

Automated Extrinsic Laser and Camera Inter-Calibration Using Triangular Targets

Stefano Debattisti, Luca Mazzei, Matteo Panciroli*
VisLab – Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Parma, ITALY
<http://www.vislab.it>
{deba, mazzei, panci*}@vislab.it

Abstract—This paper presents a method for solving the extrinsic calibration between camera and multi-layer laser scanner for outdoor multi-sensorized vehicles. The proposed method is designed for intelligent vehicles within the autonomous navigation task where usually distances between sensor and targets become relevant for safety reasons, therefore high accuracy across different measures must be kept. The calibration procedure takes advantage of triangular shapes still present in scenarios, it recovers three virtual points as target pose in the laser and camera reference frames and then compute extrinsic information of each camera sensor with respect to a laser scanner by minimizing a geometric distance in the image space. To test algorithm correctness, and accuracy a set of simulations are used reporting absolute error results and solution convergence, then tests on robustness and reliability (i.e., outliers management) are based on a wide set of datasets acquired by VIAC prototypes.

Index Terms—Calibration, Inter Calibration, Multi-Layer LIDAR, Camera, Extrinsic Calibration, Triangular calibration targets

I. INTRODUCTION

In the mobile robotics field and particularly in the intelligent vehicle context, the autonomous navigation task is achieved by heterogeneous suit of perception sensors in order to acquire more rich information content than using only active or passive sensors. For their complementary purposes LIDARs and camera are widely used together on such vehicles, Zehang et al. in [1] reviews the problem of on-road vehicles detection using optical sensors and details pros and cons by introducing active sensor in object perception. In the last year several advanced driver assistance systems coupled camera and laser devices, such as vehicle detection [2] where LIDAR provide good range information to define obstacles while it is hard to recognize vehicles and cameras that allows better recognition but does not provide good range information.

Since laser and camera are complementary sources, in order to fuse their information the extrinsic calibration of these heterogeneous sensors is required. This paper present an extrinsic calibration algorithm between slave and master sensors which are a vision system and a multi-layer LIDAR respectively. By detecting a planar triangular board beside the road, and in front of the sensors, the proposed method solves the Location Determination Problem (LDP) in the

laser frame by a 3D reconstruction of the triangular targets, than minimizing its projection in the image space. Two main steps compose such algorithm: a target detection in each sensor frame and triangle vertex estimation, and compute extrinsic in a optimization phase that minimizing a geometric distance. A preliminary comparison between different extrinsic estimation techniques based on geometric and algebraic minimization functions (i.e., SVD rigid body method or Iterative Closest Point) lead us to chose for a non-linear optimization algorithm based on geometrical constraints in the image space.

II. PREVIOUS WORKS

The extrinsic camera calibration problem can be solved using laser range finders by making correspondences between features seen by laser and the same features seen by the camera. However, produce correspondences between these heterogeneous sensors is complex and arise several problems. First of all, the collisions between laser beams and objects are not visible, since standard automotive camera systems do not operate in the same frequency spectrum of the LRF. Moreover, camera and laser have different error models, since cameras are affected by perspective projection distortion and laser range measurements provide constant precision over distance. To deal with these known problems there are different methods present in literature that estimate 3D features such as points that lying on depth edges than the camera pose is measured by minimizing a geometric, rather than an algebraic distance either projecting LRG features in the image plane or projecting camera features in the 3D space.

The procedure in [3] describe a extrinsic calibration algorithm by placing a planar chessboard at different positions and orientations in front of the sensors, the proposed method solves the problem based on 3D reconstruction of the chessboard and geometric constraints between views from the stereo vision system and the LIDAR. The three principle steps of the approach are: 3D corner points triangulation, 3D plane least-squares estimation, solving extrinsic parameters by applying a non-linear optimization algorithm based on the geometric constraints. To evaluate the performance of the algorithm, experiments based on computer simulation and real data are performed. The proposed approach is also

* Corresponding author.

compared with a popular calibration method to show its advantages.

In 2007 Li [4] propose an extrinsic calibration method based on line features detection and a Gauss Newton optimization with a geometrical distance function. The extrinsic parameters are obtained by minimizing the distance from the calculated projection of the intersection point to the projected edges of the checkerboard in the image. This procedure takes advantages by getting easily laser point using well known target object, moreover it performs better results than previous procedure (i.e.:Zhang-Pless [5], and Wasielewski-Strauss [6]). Despite this contribution fit for indoor mobile robot and calibration procedure is achieved by short range distances, it employ a camera with 1024×768 pixel resolution and a laser scanner SICK LMS221-30206 that provides 100 FoV and measurement up to 80 m with accuracy $\pm 50\text{mm}$, which are usable also for outdoor vehicles.

Another important contribution is provided by [7] in 2011 based on the same ideas of Wasielewski [6] and Li [4], in which extrinsic calibration parameters are estimated by minimizing the distance between corresponding features projected on the 2D image plane. The features are lines as mentioned in [4] although, calibration target is a particular two plane panel arranged in a v-shape. Then weighted measure of distances coupled with a penalizing function are used in order to exclude outliers and estimates extrinsic camera parameters. This procedure extends previous works and improves performances by remove bias at depth edges and removing outliers.

A similar platform to ours is considered in [8] 2008. Using multi-layer LIDAR and mono camera system, the proposed method employ a circular ring as calibration target and solve extrinsic and intrinsic camera parameters. Further results on error propagation and confidence intervals are useful to define the scope of this procedure in the automotive field.

III. PROBLEM STATEMENT

Perception system employed on vehicle is composed by a four layer laser scanner and a vision system both placed in front of the vehicle facing the road (i.e. four layer LRF and a stereo system in figure 5 detailed in V). The calibration target designed for sensor registration consists of a planar isosceles triangular shape (of size $w_t = 0.6\text{m}$, $h_t = 1\text{m}$) of a specific white bright color to easier detect the shape in image. Hence, a Target Reference Frame (TRF), Camera Reference Frame (CRF) and Laser Reference Frame (LRF) are defined, notice that a permutation have to be considered to convert from sensor frame to image frame as shown in figure 1.

Basic hypothesis is that triangular calibration target will be fully observable by a laser scanner and a camera, since its geometry is known triangle vertexes can be computed and used as virtual constraints. Partial observed object lead to errors in optimization phase. The transformation of target vertex points from LRF to CRF is represented by a rigid transformation composed by a 3×3 rotation matrix R and

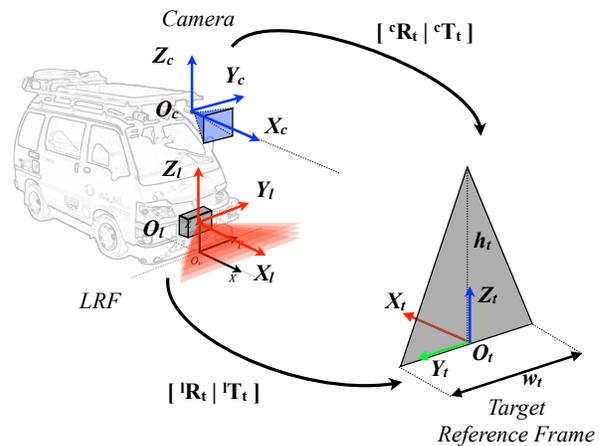


Fig. 1. Reference frames. CRF and LRF are placed in front of the vehicle facing the same region of interest and TRF is placed in the middle bottom of the calibration target. Target must be seen completely and simultaneously by laser and camera.

a translation vector t

$$P_c = {}^c R_l \cdot P_l + {}^c t_l \quad (1)$$

Given two set of N 3D points X_i and Y_i in the LRF frame and camera frame respectively, where correspondences between points are known. The problem statement consist to solve the *rigid pose transformation* $[R_\Theta | t]$ of the camera frame with respect to the LRF frame in order to minimize the overlapping error e between X and Y ,

$$e = \min_{t, \Theta} \sum_i^N \|X - s(R_\Theta Y + t)\|^2 \quad (2)$$

where $t = (t_x, t_y, t_z)^\top$ is translation vector and $\Theta = (\text{roll}, \text{pitch}, \text{yaw})^\top$ is a rotation vector.

IV. EXTRINSIC CALIBRATION

The proposed method is based on three main steps that are explained in the next three sections: acquiring synchronized data from laser and camera (i.e., with the aid of an hardware trigger), then select regions of interest and estimate position of calibration targets with respect to cameras and laser frames, in which targets pose is computed in 2D and 3D respectively. Finally, an optimization procedure based on geometric minimization involves image and laser correspondences simultaneously by fitting virtual points.

A. Target Detection in laser frame

In this section detail on the range images acquired by laser scanner is given. In this studies approaches are based on four layer LIDAR SICK or similarly IBEO Lux that have narrow elevation field of views and high scanning resolution.

The first step is to filter out far points (up to 10m), since it is not possible to detect related target objects in image with high accuracy. Then, a clusterization and recognition phase groups raw laser points by an euclidean distance policy and

Algorithm 1 Extrinsic camera registration

Require: 4-plane LIDAR.

Require: Single or stereo camera.

Ensure: Pose of camera frame w.r.t. laser frame

- 1: Acquiring simultaneously camera and laser frames
 - 2: Detect targets in laser frame and generate 3D target pose
 - 3: Detect targets in camera frame and generate 2D target pose
 - 4: Estimate rigid transformation $[R|t]$ of cameras frame w.r.t. laser frame
-

all only oportune planar thin object will be kept. Next, given a set of N potential targets T_j with $j = \dots N$ a procedure to estimate pose of each T_j is computed. RANSAC fitting is used to generate target plane π represent by a centroid C and a normal vector n model.

$$\pi : C + nx = 0 \quad (3)$$

All points X_i^j owned by T_j are projected on π , getting a new set of points X_i^π and a plane reference frame $[R^\pi|t^\pi]$ apply to T_j , where $R^\pi = [x_\pi, x_\pi \times n, n]$ and $t^\pi = X_i^j$. The virtual point O_t (shown in figure 1) in T_j will be computed on π considering only two edge point sets S_1 and S_2 closer to target segments AC and AB respectively. The procedure to obtain O_t is composed of two steps: first, an intersection point $A = (A_x, A_y)^\top$ is obtained by intersection of two polar lines r_1 and r_2

$$\begin{aligned} r_1 : x \cos(\theta_1) + y \sin(\theta_1) &= \rho_1 \\ r_2 : x \cos(\theta_2) + y \sin(\theta_2) &= \rho_2 \\ \theta_2 &= \theta_1 + \alpha \end{aligned} \quad (4)$$

solving a constrained non-linear optimization problem defined as loss function e in equation 7 with a constrained angle α ,

$$e = LS_{S_1}(r_1) + LS_{S_2}(r_2) \quad (5)$$

solution of equation 6 is guaranteed by constrain α .

$$\begin{pmatrix} \cos(\theta_1) & \sin(\theta_1) \\ \cos(\theta_2) & \sin(\theta_1) \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \quad (6)$$

Finally point O_t is obtained by shifting of h from A toward \overline{BC} then through the point cloud X_i^j . Resulting a target reference $[R^T|t^T]$ frame as follow

$$\begin{aligned} R^T &= [n, v, v \times n,] \\ t &= O_t \end{aligned} \quad (7)$$

B. Target Detection in Camera Frames

Aim of this procedure is to recover target pose in image frame with less assumptions. Two procedures have been described to recover targets position in camera reference frame based on monocular view and stereo images.

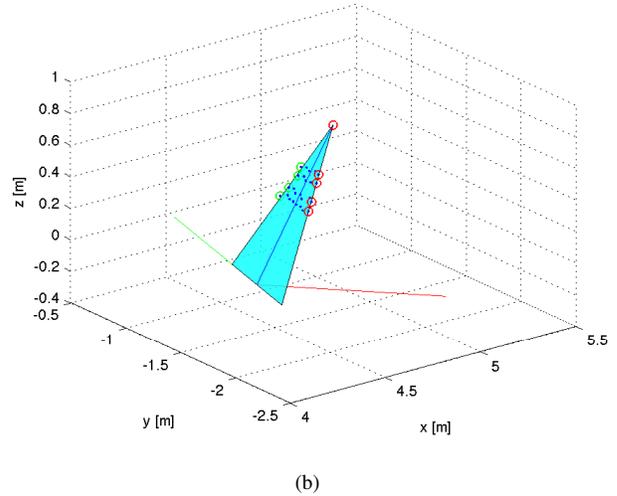
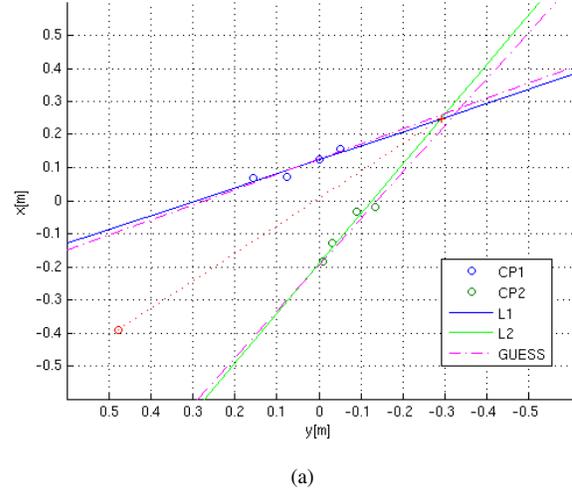


Fig. 2. Target detection with laser data. Plane projection and temporary reference frame (a). On 2D plane triangle fitting and final result in (d).

1) *Mono images:* In this section the targets identification procedure based on singular camera is detailed. To find a triangular target in the image is necessary to use a image processing algorithm. There are not assumption on the target orientation and position and moreover it is not possible to compute distances from uncalibrated sensor. Than first step of the automatic calibration procedure is to recognize all triangular shape in the scene with no constraints on the pose of such target.

The overall schema of the algorithm applied to find triangular targets on a 2D image I is showed in algorithm 2.

Firstly a preprocessing stage involves sequentially color filtering, edges extraction, and an Hough transformation that are applied to I generating image E and accumulator space $H(\rho, \theta)$ to be able to detect lines on I (as shown in *Preprocessing* function of algorithm 2). At the end of preprocessing stage parametric equations L of linear components are computed from *local maxima* extraction on

H and applied on I by relation 8.

$$x \cos(\theta) + y \sin(\theta) = \rho \quad (8)$$

Second algorithm stage is concerning match between lines L obtained by the previous stage and edges in E . Hence, a set of lines L_i are obtained as intersection between E and L where white pixel amount over each line is higher with respect to a specific threshold set with respect to the target dimensions in pixels. This stage is detailed on the *ExtractLines* function of the algorithm 2. According to the line extraction algorithm presented in [9], the *SearchTargets* procedure is able to extract N segments S_i for each line L , and compose it in a large ones with a double thresholded algorithm. Finally, a set of line S_i , each with a unique segment, is processed by a targets finder step. This step tests each possible triangle composed as line triplet $t = \{S_i, S_j, S_k\} \forall i, j, k$ with $i \neq j \neq k$ by matching couple of segment vertex, whether all vertex couples differ each other by less than a Manhattan threshold an intersection point of the lines will be computed. Only when three valid intersection points are obtained, a triangular shape is detected by algorithm.

Algorithm 2 2D triangular shape detection on mono images

Input: 2D image $I = \{x_{i,j}, i = 1 \dots w, j = 1 \dots h\}$.

Output: Set T of n vector T_i with $i = 1 \dots n$, where $T_i = [x_1, x_2, x_3]^T \in I$.

```

1: procedure SEARCH TARGETS( $I, E, H$ )
2:    $[E, H] \leftarrow \text{Preprocessing}(I)$ 
3:    $[L] \leftarrow \text{ExtractLinesfrom}(H)$ 
4:   Segments  $S \leftarrow L \cap E$ 
5:   for all tuple  $t = \{S_i, S_j, S_k\}$  where  $i \neq j$  do
6:     if Are closer segments ? then
7:        $T \leftarrow t$ 
8:     end if
9:   end for
10: end procedure

```

2) *Stereo images*: In this procedure the relative pose between left and right camera are supposed to be known in order to assume that distances in the sensor frame are correct and only the absolute pose need to be recorded. This procedure is quite similar to the laser procedure due to the same range information, however a different error distribution is present and must be consider in model extraction.

C. Pose Estimation

Point \mathbf{x} in LRF is represented on image plane with pixel \mathbf{p} , as follow

$$\mathbf{p} = F \cdot \mathbf{x}$$

F is composed by a rigid transformation

$$\mathbf{x}' = T \cdot \mathbf{x}$$

with $T = [R_\Theta | t]$ shown in equation 1 and a perspective projection K .

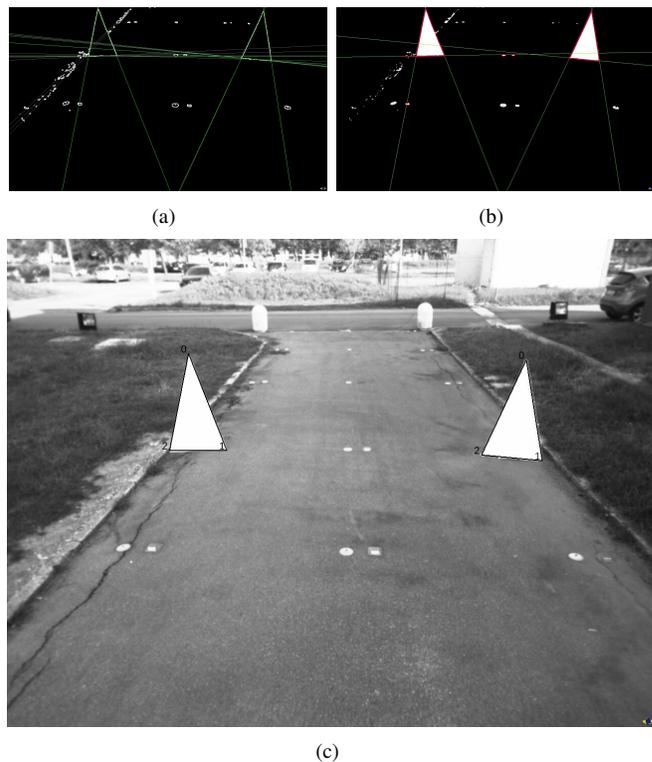


Fig. 3. Image processing phases.

$$\mathbf{p} = K \cdot \mathbf{x}'$$

where K is a pinhole model defined as follow

$$K = \begin{pmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

To be able to compute the extrinsic parameters in T intrinsic parameters pixel focal length k_u, k_v , principal point (u_0, v_0) and a skew s must be known. Extrinsic parameters registration problem can be formulated as a non-linear problem of finding the optimal solution of rotation vector Θ and translation vector t that minimize the euclidean distance between image points.

$$\min_{\Theta, t} \|\mathbf{p}_i - F\mathbf{x}_i\|_2 \quad (9)$$

between image point \mathbf{p}_i and image projection of laser points \mathbf{x}_i , where index i represent the i -th constraint pair $\langle \mathbf{p}, \mathbf{x} \rangle$ that match a triangle vertex in CRF and LRF respectively. The non-linear optimization is solved by using the Levenberg-Marquardt algorithm.

V. EXPERIMENTAL RESULTS

Experiments on synthetic and real data have been conducted to evaluate algorithms behaviors.

A. Experiments on synthetic data

To test algorithm correctness, and accuracy a set of simulations are used reporting absolute error results and solution convergence, then tests on robustness and reliability (i.e., outliers management) are based on a wide set of datasets acquired by VIAC prototypes. The simulator engine is a separated software provided by TASS, called PreScan[®] that is used coupled with Matlab Simulink[®]. Ground truth of sensors relative position $[R_{\Theta}|t]$ is used to evaluate procedure correctness, with a translation vector $t = (-2.0, 0.1, 0.98)^T$ and a rotation vector $\Theta = (0.0, 0.0, 0.0)^T$. Simulated camera had a resolution of 1280×960 pixels, 7.50 mm of focal length, 25 Hz frame rate, with a $1/2''$ (6.4×4.8 mm) CCD sensor, and without distortion. Simulated laser scanner had 85 degrees of azimuth FOV with 340 beams, and 3.2 degrees of elevation FOV with 4 beams. Laser provided data at frequency of 25 Hz. Concerning system resolution, angular azimuth cell size was 0.125 degree, and angular elevation cell size was 0.8 degrees, with a 0.04 m range cell size.

TABLE I
RESULTS ON SIMULATED DATA.

Relative Pose Error [mm]	80
Relative Orientation Error [degree]	0.3
Number of pose	25

TABLE II
TEST WORST CASE ON SIMULATION.

	Value [m]	$\sigma_1 = 4.2159$	Confidential interval	GT
t_x	-1.819	± 0.096	[-1.916, -1.723]	-2
t_y	-0.022	± 0.224	[-0.246, 0.202]	-0.1
t_z	1.080	± 0.279	[0.801, 1.360]	0.98
	Value [rad]	$\sigma_1 = 4.2159$	Confidential interval	GT
Roll	0.02578	± 0.0241386	[0.00164, 0.0499]	0
Pitch	-0.07278	± 0.0512392	[-0.12402, -0.02154]	0
Yaw	-0.01621	± 0.039511	[-0.05572, 0.02330]	0

Tables I and II refers to an average value of extrinsic parameters that is evaluated with minimum number of poses.

With 2 targets per frame as shown in figure 4.(b), position error (figure 4.(d)) decreases up to 20 cm when more 5 poses are used. In this case orientation error in figure 4.(f) decrease in a range of ± 1 degree for each angles. In the other hand, test in figure 4.(a) considers one target per frame and proves that error on yaw angle does not converge, affecting other parameters convergence. This behavior is shown in 4.(c)(e).

B. Experiments on real data

Experiments on real data were made with an intelligent vehicle (figure 5.(a)) equipped by VisLab during VIAC expedition. Vehicle was equipped with a autonomous driving state-of-art suite of sensors as detailed in [10]. Concerning computer vision navigation systems, perception tasks were made by a 3 synchronized cameras panoramic vision system, a front and rear stereo cameras and three different mounted

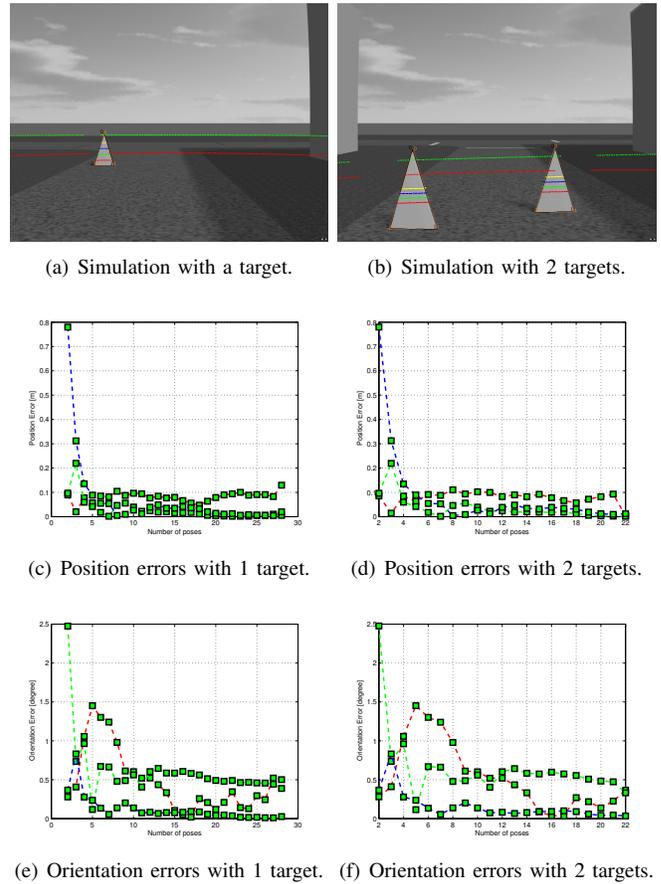


Fig. 4. Errors on position and orientation on simulated data. In the orientation error graphs dash line representing yaw is in blue, roll in red, and pitch in green.

laser scanners suites to perform 360 degrees environmental perception.

Two stereo rigs, shown in figure 5.(b), was used for experimentation, a large baseline couple of PointGrey Firefly cameras, and a PointGrey Bumblebee X3 for the smallest one. The large stereo rig had a 0.80 m baseline, each cameras had 752×480 pixels resolution. Point Grey Bumblebee X3 had a 24 cm baseline, a resolution of 1280×960 pixels, and a 3.8 mm focal length.

Laser scanner used for experiments was a 4-layers Sick LD-MRS-400001. Laser scanner had 3 different angular resolution $0.125^\circ / 0.25^\circ / 0.5^\circ$ with scanning frequency of 12.5 Hz / 25 Hz / 50 Hz respectively. Distance resolution was 40 mm, and the statistical error σ_1 was 100 mm with a range up to 50 m at 10% reflectivity.

Real test was conducted with 2 calibration targets per frame in a fixed position 5 m far from vehicle, detailed in figure 5.(c). Results on 20 frames is shown in Table III, that highlights imprecision concerning target pose estimation in LRF due to high level of noise. This problem can be solved acquiring frame with different targets pose.

TABLE III
TEST RESULTS ON REAL DATA

	Value [m]	$\sigma_1 = 6.45715$	Confidential Range
t_x	-0.162986	± 0.0416555	[-0.204642, -0.121331]
t_y	-0.0285926	± 0.0884247	[-0.117017, 0.0598321]
t_z	2.47908	± 0.0884488	[2.39063, 2.56752]
	Value [rad]	$\sigma_1 = 6.45715$	Confidential Range
Roll	-0.0108902	± 0.00814383	[-0.019034, -0.00274638]
Pitch	0.556509	± 0.0190649	[0.537444, 0.575574]
Yaw	0.00645968	± 0.0184597	[-0.012, 0.0249194]

VI. CONCLUSION AND FURTHER WORKS

This article facing the multi-sensorized intelligent vehicles extrinsic calibration problem. Cameras and laser scanners extrinsic inter-calibration is performed by the aim of triangular shape targets. Proposed method might be used both for monocular and stereo vision systems, where intrinsic parameters are considered as known. Furthermore, inter-calibration can be coupled with stereo auto-calibration to be able to recover correct distance measure.

The proposed method reports good results in simulated environments, however on real scenarios it is strongly affected by laser scanner statistical error. Therefore a multi pose calibration is preferred to come through laser uncertainty limitation as discussed in experimental results section.

Further works are focused on using this procedure in urban environment by enforcing the shape detection phase, and using such method with known objects beside the road, even with road signs whenever a laser scanner can percepts it.

ACKNOWLEDGMENT

This work has been supported by the European Research Council (ERC) within the Open intelligent systems for Future Autonomous Vehicles (OFAV) Advanced Investigators Grant. Simulations have been conducted with PreScan[®] software provided by TASS, that is used coupled with Matlab Simulink[®].

REFERENCES

- [1] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: a review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 694–711, may 2006.
- [2] L. Huang and M. Barth, "Tightly-coupled lidar and computer vision integration for vehicle detection," in *Intelligent Vehicles Symposium, 2009 IEEE*, june 2009, pp. 604–609.
- [3] Y. Li, Y. Ruichek, and C. Cappelle, "3d triangulation based extrinsic calibration between a stereo vision system and a lidar," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, oct 2011, pp. 797–802.
- [4] G. Li, Y. Liu, L. Dong, X. Cai, and D. Zhou, "An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, 29 2007–nov. 2 2007, pp. 3854–3859.
- [5] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, sept.-2 oct. 2004, pp. 2301–2306 vol.3.
- [6] S. Wasielewski and O. Strauss, "Calibration of a multi-sensor system laser rangefinder/camera," in *Intelligent Vehicles '95 Symposium, Proceedings of the*, sep 1995, pp. 472–477.

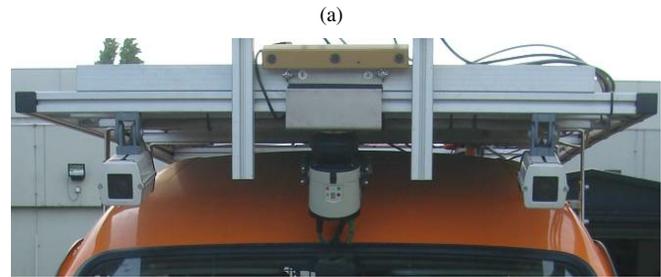


Fig. 5. VIAC prototypes used for test on real data and results. (a) Vehicle frontal perception system. (b) Detail on two stereo camera systems with both sort and wide baselines. (c) Data obtained by the perception system after sensor registration.

- [7] K. Kwak, D. Huber, H. Badino, and T. Kanade, "Extrinsic calibration of a single line scanning lidar and a camera," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, sept. 2011, pp. 3283–3289.
- [8] S. Rodriguez F, V. Fremont, and P. Bonnifait, "Extrinsic calibration between a multi-layer lidar and a camera," in *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, aug 2008, pp. 214–219.
- [9] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "On straight line segment detection," *Journal of Mathematical Imaging and Vision*, vol. 32, pp. 313–347, 2008.
- [10] L. Mazzei, P. Medici, and M. Panciroli, "A lasers and cameras calibration procedure for viac multi-sensorized vehicles," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, june 2012, pp. 548–553.