important components of any machine vision system based on a single camera, is the detection and analysis of image motion. Properly done, image motion processing can lead to the recovery of “three dimensional” (3D) information such as structure, depth, and shape from a sequence of two dimensional images. The ability to derive such information has attracted substantial interest in image motion understanding from researchers in fields ranging to robotics.

A state of the art about monocular and binocular vision is done in [Alix *et al* 03]. This paper won’t discuss of the techniques developed during this time, but on the results derived using one method based on retro-projection of image points in 3D. The retro-projection technique is a mathematical method based on projective geometry and which allows to “locate” objects in 3D from their projected points in the image. Often, such a solution utilize some form of an object/feature matching technique to calculate the speeds and trajectories of the objects in motion and perhaps even make predictions about their future motion.

To locate moving objects, several assumptions are usually made to simplify the process: a maximum velocity of the object (given some short time interval dt , possible position points of an object must be inside a circle with radius v_{max}/dt centered at the object position in the previous image frame), bounded acceleration (the rate of change of velocity must be relatively small), and common object motion (different points in an object move in the same way). Of course, these assumptions do not always hold true for all image sequences: motion due to rotation and motion of non-rigid objects are both common examples of motion which do not conform to the above assumptions. When this

is the case, the results obtained by moving object detection are usually unreliable.

A far more general approach to representing motion in images is the optic flow field [Beauchemin 95]. Unlike the image differencing techniques, optic flow fields do not assume a fixed camera position. Optic flows describe the velocities of image features at a given instant in time and an optic flow field attempts to represent the three-dimensional motion of these features in a two-dimensional image [Alix et al 03], [Irani 98], [Uchimura 98].

1.3 Binocular vision

3-D structural information is directly derived from triangulation. For a conventional parallel axis stereo geometry the world coordinates is supposed to be coincide with the coordinate axes of the left camera without loss of generality (see Figure 1).

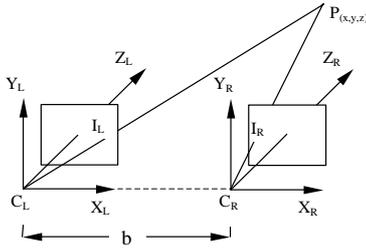


Fig. 1. Stereo Coordinate System

Fig.1 shows the imaging geometry of a stereo pair of cameras. The equivalent pinhole models represent two cameras with the projection centres C_L and C_R respectively. The origin of the world coordinates is C_L , and the stereo baseline is b . For a given 3-D scene point $P(x, y, z)$ the projection points on the left and right image are P_L and P_R respectively.

From the disparity map, the world coordinates of the scene point $P(x, y, z)$ can be easily obtained as:

$$x = \frac{bX_L}{d}, \quad y = \frac{bY_L}{d}, \quad \text{and} \quad z = \frac{bf}{d}, \quad \text{Eq.1}$$

where d is the disparity that is equal to $X_L - X_R$, b is the baseline and f is the focal length.

The detected and successfully matched object points are not necessarily all related to the object of interesting that were the pedestrians. The majority of the non-obstacles related points were on the road surface. The plane representing the road surface is found by a plane fitting procedure through all the lowest feature points in the scene, and therefore all the detected points that are on the road surface are discarded. Those points are normally representing the lane markings, shadows, texture...

2 Camera modelling

This part is dedicated to the modelling of the video sensor. We will define the transformations to go from the coordinates of some point 'P' in the 3D space to the projected point 'p' in the 2D image plane.

2.1 Integrating the vision modules in the vehicle

Fig.1 shows the positioning of cameras in the vehicle as well as the different coordinate systems that we have used. In the following, we will need to use four different coordinate systems (CS). The first one is the absolute R_a CS, linked with the road. The second is linked with the front part of the vehicle and it bears R_v . The third is the camera CS, bearing R_c . With a stereo system we will use the R_{cR} and R_{cL} notations for respectively the right and left cameras. The 4th CS correspond to the image plane, noted R_i . Assuming the cameras height is h , the tilt angle towards the ground plane is φ and d_c is the distance between the front of the vehicle and the projection of the centre of the camera on the road.

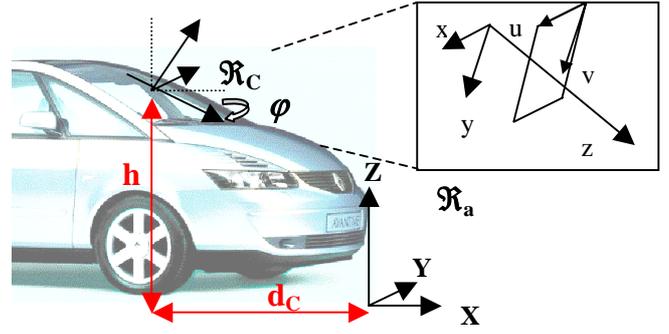


Fig. 3. Coordinate Systems used

To simplify the notations, the reference in which a vector is expressed be noted in a vector rating.

2.2 Camera parameters

Let $P = [x, y, z]^T$ denotes an object point in 3-D space and $[u, v]_{R_i}^T$ denotes its projection on the image plane. In the camera CS, the coordinates of a point are noted $[x, y, z]_{R_c}^T$. The perspective projection can be expressed as follows:

$$\begin{aligned} u &= \alpha_u \cdot x / z + u_o \\ v &= \alpha_v \cdot y / z + v_o \end{aligned} \quad \text{Eq.2}$$

where α_u , α_v are expressed as the focal reported to the image pixel size t_u and t_v in meter : $\alpha_u = f / t_u$, f is the

focal length and $[u_o, v_o]_{R_i}^T$ are image centre coordinates. With the standard video camera used, we can assume that $\alpha_u \cong \alpha_v \cong \alpha$.

2.3 Projective geometry

In the next part of this paper, we adopt the following notation: the expression of the coordinates of a point in the CS linked with the vehicle will be in capital letters, while in the CS linked with the camera, they will be in small letters.

The transformation from the vehicle CS to the camera is done through the composition of a translation vector $\vec{t} = d_c \vec{X} - h \vec{Z}$ and a rotation vector around \vec{Y} with angle φ .

In the case of a stereo perception system, it is necessary to

make a translation of \vec{LY} , with l , the half spacing between the cameras. The different transformation matrices in homogeneous coordinates are :

$$T_i = \begin{pmatrix} 1 & 0 & 0 & d_c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -h \\ 0 & 0 & 0 & 1 \end{pmatrix}, R_Y = \begin{pmatrix} \cos \varphi & 0 & -\sin \varphi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \varphi & 0 & \cos \varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. 3}$$

The transformation matrix between the vehicle CS and the camera is then $D_i = R_Y T_i$:

$$D_i = \begin{pmatrix} \cos \varphi & 0 & -\sin \varphi & d_c \cos \varphi + h \sin \varphi \\ 0 & 1 & 0 & \varepsilon_i l \\ \sin \varphi & 0 & \cos \varphi & d_c \sin \varphi - h \cos \varphi \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. 4}$$

, where 'i' is either l, m, r (Left, Mono, Right) and ε_i is respectively equivalent to $-1, 0, 1$.

To completely express the coordinates of the points in the CS linked with the camera, you have to operate a permutation of suffixes (and so get 'z' as the depth axis and (x,y) a plane parallel to the image plane), thus the expression of the projection will be simpler. Let it be M_{perm} , which allows us to have the coordinates of the point to be projected into the camera CS :

$$M_{perm} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. 5}$$

To obtain the coordinates (u, v) in the image CS, it is sufficient to use the classical projection formulas, and express M_{proj} as the perspective projection matrix :

$$M_{proj} = \begin{pmatrix} \alpha_u & 0 & u_o & 0 \\ 0 & \alpha_v & v_o & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{Eq. 6}$$

We get the transformation matrix between the vehicle CS to image plane :

$$M_i = M_{proj} \cdot M_{perm} \cdot D_i \quad \text{Eq. 7}$$

Let's call Q_t , the transformation matrix between the road CS and the vehicle at instant time t . These matrix coefficients come from proprioceptive sensors embedded in the vehicle (accelerometer, odometer, GPS, gyro,...) [Alix et al 03]. We obtain the transformation matrix between R_a and the image plane :

$$M_{i_t} = M_i \cdot Q_t = (m'_{ijk})_{j=1..3, k=1..4} \quad \text{Eq. 8}$$

Let $P = [X, Y, Z]_{R_a}^T$ be a point in R_a , then his coordinates in R_c are :

$$p = M_{i_t} P = [x, y, z]_{R_c}^T \quad \text{Eq. 9}$$

P coordinates in the image will be, according to equation 2:

$$u'_p = \frac{x}{z} = \frac{m'_{i11}X + m'_{i12}Y + m'_{i13}Z + m'_{i14}}{m'_{i31}X + m'_{i32}Y + m'_{i33}Z + m'_{i34}} \quad \text{Eq. 10}$$

$$v'_p = \frac{y}{z} = \frac{m'_{i21}X + m'_{i22}Y + m'_{i23}Z + m'_{i24}}{m'_{i31}X + m'_{i32}Y + m'_{i33}Z + m'_{i34}}$$

2.4 Back-projection

Knowing the changes to be performed to go from the coordinates of a point in the absolute CS to the coordinates of the point in the image, we are now going to perform the reverse work, using the knowledge of the point position in the image and that of the changes to perform.

2.4.1 Differences between mono and stereo vision

To locate a fixed point in space, we need to get 3 pieces of information : its coordinates X, Y and Z. In the case of a vision through a mono-camera system, we get at any moment, an image 'i' which can provide us the coordinates (u_i, v_i) of the tracked point. So we need two images corresponding to different instants, as well as the knowledge of the camera shifting (motion) between these two images. Then we call this temporal retro-projection or back-projection.

For a system based on stereovision, we get 2 images at any instant, noted i_{IR} and i_G , and then four pieces of information $(u_{IR}, v_{IR}, u_{IL}, v_{IL})$. This allows us theoretically get the position of the point in space at any moment. In this case this is spatial retro-projection. Nevertheless, with a stereo system, we can combine these two methods to follow a point both temporally and spatially. The recognition of the followed point between images resorts to matching methods which won't be developed in this article. A strong constraint common to these two types of retro-projection is the precise knowledge of the transformation between these two images.

2.4.2 Temporal retro-projection

As it can be shown in Fig. 4, to reconstruct a point in the 3D space from an image sequence (2 or more).

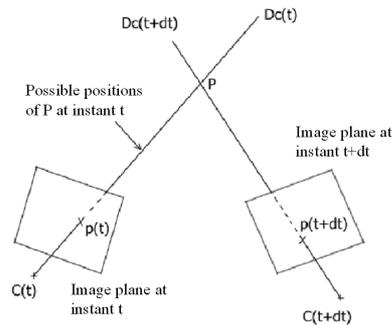


Fig. 4. Theoretical temporal retro-projection

Nevertheless, the pixel of an image has a non-zero size, then it exists an infinity of lines passing through this pixel. (Fig. 5).

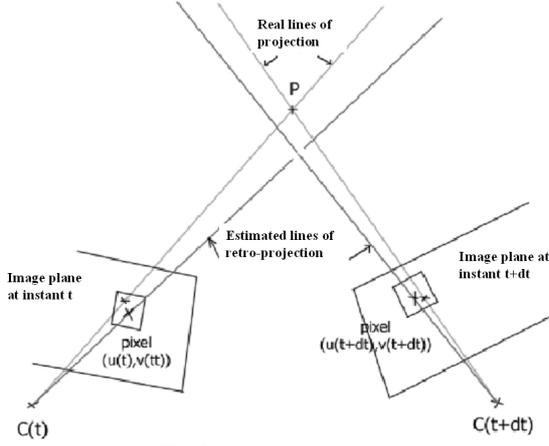


Fig. 5. Error due to pixel size

Assuming the same probability repartition of these lines, we are using a statistical method to search for the average position of point P [Olague 98]. This type of method allows us to obtain at the same time, an estimation of the error committed on the positioning of the point in space.

Position estimation

Let be P the point we are searching for the coordinates in Ra : $P = [X, Y, Z]_{Ra}^T$. Supposing that we know its projections $(u_p^t, v_p^t)_{t=1..n}$ in different images. We have also for each image, from equation 10 :

$$(u_p^t m_{i31}^t - m_{i11}^t)X + (u_p^t m_{i32}^t - m_{i12}^t)Y + (u_p^t m_{i33}^t - m_{i13}^t)Z = m_{i14}^t - u_p^t m_{i34}^t \quad \text{Eq. 11}$$

$$(v_p^t m_{i31}^t - m_{i21}^t)X + (v_p^t m_{i32}^t - m_{i22}^t)Y + (v_p^t m_{i33}^t - m_{i23}^t)Z = m_{i24}^t - v_p^t m_{i34}^t$$

We obtain a system which has the following form:

$$AP = b \quad \text{Eq. 12}$$

A is a $2t \times 3$ dimensions matrix in the case of monocular vision and $4t \times 3$ for stereovision.

In the stereovision case, this matrix contains the information provided by the left and right images at each time t. The solution obtained by a mean square method is possible from two images. Then equation 12 becomes:

$$A^T AP = A^T b \quad \text{Eq. 13}$$

Also, if $A^T A$ is inversible (this implies a non-zero movement between at least two images in the case of mono-vision), it comes:

$$P = (A^T A)^{-1} A^T b \quad \text{Eq. 14}$$

Error estimation

The inherent system error is due to the pixel size. It gives a 3D cone-shaped (Fig. 5) of probable positions of point P from a pixel. The error committed passing through a pixel is $\max(t_u/2, t_v/2)$.

The retro-projection lines repartition through a pixel is the same. This bring us to assume that the error could be

express like $X = \bar{X} + \varepsilon_x$.

By developing the equation 12 and neglecting errors of order 2, we obtain:

$$\bar{A} \varepsilon_p = \varepsilon_b + \varepsilon_A \bar{P} \quad \text{Eq. 15}$$

If we set $D = (\bar{A}^T \bar{A})^{-1} \bar{A}^T$, it comes :

$$\varepsilon_p = D(\varepsilon_b + \varepsilon_A \bar{P}) \quad \text{Eq. 16}$$

The errors committed on the main axes are obtained by extracting the elements on the covariance matrix diagonal on P, noted by Cov_p :

$$Cov_p = \varepsilon_p \varepsilon_p^T = D(\varepsilon_b + \varepsilon_A \bar{P})(\varepsilon_b + \varepsilon_A \bar{P})^T D^T \quad \text{Eq. 17}$$

This method exposed above is simple, nevertheless it allows to estimate both a target position and the error committed during the estimation process. As we will see it in the next parts, if we want to rise the accuracy in position, we must use more than two images.

3 Stereo and mono-vision compared accuracies

In the general case, we will take a fixed point to be ‘‘retro-projected’’. This point is at altitude 0.5m and has a lateral distance of 1.5m from the main axis. The parameters relative to the camera and its positioning are:

- $t_u = t_v = 8.3e-6$ m : size of the pixels
- $f = 8.5$ mm : focal distance
- $h = 1$ m : camera height
- $dc = 1$ m : distance between camera and the front of the vehicle
- $l = 1$ m : distance between the cameras (stereo base)
- $\alpha = 5.4^\circ$: angle between the camera optical axis and horizontal axis
- $v = 14$ m/s : vehicle speed
- $n = 10$: number of images
- $dt = 1/25$ s : video frequency

Number of images variation

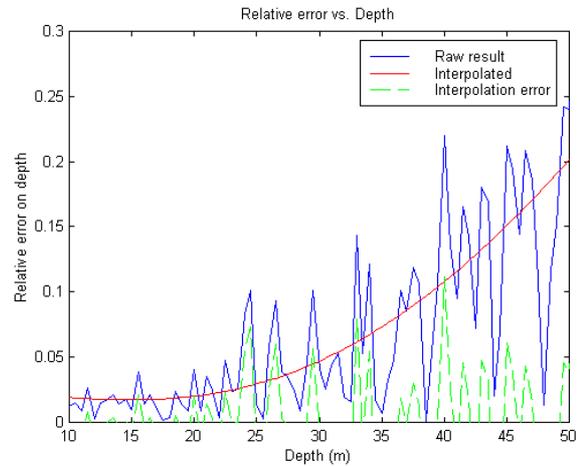


Fig. 6. Comparison between raw result and 2order interpolation, for a sequence of 10 images

In figure 6, we have characterized the relative error on target’s depth estimation in the scene (distance to the obstacle). The error is here only due to the pixel size on the CCD matrix. We can observe a global behavior of the

error evolution provided by the 2order interpolation. In the next parts we will assume that the error graphics are associated to those of interpolation.

On figure7, the error on estimation of depth decrease with the number of images, but concerning the obstacle detection application, it is not possible to wait for 25 images to detect precisely a fixed obstacle. For a retro-projection using 10 or 15 images, the ratio between accuracy and detection speed is suitable (the error is under 10 % at 35 m by using 10 images, and it is under this threshold at 45 m by using 15 images).

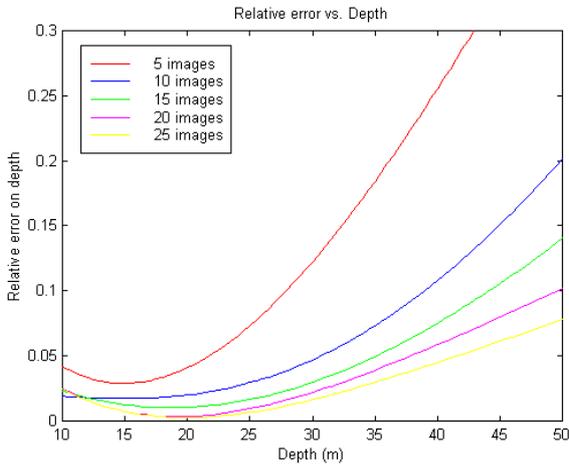


Fig. 7. Depth error variation for a monocular sequence of 10 images

For the graphs using 15, 20 and 25 images, we can observe a minimum of the error between 15 and 20 meters of depth (Fig. 7).

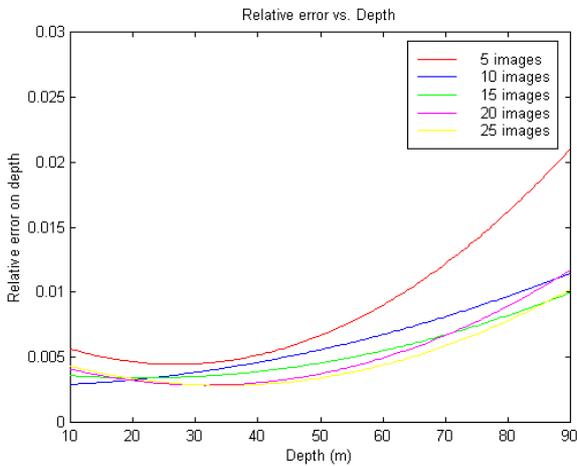


Fig. 8. Depth error variation for a binocular sequence of 10 images

But knowing that the vehicle speed is not zero and that the dimensions of the CCD matrix are not zero too, the physical point can not be visible in all the images of the sequence. This minimum is observed in the short depths more the number of images is decreasing. The part below this minimum has not any physical signification. We must consider the part greater than this minimum.

In the case of a binocular detection (Fig.8) we can not extract a global behavior of the graphs because of the small relative errors. Figure 9 shows the mean behavior of the retro-projection error. The maximum error observed at 90 m is below 5 %. The error is also really lower in the case of stereo than in the case of mono-vision, for two reasons :

- First of all, there are twice more information in

stereo than in monocular vision, for an identical number of samples.

- We can assume that the equipped vehicle is moving in the direction of the obstacles. As you can see it on figure 9, in the case of mono-vision, the region of the probable target point positions is greater than in the case of stereovision. This is due to the motion of image planes which is longitudinal in the case of monocular against a lateral plus longitudinal in the case of binocular.

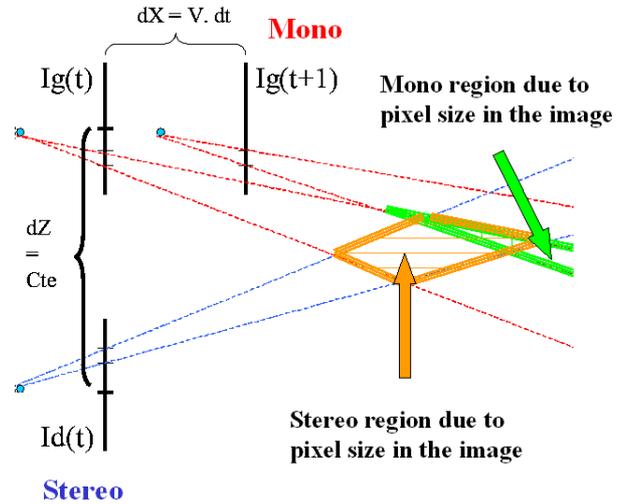


Fig. 9. Regions of the probable target point positions in stereo and mono

We have presented some results in the case of a fixed distance between cameras. Figures 10 (2D map) and 11 (3D map) show maps of the relative error projected on the estimation of the point to be retro-projected position (the height of this point is fixed to 0.5 m).

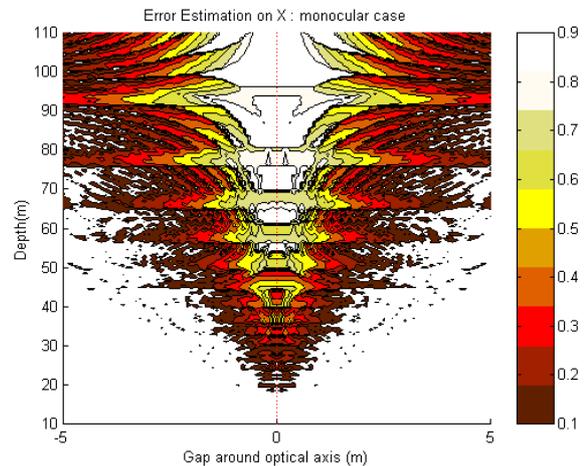


Fig. 10. Depth error variation around the optical axis for 10 monocular images (2D map)

Figure 10 presents two distinct regions. The first for which the error is greater than 10 %, corresponds to the camera's optical axis and its closest area (around 10 pixels on each side of the optical axis which is represented by the red line at $x = 0$, on Fig. 10). The errors are very important (100 %) for distances greater than 50m. The errors have also no sense. The second region (error lower than 10 %) corresponds to projected points which will have a more important motion in the sequence.

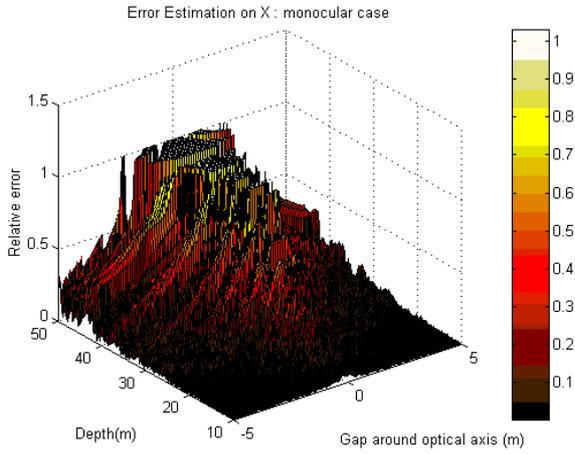


Fig. 11. Depth error variation around the optical axis for 10 monocular images (3D map)

We can observe this areas but more attenuated, on the figure 12 and 13. These results are valid in the case of a perfect detection. But the different particular feature detectors [Harris 88], [Schmid 98], [Achard-Rouquet 00]) are generally accurate at one pixel size. This error is introduced in the next section.

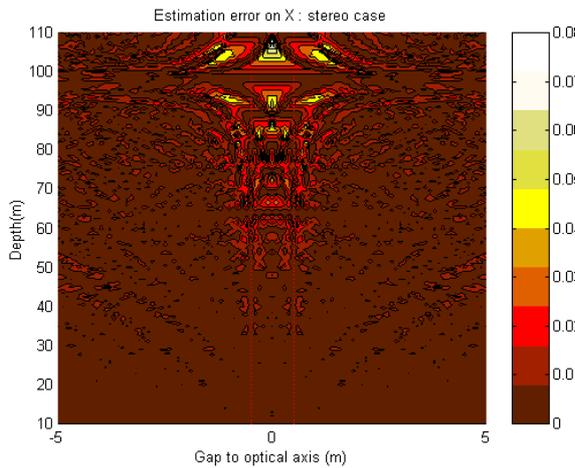


Fig. 12. Depth error variation around the optical axis for 10 binocular images (2D map)

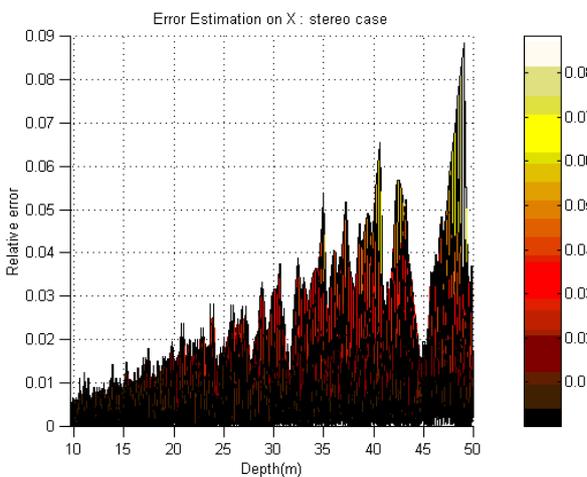


Fig. 13. Depth error variation around the optical axis for 10 binocular images (2D profile map)

When taking into account the error on detection

To simulate this error, we introduce a white centered noise. Thus, each of the 8 pixels around the projected real point

can be detected as a particular point. In the monocular case, the average relative error is multiplied by a factor 3 (Fig. 14). Nevertheless, in an obstacle detection application at short distances (less than 20 m), the error is about 10 %. The results keep being exploitable.

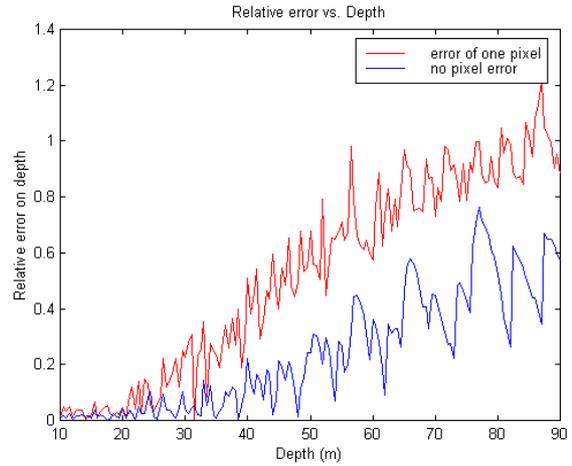


Fig. 14. Comparisons of detections with error or not – Monocular case

In the stereo domain, one pixel error at detection, multiplies the relative averaged error on depth by a factor 4 (Fig. 15). But it keeps being less than 4 % at 90 m (with a temporal tracking). The robustness of a stereo system beside the detection error is greater than those of a monovision system. In fact, we take more information in the stereo case than in the monocular case. But the noise is centered, so the mean of the measures is closer to the real data.

Reconstruction accuracy vs. intrinsic parameters

Other parameters are implied in the precision reconstruction. The first is the vehicle speed (this is not an intrinsic parameter) : the more greater it is, the more accurate is the reconstruction. This is because that the movement between images is taller, so the intersection of the retro-projection cone shaped is smaller. The focal distance is implied in the process, but in a less consequence on the reconstruction accuracy (Fig. 16).

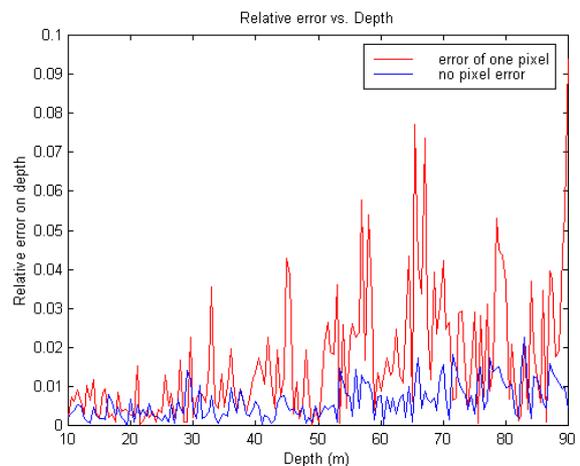


Fig. 15. Comparisons of detections with error or not – Binocular case

When the focal lengths are different (we took $f = 6.5$ mm and $f = 16$ mm), the relative error is quite identical below 45 m, and after that different of a factor 2.

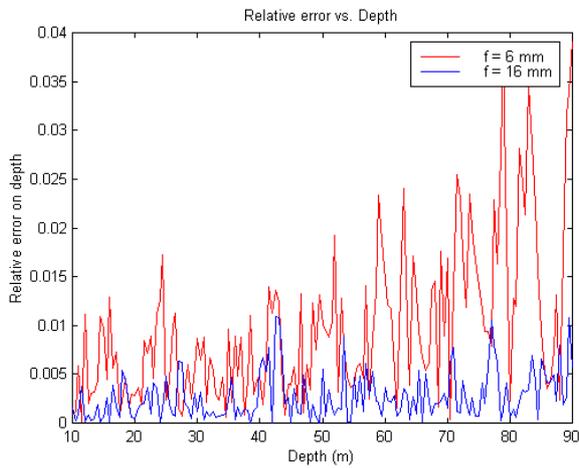


Fig. 16. Focal distance variation – Binocular case

The pixel size is quite important. By dividing the pixel sizes of a factor 2, the relative error on depth is divided by a factor 3 after 30 m (Fig. 17). It is also important to consider the sub-pixel feature detectors ([Devernay 95], [Achard-Rouquet 00]) and / or high resolution video cameras.

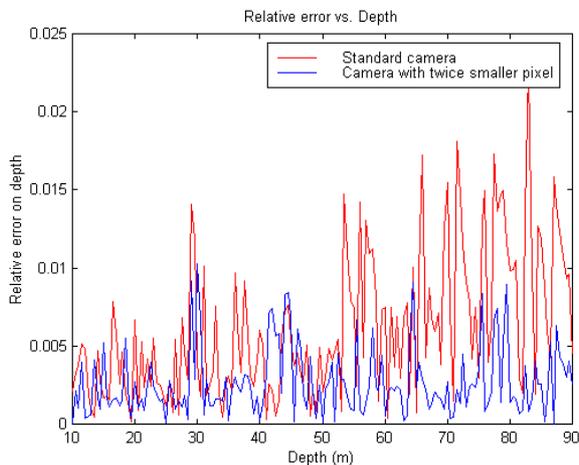


Fig. 17. Pixel size variation – Binocular case

In figure 18, the distance is computed and compared to the real distance to different targets at different distances. The cone shaped of uncertainty is represented in dot points on the graphs. The calculation is done on a pair of images, the retro-projection is this case only spatial. The images format is 1/4 PAL (384 * 288 pixels), to assume a real time calculation of the distances. Thus, in the worse case, the accuracy is equal to 0.7 % at 3 m and 14 % at 40 m. These calculations of accuracy must be correlated to figures 8 and 17. The image format used for our simulations is the PAL format (768 * 576 pixels), the precision is also 3 times lower. Regards to the simulations, the accuracy at 40 m is below 1 % (Fig. 8). Taking into account the chosen resolution, the error would be around 9 %, if the detection would be without errors. Figure 15 shows a detection with and without an error of one pixel. It shows also that it multiplies the reconstruction error with a factor 2 or 3, when adding one pixel error in the detection. By taking into account this parameter, the error committed on reconstruction should vary between 18 % and 27 %. The reconstruction method is also relatively precise.

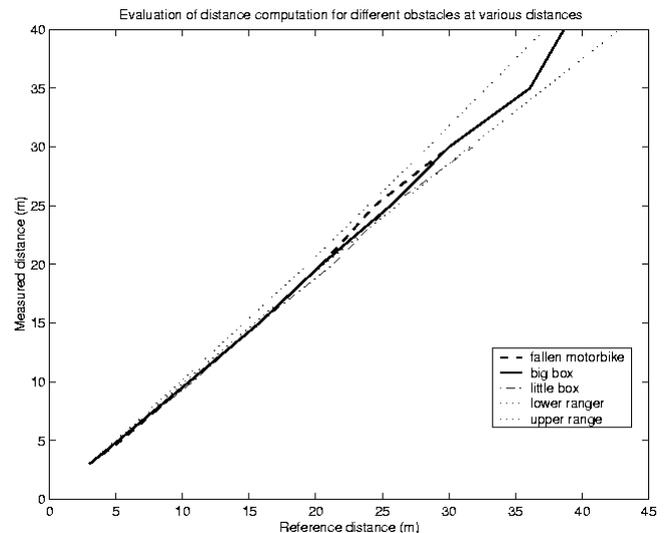
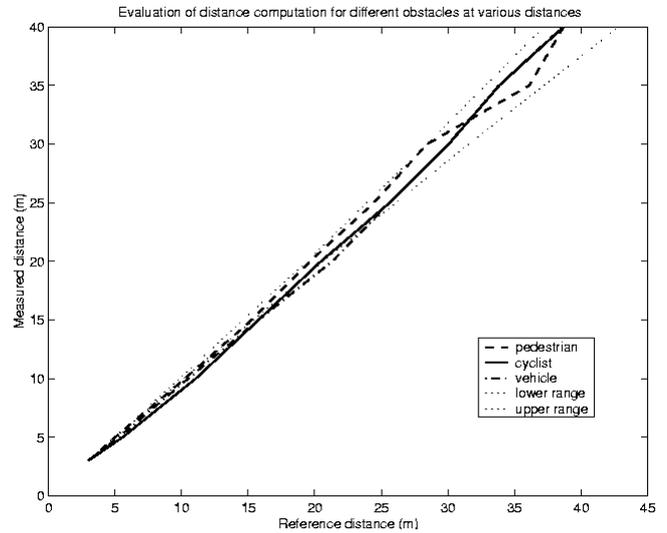


Fig. 18. Distance estimation accuracy for different targets and different distances – Real binocular case

Acknowledgements

The authors would like to acknowledge M. R. Labayrade [Labayrade *et al* 02], PhD student at the LIVIC Laboratory (Versailles, France) for his contribution in giving us information about the stereovision reconstruction performances.

4 Conclusion

We presented in this study only the case of fixed obstacles. For mobile objects, mono-vision can not give information on positioning without any strong constraints about the object to be detected (template, geometry, speed...). Thus, if these variables are supposed to be unknown, it is nevertheless possible to obtain a time to collision [Barron *et al* 94], which is independent of the depth scale factor (distance to target in monocular vision). A localization system based on stereovision [Koller *et al* 94] has not those problems, we have seen it in this analysis. This work could be useful for a video sensor specification which could in particular be usable in our research context (obstacle detection).

5 References

[Achard-Rouquet 00] Achard-Rouquet C., Bigorgne E., Devars J., "A sub-pixel and multi-spectral corner detector", ICPR, 2000.

[Alix *et al* 03] R.Alix, F.Lecoat, D.Aubert, "Flat World Homography for Non-Flat World On-Road Obstacle Detection", *to appear* in IV – Intelligent Vehicle Conference, Columbus, USA, 2003.

[Barron *et al* 94] J. L. Barron, S. S. Beauchemin and D. J. Fleet, "On Optical Flow", 6th Int. Conf. on Artificial Intelligence and Information-Control Systems of Robots (AIICSR), pp. 3-14, Sept. 12-16, 1994, Smolenice Castle, Slovakia.

[Beauchemin 95] Beauchemin S. S., Barron J. L. The Computation of Optical Flow. *ACM Computing Surveys, Vol 27, No 3*, pp. 433-467, 1995, Sept.

[Chausse *et al* 00] F. Chausse, R. Aufrère and R. Chapuis, "Recovering the 3D Shape of a Road by On-Board Monocular Vision", ICPR, 2000.

[Devernay 95] Devernay F., "A non-maxima suppression method for edge detection with sub-pixel accuracy", INRIA Technical report, 1995.

[Dickmanns 92] E.D. Dickmanns and B.D. Mysliwetz, "Recursive 3D Road and Relative Ego-State Recognition", proceedings IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992.

[Franke 00] U. Franke and A. Joos, "Real-Time stereo vision for urban traffic scene understanding", DaimlerChrysler AG. D-70546 Stuttgart. HPC : T728. IEEE, 2000.

[Harris 88] Harris C., Stephens M., "A combined corner and edge detector", Proceedings of the 4th Avley Vision Conference, pp. 147-151, 1988.

[Irani 98] M. Irani and P. Anandan. "A Unified Approach to Moving Object Detection in 2D and 3D Scenes". IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), June 1998.

[Koller *et al* 94] D. Koller, T. Luong, and J. Malik, "Binocular stereopsis and lane marker flow for vehicle navigation : lateral and longitudinal control", Technical report of UCB/CSD 94-804, University of California at Berkeley, Computer Science Division, 1994.

[Labayrade *et al* 02a] R.Labayrade, D.Aubert, J.P.Tarel, "Real Time Obstacle Detection on Non Flat Road Geometry through 'v-disparity' Representation", IEEE Intelligent Vehicles Symposium, Versailles, June 2002.

[Labayrade *et al* 02b] Labayrade R. and Aubert D., "Robust and Fast Stereovision Based Road Obstacles Detection for Driving Safety Assistance", Machine Vision and Application 2002 (MVA 2002), Nara, Japan, 11-13 December 2002.

[Labayrade *et al* 03] Labayrade R. and Aubert D., "Onboard Road Obstacles Detection in Night Condition Using Binocular CCD Cameras", to be published in ESV 2003 Proceedings, Nagoya, Japan, May 2003.

[Olague 98] Olague G., Mohr R., "Optimal Camera Placement TO obtain Accurate 3D Point Positions", ICPR 98, pp. 8-10, 1998.

[Schmid 98] Schmid C., Mohr R., Bauckage C., "Comparing and Evaluating Interest Points", ICCV, pp. 230-235, 1998.

[Tarel 00] J. P. Tarel and F. Guichard, "Combined Dynamic Tracking of Lane Markings and Boundaries in Curved Road", IEEE International Conference on Image Processing (ICIP'00), September 2000.

[Uchimura 98] Uchimura K., Hu Z., « Lane Detection and Tracking Using Estimated Camera Parameters for Intelligent Vehicles », Proc. IEEE ICIV, pp. 11-16 (1998).

[Williamson 98] T. A. Williamson, "A high-performance stereo vision system for obstacle detection", PhD CMU-RI-TR-98-24 Robotics Institute Carnegie Mellon University. Pittsburg, PA, September 1998.