Rodríguez-Garavito, C. H.; Ponz, A.; García, F.; Martín, M.; Escalera, A.; Armingol, J. M. (2014). Automatic laser and camera extrinsic calibration for data fusion using road plane. *Information Fusion (FUSION), 2014 17th International Conference on*. Salamanca, pp. 1-6.

Automatic Laser And Camera Extrinsic Calibration for Data Fusion Using Road Plane

C. H. Rodríguez-Garavito[+], A. Ponz[*], F. García[*], D. Martín[*], A. de la Escalera[*] and J.M. Armingol[*]

[+]Automation Engineering Department
Universidad de La Salle
Bogotá-Colombia
cerodriguez@unisalle.edu.co

[*]Universidad Carlos III de Madrid
Intelligent Systems Lab.
Spain,
{apv,fegarcia, dmgomez,escalera,armingol}@ing.uc3m.es

***Abstract*—Driving Assistance Systems and Autonomous Driving applications require trustable detections. These demanding requirements need sensor fusion to provide information reliable enough. But data fusion presents the problem of data alignment in both rotation and translation. Laser scanner and video cameras are widely used in sensor fusion. Laser provides operation in darkness, long range detection and accurate measurement but lacks the means for reliable classification due to the limited information provided. The camera provides classification thanks to the amount of data provided but lacks accuracy for measurements and is sensitive to illumination conditions. Data alignment processes require supervised and accurate measurements, that should be performed by experts, or require specific patterns or shapes. This paper presents an algorithm for inter-calibration between the two sensors of our system, requiring only a flat surface for pitch and roll calibration and an obstacle visible for both sensors for determining the yaw. The advantage of this system is that it does not need any particular shape to be located in front of the vehicle apart from a flat surface, which is usually the road. This way, calibration can be achieved at virtually any time without human intervention.**

Keywords—Calibration, Data Alignment, LIDAR, Extrinsic, RANSAC, Fusion

I. INTRODUCTION

Information fusion between Laser scanner and video camera is widely used in both autonomous driving and driving assistance systems, in order to complement the virtues of each sensor and mitigate its weaknesses. While the laser is able to accurately determine distances in any light condition, obstacle classification is very inaccurate without the help of a camera, which in turn can't operate in poor light conditions. The complementary capabilities of these two sensors make them a very suitable solution for obstacle detection and classification in automotive applications.

Different sensors have different field of view and different characteristics, but the reality they sense must match in order to get useful information from them. We must find the relation between each of the sensors and the world, and then the relation between the sensors. The extrinsic parameters of the sensors are rotation and translation. Determining the translation with respect to the reference point of the vehicle is a laborious task, but usually is done just once and is relatively easy, accurate, and small errors does not affect significantly to the precision of the system. Rotation, in the other hand, is more difficult to measure and is more prone to involuntary changes.

The article is divided in the following sections: section II provides scientific context through the state of the art description. Section III gives a general description of the approach. Section IV explains the single sensor parameter estimation based on a point cloud. Section V details the automatic data alignment process. Section VI describes the tests performed and the results obtained and VII provides some conclusions to the presented work.

II. RELATED WORK

Laser scanner and video camera does not share the spectrum, so it is not possible to match the rotations without the use of computer vision and algebraic algorithms.

Several possibilities have been proposed in literature to provide data alignment: In [1] Li et al. proposed a method for calibration based on planar chessboard in different positions and the physical constraints related to the chessboard. In [2], [3] and [4] projects the features into a 2D plane to minimize distances among the features in the different sensors. Other approaches based the calibration process in obstacles with specific patterns that allows thanks to the physical constraints to match the detection in the different sensors. In [5] Lişca et al. used a Calibration Object with a CAD model that allows to perform the matching by a single frame. An advanced version based on triangular patterns is provided in [6]. In [7] authors provide a similar system for a monocular camera based on a circular shape pattern for calibration.

Most of the aforementioned works deal with the calibration process by the use of specific patterns and/or cumbersome procedures, that involves the implication of the driver or user. This need makes impossible to perform the calibration at any moment and in any place, restricting the calibration to some particular moments and places, and, most important, demanding very special conditions or requiring human intervention. Thus, it involves difficulties for generalization and for commercial applications, in case of eventual changes in the configuration. Other approaches, such as [2], [3] and [4] require conversion to 2D which implies assumptions. Finally, other approaches such as [8], require the supervision of users to

select specific parameters e.g. selection of correspondence points, or [9] which calibration is manually performed on line.

The algorithm presented in this paper performs the calibration of the extrinsic rotations and sensor elevation for both laser scanner and a stereo video camera, obtaining then the calibration between both sensors. This method needs only a flat surface in front of the vehicle to find two of the rotation angles and elevation between sensors, and one identifiable obstacle different from the road surface and visible by both sensors in order to compute the remaining angle.

III. GENERAL DESCRIPTION

The simultaneous use of several sensors is common in Advanced Driving Assistance Systems (ADAS), in order to obtain higher reliability in the algorithms while detecting, for example, pedestrians, vehicles, or even the topology of the road we are driving on. Therefore, the need for synchronization of several sources with a common reference system arises. This is the case that this work deals with, three-dimensional data captured from a stereo camera and a laser scanner.

The present work is included in IVVI 2.0 project. IVVI 2.0 is the second platform for development and research of ADAS technologies of the Intelligent Systems Lab, at Universidad Carlos III de Madrid.

Fig. 1. IVVI 2.0 vehicle with the sensors used in detail.

The solution presented obtains a point cloud (PC) from both the laser scanner and the stereo camera. These point clouds must present a significant amount of points belonging to a flat surface in front of the vehicle, in this specific application it is the road. Applying to both point clouds the RANSAC [10] algorithm to the detection of planes in the space, we get a unitary vector normal to the surface for every sensor. These vectors determine the two orientation angles for the sensor (pitch and roll), and the height from the plane found is obtained from the projection of the origin of the point cloud in the plane of the road. The last orientation angle (yaw) is obtained by obtaining the projection of the distances to the obstacles within the road, calculating their angle signatures and correlating among the different clouds of points from the different sensors.

IV. EXTRINSIC PARAMETERS ESTIMATION FROM A POINT CLOUD

The process starts from the three-dimensional reconstruction, based on a PC, of the environment with respect to the camera system reference {c}. This plane can be represented in the world system reference {m}, attached at ground plane, Fig. 2.

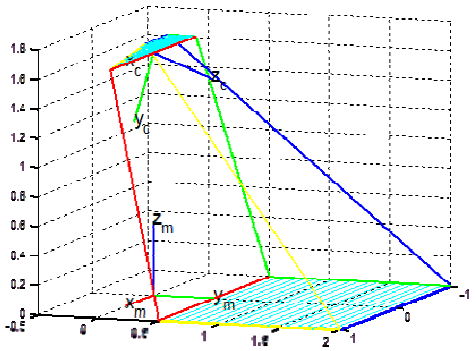

Fig. 2. Framework of the images system: Camera {c}, World {m}

Using the M-estimator-SAmple-Consensus (MSAC) algorithm [11], applied to plane detection in the space, it is possible to generate a $[a, b, c, d]$ vector defining the $\pi_{(X)}$ plane (1), as the most populated plane in the PC.

$$\pi_{(X)}: ax_c + by_c + cz_c + d = 0 \quad (1)$$

In its Hessian form, (1) (2) can be written as:

$$\pi_{(X)}: \vec{n}.\vec{p} = h \quad (2)$$

From (2) we deduct that the vector $\vec{n}$ is normal to the $\pi_{(X)}$ plane we found, as the projection of $\vec{n}$ of any point $\vec{p}$ located in the plane always generates a fixed distance. This distance is the minimal from the plane to the PC's origin of coordinates, thus the height h of the sensor.

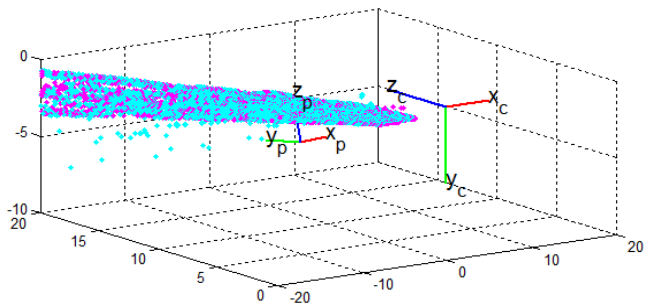

Fig. 3. Base obtained on the road plane {p}. Cyan points are inliers and magenta points are its projections on the plane.

The direction of the $\vec{n}$ vector is then defined as the cross-product between u_p and v_p, which corresponds to the direction of Z_p. Fig. 3 depicts the base generated on the road plane, system reference {p}.

A. *Rotation extraction with respect to the camera on X and Y'*

The vector ${}^c\omega_p = [{}^c\omega_{px} \quad {}^c\omega_{py} \quad {}^c\omega_{pz}]^T$ normal to the road plane as seen from the camera in Fig. 4, correspondent to the direction of Z_p is related to the orientation of the Z_c axis of the camera according to (3).

$$ {}^c\omega_p = Rot_x\left(\frac{\pi}{2}\right) * Rot_x(-\alpha) * Rot_y(-\sigma) * {}^p\omega_p \tag{3}$$

The rotation on Z'' is not taken into account, as its application does not change the road plane Z_p axis with respect to the Z_c axis of the camera.

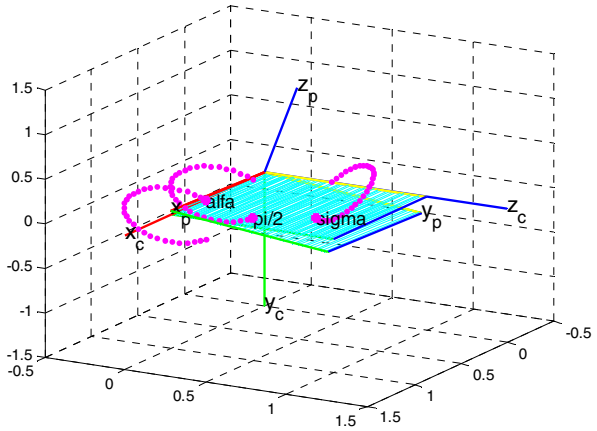

Fig. 4. Road plane rotation respect to camera reference system

Expanding (3) we get:

$$ {}^c\omega = {}^cR_p * {}^p\omega $$

$$ \begin{bmatrix} {}^c\omega_x \\ {}^c\omega_y \\ {}^c\omega_z \end{bmatrix} = \begin{bmatrix} c\sigma & 0 & -s\sigma \\ -c\alpha s\sigma & s\alpha & -c\alpha c\sigma \\ s\alpha s\sigma & c\alpha & s\alpha c\sigma \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4} $$

$$ \sigma = \sin^{-1}(-{}^c\omega_x) $$

$$ \alpha = \cos^{-1}\left(-\frac{{}^c\omega_y}{c\sigma}\right) \tag{5} $$

Last, relating (1) and (2), sensor height is computed as

$$ \mathrm{h} = -\mathrm{d} \tag{6} $$

B. *Algorithm optimization for large PC*

Two strategies were implemented to accelerate the MSAC algorithm, RANSAC variant. One consists on eliminating all the points allegedly not located in the ground plane, this is, in the upper half of the image, as far as it is a PC obtained from a stereo reconstruction. The second one consists on extracting a uniform sample of the point cloud in order to get a limited amount of data representing the environment, thus the PC can be processed in real time, less than 100ms.

V. DATA ALIGNMENT BETWEEN TWO THREE-DIMENSIONAL CAPTURE SENSORS

The problem to solve then is finding the spatial position of a sensor reference system{c} (for the stereo-camera) respect to a sensor reference system{l} (for the laser), both attached to a mobile system i.e. the vehicle. Each sensor has its own position, and its Field of View (FOV) to see the road as the mobile system moves on, as shown in Fig. 5.

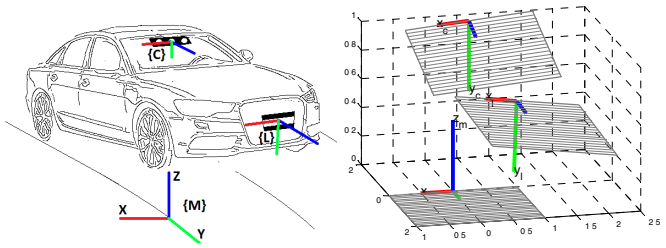

Fig. 5. Sensor array configuration, stereo camera located in the windshield, and laser scanner located in the bumper.

The next step is based on the point clouds PC_c and PC_l, captured from each of the sensors. Each cloud is transformed until the most populated plane matches the $X_m - Y_m$ plane. In each transform ${}^MT_C$ and ${}^MT_L$, the respective extrinsic parameters $[\alpha_c \quad \beta_c \quad h_c]$ and $[\alpha_l \quad \beta_l \quad h_l]$ are estimated, as explained in Section III.

Fig. 6 shows PC from each sensor, green-red are PC_c inliers-outliers respectively and purple-blue are PC_l inliers-outliers.

Transforms from {c} and {l} into {m} are presented in (7) and (8).

$$ {}^lP = {}^lT_m\,{}^mT_c\,{}^cP \tag{7} $$

$$ {}^lT_c = {}^lT_m\,{}^mT_c $$

$$ D_Z(h)R_Z(\gamma)R_Y(\beta)R_X(\alpha) = R_X(-\alpha_l)R_Y(-\beta_l)\cdot R_Z(-\gamma_{lc})D_{XYZ}([-dx_{lc} \quad -dy_{lc} \quad -h_{lc}])\cdot R_Y(\beta_c)R_X(\alpha_c) \quad , $$

$$ \begin{bmatrix} c\beta c\gamma & s\alpha s\beta c\gamma + c\alpha s\gamma & -c\alpha s\beta c\gamma + s\alpha s\gamma & 0 \\ -c\beta s\gamma & -s\alpha s\beta s\gamma + c\alpha c\gamma & c\alpha s\beta s\gamma + s\alpha c\gamma & 0 \\ s\beta & -s\alpha c\beta & c\alpha c\beta & -h \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8} $$

$$ = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ T_{41} & T_{42} & T_{43} & T_{44} \end{bmatrix} $$

Decomposition matrix ${}^lT_c$ through world reference system {m} and its relationship to the global transformation element is shown in (8) .

As the plane alignment lets free the rotation angle γ around Z_m, a rotation of ${}^{m}PC_c$ is made with respect to ${}^{m}PC_l$ in an angle γ_{cl}, assuming that the translation of the sensors in the X_m i.e. dx_{cl} and Y_m i.e. dy_{cl} axis is known.

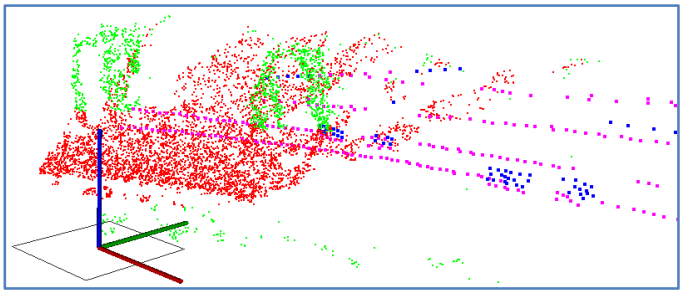

Fig. 6. Alignment and segmentation of the road plane in point clouds PCL_c and PCL_l green-red are PCL_c inliers-outliers and purple-blue are PCL_l inliers-outliers.

As the point clouds are different, in order to find the γ_{cl} angle, it is necessary to adjust the data by looking for the highest similarity. To do so, outliers from ${}^{m}PC_l$ further than 10m from the Z_m axis are removed, obtaining ${}^{m}PC_{out\_l}$. Then, the minimal and maximal distance to Z_m, ${}^{m}PC_{out\_l}$: $[d_{min} \quad d_{max}]_{\perp Z_m}$ are calculated among the filtered points, and then the minimal and maximal distance along the Z_m, ${}^{m}PC_{out\_l}$: $[h_{min} \quad h_{max}]_{Z_m}$. Using the computed boundaries in cylindrical coordinates, information from ${}^{m}PC_c$ is filtered in order to extract the ${}^{m}PC_{out\_c}$ cloud. Both filtered clouds are shown on the right of Fig. 7, ${}^{m}PC_{out\_l}$ in blue and ${}^{m}PC_{out\_c}$ in green.

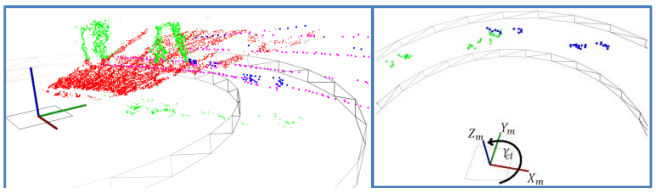

Fig. 7. ${}^{m}PCL_{out\_l}$ and ${}^{m}PCL_{out\_c}$ in the Region Of Interest $[d_{min} \quad d_{max}]_{\perp Z_m}$ and $[h_{min} \quad h_{max}]_{Z_m}$.

The final procedure for finding γ_{cl} consists on obtaining the projections from each cloud $Proy_{XY}({}^{m}PC_{out\_c})$ and $Proy_{XY}({}^{m}PC_{out\_l})$, or ${}^{XY}PC_{out\_c}$ y ${}^{XY}PC_{out\_l}$. To do so, the Z coordinate is eliminated from the definition of every point, as show in Fig. 8.

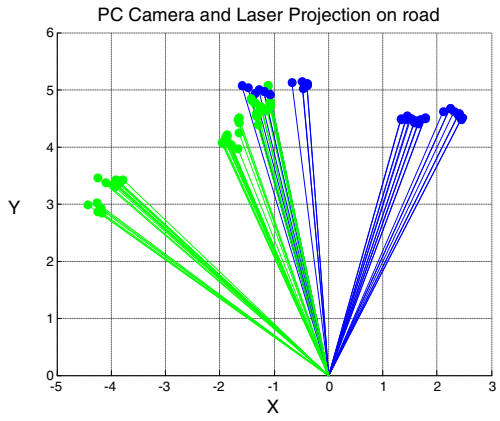

Fig. 8. PC camera and PC laser projections.

Next, we obtain the signature from each projection, $signature(PC_{out\_c_{XY}})$ and $signature(PC_{out\_l_{XY}})$ or s_{out_c} and s_{out_l}. To do so, the bi-dimensional PC is translated into polar coordinates as magnitude and angle of a vector S, meaning the angle of every point the position n, where the magnitude S of the point will be stored, as shown in Fig. 9.

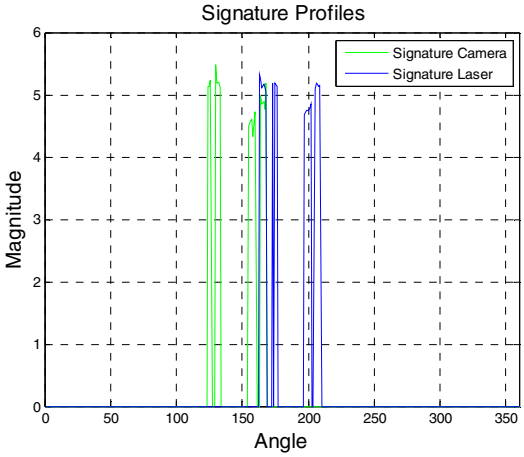

Fig. 9. Profile signatures for PC, camera and laser projections in XY.

$$s(n) = s\big(ang(PC(i)_{XY})\big) = mag(PC(i)_{XY}) \tag{9}$$

Next, the correlation between both signatures is found.

$$E(m) = \sum_{n} \left(s_{out\_c}(m-n) - s_{out\_l}(n)\right)^2 \tag{10}$$

Fig. 10 shows the correlation between profile signatures for PC, laser and camera projection.

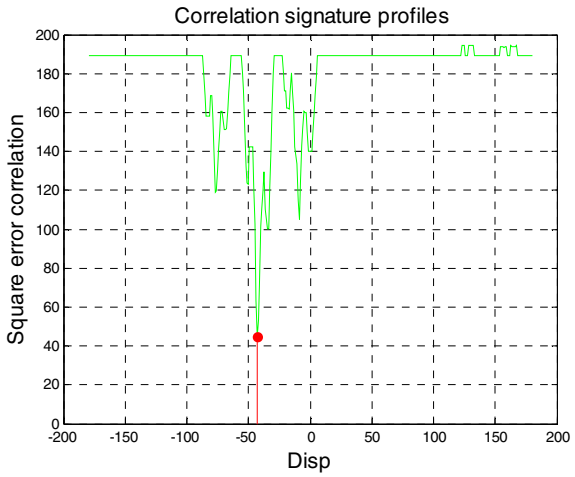

Fig. 10. Correlation between profile signatures on camera and laser PC projections.

The minimal value of E is found on a m^* shifting, whose value matches the rotation γ_{cl} we was looking for.

$$\gamma_{cl} = m^* : min\big(E(m)\big)_{\lfloor m^*} \tag{11}$$

After finding γ_{cl} we can get rotation angles between sensors, and sensors' height from (8) . So,

$$\begin{gathered} -c\beta s\gamma = T_{12} \\ s\beta = T_{13} \\ -s\alpha c\beta = T_{23} \\ c\alpha c\beta = T_{33} \\ h = h_{CL} \\ \beta = sin^{-1}(T_{13}) \\ \alpha = cos^{-1}\left(\frac{T_{33}}{c\beta}\right), \alpha = sin^{-1}\left(-\frac{T_{23}}{c\beta}\right) \\ \gamma = sin^{-1}\left(-\frac{T_{12}}{c\beta}\right) \end{gathered} \tag{12}$$

VI. TEST

A. Explanation

Two different sets of tests are performed to the presented algorithms. The first set of tests measure the accuracy of the single sensor extrinsic parameters identification algorithm, explained in section IV. Once the performance for the single sensor algorithm is proved, the next set of tests intend to check the accuracy of the relative extrinsic parameters estimation algorithm, detailed in section V. To do so, the first tests use a single frame, and later the algorithm is tested in a recorded sequence.

B. Single sensor Ground Truth extrinsic comparison

As the method for data alignment between two sensors is based on the extrinsic measurement for each of the sensors, that is, height and roll and pitch rotation angles with respect to the ground plane, the first step on the check for the final result is to quantify the precision and accuracy of these three measurements. For that purpose, data from a reconstruction coming from a Bumblebee XB3 stereo camera are compared to data obtained from a MicroStrain 3DM-GX2 IMU, attached to the camera.

Accuracy is defined as the difference between the mean of the measurements and the mean of the ground truth or real value, while the precision is calculated as the standard deviation of the measurement, i.e. its repeatability. So, accuracy for the measurement of a rotation on the X axis –roll- is 0.43 degrees, and the precision is 0.12 degrees. For a Y axis –pitch- rotation measurement, the accuracy is 0.53 degrees, and precision is 0.08 degrees, as shown on Table 1.

Param\Sensor	Camera		IMU	
	Mean	Std D	Mean	Std D
Pitch	-17.82	0.08	-18.33	0.09
Roll	5.15	0.12	5.52	0.28

Table 1. Graph depicting Roll and Pitch measurements from a ground plane estimated from a camera point cloud, compared to the reading from a MicroStrain 3DM-GX2 IMU.

Next, after metrologic especifications of the sensor have been determined, its behaviour will be checked in the range of interest, i.e., between -20 and 20 degrees in roll and between -5 and -20 degrees in pitch, as shown in Fig. 11.

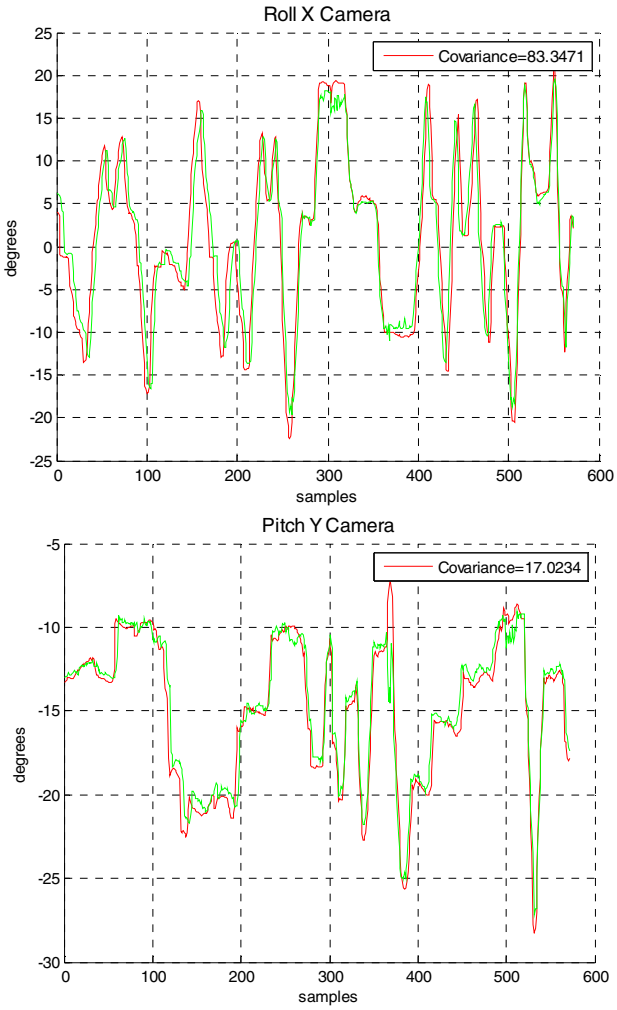

Fig. 11. Global observation of the extrinsic sensor behaviour based on the determination of the ground plane with respect to the readings from a reference IMU.

C. *Relative extrinsic parameters measurement*

Absolute extrinsic parameters measurement for each sensor camera and laser, and relative measurement from camera to laser, have been done from the configuration presented in Fig. 5 and considering a scene like the one in Fig. 12.

Fig. 12. Laser projection on the image.

The result of the ground plane detection and the subsequent alignment in the Z axis is shown in Fig. 13.

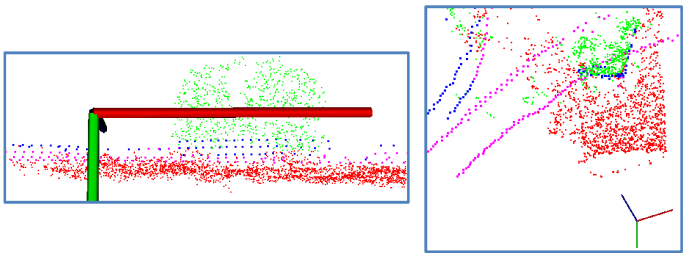

Fig. 13. Ground plane alignment for the point clouds obtained from a stereo camera and a laser scanner.

The experiment is now repeated, this time for a sequence of images and laser captures similar to the depicted in Fig. 13. Data are shown in Table 2.

Parám\Sensor	Camera		Laser		Camera-Laser	
	Mean	Std D	Mean	Std D	Mean	Std D
High	1.27	0.007	0.267	0.04	-0.937	0.048
Pitch	-7.3	0.06	-2.22	0.29	-5.14	0.29
Roll	1.08	0.18	-0.71	0.23	4.12	0.33
Yaw	-4.39	0.34	0	0	2.34	0.34

Table 2. Extrinsic parameters measured from a sequence of synchronized images and laser captures.

VII. CONCLUSIONS

The estimation method for extrinsic parameters based on the road plane detection from a point cloud shows a pitch angle difference respect to the ground truth of 0.5 degrees and 0.4 degrees for roll angle. Furthermore, the reference sensor (IMU) exhibits the same standard deviation of 0.08 degrees for the pitch angle than the proposed measurement method. As far as the roll angle is concerned, the IMU sensor shows a standard deviation of 0.28 against 0.12 in the proposed method.

After the transformation of the point clouds to the road reference system, i.e., the planes estimated for each point cloud coming from each sensor are coplanar and superimposed, the final alignment produced good results. This alignment was achieved by rotating the camera point cloud until the objects, out of road, matched in both point clouds. In a static sequence, standard deviation of the rotation with respect to the Z axis of the road is 0.33 degrees.

Relative extrinsic estimation between camera and laser was tested by projecting the laser on the image and checking quantitatively the match, as seen in Fig. 12.

ACKNOWLEDGMENT

This work was supported by Automation Engineering Department from de La Salle University, Bogotá-Colombia; Administrative Department of Science, Technology and Innovation (COLCIENCIAS), Bogotá-Colombia and the Spanish Government through the Cicyt projects (GRANT TRA2010-20225-C03-01) and (GRANT TRA 2011-29454-C03-02).

REFERENCES

[1] Y. Li, Y. Ruichek, and C. Cappelle, “3D triangulation based extrinsic calibration between a stereo vision system and a LIDAR,” 2011 14th Int. IEEE Conf. Intell. Transp. Syst., pp. 797–802, 2011.

[2] G. L. G. Li, Y. L. Y. Liu, L. D. L. Dong, X. C. X. Cai, and D. Z. D. Zhou, “An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features,” 2007 IEEE/RSJ Int. Conf. Intell. Robot. Syst., 2007.

[3] S. Wasielewski and O. Strauss, “Calibration of a multi-sensor system laser rangefinder/camera,” in Intelligent Vehicles ’95 Symposium., Proceedings of the, 1995, pp. 472–477.

[4] K. Kwak, D. F. Huber, H. Badino, and T. Kanade, “Extrinsic calibration of a single line scanning lidar and a camera,” 2011 IEEE/RSJ Int. Conf. Intell. Robot. Syst., pp. 3283–3289, 2011.

[5] G. Lisca, P. J. P. Jeong, and S. Nedevschi, “Automatic one step extrinsic calibration of a multi layer laser scanner relative to a stereo camera,” Intell. Comput. Commun. Process. (ICCP), 2010 IEEE Int. Conf., 2010.

[6] S. Debattisti, L. Mazzei, and M. Panciroli, “Automated extrinsic laser and camera inter-calibration using triangular targets,” in Intelligent Vehicles Symposium (IV), 2013 IEEE, 2013, pp. 696–701.

[7] S. A. R. F, V. Fremont, and P. Bonnifait, “Extrinsic calibration between a multi-layer lidar and a camera,” 2008 IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst., 2008.

[8] D. Scaramuzza, A. Harati, and R. Siegwart, “Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes,” 2007 IEEE/RSJ Int. Conf. Intell. Robot. Syst., 2007.

[9] F. García, F. Jiménez, J. J. Anaya, J. M. Armingol, J. E. Naranjo, and A. de la Escalera, “Distributed pedestrian detection alerts based on data fusion with accurate localization.,” Sensors (Basel)., vol. 13, no. 9, pp. 11687–708, 2013.

[10] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with,” Commun. ACM, vol. 24, no. 6, pp. 381–395, 1981.

[11] Torr, Philip HS, and Andrew Zisserman. "MLESAC: A new robust estimator with application to estimating image geometry." Computer Vision and Image Understanding 78.1 (2000): 138-156.