A multi-sensor calibration toolbox for Kinect : Application to Kinect and laser range finder fusion.

Jean-Clement Devaux, Hicham Hadj-Abdelkader, Etienne Colle

A multi-sensor calibration toolbox for Kinect : Application to Kinect and laser range finder fusion

Jean-Clément Devaux, Hicham Hadj-Abdelkader and Etienne Colle[1]

***Abstract*— In most of indoor mobile robot navigation, obstacle avoidance is a crucial task and should be reliable. Fusion of sources of data can be used to detect any obstacle shapes. Laser range finder is usually used to deal with this task in presence of simple obstacles. 3D sensors as Kinect provide 3D information which can be used for more complex obstacle detection. However, opposed to laser range finder, Kinect has strong limitations like measuring range or the field of view. This paper proposes a full calibration of different sensors which can be coupled with the Kinect sensor by Microsoft. The approach can also be applied to a large variety of 3D depth sensors like time-of-flight cameras, 3D LIDAR or RADAR. The basic idea is to compute the Euclidian transformation between each sensor. In this paper, we show that chessboard methods used to calibrate color cameras can be extended to deal with 3D depth sensors like Kinect. We show in a real experiment, the benefit of the fusion based on the calibration results, in order to detect complex obstacles reliably.**

I. INTRODUCTION

A good knowledge of the environment is essential for most mobile robotic applications to allow robots to move safely. For example, obstacle avoidance is often based on distance sensors like laser range finders or ultrasound belts. They provide very accurate and reliable depth data. However, the efficiency of those sensors depends a lot on material and shape of obstacles. In fact, obstacles like chair or table can not be well detected by such sensors. Mostly, only their legs are detected and the area between legs are exploited as a free space by the robot navigation system. Several and various solutions are proposed in the literature. The authors in [1] explored the environment using a 3D laser range finder which is built by mounting a basis 2D range finder with a servomotor. Other works attempt to build a 3D representation of the environment using, at least, two 2D laser range finders [2], [3], [4] or 3D laser scanners. Other researchers are interested in fusing 2D laser data and stereo vision camera which is less expensive and easier than using 3D laser technologies. In [5], 2D laser range finder and stereo cameras data are exploited together to create an occupancy grid for obstacle avoidance and trajectory planning.

Launched on November 4, 2010 in North America, Microsoft Kinect sensor is an interesting sensor for robotic applications since it offers at a very low cost a solution to get color and 3D description of the environment [6], [7], [8]. However, compared to the laser range finders with a horizontal field of view of more than 200°, Kinect has a limited field of view (around 57°). Furthermore, Kinect is able to provide depth measurements between 80 and 500 cm whereas laser sensors can provide data from almost 0 to hundreds of meters. In many indoor robotic applications, the Kinect would be more interesting to look forward in order to enhance the data describing the environment all around the robot and which is delivered by the laser range finder.

Fusing different sources of data is a way to get a better knowledge of the environment : both in the sense of field of view and of accuracy. The main condition to deal with fusion is to know the Euclidian transformation between sensor frames where data are represented.

Many works have dealt with intrinsic and extrinsic color camera calibration. Z. Zhang in [9] and [10] releases a method widely used in calibration tools especially in Matlab calibration toolboxes like *Camera Calibration Toolbox for Matlab* by J.-Y. Bouguet [11].

Some works have proposed a solution to calibrate a laser range finder with a camera when the laser beam is visible by the camera [12], [13]. In the case when the laser beamer is invisible in the image, Q. Zhang and R. Pless proposed, in [14], a laser-camera extrinsic calibration process using a checkerboard pattern plane. This paper extends their method in section II-A and adapts it to the Kinect sensor.

The paper is organized as follow. Section II presents the data modeling and the calibration process of the laser-kinect system. Section III presents and analyzes results with real world data for laser-Kinect calibration methods. Finally, an application of data fusing for an efficient obstacle detection is presented in section IV.

II. LASER - KINECT CALIBRATION

A. Sensors modeling

As Kinect depth sensor is based on IR camera, we assume that it can be described by the usual pin-hole model. Consider the rotation matrix $\mathbf{R}_K$ and the translation vector $\mathbf{t}_K$ between the Kinect frame $\mathcal{F}_K$ and the world frame $\mathcal{F}_W$. An observed 3D point of coordinates in the world frame $\mathbf{p}_W = [x_W, y_W, z_W]^\top$, can be expressed in the Kinect-depth image plane into an image point of coordinates $\mathbf{m}_K = [u_K, v_K]^\top$ using the known pin-hole projection:

$$\mathbf{m}_K \sim \mathbf{K}_K(\mathbf{R}_K.\mathbf{p}_W + \mathbf{t}_K) \tag{1}$$

with

$$\mathbf{K}_K = \begin{pmatrix} f_u & s_{uv} & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix}$$

[1]J.-C. Devaux, H. Hadj-Abdelkader and E. Colle are with IRA2 team in IBISC laboratory of University of Evry, 40 Rue du Pelvoux, CE 1455 Courcouronnes, 91020 Evry Cedex, France. E-mail : `givenname.name at ibisc.univ-evry.fr`

is intrinsic parameters matrix of IR camera. c_u and c_v are the coordinates of the principal point in the image plane. f_u and f_v are the focal lengths in u and v axes. We assume that the skew s_{uv} is neglected (equal zero) since pixels are square or rectangular. Note that the exploited model does not take into account lens distortion which is not significant for the Kinect sensor. We will assume, without loss of generality, that the Kinect and the world frames are superposed ($\mathbf{R}_K = \mathbf{I_{3X3}}$ and $\mathbf{t}_K = [0,0,0]^\top$). Equation (1) becomes:

$$\mathbf{m}_K \sim \mathbf{K}_K.\mathbf{p}_K$$

where $\mathbf{p}_K = [x_K, y_K, z_K]^\top$ is the coordinate vector of the observed 3D point. Since the z_K component is provided by the Kinect sensor ($z_K = \text{depth}(u_K, v_K)$), the remaining components can be computed as $x_K = z_K \frac{u_K - c_u}{f_u}$ and $y_K = z_K \frac{v_K - c_v}{f_v}$.

Let $\mathbf{R}$ and $\mathbf{t}$ be respectively the rotation matrix and the translation vector, linking the Kinect frame $\mathcal{F}_K$ and the laser frame $\mathcal{F}_L$. The aim of the calibration between the kinect and the laser sensors is to estimate $\mathbf{R}$ and $\mathbf{t}$ which are known as the extrinsic parameters in the laser-Kinect calibration.

The relationship between the coordinate vectors $\mathbf{p}_K$ and $\mathbf{p}_L$, expressed in the kinect and the laser frames respectively, of the observed 3D point is given by:

$$\mathbf{p}_K = \mathbf{R}.\mathbf{p}_L + \mathbf{t} \tag{2}$$

Let (π) be a plane defined by the normal unit vector $\mathbf{n}$ and the distance d expressed in the kinect frame $\mathcal{F}_K$. When the observed 3D point lies onto the plane (π), the following constraint can be derived using (2):

$$\mathbf{n}^\top \mathbf{p}_K - d = \mathbf{n}^\top (\mathbf{R}\mathbf{p}_L + \mathbf{t}) - d = 0 \tag{3}$$

Consider now that the observed 3D point lies onto the line defined by the intersection between the plane (π) and the laser plane. The coordinate vector $\mathbf{p}_L$ can be obtained from the laser data. The plane parameters $\mathbf{n}$ and d can be estimated from the kinect data as it will be described in the next section. $\mathbf{R}$ and $\mathbf{t}$ can thus be estimated using constraints in (3).

As laser range finder is only measuring obstacle distances on a plane, we fix the laser frame $\mathcal{F}_L$ such as the laser plane is defined by the plane containing the axis x and y. Consequently, the vector $\mathbf{p_L}$ can be simplified to a $\mathbb{R}^2$ homogeneous coordinates system $\tilde{\mathbf{p}}_L = [X_L, Y_L, 1]^\top$ and (3) becomes:

$$\mathbf{n}^\top \mathbf{H} \tilde{\mathbf{p}}_L - d = 0 \tag{4}$$

Where $\mathbf{H} = \mathbf{R} \begin{pmatrix} 1 & 0 & \\ 0 & 1 & \mathbf{t} \\ 0 & 0 & \end{pmatrix}$.

B. Modeling data

In order to solve the calibration problem, depth image and laser scan data are used. The plane parameters ($\mathbf{n}$ and d) can be extracted from the depth image provided by the Kinect sensor. The coordinates vector $\tilde{\mathbf{p}}_L$ can be computed using the laser scan data. We assume that the planar regions of the calibration pattern are selected by the user from the depth images and the laser scan. Usually, some points of the selected region by the user are completely outside of the calibration pattern. These outliers can be rejected using a robust fitting based on RANSAC and a good estimation of the plane parameters can be obtained. Figure 1 shows a representation of an extracted plane from 3D point clouds provided by the Kinect sensor (scale in cm). Blue dots represent projection of the selected points with the plane model using RANSAC. On the right part of the figure, the same point cloud is represented with another point of view to better see the accuracy of plane fitting. Then, robust fitting allows user to quickly select the calibration pattern area in the depth image without worrying about the effect of outliers. Furthermore, RANSAC based fitting filters efficiently the Kinect noise and improve the depth resolution (e.g. depth is discrete and has 1 cm of accuracy as shown on figure 2).

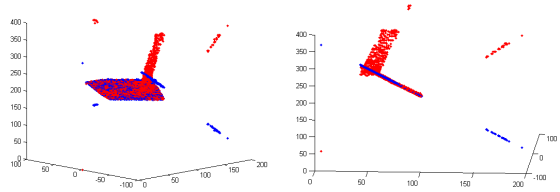


Fig. 1. RANSAC plane fitting (blue dots) on Kinect 3D point cloud (red dots) with great amount of outliers

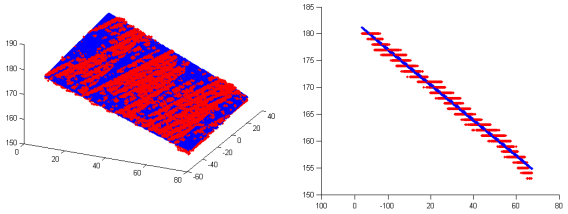


Fig. 2. RANSAC plane fitting (blue dots) on Kinect 3D point cloud (red dots) showing 1 cm max precision in depth

Laser range finder is also sensitive to noise which depends on color and material of the detected object. Black and white objects situated at the same distance with respect to the kinect, are detected at different depths. So, it is better to use a mono-color calibration plane to deal with that problem. However, we will be not able to calibrate an extended system including kinect, laser range finder and conventional camera sensors. When the calibration plane is a chessboard pattern, laser data has a particular shape which looks like battlements since the chessboard is seen as alternate white and black cells. Figure 4 represents a sample of different calibration planes at several orientations and distances. As can be shown on the left side of the figure, the *battlement phenomena* appears and increases proportionally to the distance between the laser and the plane. Further, a low incident angle (e.g. 30°) will imply loss of data along the segment (as shown in the bottom-left figure). Roughly, we can say that the quality of laser scans decreases when increasing the distance and decreasing the incident angle.

Battlements show that the color of material also has some effects on sensor results. We propose a modified chessboard pattern (right part of the figure 3) in order to deal with this problem and improve the quality of the laser scans as shown in the right side of figure 4. A white band has been added all along the width to improve the laser data. The height of this band has to be chosen depending on the maximum leaning angle (pitch). As it is shown in figure 3, black crosses are added in the white band to keep the possibility to calibrate the conventional cameras. Furthermore, the chessboard pattern is kept on the bottom of the calibration plane since the laser will never impact this part. It should improve the accuracy of the concentional camera calibration by adding more points along the vertical axis.

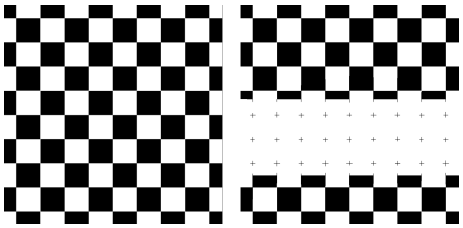


Fig. 3. Chessboard plane to calibrate conventional cameras (left) and modified chessboard pattern to be compliant with laser range finder (right)

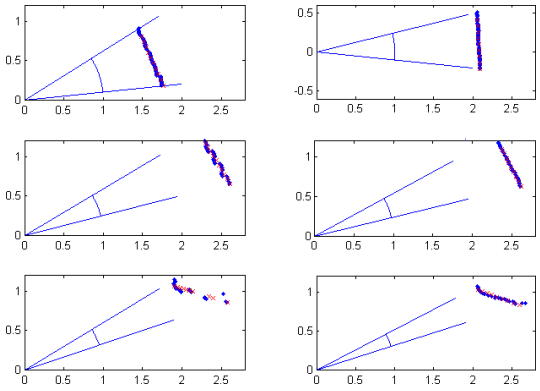


Fig. 4. Laser scan of calibration plane (blue dots) with fitted data (red crosses) (representation in cartesian plane, distance in meters), left column with traditional chessboard plane, right column with modified calibration plane

Since the calibration chessboard is a plane, the laser range finder detects that plane as a segment. A least square method is used to fit a line containing that segment. The fitted line is then re-sampled to generate a set of n points $\tilde{\mathbf{p}}_L$ in order to solve (4).

C. Calibration

This section explains and compares linear and several nonlinear solutions to (4). We use the linear solution to initialize the nonlinear least square Levenberg-Macquardt optimizations.

1) Linear solution: So as to solve the constraint (4), a set of data is made up of n points (re-sampled laser segment) of each m samples taken at several positions and orientations of the calibration plane. Since 9 unknown parameters have to be estimated, at least 9 equations are needed to solve the problem. Usually, an over determined system is used to deal with the noise measurements. The over determined linear problem can be written as m equations of:

$$(\mathbf{n}_j^\top \otimes \mathbf{P}_{L_{ij}}^\top)\mathbf{h} = \mathbf{d}_j \otimes \mathbb{1}_n \tag{5}$$

where $\otimes$ is the Kronecker product, $i = 1 \cdots n$ is the point number, $j = 1 \cdots m$ is the plane number and $\mathbb{1}_n$ is a column vector of n elements all equal to 1. $\mathbf{P}_{L_{ij}}$ is the matrix of n coordinates vectors $\tilde{\mathbf{p}}_{L_i}$ stacked on n rows. The column vector solution $\mathbf{h} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^\top$ contains elements of an estimated $\widehat{\mathbf{H}}$ matrix.

Thanks to the construction of $\mathbf{H}$, we can easily extract $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{t}}$ from the estimated vector columns $\widehat{\mathbf{h}}_\mathbf{k}$, $(k = 1 \cdots 3)$, as:

$$\widehat{\mathbf{R}} = [\widehat{\mathbf{h}}_\mathbf{1}, \widehat{\mathbf{h}}_\mathbf{2}, \widehat{\mathbf{h}}_\mathbf{1} \times \widehat{\mathbf{h}}_\mathbf{2}] \tag{6}$$

$$\widehat{\mathbf{t}} = \widehat{\mathbf{R}}\widehat{\mathbf{h}}_\mathbf{3} \tag{7}$$

However, (4) does not include any constraint on the H matrix where the first two columns should be orthogonal (rotation group $SO(3)$). However, the estimated $\widehat{\mathbf{R}}$ will not be a rotation matrix.

In [15], the rotation matrix $\mathbf{R}$ can be approximated using badly formed estimated matrix $\widehat{\mathbf{R}}$ by the rotation matrix $\widetilde{\mathbf{R}}$ which minimizing Frobenius norm $\left\|\mathbf{R} - \widetilde{\mathbf{R}}\right\|_f$ under constraint $\widetilde{\mathbf{R}}\widetilde{\mathbf{R}}^\top = \mathbb{I}_\mathbf{3}$ using singular value decomposition:

$$\widehat{\mathbf{R}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$
$$\mathbf{R} \approx \widetilde{\mathbf{R}} = \mathbf{U}\mathbf{V}^\top$$
$$\mathbf{t} \approx \widetilde{\mathbf{R}}\mathbf{h}_\mathbf{3}$$

Note that other techniques can be used to extract $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{t}}$. Any of these techniques can be used since the extracted parameters are only used to initialize the non-linear methods.

Even if $\mathbf{H}$ is representing a rotation and a translation, because of equation form we have to estimate 9 parameters instead of 6 (3 for rotation and 3 for translation). Each laser segment has two degrees of freedom. Then, we need at least five different plane positions to solve (5).

2) Nonlinear optimization of $\mathbf{t}$*:* The linear solution is not sufficiently accurate but provides a good initialization for the nonlinear optimizations. After observation of reprojection data from the kinect frame to the laser frame using the linear estimation, we made assumption that rotation matrix $\widetilde{\mathbf{R}}$ was not so bad. In this section, our optimization focuses only on the nonlinear estimation of the translation $\widetilde{\mathbf{t}}$.

The aim here involves to minimize the projection error of the laser points in the Kinect coordinate system. The orthogonal distance to the plane is used as a metric to optimize $\breve{\mathbf{t}}$:

$$min \sum_{j=1}^{m} \sum_{i=1}^{n} \left\| \mathbf{n}_j.(\widetilde{\mathbf{R}}\mathbf{p}_{L_{ij}} + \breve{\mathbf{t}}) - d_j \right\|^2 \tag{8}$$

3) Nonlinear optimization of $\mathbf{R}$ *and* $\mathbf{t}$ *using* $\mathbf{H}$*:* To confirm if the linear solution of the rotation matrix was correct, we both optimize rotation and translation at the same time. Using the same metric as in (8) and the same decomposition of $\mathbf{H}$ as in (4), the problem becomes a constrained minimization problem:

$$\min_{\breve{\mathbf{H}}} \sum_{j=1}^{m}\sum_{i=1}^{n} \left\| \mathbf{n}_j . \breve{\mathbf{H}} \mathbf{p}_{Lij} - d_j \right\|^2$$
$$s.c. \left\| \breve{\mathbf{h}}_\mathbf{1} \right\| = \left\| \breve{\mathbf{h}}_\mathbf{2} \right\| = 1 \; and \; \breve{\mathbf{h}}_\mathbf{1} . \breve{\mathbf{h}}_\mathbf{2} = 0$$

Those constraints imply to finally have a correct rotation matrix from (6) because it forces $\breve{\mathbf{h}}_\mathbf{k}$ columns to be unit orthogonal vectors. So $\breve{\mathbf{R}}$ and $\breve{\mathbf{t}}$ can be extracted from $\breve{\mathbf{H}}$ in the same way as in (6) and (7).

4) Beam metric: Dupont et al. in [16] introduce a different metric to improve calibration in outdoor context. Instead of using orthogonal distance between the point re-projection and the plane, they propose to use residual distance along laser beam. We used this metric in the non-linear optimizations of calibration to compare it with the orthogonal distance. The distance d_b along the laser beam is given by:

$$d_b = \left\| \mathbf{p}_K - \mathbf{s} + \overrightarrow{\mathbf{r}} e \right\|$$

where $\mathbf{s}$ is the source of beam position (in other words, the origin of $\mathcal{F}_L$ expressed in the Kinect frame $\mathcal{F}_K$), $\overrightarrow{\mathbf{r}}$ is the beam vector and e is the time of collision between plane and source:

$$\mathbf{s} = \mathbf{R}\mathbf{t}$$
$$\overrightarrow{\mathbf{r}} = \mathbf{p_K} - \mathbf{s}$$
$$e = \frac{d - \mathbf{n}.\mathbf{s}}{\mathbf{n}.\overrightarrow{\mathbf{r}}}$$

Minimization problem (8) becomes now:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{j=1}^{m}\sum_{i=1}^{n} d_{b_{ij}}^2(\mathbf{R},\mathbf{t}).$$

III. SENSOR NETWORK CALIBRATION RESULTS

A. Sensor network setup

Affordable service robots in indoor environments usually have low cost on-board sensors. An example of such robot is given by figure 5. In this experiment, we calibrate a low cost laser range finder with a low range (maximum 4 meters), a Kinect sensor and a color camera mounted on our mobile robot (see figure 5). Our data set was taken in a real use case configuration.

Two data sets are considered in the calibration experiment. The first one is composed of 16 samples of a 80×80 cm traditional chessboard pattern (with 89×89 mm cells) taken at different positions and orientations (mainly around pitch and yaw axis). The second data set is produced, in the same configuration as the first one, using the modified chessboard pattern.

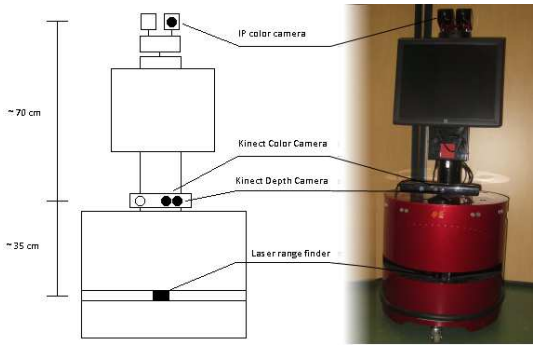


Fig. 5. Sensor positions on Lina robot (from Droïds Company)

B. Kinect and laser range finder calibration results

To compare the calibration methods presented in the previous section, means and standard deviations of two errors, orthogonal and beam distances, are presented in Figure 6. These statistics parameters were produced for each calibration method (linear solution, optimization of $\mathbf{t}$, optimization of $\mathbf{R}$ and $\mathbf{t}$ using orthogonal distance and beam metric) and for each data set (using traditional and modified chessboard patterns).

The figure 6 shows that the calibrations using the orthogonal distance and the beam metric roughly provide the same results. In both cases, linear solution produces a higher error than the non-linear methods which is typically expected.

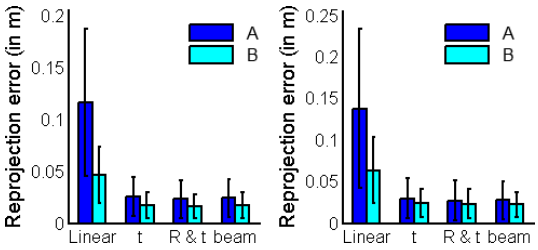


Fig. 6. Mean and standard deviation of reprojection error. Computed with orthogonal distance (left) and beam metric (right) for the chessboard plane data set (A) and the modified chessboard plane data set (B)

Even if linear results are not enough for sensor network calibration, the rotation is always well estimated before the non-linear optimization since its estimation from the matrix $\mathbf{H}$ is constrained by the $SO(3)$ properties. However, the translation is in $\mathbb{R}^3$ which is less restrictive and its estimation should be improved by the non-linear optimization. It is also important to note that the modified calibration pattern improves significantly the linear estimation and thus increases the performances of the non-linear optimization. In fact, a good linear solution reduces the risk of local minimum and increases the speed of the optimization algorithms.

Table I summarizes (in cm) the mean, the standard deviation, the minimum and the maximum of orthogonal and beam re-projection errors, to be able to compare, in the same error space, all presented estimation methods for the both calibration patterns. As can be observed on figure 6, the differences between the translation optimization and the

global optimizations based on the orthogonal distance and beam metric, are lower than 1cm. Green cells represent the best values in their categories.

TABLE I

REPROJECTION ERRORS FOR LASER-KINECT CALIBRATION IN CM, CALIBRATION WITH TRADITIONAL AND MODIFIED CHESSBOARD PLANE

Traditional chessboard calibration plane

Stat	Reproj. err	Linear	t opt.	R & t opt.	beam opt.
Mean	Orthogonal	11.62	2.55	2.33	2.42
	Beam	13.80	3.00	2.71	2.79
Std	Orthogonal	7.11	1.92	1.85	1.79
	Beam	9.54	2.43	2.45	2.29
Min	Orthogonal	0.03	0.03	0.03	0.00
	Beam	0.03	0.03	0.03	0.00
Max	Orthogonal	26.01	7.92	10.58	8.81
	Beam	41.01	12.49	15.64	13.23

Modified chessboard calibration plane

Stat	Reproj. err	Linear	t opt.	R & t opt.	beam opt.
Mean	Orthogonal	4.66	1.75	1.64	1.71
	Beam	6.37	2.37	2.30	2.23
Std	Orthogonal	2.74	1.27	1.16	1.27
	Beam	4.01	1.69	1.77	1.54
Min	Orthogonal	0.51	0.02	0.04	0.02
	Beam	0.15	0.02	0.04	0.02
Max	Orthogonal	12.35	6.60	4.89	6.66
	Beam	16.91	8.05	7.71	6.80

The Kruskal-Wallis test results in the table II presents the significance of the comparison. We note that the P-value is always lower than 10^{-6} when the test is performed between the linear method and any non-linear optimization. It means optimization of $\mathbf{t}$, optimization of $\mathbf{R}$ and $\mathbf{t}$ using orthogonal distance and beam metric are all very significantly more accurate than the linear solution neither with traditional chessboard pattern nor with the modified one. Overall, and as it is shown in the table III, the difference between using a traditional chessboard pattern and the modified one is statistically significant for all calibration methods when the orthogonal distance error is minimized, and for the linear and the non-linear minimization of the beam metric when the beam distance error is considered.

TABLE II

KRUSKAL-WALLIS TEST RESULTS BETWEEN THE THREE NON-LINEAR OPTIMIZATION METHODS AND THE LINEAR SOLUTION

Traditional chessboard calibration plane

p > KW	Linear	t opt.	R & t opt.	beam opt.
Linear	X	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
t opt.	X	X	0.3428	0.6856
R & t opt.	X	X	X	0.5203

Modified chessboard calibration plane

p > KW	Linear	t opt.	R & t opt.	beam opt.
Linear	X	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
t opt.	X	X	0.6999	0.8438
R & t opt.	X	X	X	0.8514

Even if we cannot conclude on accuracy (the three non-linear methods are different with no statistical significance in term of residual error), we need to choose one optimization method to complete the calibration process. In the sequel, the global optimization and calibration both refer to non-linear optimization of $\mathbf{R}$ and $\mathbf{t}$.

TABLE III

KRUSKAL-WALLIS TEST RESULTS BETWEEN CHESSBOARD AND MODIFIED CALIBRATION PLANE

p > KW	Linear	t opt.	R & t opt.	beam opt.
Orthogonal	$< 10^{-6}$	$< 10^{-3}$	$< 10^{-2}$	$< 10^{-3}$
Beam	$< 10^{-6}$	0.7125	0.1242	$< 10^{-1}$

Since the ground truth to evaluate the rotation and translation estimation is hard to be obtained, we can only use the re-projection error. This error seems compliant with the final aim of fusing data sensors for most robotic applications. In particular, the laser-Kinect mean re-projection error (2.33 cm with global optimization method) is lower than the battlement depth (5.76 cm on battlements on figure 4). That means calibration accuracy can be better than our laser range finder accuracy for black & white objects. It also means that the Kinect sensor is less sensitive to colors (there is no battlement on kinect point cloud) and that a smart fusion with Kinect data will be able to improve laser data.

IV. KINECT - LASER RANGE FINDER FUSION FOR RELIABLE OBSTACLE AVOIDANCE

A. Fusion model

In order to fuse laser and Kinect data to detect difficult obstacles such as tables and chairs, the ground has to be removed from Kinect point cloud. We rotate and translate kinect point cloud to bring the ground parallel to the laser plane and then we remove all the points under that plane.

Once the ground and the ceiling are removed, all objects detected by the Kinect sensor are a potential obstacle. The kinect data can be fused with the laser range data to improve the obstacle detection efficiency. In order to realize that fusion, each 3D point provided by the kinect sensor, of coordinates $(x_i \ y_i \ z_i)$ expressed in the laser frame, is projected onto the laser plane. For the data consistency, projected points are given in the polar coordinates $(l_K, \ \theta_K)$ in the laser frame with:

$$\theta_K = \tfrac{7\pi}{6} - \tan^{-1}(z_i/x_i) \quad and \quad l_K = \sqrt{x_i^2 + z_i^2}$$

The fusion is simply realized by considering the closer obstacles:

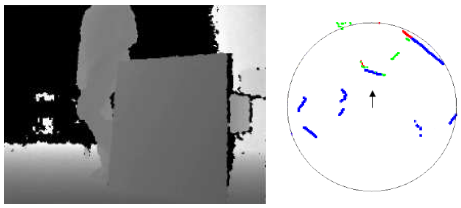
$$\mathbf{d}(\theta) = \min\left(\mathbf{l}(\theta), l_K(\theta)\right)$$

with $\mathbf{l}$ and $\mathbf{d}$ maps are respectively the laser data and the fusion result distances. Missing data or errors are supposed at $+\infty$ in $\mathbf{l}$ and $\mathbf{d}$.

B. Experimental results

Improvements of the laser data with the Kinect sensor is independent to the laser range finder accuracy and can be observed in many situations. In figure 7, the Kinect depth image is shown besides the range finder map. This map contains a black circle of 4m radius (corresponding to the

maximum detection range of the low cost laser) and color dots. Blue dots represent the detected obstacles by the laser range finder which have not been modified by the Kinect-Laser fusion (either because no data are available from the Kinect, or laser and Kinect distance data are equal). Green dots represent obstacles which have been observed only by the Kinect sensor whereas red dots show what the laser range finder would have detected in that same direction.

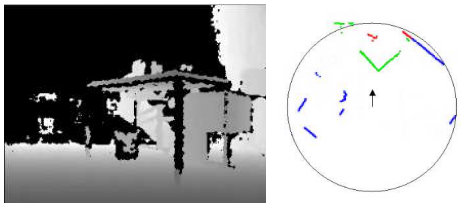


(a) Kinect depth image (b) Fusion

Fig. 7. Vertical obtacle in front of Kinect and Laser sensors. The black arrow is centered on robot position and represents its orientation

Figure 7 shows a vertical panel detected by both the Kinect and the laser range finder. The panel is drawn in two colors (blue and green) since it has been detected by both sensors. However, the person standing on the left side of the panel is detected by the laser range finder at only the level of his legs whereas his body is larger than his legs. Fortunately, the kinect-laser fusion enhances the detection by adding the detected body in the obstacle map (green dots).

Another interesting situation is presented in Figure 8 where a table is completely insensible by the laser. Thanks to the Kinect, the top of the table is detected and visible in the green color. Obviously, an obstacle avoidance algorithm using only laser would have a collision with the table. This collision can be avoided easily when fusing Kinect and laser detection data. This experimental situation is provided in the attached video to this paper. It shows real experiment of complex obstacle (like table) avoidance.



(a) Kinect depth image (b) Fusion

Fig. 8. A table in front of Kinect and Laser sensors. The black arrow is centered on robot position and represents its orientation

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a method to calibrate a laser range finder and 3D sensors like Kinect. The proposed method requires the use of a large enough chessboard plane pattern. We have shown through experiments that the proposed calibration technique is enough for fusing Kinect data with other sensors data to improve the complex obstacles detection such as table and chair. This fusion is helpful to pass beyond the Kinect limitations (field and range of view) and insure a reliable obstacle avoidance. Concerning the kinect and laser calibration, we have proposed a modified calibration pattern and shown that it significantly increases the calibration accuracy with the low cost laser sensors.

We have extended the *Camera calibration toolbox for Matlab* [11] to propose a multi-sensor calibration toolbox including conventional camera, Laser range finder and Kinect. Our toolbox is available at `http://goo.gl/OlJmGX`.

In the future works, we would like to calibrate the internal sensors of the Kinect without accessing to the IR raw images and improve the fusion of the two sensors.

REFERENCES

[1] A. N. H. Surmann and J. Hertzberg, "An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments," *Robotics and Autonomous Systems*, vol. 45, no. 3–4, pp. 181–198, December 2003.

[2] S. Thrun, W. Burgard, and D. Fox, "A real-time algorithm for mobile robot mapping with applications to multi-robot and 3d mapping," in *International Conference on Robotics and Automation*, ser. ICRA'00, 2000, pp. 321–328.

[3] C. Frh and A. Zakhor, "3d model generation for cities using aerial photographs and ground level laser scans," in *Conference on Computer Vision and Pattern Recognition*, ser. CVPR'01. Kauai, HI, USA: IEEE Computer Society, 2001, pp. 31–38.

[4] H. Zhao and R. Shibasaki, "Reconstructing textured cad model of urban environment using vehicle-borne laser range scanners and line cameras," in *International Workshop on Computer Vision Systems*, ser. ICVS '01, London, UK, UK, 2001, pp. 284–297.

[5] S. Kumar, D. Gupta, and S. Yadav, "Sensor fusion of laser and stereo vision camera for depth estimation and obstacle avoidance," *International Journal of Computer Applications*, vol. 1, no. 25, pp. 20–25, February 2010.

[6] N. Ganganath and H. Leung, "Mobile robot localization using odometry and kinect sensor," in *Emerging Signal Processing Applications*, Las Vegas, 12–14 January 2012, pp. 91–94.

[7] D. Correa, D. Sciotti, M. Prado, D. Sales, D. Wolf, and F. Osorio, "Mobile robots navigation in indoor environments using kinect sensor," in *Critical Embeded Systems*, Sao Paolo, 2012, pp. 36–41.

[8] P. Benavidez and M. Jamshidi, "Mobile robot navigation and target tracking system," in *International Conference on System of Systems Engineering*, June 2011, pp. 299–304.

[9] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *International Conference on Computer Vision*, vol. 1, Corfu, Greece, September, 20-25 1999, p. 666.

[10] ——, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, 2000.

[11] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2003.

[12] O. Jokinen, "Self-calibration of a light striping system by matching multiple 3-d profile maps," in *International conference on 3-D digital imaging and modeling*, ser. 3DIM'99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 180–190.

[13] A. M. McIvor, "Calibration of a laser stripe profiler," in *International conference on 3-D digital imaging and modeling*, ser. 3DIM'99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 92–98.

[14] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *International Conference on Intelligent Robots and Systems*, ser. IROS '04, vol. 3, Oct. 2004, pp. 2301–2306.

[15] G. H. Golub and C. F. Van Loan, *Matrix computations (3rd ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[16] R. Dupont, R. Keriven, and P. Fuchs, "An improved calibration technique for coupled single-row telemeter and ccd camera," in *International Conference on 3-D Digital Imaging and Modeling*, ser. 3DIM '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 89–94.