2015 IEEE International Conference on Robotics and Automation (ICRA)
Washington State Convention Center
Seattle, Washington, May 26-30, 2015

Radar and vision sensors calibration for outdoor 3D reconstruction

Ghina El Natour [1], Omar Ait Aider[1], Raphael Rouveure[2], François Berry[1] and Patrice Faure[2]

Abstract— **In this paper we introduce a new geometric calibration algorithm, and a geometric method of 3D reconstruction using a panoramic microwave radar and a camera. These two sensors are complementary, considering the robustness to environmental conditions and depth detection ability of the radar on one hand, and the high spatial resolution of a vision sensor on the other hand. This makes the approach well adapted for large scale outdoor cartography. Firstly, we address the global calibration problem which consists in finding the exact transformation between radar and camera coordinate systems. The method is based on the optimization of a non-linear criterion obtained from a set of radar-to-image target correspondences. Unlike existing methods, no special configuration of the 3D points is required, only the knowledge of inter-targets distance is needed. This makes the method flexible and easy to use by a non expert operator. Secondly, we present a 3D reconstruction method based on sensors geometry. Both methods have been validated with synthetic and real data.**

I. INTRODUCTION

The conscience of the surrounding environment is an inevitable task for several applications such as mapping, autonomous navigation and localization. Therefore, data acquisition through sensors is required. A survey of 3d reconstruction of urban environments is represented in [1]. Despite the large number of studies and researches in this field ([2], [3]), there are still many challenges for fully automatic and real time modeling process together with high quality results, because of acquisition and matching constraints, calling for more contributions. The methods can be classified according to sensors used: vision sensor, range sensor, odometers or a set of sensors. To overcome the limitations of single sensor approaches, multi-sensory fusion have been recently a point of interest in widespread applications and researches especially for 3D mapping applications.

Regarding the low cost and high spatial resolution of vision sensors, a huge number of vision based approaches for 3D reconstruction have been proposed. Some examples can be found in [4], [5] and [6]. Methods for 3D scene reconstruction from an image sequence can be grouped in two classes: Structure-from-Motion and dense stereo. In the last years, many works tend to fill the gap between the two approaches in order to propose methods which may handle very large scale outdoor scenes. Results seems to be of good quality though it recommends a large amount of input data and heavy algorithms which make it not quite suitable for real time processing. It is also known that techniques for large scene reconstruction by vision generally suffers from scale factor drift and loop closing problems. In addition, vision sensors present several drawbacks due to the influence of image quality when illumination and weather conditions are deteriorated. For this reason, tapping into active sensors has become essential.

The capability of a range sensor to work in difficult atmospheric conditions and its decreasing cost, has made it well suited for extended outdoor robotic applications. For exemple, [7] et al. used RAdio Detection And Ranging (RADAR) sensor for simultaneous localization and mapping (R-SLAM algorithm) applications in agriculture. However, range sensors fail to recognize elevation, shape, texture, and size of a target. Many solutions based on the combination of depth and vision sensors are described in the literature. An example of this fusion can be found in [8] and [9] the authors combine a 3D LIght Detection And Ranging (LIDAR) and image sensing to create geometric 3D models of the world. The Lidar provides a large number of 3D points which need data processing algorithms, and can be memory and time consuming. SLAM applications with Kinect are also numerous ([10], [11]). Yet, the performances for outdoor applications are generally limited due to the small depth range and sensitivity to illumination conditions.

In this paper we are interested by the combination of a panoramic millimeter wave (MMW) Radar and a camera, in order to achieve a sparse 3D reconstruction of large scale outdoor environments. Recently, this combination has been studied for on-road obstacle detection and vehicle tracking: in [12], camera and radar were integrated with an inertial sensor to perform road obstacle detection and classification. Other works on radar-vision fusion for obstacle detection can be found in the literature ([13]–[16]). However, we are not aware of any work using radar and camera for outdoor 3D reconstruction. More than data fusion, our aim is to build a 3D sensor which provides textured elevation maps as illustrated in fig.1. Therefore, a geometrical model of the sensors and a calibration technique should be provided. These two sensors are complementary: we rely on the fact that the distance of a 3D object to the system is given by the radar measurements having a constant range error with increasing distance. While its altitude and size can readily be extracted from the image. In multi-sensors systems, each sensor performs measurements in its own coordinate system. Thus, one needs to transform these measurements into a global coordinate system. Generally, a calibration step enables to compute this transformation in order to make the reconstruction simpler. In related works, this calibration method is not explicitly described. Sugimoto et al. [17]

[1]Institut Pascal UMR CNRS/UBP/IFMA 6602, Université Blaise Pascal, FRANCE, Ghina.EL_NATOUR@univ-bpclermont.fr

[2]IRSTEA, Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture, FRANCE, raphael.rouveure@irstea.fr

used the reflection intensity from MMW radar and image sequences to achieve a radar-vision spatial calibration. This method is hard to implement, because all the points should be positioned exactly on the radar plane as explained in section III-A. Our goal is to simplify this tricky and important step,

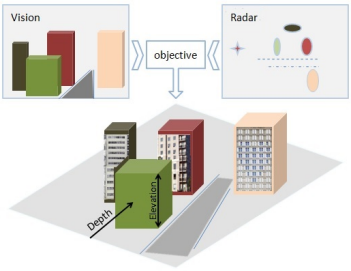


Fig. 1. An illustration of elevation map generation exploiting radar and vision complementarity.

which is crucial for the matching process and the reconstruction accuracy so that it can be carried out, easily and anywhere, by a non expert operator. Therefore we propose a technique which uses only a set of radar-to-image point matches. These points are simple targets positioned in front of the camera/radar system and the distances between them are measured. Then, the acquisitions from the two sensors are done simultaneously. A non linear geometrical constraint is derived from each match and a cost function is built. Finally, the transformation between the radar and the camera frames is recovered by a LM-based optimization (Levenberg-Marquardt). Once the calibration parameters are computed, 3D reconstruction of any radar-vision matched target can be achieved. Indeed, the intersection point of a sphere centered on the radar frame origin and a light ray passing through the camera optical centre is the 3D position of the object. So, a small amount of input data (single image and panoramic frame) is sufficient to achieve a sparse 3D map allowing thereby the real time processing. The paper is organized as follows : In section II we describe the camera and radar geometrical models. Section III focuses on the calibration method. 3D reconstruction problem is addressed in section IV. And finally, results obtained with both synthetic and real data are discussed in section V.

II. System Model

The system model is formed by a camera and a radar rigidly linked. A standard pinhole model is assumed for the camera. The camera frame and centre are denoted R_c and O_c respectively. Similarly R_r and O_r are respectively the radar's frame and centre. The sensors system is illustrated in fig.2. The radar performs acquisitions over 360° per second thanks to its 1° step rotating antenna. It generates each second a panoramic image, where detected targets are localized in 2D polar coordinates. The radar-target distance measurement is based on the FMCW principle [18]. It can be shown that the frequency difference (called beat frequency) between the transmitted signal and the signal received from a target is proportional to the sought distance. The reflected signal has

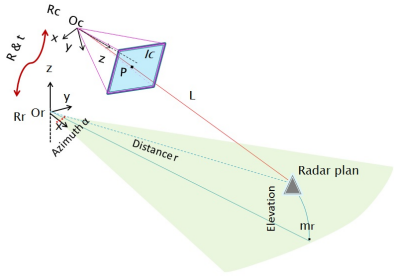


Fig. 2. Sensors system geometry: R_c and R_r are the camera and radar frames respectively. Polar coordinates $m_r(\alpha, r)$ of the target are provided by radar data but not the elevation angle. The Light ray L and the projected point p in the image I_c are shown together with the horizontal radar plane.

a different frequency because of the continuous variation of the transmitted signal around a fixed reference frequency. Therefore, the radar performs a circular projection on a horizontal plane passing through the centre of the antenna first lobe. So, the real depth and azimuth of a detected target is provided without any altitude information. The projected point is denoted $m_r(\alpha, r)$ where α and r are the real azimuth angle and depth of a target in the 3D space.

For the camera, we assume a pinhole model consisting of two transformations: first transformation projects a 3D point $\widetilde{M}(X_c, Y_c, Z_c, 1)^T$ into $\widetilde{p}(u, v, 1)^T$ (in homogeneous coordinates system) of the image plane Ic and it is written as follows:

$$w\widetilde{p} = [K|0]I_{4x4}\widetilde{M} \tag{1}$$

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & |0 \\ 0 & f_y & v_0 & |0 \\ 0 & 0 & 1 & |0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{2}$$

Where w is a scale factor and K is the matrix of intrinsic parameters, assumed to be known, since the camera is calibrated [19]. The calibration parameters are described by a 3D transformation (rotation R and translation t) mapping any point $\widetilde{M}$ from the camera frame R_c to a point $\widetilde{Q}(X_r, Y_r, Z_r, 1)^T$ in the radar frame R_r such as:

$$\widetilde{M} = A\widetilde{Q} \tag{3}$$

with A the extrinsic matrix parameters.

$$A = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{4}$$

Replacing $\widetilde{M}$ in equation (1) by the formula in (3), provides the final transformation mapping 3D to 2D point as follows:

$$w\widetilde{p} = [K|0]A\widetilde{Q} \tag{5}$$

III. System Calibration

A. Related work

The closest work on camera-radar system calibration is the work of S.Sugimoto et al. [17]. Radar's acquisitions are considered to be coplanar, since it perform a planar projection on its a horizontal plane. Therefore, the transformation A is a homography H between image and radar

planes. In spite of its theoretical simplicity, this method is hard to be implemented. Indeed, while the canonical target is being continuously moved up and down by a mechanical system, it should be simultaneously acquired by radar and camera. Then, pairs of matches (4 pairs at least) corresponding to the exact intersection of the target with the horizontal plane of the radar, are extracted. Moreover, due to sampling frequency, the exact positions are determined from the maximum of the intensity reflected by the target using bilinear interpolation of the measurement samples along the vertical trajectory of each target.

B. The proposed calibration method

The calibration stage is an important factor affecting the reconstruction accuracy. Hence, it is required to develop an accurate calibration method. For application such as Slam, one might need to recalibrate the system due to mechanical vibrations or to enable free positioning of the sensors thus it should be rather simple to implement. The only constraint for the proposed method is the *a priori* knowledge of distances between the targets used for calibration. Our goal is to determine the rotation and translation transformations. Radar provides the depth r and the azimuth angle α of a 3D object (fig.3). This object is considered to be lying on a sphere C centred on radar with radius r whose equation is:

$$(C)\,(X)^2+(Y)^2+(Z)^2=r^2 \tag{6}$$

And we have for an azimuth α, the normal to the plane VP, $\vec{n}=(\sin(\alpha),-\cos(\alpha),0)$. Since VP is a vertical plane passing by O_r we have the following equation:

$$X\sin(\alpha)-Y\cos(\alpha)=0 \tag{7}$$

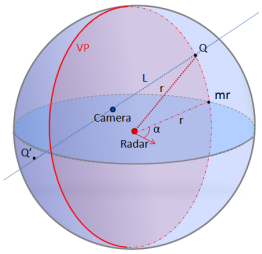


Fig. 3. The 3D point Q is the intersection of light ray L and the sphere C at α. m_r is the projected 2D point on the horizontal radar plane and VP is the vertical plane of the target at α.

For n matches, system (S_1) is obtained, with $i=1\rightarrow n$ and ε the residuals:

$$(S_1)\begin{cases} X_i^2+Y_i^2+Z_i^2-r_i^2=\varepsilon_1^i \\ X_i\sin(\alpha_i)-Y_i\cos(\alpha_i)=\varepsilon_2^i \end{cases}$$

From equation (5) a 3D point $\widetilde{Q}(X_r,Y_r,Z_r,1)^T$ in radar frame can be expressed in terms of rotation R, translation t and scale factor w, as:

$$\begin{aligned} \widetilde{Q} &= A^{-1}\begin{bmatrix}K^{-1}w\widetilde{p}\\ 1\end{bmatrix}=\begin{bmatrix}R^T & -R^Tt\\ 0 & 1\end{bmatrix}\begin{bmatrix}wJ\\ 1\end{bmatrix} \\ &= \begin{bmatrix}R^TwJ & -R^Tt\\ 0 & 1\end{bmatrix} \end{aligned} \tag{8}$$

and

$$J=K^{-1}\widetilde{p}=\begin{bmatrix}J_1 & J_2 & J_3\end{bmatrix}^T \tag{9}$$

Then $\widetilde{Q}(X_r,Y_r,Z_r,1)$ can be written as:

$$\begin{cases} X_r=A_{11}^{-1}wJ_1+A_{12}^{-1}wJ_2+A_{13}^{-1}wJ_3+A_{14}^{-1} \\ Y_r=A_{21}^{-1}wJ_1+A_{22}^{-1}wJ_2+A_{23}^{-1}wJ_3+A_{24}^{-1} \\ Z_r=A_{31}^{-1}wJ_1+A_{32}^{-1}wJ_2+A_{33}^{-1}wJ_3+A_{34}^{-1} \end{cases} \tag{10}$$

The equations are expressed with respect to a parameter vector $[\gamma_x,\gamma_y,\gamma_z,t_x,t_y,t_z,w_i]$, γ are the three rotational angles relative to Ox, Oy and Oz.

In order to calculate the scale factor w for each target, we use the theorem of Al-Kashi [20] known as the "law of cosines" that generalizes the Pythagorean theorem for an unspecified triangle. We applied this theorem to the triangle formed by two 3D points M_1, M_2 with O_c as illustrated in fig.4, gives the following equations:

$$D_1^2+D_2^2-2L_{12}=d_{12}^2 \tag{11}$$

where

$$L_{12}=D_1D_2\cos(\theta_{12}) \tag{12}$$

For n matches with $i=1\rightarrow n$, D_i is the depth of the point relative to O_c and it is related to the scale factor w_i, and the angle β_i formed between the principle point p_c and pixel p_i by the formula:

$$D_i=\frac{w_i}{\cos(\beta_i)} \tag{13}$$

with

$$\cos(\beta_i)=\frac{p_c^T(KK^T)^{-1}p_i}{\sqrt{(p_c^T(KK^T)^{-1}p_c)(p_i^T(KK^T)^{-1}p_i)}} \tag{14}$$

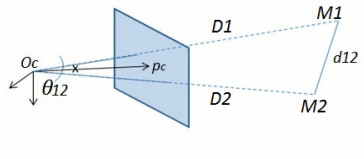


Fig. 4. The triangle formed by $M1$, $M2$ (3D points in the camera frame) and O_c is shown. d_{12} is the euclidean distance between $M1$ and $M2$ supposed to be known, and D_1, D_2 are their depths relative to O_c.

d_{ij} is the known distance between points and θ_{ij} is the angle between two rays lining up the 3D points with O_c. Cosine of θ_{ij} is calculated from corresponding pixels in the image, and intrinsic matrix K.

Since we have six degrees of freedom (DOF): three for rotation angles and three for the translation, both relative to Ox, Oy and Oz, we need at least six points. With six 3D points, we have 15 inter-distances so we obtain a system (S_2) of 15 equations in terms of $w_{i=1\rightarrow 6}$ and ε are the residuals:

$$(S_2)\left\{D_i^2+D_j^2-2L_{ij}-d_{ij}^2=\varepsilon_3^{ij}\right.$$

The system is solved by the algorithm of Levenberg-Marquardt, based on non-linear least squares optimization of the sum of squared residuals $(\varepsilon)^2$, in order to determine the approximate solution as shown hereafter:

$$\sum(\varepsilon_1^i)^2+(\varepsilon_2^i)^2 \text{and} \sum(\varepsilon_3^{ij})^2 \tag{15}$$

IV. Three-dimensional Reconstruction

Because of the projected geometry of vision and radar sensors, part of informations are lost while acquisition. In order to recover the third dimension we proceed as follows: a 3D point Q detected by both camera and radar is the intersection of the light ray L passing by the optical centre and the sphere C centred on radar as shown in fig.3. Therefore, its 3D coordinates are obtained by estimating the intersection point. Our method consists of three steps: First the scale factor w is computed, from equation (1). One can write:

$$\widetilde{M} = \begin{bmatrix} K^{-1}w\widetilde{p} \\ 1 \end{bmatrix} = \begin{bmatrix} wJ \\ 1 \end{bmatrix} \tag{16}$$

The equation of the sphere (C) expressed in the coordinate system of the camera:

$$(C)\ (X_c - x_{O_r})^2 + (Y_c - y_{O_r})^2 + (Z_c - z_{O_r})^2 = r^2 \tag{17}$$

Where $O_r(x_{O_r}, y_{O_r}, z_{O_r})$ is the radar origin in camera frame. Therefore $X_c = wJ_1$, $Y_c = wJ_2$ and $Z_c = wJ_3$, leading to a quadratic equation in w:

$$\begin{aligned} w^2(J_1^2 + J_2^2 + J_3^2) - 2w(J_1 x_{O_r} + J_2 y_{O_r} + J_3 z_{O_r}) \\ + (x_{O_r}^2 + y_{O_r}^2 + z_{O_r}^2 - r^2) = 0 \end{aligned} \tag{18}$$

Since we are working in large scale environment, the targets are usually too far compared to the baseline (the distance between the radar and camera frames). Then, the camera is always inside the sphere C, so theoretically, two solutions exist, w and w'. From these solutions, two points $\widetilde{M}(x,y,z,1)^T$ and $\widetilde{M'}(x,y,z,1)^T$ relative to the vision sensor are deduced. Secondly, the transformation is applied to these latter in order to determine their coordinates in the radar frame system from equation (3):

$$\widetilde{Q} = A^{-1}\widetilde{M} \text{ and } \widetilde{Q'} = A^{-1}\widetilde{M'} \tag{19}$$

Finally, the correct solution is selected by comparing the computed azimuth angle and the one measured by radar.

V. Simulation and Experimental Results

A. Simulation results

Simulations with synthetic data where carried out in order to test the efficiency and the robustness of the new methods with respect to numerous parameters such as number of points and noise level. Sets of 3D points are randomly generated following a uniform random distribution within a cubic work space in front of the camera-radar system. The projected pixel of each 3D point is computed using the pinhole model of the camera, and its spherical coordinates are computed. At first, both algorithms where tested without additional noise and a very low error levels were obtained ($1.180\ 10^{-6}\ m$ on translation, $1.269\ 10^{-12}\ ^\circ$ on rotation and $3.671\ 10^{-14}\ m$ on reconstruction results). Afterwards, the simulations were extended emulating realistic cases. Therefore synthetic data are perturbed by uniformly distributed, random noise. First, we added noise corresponding to ± 2 pixels, $\pm 2^\circ$ on azimuth angle, $\pm 2cm$ on distance according to the radar precision (table I).The number of matches used for the calibration process is increased by steps of 1 from 5 to 40 points. Results are shown in fig.5. It is noticeable that

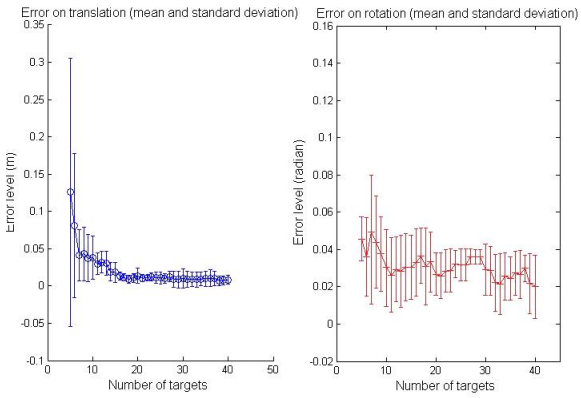


Fig. 5. Calibration error with respect to the number of points. Left: translation error in meter. Right: rotation error in radian. The graphs show the mean and the standard deviation of RMSE upon 6 specimen.

the error decreases starting from 6 calibration points. This is due to the non linear problem that converges more precisely to the correct solution when the noisy equation system is overdetermined.Secondly, in order to test the accuracy of the calibration and reconstruction methods, linear increasing of noise level is applied on data, starting from level 1 corresponding to ± 0.5 pixels, $\pm 0.5^\circ$ on azimuth angle, $\pm 0.5cm$ on distance up to level 25 corresponding to ± 2.5 pixels, $\pm 5^\circ$ on azimuth angle, $\pm 50cm$ on distance and $\pm 5mm$ on inter-distance. This multi level noise is added progressively on data. Both calibration and reconstruction errors graphs are shown in fig.6 and 7. The number of matches used for the calibration process is 10. It should be noticed that the relative

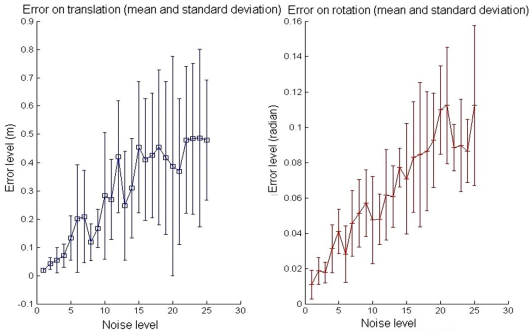


Fig. 6. calibration error with 10 matches with respect to the noise level. Left: translation error in *meter*. Right: rotation error in *radian*. The graphs show the mean and the standard deviation of RMSE upon 6 specimen.

effects of the increasing noise on the rotation and translation increases the errors: non-linear algorithms are affected by noise and yet our algorithm shows an acceptable behaviour in the presence of noisy data. The graph in fig.7 shows the mean and standard deviation of the RMSE upon 50 reconstructed points. Despite of the slight raising of error with increasing noise level, it is quite clear that the method is very robust in the presence of a realistic noise level.

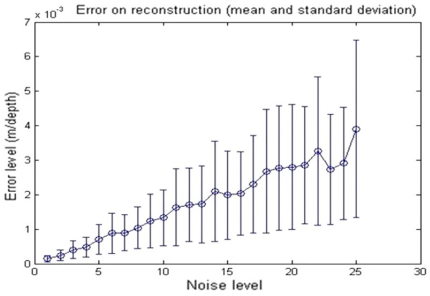
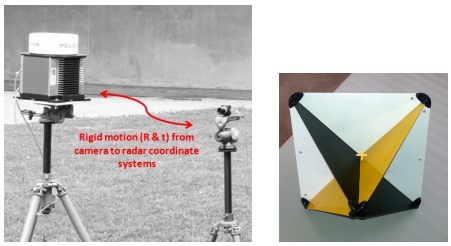
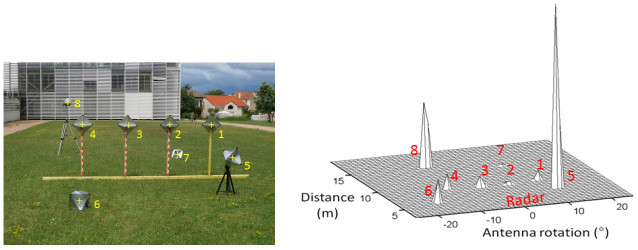


Fig. 7. Reconstruction error with respect to the noise level. The error is in *meter* relative to the points depths (r). The mean and standard deviation over 50 reconstructed points are shown.



(a) (b)

Fig. 8. Radar and camera system are shown in (a). The canonical targets used were painted in order to add contrast to their image as shown in (b). The targets centre (yellow cross) can then be readily detected.



(a) (b)

Fig. 9. An image and a panoramic of targets. The targets are numbered from 1 to 8. (a) Camera image of the eight canonical targets: one Luneburg lens and seven trihedral corners. The yellow crosses indicate the centres of the targets. (b) Radar image with eight canonical targets. The variations of amplitude are introduced by the nature and orientation of each target.

B. Experimental results

The radar and the camera were mounted in a fixed configuration in order to carry out real data acquisitions (for the current stage the radar antenna rotates 360° but the camera is stable). The system is shown in fig. 8 (a). The radar is called K2Pi. It has been developed by Irstea Institute. The optic sensor used is uEye by IDS (Imaging Development Systems). Camera and radar's characteristics are listed in table I. Eight

TABLE I
CAMERA AND RADAR CHARACTERISTICS

Camera characteristics	
Sensor technology	CMOS
Sensor size	$4.512 \times 2.880mm$
Pixel size	$0.006mm$
Resolution $(h \times v)$	752×480
Focal distance	$8mm$
Radar characteristics	
Carrier frequency	$24GHz$
Antenna gain	$20dB$
Range	$3-100m$
Angular resolution	$4°$
Distance resolution	$1m$
Distance precision	$0.02m$

canonical targets were placed in front of the sensors system. Metallic targets are highly reflective. The depth of targets is chosen to be slightly close (between 6 and $14m$) and targets were painted to increase the contrast and thus facilitate the features (targets centres) extraction from the image. The fig. 8 (b) shows an example of these targets. First the system is calibrated using our algorithm. The inter-distances between the targets centres are measured precisely, and an image and a panoramic of the 8 targets in random configurations are used. image and radar targets are extracted and matched manually. In order to assess our calibration results and to validate the reconstruction method, the 8 targets were placed at different heights and depths. The matches were also extracted and the reconstruction is done in radar frame. Fig.11 (a) and (b) shows the corresponding pixels and radar points extracted in the image of the camera and the panoramic image. Fig.10 and table II represent the results of the reconstruction technique, using the method of section III-B and reconstruction from a stereo head, used as ground truth. The ground truth point set and the reconstructed one were registered using ICP (Iterative Closest Point) algorithm. The RMS of the reconstruction error is about $0.1939m$ with a standard deviation of $0.015m$. The results show a realistic error for the 3D reconstruction of targets at a mean depth of $12m$ and prove the accuracy of the calibration method. A qualitative result of the reconstruction of an urban scene is represented in fig. 11. The segmentation and matching steps are done manually. The matched paires are then reconstructed and enhanced with texture.

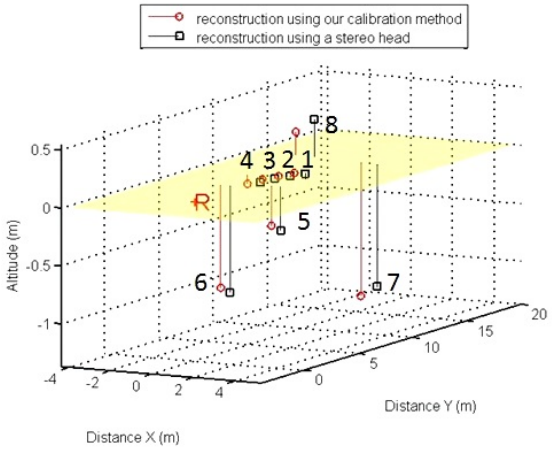


Fig. 10. The reconstruction results using results from both, our reconstruction methods (circular points) and the stereo head method as a ground truth (squared points). Radar position is notified by the letter R.

TABLE II

RECONSTRUCTION RESULTS IN (M)

	Targets coordinates in meter with the stereo head (ground truth)							
X	1.32	1.45	0.70	-0.03	-0.78	-0.68	2.24	-1.03
Y	5.41	7.45	7.47	7.44	7.53	4.59	12.53	13.01
Z	-0.38	0.05	0.02	-0.01	-0.06	-0.92	-1.05	0.33
	targets coordinates in meter with the developed method							
X	0.83	0.86	0.11	-0.62	-1.37	-1.12	1.40	-1.89
Y	5.50	7.55	7.53	7.46	7.51	4.57	12.64	12.98
Z	-0.35	0.05	0.01	-0.03	-0.08	-0.89	-1.15	0.21
	Error in meter (Euclidean distance)							
	0.49	0.60	0.59	0.59	0.59	0.44	0.85	0.87
	Error - Mean in meter= 0.63							
	Error - Standard deviation (m)=0.15							

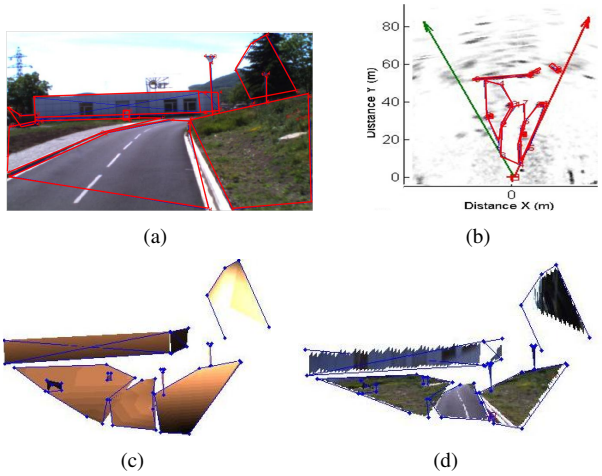


Fig. 11. Results of the reconstructed urban scene using the camera-radar system, the segmented images (red polygons in (a) and (b)) and a view of the resuslting reconstructed scene is shown in (c). The result is enhanced with texture mapping in (d).

VI. CONCLUSIONS AND PERSPECTIVES

In this paper, we presented a geometrical algorithm for spatial calibration of a radar and a camera and a 3D reconstruction method. Results show the accuracy of the calibration and proves the feasibility and the interest of the application of these two sensors for 3D reconstruction of outdoor scenes. To our knowledge, this type of sensors have not been used for large outdoor reconstruction applications. Our methods are both validated by simulated data, and real experiments with sparse data. Results show their feasibility and quite a good behaviour in the presence of noise. For the current stage, sparse point matches are extracted manually, therefore, RANSAC-like algorithms must be setted up to make automatic matching and to achieve dense 3D reconstruction. On the other hand, camera rotation is an interesting step since the viewin angle of the radar allow for a large covering of the area. Further real time reconstruction experiments of urban and semi-urban scenes should be carried out, and then adding texture to enhance the resulting map.

ACKNOWLEDGMENT

This work has been sponsored by the French government research programm "Investissements d'avenir" through the IMobS3 Laboratory of Excellence (ANR-10-LABX-16-01), by the European Union through the programm Regional competitiveness and employment 2007-2013 (ERDF Auvergne region), and by the Auvergne region.

REFERENCES

[1] P. Musialski, P. Wonka, D. G. Aliaga, M. Wimmer, L. van Gool, and W. Purgathofer, "A survey of urban reconstruction."

[2] D. Bhagawati, "Photogrammetry and 3-d reconstruction-the state of the art," *ASPRS Proceedings, Washington, DC*, 2000.

[3] G. Kordelas, J. Perez-Moneo Agapito, J. Vegas Hernandez, and P. Daras, "State-of-the-art algorithms for complete 3d model reconstruction." *Engage Summer School*, 2010.

[4] D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, "Real-time plane-sweeping stereo with multiple sweeping directions," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[5] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, *et al.*, "Detailed real-time urban 3d reconstruction from video," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 143–167, 2008.

[6] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 237–260, 2007.

[7] R. Rouveure, M. Monod, and P. Faure, "High resolution mapping of the environment with a ground-based radar imager," in *Radar Conference-Surveillance for a Safer World, 2009. RADAR. International*. IEEE, 2009, pp. 1–6.

[8] G. Forlani, C. Nardinocchi, M. Scaioni, and P. Zingaretti, "Complete classification of raw lidar data and 3d reconstruction of buildings," *Pattern Analysis and Applications*, vol. 8, no. 4, pp. 357–374, 2006.

[9] I. Stamos and P. K. Allen, "3-d model construction using range and image data," in *In CVPR*, 2000.

[10] J. Smisek, M. Jancosek, and T. Pajdla, "3d with kinect," in *Consumer Depth Cameras for Computer Vision*. Springer, 2013, pp. 3–25.

[11] C. K. Schindhelm, "Evaluating slam approaches for microsoft kinect," in *ICWMC 2012, The Eighth International Conference on Wireless and Mobile Communications*, 2012, pp. 402–407.

[12] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. C. Antonello, and M. Miglietta, "Obstacle detection and classification fusing radar and vision," in *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE, 2008, pp. 608–613.

[13] A. Roy, N. Gale, and L. Hong, "Fusion of doppler radar and video information for automated traffic surveillance," in *Information Fusion, 2009. FUSION'09. 12th International Conference on*. IEEE, 2009, pp. 1989–1996.

[14] U. Hofmann, A. Rieder, and E. D. Dickmanns, "Radar and vision data fusion for hybrid adaptive cruise control on highways," *Machine Vision and Applications*, vol. 14, no. 1, pp. 42–49, 2003.

[15] T. Wang, N. Zheng, J. Xin, and Z. Ma, "Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications," *Sensors*, vol. 11, no. 9, pp. 8992–9008, 2011.

[16] L. Bombini, P. Cerri, P. Medici, and G. Alessandretti, "Radar-vision fusion for vehicle detection," in *Proceedings of International Workshop on Intelligent Transportation*, 2006, pp. 65–70.

[17] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave radar and its visualization on image sequence," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3. IEEE, 2004, pp. 342–345.

[18] M. I. Skolnik, "Introduction to radar systems," 2001.

[19] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2004.

[20] E. S. Kennedy, "A fifteenth-century planetary computer: al-kāshī's tabaq al-manāteq. ii. longitudes, distances, and equations of the planets," *Isis*, vol. 43, no. 1, pp. 42–50, 1952.