3D Modeling of Urban Environments for Enhancing GPS Localization's Accuracy

Julien Moreau, Sébastien Ambellouis, Yassine Ruichek

3D modelling of urban environments for enhancing GPS localization's accuracy

Julien Moreau (a), Sébastien Ambellouis (b), Yassine Ruichek (a)

(a) IRTES-SET, UTBM, 90010 Belfort cedex, France
(b) IFSTTAR, LEOST, Univ Lille Nord de France - F-59000 Lille, F-59650 Villeneuve d'Ascq, France

Abstract

Described project aims at accessing accurate localization. It consists in improving accuracy of localization systems with video perception of the environment around equipped vehicles. The environment corresponds to the propagation area of GNSS receiver installed on top of the vehicle. Accuracy problems are linked to obstacles density, especially in urban environments. The objective is to reduce multipath effect on the measurements, without lowering the availability of the service, which is typically the case for errors detection procedures. The presented work aims at using two fisheye cameras in a stereoscopic configuration in order to compute 3D coordinates of the scene's points. A reconstructed 3D model allows then to correct the propagation error thanks to the knowledge of the satellites' reception status.

Keywords: localization ; computer vision; urban environments ; stereovision ; 3D.

Résumé

Le projet décrit ici a pour but d'accéder à une localisation précise. Il s'agit d'améliorer la précision des systèmes de localisation avec la perception vidéo de l'environnement autour du véhicule. L'environnement correspond à la zone de propagation du récepteur GNSS installé sur le toit du véhicule. Les problèmes de précision sont liés à la densité en obstacles, en particulier dans les environnements urbains. L'objectif est de réduire l'action de l'effet multi-trajet sur les mesures, sans diminuer la disponibilité du service, ce qui est généralement le cas avec les procédures de détection des erreurs. Le travail présenté a pour but d'utiliser deux caméras fisheye dans une configuration stéréo pour calculer les coordonnées 3D des points de la scène. Un modèle 3D reconstruit permettant ensuite de rectifier les erreurs de propagation grâce à la connaissance de l'état des satellites perçus.

Mots-clé : localisation ; vision par ordinateur ; environnement urbain ; stéréovision ; 3D.

1. Introduction

Multipath effect is mainly caused by urban structures. Signals are reflected on objects and take more time to reach the GNSS receiver. The receiver processes them as provided by line of sight satellite and thus generates errors of few meters. Fig. 1 shows GPS positions in red of a car moving straightly in a small street.

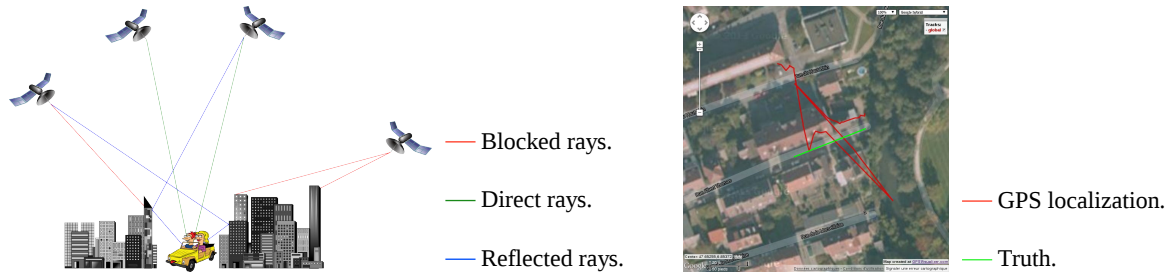

Fig. 1. (a) multipath effect description; (b) multipath effect consequence.

To solve this problem, the idea is to generate a 3D model of the environment in order to estimate the delays of reflected rays between the satellites and the GPS antenna. By doing this dynamically when the vehicle moves, the errors can be reduced by correcting the pseudorange used to estimate the position.

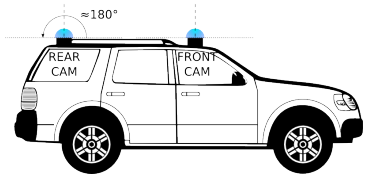

Fig. 2. fisheye stereovision configuration on a car.

A way to compute a 3D model is to use the principle of stereovision. 3D information can be recovered by using two views of the same scene acquired at different positions. As we need to know the structures from all around the vehicle, we use fisheye cameras, that acquire 180° field of view, but with highly distorted images. These cameras are installed on top of the vehicle with longitudinal alignment, see Fig. 2.
The 3D point clouds are computed by matching the contents of the stereo images that required a calibration step.

2. Calibration

2.1. Epipolar geometry for the spherical model

An omnidirectional camera can be modeled as a unit sphere as described in the Fig. 3. If *P* is a point in the real world reference frame, it can be projected on the sphere's surface and then on a sensor according to the lens projection function. With two views, *P* is represented by two points at different positions on the sensors. To compute the disparity of these points i.e. the position difference allows getting the 3D information. Epipolar geometry has an interesting property: all 3D points belonging to the same epipolar plane are projected along the same epipolar curve on each of both sphere and both sensor; both 3D or 2D curves are conjugate. For fisheye cameras, one shall consider the difference between angular positions of the 3D point projections on both spheres inside conjugate 3D epipolar curves (α_r and α_l in Fig. 3), that is the angular disparity.

The relation from an image to another is described by a 3x3 matrix called the Fundamental matrix up to scale. A calibration step yields this matrix. It aims at determining cameras' internal (fisheye distortions in the projection function, optical center, skew) and external (rotation and translation between both sensors) parameters. The Fundamental matrix is often used as it contains external parameters. Scale factor may be retrieved an other way.

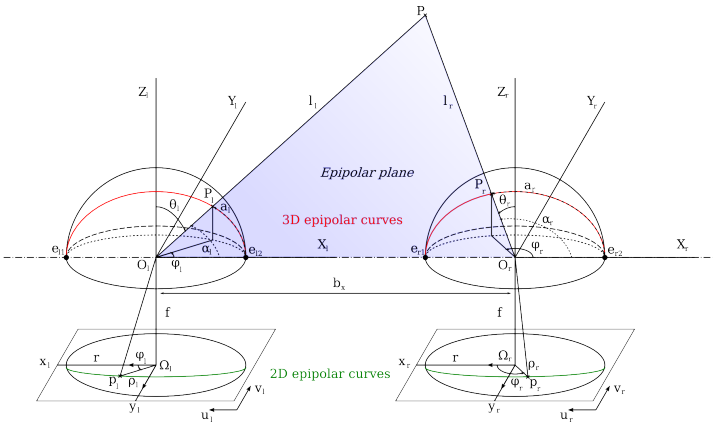

Fig. 3. epipolar geometry of fisheye stereovision setup.

2.2. Calibration method based on feature points

Two categories of calibration procedures exist for a stereoscope.
Strong calibration is based on a pattern, usually a checkerboard. It consists in acquiring different images of the pattern (or the cameras) at different positions and computing parameters by minimizing an error function. Scale is known thanks to the pattern dimensions. Some authors use a classic planar pattern (Gehrig et al., 2008a), while others use an half-box pattern (Li, 2008b) or develop a specific one (Ragot, 2009b, creates a cylindrical pattern). Half-box pattern needs less acquisitions as it provides points inside 3 orthogonal planes, but it is usable only with short baseline stereoscopes.
Self-calibration methods use features automatically detected and matched in both images. Mičušìk (2004b) and Kawanishi et al. (2009a) use features points. These methods are based on the RANSAC algorithm, in order to remove outliers and consider only detected features couples with a high reliability for the final parameters estimation. As is, self-calibration permits computing the Fundamental matrix but does not provide the scale factor. The advantage is that the process is automatic and fast. Self-calibration may be applied several times during acquisitions in order to ensure keeping good parameters, and it is necessary for a mono-camera stereo setup when images are acquired from the same moving sensor.

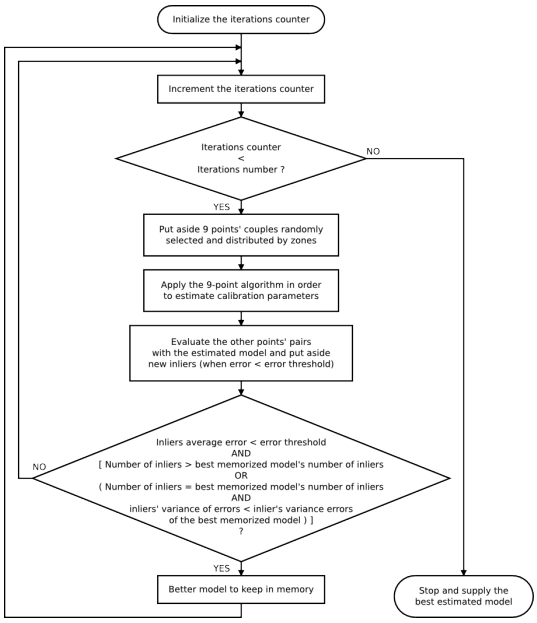

Fig. 4. RANSAC algorithm used for self-calibration.

As it is more flexible, retained method is based on self-calibration. Mičušìk's 9-point algorithm (2004b) is used and adapted to the fisheye lens used in our experiments. We use the RANSAC as described in algorithm Fig. 4 and compute the Fundamental matrix. Scale factor can be retrieved with the knowledge of the baseline between both cameras.

2.3. Calibration results

Calibration evaluation is estimated by the stability of the calibrations for the same pictures. Fig. 5 & 6 present an example of calibration images for a real scene. Displayed results are obtained for 200 iterations of the calibration process for the couple of real images. Evaluated parameter is called a and is linked to the lens field of view. Initial a_0 value is 0.7071 and corresponds to a 180° field of view given by manufacturer specifications. The standard deviation of a estimation is 0.0034. This standard deviation is very low, meaning that estimations are precise at $3x10^{-3}$. The average of the estimated a value is $\hat{a}$ = 0.7309. We can consider this value as the real a value for this lens. The difference between $\hat{a}$ and a_0 is 0.0238, a high value compared to the scale of a estimations standard deviation. According to the $\hat{a}$ value and the lens projection model, we can conclude that the lens field of view is about 188° (8° difference with the specifications!).

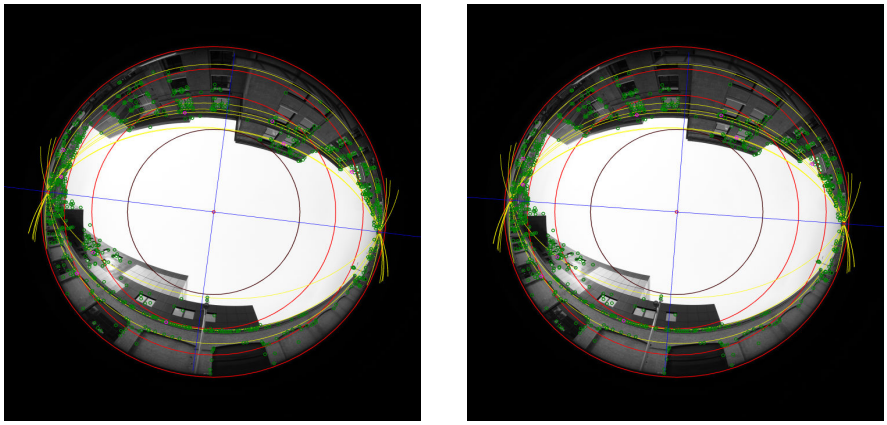

Fig. 5. Calibration images couple example.

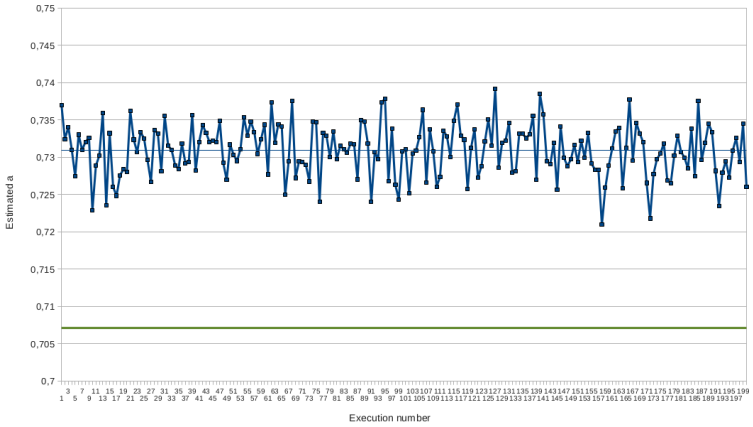

Fig. 6. Stability of the parameter issued from the calibration for 200 executions.

With the knowledge of calibration parameters, one can apply a matching process in order to measure points' disparities and to compute 3D positions.

3. Matching and 3D points computation

3.1. Matching points by dynamic programming

There exists a large variety of matching point techniques. It is possible to sort them according to their local or global properties or according to the dense or sparse computed map.
Local methods are looking for corresponding points around the same area in both images and compare data locally. Global methods are intended to minimize a cost function linked to a matching solution within a global optimization process that can takes into account a complete set of points couples at the same time. Global methods are slower to compute than local ones, but provides better results. The reader can find a local and a global method with fisheye images transformed into cylindrical views in Gehrig et al. (2008a). All methods may exploit the epipolar constraint, i.e. projections of 3D points are lying in conjugate epipolar curves (see Fig. 3), in order to reduce the research area to conjugated curves and improve computation time and matching efficiency.
Dense methods match all the points (Li, 2008b), whereas sparse method only match well established feature points (Mičušîk, 2004b) with a high confidence measure. Dense data is complete but subject to more matching errors than sparse content.
Chambon (2005) gives a complete taxonomy of matching algorithms for standard images.

(← j) left \ right (i →)	13	14	15	10	11
10	3	4+3	5+7	0+12	1+12
12	1+3	2+3	3+5	2+8	1+10
14	1+4	0+4	1+4	4+5	3+9
15	2+5	1+4	0+4	5+4	4+9
10	3+7	4+5	5+4	0+4	1+4

(← j) left \ right (i →)	13	14	15	10	11
10	3	7	12	12	13
12	**4**	5	8	10	11
14	5	**4**	5	9	12
15	7	5	**4**	9	13
10	10	9	9	**4**	**5**

Fig. 7. Matching by dynamic programming: (a) graph construction; (b) shortest path in the graph.

An interesting matching technique is the dynamic programming, as it gives a good trade-off between computation time and accuracy (Forstmann et al., 2004a). We implement and adapt this technique for fisheye images. This is a global and dense matching process, based on a graph built for each couple of conjugate epipolar curves (epipolar constraint). Graphs are based on dissimilarity measures between points of both curves, with a propagation of lowest costs. The lowest-cost path provides the matching solution for the points of conjugate curves.
For pinhole images, epipolar curves are straight lines. For fisheye images, epipolar curves are represented by arcs of a circle. The need is to compute curves and their points and to stretch them in order to construct graphs. For same size images, a stretched fisheye epipolar curve is significantly longer than a standard epipolar line (see Fig. 8). The result is that matching computation time is a lot longer for omnidirectional stereovision. This observation remains true for all matching techniques based on the epipolar constraint.

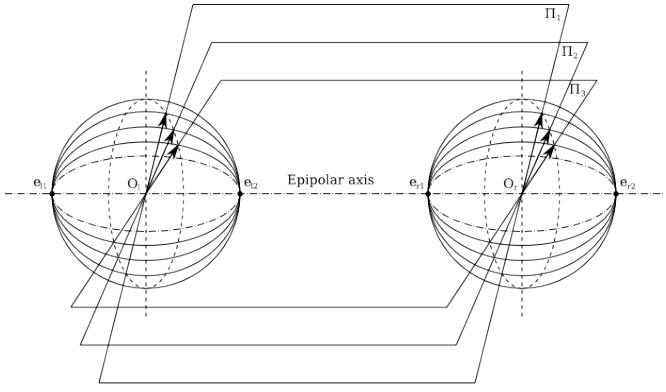

Fig. 8. Scanning epipolar planes to get conjugated epipolar curves.

After matching the images points, 3D points' positions can be determined.

3.2. 3D points computation

3D points coordinates computation can be done using points' angular position in 3D epipolar curves. Li (2008b) gives the first step to this method. Firstly, it is possible to determine distances of each 3D points to both cameras centers. Let be l_l and l_r the distances of a point P to left and right cameras (see Fig. 3):

$$l_l = \frac{b_x \sin(\alpha_r)}{\sin(\alpha_r - \alpha_l)} \qquad \textit{and} \qquad l_r = \frac{b_x \sin(\alpha_l)}{\sin(\alpha_r - \alpha_l)} \tag{1}$$

With the distances, it is possible to compute 3D coordinates of scene's points, centered on each camera frame. Following equation gives the 3D position of P in the left camera frame:

$$P\,(\, l_l \cos(\alpha_l),\, l_l \sin(\alpha_l) \cos(\eta),\, l_l \sin(\alpha_l) \cos(\eta)\,), \tag{2}$$

where η designs the rotation angle of the epipolar plane containing P to the Oxy plane.

To illustrate matching results (Fig. 9), we calculate the distance maps centered on each view, we build the 3D point clouds and the mesh using a triangulation method.

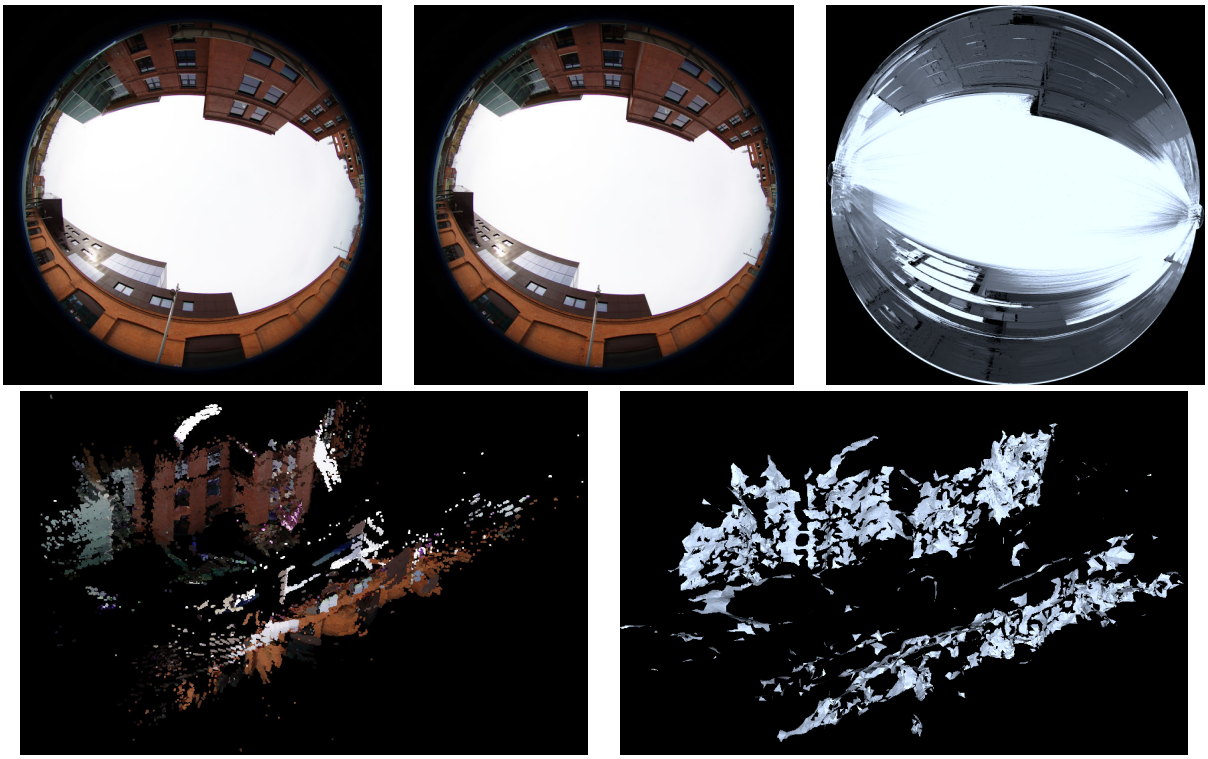

Fig. 9. Result for a real scene: (a) & (b) original images; (c) left distance map; (d) 3D point cloud; (e) smoothed mesh.

3D points computation allows getting distance maps and 3D models. Results depend on the calibration and on the matching. It is possible to do an evaluation that shows the quality of the two steps process. We can not evaluate the matching separately from the calibration, a good matching depends on a good calibration.

4. Point cloud evaluation

4.1. Simulated scene evaluation with ground truth

Proposed evaluation is based on comparison between the computed distances maps and distances ground truth. Because, ground truth is impossible to obtain for real scene, we propose to use synthetic images. The synthetic scene is composed of buildings, represented by cubes with textured surfaces. Synthesis configuration is a

stereoscope with a 2 meters baseline between both sensors. In order to approach the reality, lenses and cameras real distortions or characteristics are used during the synthesis: chromatic aberrations, image circle's edge attenuation and blue fringe, flare, low amount of optical dispersion blur, field of view of 181.8° (and considered as 180° in calibration step's initialization). The cameras used for synthesis are based on the same fisheye projection model as the cameras of the real experimental setup. An additional rotation (around 7° to each axes) and a translation (about 9 cm) are added to the rear camera to approach the positioning error we could get in a real case.

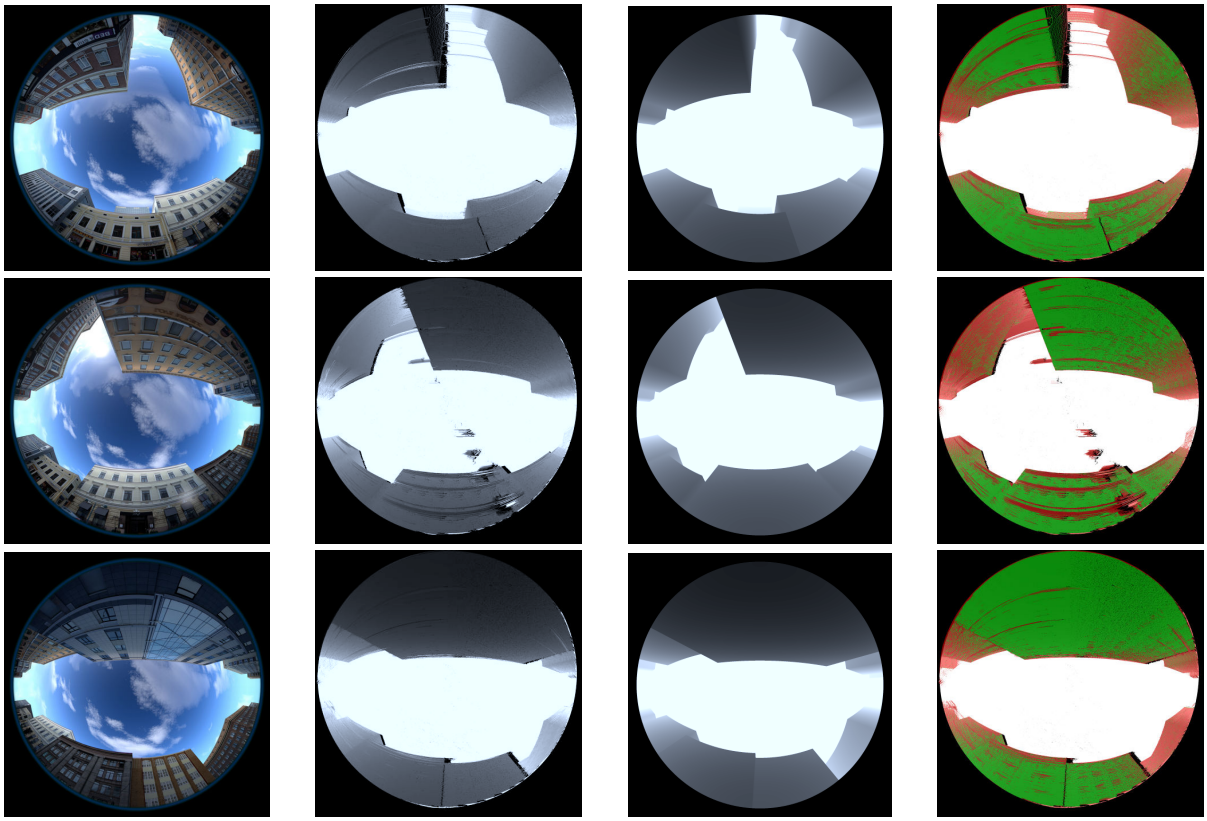

Fig. 10. First column: front image; 2nd: matching result's distance map; 3rd: ground truth; 4th: evaluation map

Proposed evaluation compares computed distance maps with ground truth. Evaluation is done for a set of 30 images couples, the stereoscope moving in the simulated street. Fig. 10. shows three results. The first and the third row are the starting and the ending positions of the trajectory. The second row contains flare that generates more errors. Distance maps are drawn with grayscale values from 0 to 255, in order to represent distances from 0 to 100 meters. A correct pixel is a pixel estimated with a gray level difference of at most 3 levels (1.2 meters). Correct pixels are drawn in green, and incorrect ones are drawn in red. Presented matching procedure matches one left pixel to one right pixel, and no multiple matching is allowed. Sometimes, a surface can be represented by far more pixels in an image than in the other one. In this case, some pixels cannot be matched, they are called the occulted pixels. These pixels are drawn in black and are not considered as false matching, as they do not give any information instead of giving erroneous content. We can see that erroneous points are mostly located at the most far distances. Error can propagate along an epipolar curve. The lens flare generates error in the sky and in the structures. Table 1. gives evaluation's values. Compared values are gray levels.

Proposed measures are the following:

- Concerning pixels that are nearer than 100 meters:
  - Percentage of occulted pixels.....%OC
  - Percentage of correct pixels among non-occulted pixels.....%OK
  - Percentage of incorrect pixels among non-occulted pixels.....%WR

- Average error among non-occulted pixels (gray levels).....ERR
- Concerning pixels that are at least at 100 meters (essentially the sky):
- Percentage of correct pixels.....%sOK
- Percentage of incorrect pixels.....%sWR

Result values are average values for the evaluations of this complete set. Currently, we are able to compute points' distances with an average error of approximately 2.67 meters. As shown in Fig. 10. & 11. the errors are mostly located on distant points. These points are the less relevant data for the presented application. It appears that the 100 meters limit may be too large, it could be better to generate smaller models for which the accuracy is good enough.

Table 1. Ground truth evaluation results.

%OC	%OK	%WR	ERR	%sOK	%sWR
12.61%	58.98%	28.41%	6.80 (2.67 m)	95.70%	1.73%

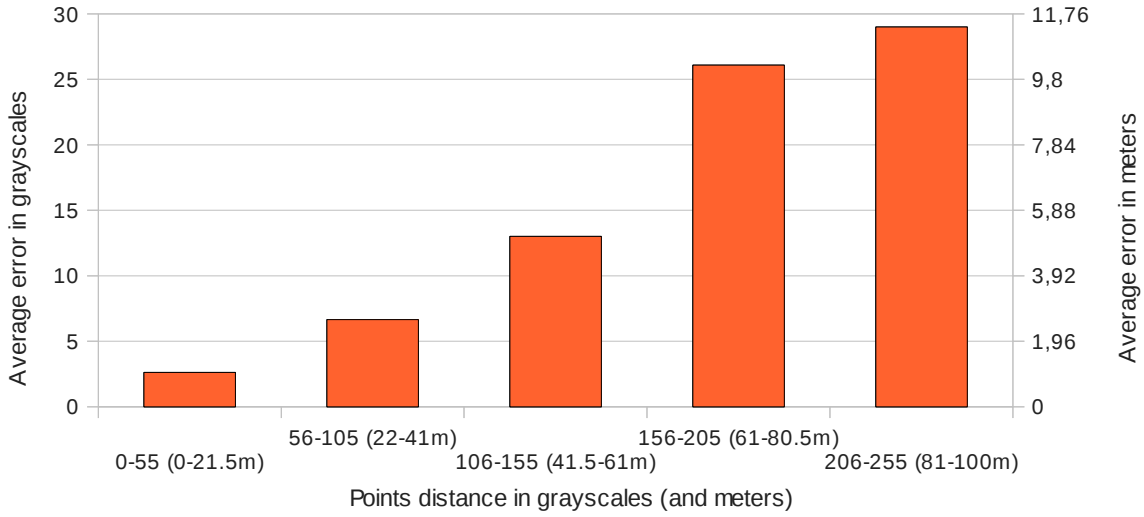

Fig. 11. Average distance error sampled according to the distance to the sensor.

5. Conclusion

The goal of this work is to improve GPS localization of a mobile moving in a urban environment by estimating the 3D structure of buildings around the vehicle and by reducing the errors due to reflections.

In this paper, we propose a 3D reconstruction of urban scene based on a stereo fisheye setup. The 3D model is computed from the point clouds thanks to a triangulation meshing method. The 3D point clouds are obtained by matching the points of conjugate epipolar curves. The matching process is based on dynamic programming techniques. The epipolar constraint is reached thanks to a self-calibration step that yields fundamental matrix. The proposed calibration method is evaluated. Its results are accurate and stable. The matching process has been applied on synthetic and real images. We propose an evaluation on synthetic images to take advantage of a ground truth. It shows quite accurate positions for points up to 50 meters from the vehicle: the average error is less than 2.6 meters.

Future works aim to finalize a 3D model with a mesh or planes fitting walls. With the knowledge of satellites positions and GNSS signals SNR used jointly with the built model, it will become possible to estimate GNSS rays trajectories including multipath effect and pseudodistances errors. Hence, this precise representation of real events can lead to a rectification of errors and to a better GPS localization accuracy in urban environment.

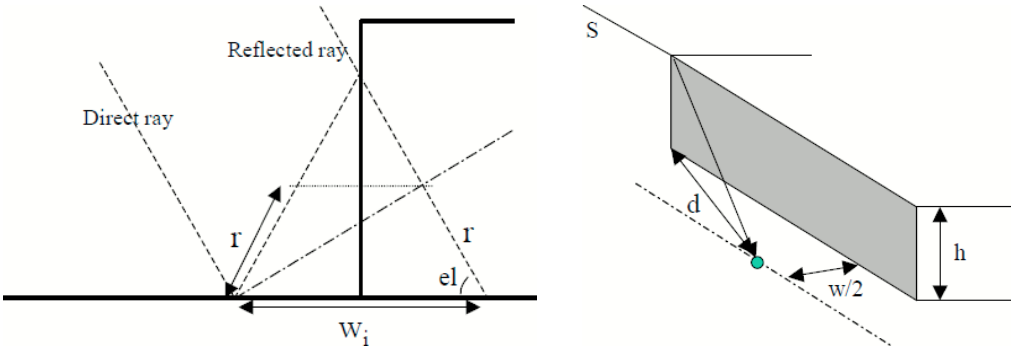

Fig. 12. Urban canyon and delay *r* caused by the reflection: (a) 2D cross-section; (b) 3D (Marais et al., 2013).

Rectification principles are described by Marais et al. (2013) and illustrated in Fig. 12. The idea is to compute the delay of the propagated signal caused by the reflection *r* in order to correct the measure $\rho_{measure}$ performed by the receiver. Then, the position could be computed using the value $\rho_{corrected}$:

$$\rho_{corrected} = \rho_{measure} - r \quad (3)$$

Delay's length *r* depends on the urban environment and on the satellite position:

$$r = 2\,d\cos(el) \quad (4)$$

Where *d* is the distance between the GPS receiver and the impact point projection on the ground, and *el* is the satellite elevation.

Acknowledgements

This work is carried out as part of the project CAPLOC (*Combinaison de l'Analyse d'image et de la connaissance de la Propagation des signaux pour la LOCalisation*), supported by ANR and funded by MEDDTL.

References

Chambon, S. (2005). Mise en correspondance stéréoscopique d'images couleur en présence d'occultations. *PhD thesis*, École Doctorale Informatique et Télécommunications, Université Paul Sabatier – Toulouse III, Institut de Recherche en Informatique de Toulouse.

Forstmann, S., Kanou, Y., Ohya, J., Thuering, S. & Schmitt, A. (2004a). Real-time stereo by using dynamic programming. *In Computer Vision and Pattern Recognition Workshop*.

Gehrig, S. K., Rabe, C. & Krüger, L. (2008a). 6d vision goes fisheye for intersection assistance. *In Canadian Conference on Computer and Robot Vision*, pp. 34–41.

Kawanishi, R., Yamashita, A. & Kaneko, T. (2009a). Three-dimensional environment model construction from an omnidirectional image sequence, Department of Mechanical Engineering, Shizuoka University, Shizuoka, Japan.

Li, S. (2008b). Binocular spherical stereo, *IEEE Transactions On Intelligent Transportation Systems*, vol. 9, pp. 589–600.

Marais, J., Attia, D., Meurie, C., Tay, S. & Ambellouis, S. (2013). CAPLOC, Livrable d'avancement de la tâche 3. Utilisation de l'image pour la caractérisation des signaux reçus et de la disponibilité globale de la constellation. *Technical report*, IFSTTAR/LEOST and UTBM/IRTES-SET.

Mičušìk, B. (2004b). Two-View Geometry of Omnidirectional Cameras. *PhD thesis*, Czech Technical University in Prague.

Ragot, N. (2009b). Conception d'un capteur de stéréovision omnidirectionnelle : architecture, étalonnage et applications à la reconstruction de scènes 3D. *PhD thesis*, École Doctorale Sciences Physiques, Mathématiques et de l'Information pour l'Ingénieur, Université de Rouen, Institut de Recherche en Systèmes Electroniques EMbarqués.