Extrinsic calibration of a set of range cameras in 5 seconds without pattern

Eduardo Fernandez-Moral, Javier González-Jiménez, Patrick Rives, Vicente Arévalo

Extrinsic calibration of a set of range cameras in 5 seconds without pattern

Eduardo Fernández-Moral[1], Javier González-Jiménez[1], Patrick Rives[2] and Vicente Arévalo[1]

***Abstract*— The integration of several range cameras in a mobile platform is useful for applications in mobile robotics and autonomous vehicles that require a large field of view. This situation is increasingly interesting with the advent of low cost range cameras like those developed by Primesense. Calibrating such combination of sensors for any geometric configuration is a problem that has been recently solved through visual odometry (VO) and SLAM. However, this kind of solution is laborious to apply, requiring robust SLAM or VO in controlled environments. In this paper we propose a new uncomplicated technique for extrinsic calibration of range cameras that relies on finding and matching planes. The method that we present serves to calibrate two or more range cameras in an arbitrary configuration, requiring only to observe one plane from different viewpoints. The conditions to solve the problem are studied, and several practical examples are presented covering different geometric configurations, including an omnidirectional RGB-D sensor composed of 8 range cameras. The quality of this calibration is evaluated with several experiments that demonstrate an improvement of accuracy over design parameters, while providing a versatile solution that is extremely fast and easy to apply.**

I. INTRODUCTION

The extrinsic calibration of different sensors is of very practical interest in robotics. This problem has been widely studied and different solutions have been presented for a variety of sensor configurations [1], [2], [3], [4]. In this paper we address the problem of extrinsic calibration of range cameras. This problem has gained importance in the last years, with the arrival of both low cost range and range-intensity cameras. In this context, the available methods for extrinsic calibration present mainly two types of disadvantages: they have restrictions on the camera positioning (e.g. requirement of overlapping), or they rely on the tracking of the camera trajectory, which can be tedious to obtain, besides having issues of robustness and accuracy.

The disadvantages of previous calibration approaches were clear after the construction of a device for omnidirectional intensity and range image acquisition based on a rig of RGB-D cameras (figure 5) [5]. This new sensor prompted in us the need of a robust and easy calibration method, since the accuracy of the parameters from the construction design were not satisfactory, and the solutions proposed in the literature were not suitable for our problem.

To put our work into context, we review first some relevant approaches to this problem. A classical strategy for extrinsic camera calibration is through the detection and matching of control points that are detected in the overlapping regions of the different cameras [6]. However, the overlap requirement constitutes a very strong constraint. Besides, even when some overlap exists, it is generally more complicated to match features in range images than in intensity images. With a different perspective, the use of a calibration pattern is a resource that has been employed as an *ad hoc* solution for very specific problems [2], [7]. The lack of generality of this solution for any configuration of cameras is indeed an important limitation. Plus the need to create the 3D calibration pattern itself is wearisome.

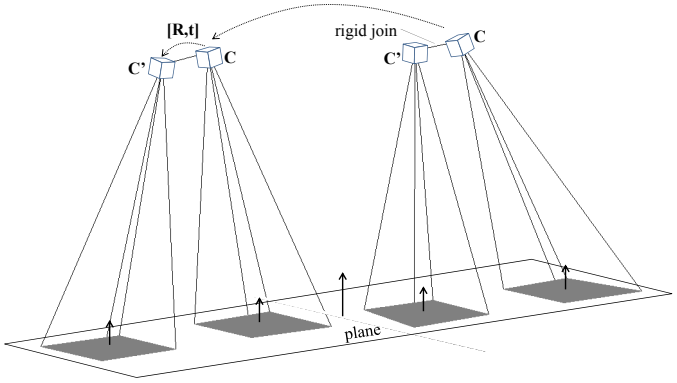

Fig. 1. A planar surface is observed by two range cameras rigidly joined from different positions. In this way, plane correspondences are captured from different orientations with respect to the rig's reference system to perform extrinsic calibration.

A more general approach which does not depend on the camera set-up is based on ego-motion to match the camera trajectories, which are tracked independently. For that, simultaneous localization and mapping (SLAM) or visual odometry (VO) techniques are applied [8], [3], [4]. Such approaches rely on the SLAM or VO robustness, which depends highly on the environment, especially for range-only cameras. Besides, obtaining a useful trajectory is far more tedious and inconvenient than taking a few images from different positions as the technique we propose here.

Concretely, we present a new method for extrinsic calibration of range cameras (including RGB-D cameras) that avoids the above-mentioned problems. Our method relies on matching planes that can be observed simultaneously from different cameras. For that, only one plane has to be observed from different camera locations. This approach has several advantages as the calibration can be performed very quickly and robustly, it does not require any calibration pattern but

[1]Mapir Group. Universidad de Málaga. E.T.S. Ingeniería de Informática-Telecomunicación. Campus de Teatinos, 29071, Málaga, Spain. E-mail: eduardofernandez@uma.es

[2]Lagadic team. INRIA Sophia Antipolis - Méditerranée. 2004 route des Lucioles - BP 93, 06902 Sophia Antipolis, France.

a single plane from the environment (the floor, the ceiling, a wall, ...), and supervision is not required. In this work, we test the performance of the method by calibrating two typical configurations of range cameras, demonstrating very satisfactory results in all the cases.

In the following we give the details of our calibration approach, the segmentation and parameterization of the planes and their matching. The observability of the problem is studied next. Then, the equations for extrinsic calibration are derived for a pair of cameras (section IV) and for an arbitrary number of cameras (section VI). For both cases, calibration results for different camera configurations are presented. Finally, the conclusions are outlined.

II. CALIBRATION APPROACH

The problem that we address in this paper is that of finding the extrinsic calibration (i.e. relative poses) between several "a-priori" non overlapping range cameras. We propose to solve this problem by matching planar features that are seen from different viewpoints. Thus, we take advantage of the fact that structured environments contain large planes (e.g. the floor, walls, ceiling) that can be reliably observed by the different sensors simultaneously. We make use of such planes to establish correspondences, see figure 1. With this strategy we avoid the need of creating an specific calibration pattern for the sensor set-up. Also, no SLAM neither odometry are needed, avoiding robustness issues and making the procedure much more accessible and easy to use.

Before addressing the extrinsic calibration itself, related issues like the intrinsic calibration of the range cameras, and the plane segmentation, parameterization and matching are briefly described next.

A. Intrinsic model of the range cameras

Measurements from range cameras are affected by noise, which can be classified in two types of errors: in accuracy (variance) and also a bias or correctness (mean error). It has been reported elsewhere [9] that the depth error variance as well as the bias of structure-light sensors from PrimeSense (including Kinect and Asus Xtion) increases with distance. Such bias is quite evident when we see the deformation of a flat surface as it is observed from increasing distance. This intrinsic error is studied in [10], where the authors propose to model the intrinsic parameters using a discrete image of multipliers to update every pixel depending on its depth. An implementation of that work, which relies on a cutting edge SLAM solution [11] is publicly available[1]. This correction has been applied to the Asus XPL sensors employed here. Thus, in the rest of this paper we will assume that they are only affected by unbiased, uncorrelated Gaussian noise.

B. Plane segmentation and parameterization

In order to obtain planes (planar patches to be precise), the depth images are segmented with a region growing approach [12]. This technique is used here due to its efficiency to segment organized images, however other methods for plane segmentation can be used equally [13], [14].

A planar patch is represented by its normal vector $\mathbf{n}$, with $\|\mathbf{n}\| = 1$, and the distance d to the optical center of the camera. In this way, a point $\mathbf{p}$ lying on the plane fulfills the equation

$$\mathbf{n} \cdot \mathbf{p} + d = 0 \tag{1}$$

This overparameterization is very convenient for the formulation of the calibration errors in the next sections.

The plane parameters and their covariances are estimated following [15], assuming accurate directions of measurements $\mathbf{m}_i$, where the noise only affects the range measurements ρ_i. After the intrinsic calibration has been performed, we can assume that $\rho_i \sim N(\hat{\rho}_i, \sigma_i)$, where $\hat{\rho}_i = d/\mathbf{n} \cdot \mathbf{m}_i$ is the true range of the i-th measurement. The standard deviation σ_i is generally a function that depends on the range ρ_i and on the incidence angle $\sigma(\rho_i, \mathbf{n} \cdot \mathbf{m}_i)$. However, in this work we make the same simplification as in [15] to assume the standard deviation σ_i independent on $\{\mathbf{n}, d\}$, and estimate σ_i in a conservative way: $\sigma(\rho_i, \mathbf{n} \cdot \mathbf{m}_i) < \sigma$. From this simplification the plane parameters and their covariances can be analytically defined. Thus, the optimal $\mathbf{n}^*$ is the eigenvector corresponding to the smallest eigenvalue of the matrix

$$M = \sum_{i=1}^{N} (r_i - r_G)(r_i - r_G)^T \tag{2}$$

where $r_G = \frac{1}{N} \sum_{i=1}^{N} r_i$ is the gravity center of the plane pixels. The optimal d^* is given by

$$d^* = \mathbf{n}^* r_G \tag{3}$$

and the covariance of the plane parameters $\Sigma^* = (H)^+$ is calculated as the Moore-Penrose generalized inverse of

$$H = \frac{1}{\sigma^2} \sum_{i=1}^{N} \begin{bmatrix} r_i r_i^T & -r_i \\ -r_i^T & 1 \end{bmatrix} \tag{4}$$

The simplification of considering constant variance (i.i.d.) assumed above can be substituted for a more realistic model [16] to obtain more accurate results. But this requires a complex numeric calculation of the plane parameters and their covariances, which is out of the scope of this paper.

C. Obtaining plane correspondences

To establish the plane correspondences, a rough guess of the camera relative poses is provided by the user. Then, plane correspondences are established between each pair of sensors (if more than two are being calibrated) applying some simple heuristics to constraint their geometric consistency, namely: a) the angle between the normal vectors of both planar regions is smaller than a threshold; b) the distances from both planes to the camera center are under a threshold; and c) both regions are large enough (in our implementation: the number of pixels in each region is bigger than 20% of the image pixels). Note that giving an initial estimation for the relative position of the cameras is an easy and straightforward way to facilitate the matching of plane observations, though

[1] http://cs.stanford.edu/people/teichman/octo/clams/

other plane matching strategies can be applied if we want to avoid the need of this initial estimate [17], [18].

The process for gathering correspondences is performed automatically while the camera rig is moving until the problem is well conditioned according to the Fisher Information Matrix, as explained in section III. For that, having a large number of correspondences contributes to reduce the error in calibration due to sensor noise. Also, to be robust against possible outliers, after enough correspondences have been gathered, a RANSAC [19] process is launched to detect and discard possible outliers.

The range cameras synchronization effect is neglected in this work since the images are captured at a minimum frame rate of 30 Hz, and the camera rigs are never moved abruptly.

D. RANSAC outlier rejection

The set of plane correspondences obtained is cleaned up of outliers by running RANSAC. This procedure is carried out in two steps: first, the outliers showing a large error in the orientation are discarded, and second, those outliers in distance are removed (this order is chosen since the noise in the orientation of the normal vectors is typically smaller than that in the plane position). The relative poses between the pair of cameras are calculated from a sample of 3 non-degenerate plane correspondences (section III). The models used by the two RANSAC processes are defined in section IV.

III. OBSERVABILITY

It can be seen that each plane correspondence imposes three new constraints between the pair of sensors: two for the relative rotation and one for the relative translation. Thus, we need at least 3 measurements from linearly independent plane observations (i.e the normal vectors of the planes are linearly independent) to compute the relative position of a pair of sensors (only two measurements are needed to compute the rotation), and a minimum of $3(N-1)$ correspondences to calibrate a rig with N sensors. An example to get insight into this is by considering that we have a single sensor observing the corner of a room. The observation of the three perpendicular planes gives us enough information to localize the camera and its relative motion with respect to a previous pose. In analogy, the relative pose between two cameras can be obtained if they observe 3 plane correspondences with linearly independent normal vectors (either observed from one view (fig. 2) or several ones (fig. 1)). The most simple and convenient procedure would be to take a short sequence of several images (in motion) of one big plane from different orientations of the rig (see the video attached).

In order to detect the degenerate cases for which the calibration cannot be calculated we evaluate the Fisher Information Matrix (FIM) for the parameters of the maximum likelihood estimator (MLE) of the calibration, which is presented in the following section. As we will see, the probability of the MLE is given by an unbiased Gaussian distribution (this assumption is realistic only after intrinsic correction). For this estimator (called efficient [20]), the FIM

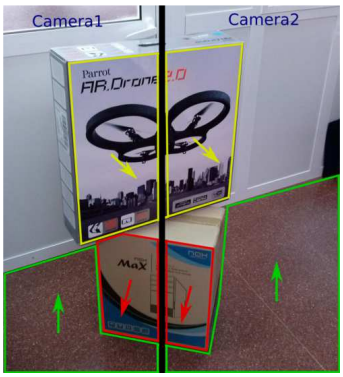

Fig. 2. A particular set-up from which we can calibrate the cameras with a single observation. The planar patches on the left and right are those extracted from the two cameras.

coincides with the Hessian of the least squares problem resulting from the MLE, and its inverse is the covariance of the resulting calibration (Cramér-Rao [20]).

When the FIM is singular, the information provided is not sufficient and the MLE does not exist. For a pair of cameras, it can be verified that when the FIM is not singular, then

$$rank(\sum_{i=1}^{N} \mathbf{n}_i \mathbf{n}_i^T) = 3 \tag{5}$$

where n_i is the normal vector of the plane i as seen from one of the cameras in the pair.

The ratio $\eta = \mu_3/\mu_1$ between the smallest and largest eigenvalues of the matrix from eq. 5 is also an indicator of how well distributed the measurements are along the different directions of the space. So, in the best case $\eta = 1$ which means that all plane observations are equally distributed in the space, while when $\eta \to 0$, the system is ill conditioned.

From our experiments, we have verified that the covariance of both the rotation and the translation estimations decrease asymptotically as the number of plane correspondences increases. The covariance is used as the condition to control the calibration convergence, and hence, to stop gathering plane correspondences. In our tests, we stop this calibration when the maximum eigenvalue of the covariance is under 10^{-3}, which has shown to be a good compromise between accuracy and effort to obtain plane correspondences.

IV. EXTRINSIC CALIBRATION OF A PAIR OF RANGE CAMERAS

A. Problem statement

Given a set of plane correspondences gathered from two rigidly joint range cameras C and C', where the camera C represents the system of reference, and C' is located with a relative transformation $[\mathrm{R}|\mathbf{t}] \in \mathbb{SE}(3)$ with respect to C, where the rotation $\mathrm{R} \in \mathbb{SO}(3)$ is represented with a 3×3 matrix and the translation $\mathbf{t} \in \mathbb{R}^3$. Provided that the correspondences fulfill the observability condition (concretely the one represented by eq. 5), we want to estimate the optimal $[\mathrm{R}|\mathbf{t}]$ assuming that the measurements are affected by unbiased Gaussian noise as modelled in II-B. This problem can be

divided into two separate ones since the rotation and the translation restrictions are decoupled.

B. Solving for the rotation

The maximum likelihood estimation (MLE) of the relative rotation R is given by the maximization of the log-likelihood

$$\underset{\mathrm{R}}{\operatorname{argmax}} \left(\ln \prod_{i=1}^{N} p(\mathbf{n}_i, \mathbf{n}'_i | \mathrm{R}) \right) \tag{6}$$

for N plane correspondences, where the likelihood of the rotation for the i-th correspondence is expressed as

$$p(\mathbf{n}_i, \mathbf{n}'_i | \mathrm{R}) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma_i|}} \exp\left(-\frac{1}{2} (\mathbf{n}_i - \mathrm{R}\mathbf{n}'_i)^T \Sigma_i^{-1} (\mathbf{n}_i - \mathrm{R}\mathbf{n}'_i) \right) \tag{7}$$

being n_i and n'_i the observed normal vectors from the plane i as seen by the cameras C and C' respectively; R is the rotation matrix in $\mathbb{SO}(3)$, and Σ_i is the 3×3 covariance block corresponding to the normal vector of the plane correspondence (calculated from the fusion of both observations [16]). Considering independent errors of the plane correspondences, the derivation of this MLE coincides with the solution of the least squares problem expressed as

$$\underset{\mathrm{R}}{\operatorname{argmin}} \sum_{i=1}^{N} \omega_i \| \mathbf{n}_i - \mathrm{R}\mathbf{n}'_i \|^2 \tag{8}$$

where ω_i is the weight of the plane correspondence

$$\omega_i = \frac{1}{|\Sigma_i|} \tag{9}$$

This problem is similar to the one of estimating the rotation of a registered set of 3D points [21]. Thus, employing the same procedure, the above equation can be expressed as

$$\begin{aligned} \mathrm{R} &= \underset{\mathrm{R}}{\operatorname{argmin}} \left(\sum_{i=1}^{N} \omega_i \mathbf{n}_i^T \mathbf{n}_i - 2 \sum_{i=1}^{N} \omega_i \mathbf{n}_i^T \mathrm{R} \mathbf{n}'_i + \sum_{i=1}^{N} \omega_i \mathbf{n}_i'^T \mathbf{n}'_i \right) \\ &= \underset{\mathrm{R}}{\operatorname{argmin}} \left(-2 \sum_{i=1}^{N} \omega_i \mathbf{n}_i^T \mathrm{R} \mathbf{n}'_i \right) \\ &= \underset{\mathrm{R}}{\operatorname{argmax}} \sum_{i=1}^{N} \omega_i \mathbf{n}_i^T \mathrm{R} \mathbf{n}'_i \end{aligned} \tag{10}$$

that can be denoted as

$$\sum_{i=1}^{N} \omega_i \mathbf{n}_i'^T \mathrm{R} \mathbf{n}_i = tr(W Y^T R X) \tag{11}$$

where $tr(\cdot)$ represents the trace of a square matrix; $W = diag(\omega_1, ..., \omega_n)$ is an $n \times n$ diagonal matrix containing the weights ω_i; and Y and X are $3 \times n$ matrices with the normal vectors $\mathbf{n}'_i$ and $\mathbf{n}_i$ as their columns, respectively. This problem is solved with singular value decomposition (SVD) over the 3×3 covariance matrix

$$S = X W Y^T \tag{12}$$

From the singular value decomposition $S = U \Sigma V^T$, the rotation is obtained as

$$\mathrm{R} = V \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & det(VU^T) \end{pmatrix}}_{A} U^T \tag{13}$$

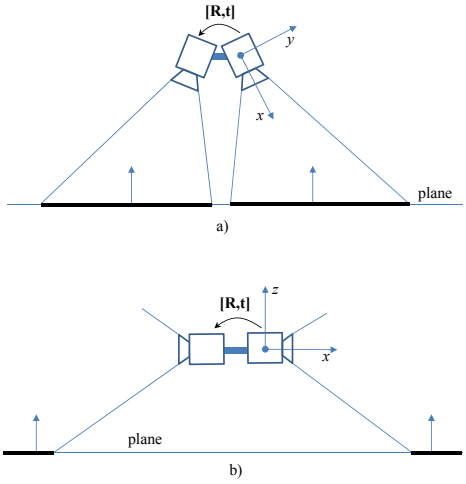

Fig. 3. Different sensor configurations with a pair of cameras: a) Adjacent cameras, b) Opposite cameras.

where the matrix A is used to convert the degenerate case of a reflection ($det(VU^T) = -1$) into a valid rotation in $\mathbb{SO}(3)$. For further details on the mathematics, please refer to [21], [22].

C. Solving for the translation

The MLE of the translation is obtained by maximizing the log-likelihood associated to the probability

$$p(\mathbf{n}_i, \mathbf{n}'_i, d_i, d'_i | \mathbf{t}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \frac{(d_i - d'_i + \mathbf{n}_i \cdot \mathbf{t})^2}{\sigma_i^2} \right) \tag{14}$$

where d_i and d'_i are the observed distances from the plane i to the optical centers of the depth cameras C and C' respectively, σ_i^2 is the error variance, and $\mathbf{t}$ is the relative translation we are looking for. This is equivalent to the least squares problem

$$\underset{t}{\operatorname{argmin}} \sum_{i=1}^{N} \omega_i (d_i - d'_i + \mathbf{t} \cdot \mathbf{n}_i)^2 \tag{15}$$

with the weight given by $\omega_i = 1/\sigma_i^2$. This has a closed form solution given by

$$\mathbf{t} = -H^{-1} g \tag{16}$$

where H and g are the Hessian and the Gradient of the error function respectively, which are calculated as

$$H = \sum_{i=1}^{N} J_i^T W_i J_i \ , \quad g = \sum_{i=1}^{N} J_i^T W_i \mathbf{r}_i \tag{17}$$

where the Jacobians, the weights and the residuals are calculated from

$$J_i = n_i^T \ , \quad W_i = \frac{1}{\sigma_i^2} \ , \quad \mathbf{r}_i = d_i - d'_i \tag{18}$$

D. Practical study cases

1. Adjacent cameras

This case is interesting to provide a larger field of view of the scene, being specially practical for low cost sensors like Asus Xtion (see figure 3.a). This case serves us to illustrate the conditioning of the problem, and so to show different possibilities for calibration. One of this situations is the calibration of the pair from one single observation, i.e. without moving the rig. This is only possible if three planar patches whose normal vectors span through the different directions of the space are visible at the same time by both cameras. This case can be easily set-up, as the example shown in figure 2.

In practice, however, it is even more convenient to take several images from different orientations pointing to one single plane (the floor, for example), since we can gather more quickly enough plane correspondences that help to reduce the error from the measurement noise. This may not take longer than 2 or 3 seconds.

In table I we show an example of how the average residual error is reduced when raising the number of plane correspondences. The alignment errors in rotation and translation are measured in a dataset containing 2K correct plane correspondences for the five calibrations, the plane correspondences were taken in all directions of the space.

TABLE I

RESIDUAL ERRORS FOR DIFFERENT CALIBRATIONS USING A DIFFERENT NUMBER PLANE CORRESPONDENCES.

Correspondences	Av rot error (deg)	Av trans error (cm)
3	1.12	1.89
10	0.68	1.01
30	0.52	0.82
60	0.49	0.74
100	0.49	0.61

2. Cameras in opposite directions

This case is interesting, for instance, for vehicles that need to observe the scene forward and backward. We address this case here also since it probably represents the most challenging case to obtain plane correspondences in different directions (notice that the further the viewing directions of the cameras are, the more difficult is to find plane correspondences). Figure 3.b shows how the plane correspondences can still be obtained to add constraints in the different directions of the space, for example, by rotating the camera rig. Calibration was performed automatically while the user waved the camera near the floor. After 5 seconds from the start of the experiment, the calibration finishes with 29 plane correspondences, see the attached video. In this case the deviation with respect to the rig parameters is less than 1 deg for the rotation, and in the order of millimeters for the translation.

3. Sensors of different types

Though most of our experiments are carried out with structured light Primesense cameras, other range sensors can also be calibrated with our method. Concretely, a time-of-flight camera and a Kinect sensor mounted on a robot are calibrated by moving the robot (figure 4, left) around to gather plane correspondences. The errors in the plane observations from both sensors will follow different distributions, so that they are weighted accordingly as said in section II-B.

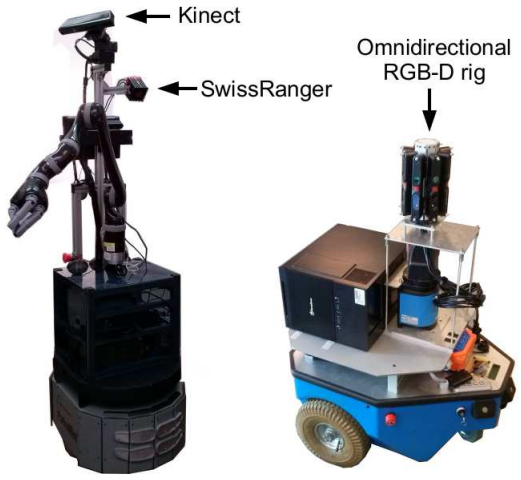

Fig. 4. Robots which mount rigs of range and RGB-D cameras.

V. EXTRINSIC CALIBRATION OF AN ARBITRARY NUMBER OF RANGE CAMERAS

This section extends the previous formulation for an arbitrary number M of range cameras. Note that for the case when there are no loop closures between the sensors, i.e. there is only one possible way to correlate the relative pose of any pair of sensors. The extrinsic calibration can be calculated as in the previous section by estimating the relative pose between each pair of adjacent sensors, and performing pose composition to place them in a common reference. Instead, this section is dedicated to the particular case in which there are plane correspondences that create loop closures between sensors. For the sake of space, we present directly the least squares equations, which as in the previous section, derive from the ML estimation.

A. Solving for the rotation

The relative rotation between the different sensors can be formulated as

$$\underset{[\cancel{\mathrm{R}_1}\mathrm{R}_2 \ldots \mathrm{R}_M]}{\operatorname{argmin}} \sum_{j=1}^{M} \sum_{k=j+1}^{M} \sum_{i=1}^{N} \lambda_i(j,k)\omega_i(j,k) \left\| \mathrm{R}_j \mathbf{n}_i^j - \mathrm{R}_k \mathbf{n}_i^k \right\|^2 \quad (19)$$

where j and k are indices of the M sensors and i is the index of each one of the N planes observed; $\lambda_i(j,k)$ is a binary variable that equals 1 when the plane i is observed by sensors j and k, being 0 otherwise; $\mathbf{n}_i^j$ and $\mathbf{n}_i^k$ are the normal vectors of the plane i observed from sensors j and k, respectively; and the rotation of each sensor is represented by $\mathrm{R}_1, \mathrm{R}_2 \ldots \mathrm{R}_M \in \mathbb{SO}(3)$, where the rotation R_1 has been arbitrarily fixed to the identity I_3 to choose a system of reference.

This least squares system has a different structure from the one in the previous section, that can not be solved with

Fig. 5. Omnidirectional RGB-D camera rig.

the strategy used from equations 10 to 13. Thus, we rewrite the problem to represent the relative rotations in minimal parameterization with the exponential map from Lie algebra [23]

$$\underset{\mu_2 \cdots \mu_M}{\operatorname{argmin}} \sum_{j=1}^{M} \sum_{k=j+1}^{M} \sum_{i=1}^{N} \lambda_i(j,k)\omega_i(j,k) \| e^{\mu_j} \mathrm{R}_j \mathbf{n}_i^j - e^{\mu_k} \mathrm{R}_k \mathbf{n}_i^k \|^2 \tag{20}$$

where e^{μ_j} is the exponential map of the increment of rotation μ_j on the rotation R_j, which is a 3D vector on a manifold space of $\mathbb{SO}(3)$, and similarly for e^{μ_k}.

This is a non-linear least squares system that is solved iteratively with Gauss-Newton using equations 16-17, where the Jacobian and the vector of residuals are given by

$$J_i = [0 \ \ldots \ 0 \ J_i^{(j)} \ 0 \ \ldots \ 0 \ J_i^{(k)} \ 0 \ \ldots \ 0]$$

$$J_i^{(j)} = skew(\mathbf{n}_i^j) \ , \ J_i^{(k)} = skew(-\mathbf{n}_i^k) \tag{21}$$

thus, the Hessian H and the gradient $\mathbf{g}$ are calculated incrementally as

$$H = \sum_{i=1}^{N} H_i \ , \quad \mathbf{g} = \sum_{i=1}^{N} \mathbf{g}_i \tag{22}$$

which have the form

$$H_i = \begin{bmatrix} J_i^{jT} \omega_i J_i^j & & J_i^{kT} \omega_i J_i^j \\ & \ddots & \\ J_i^{jT} \omega_i J_i^k & & J_i^{kT} \omega_i J_i^k \end{bmatrix} , \ \mathbf{g}_i = \begin{bmatrix} J_i^{jT} \omega_i r_i \\ \vdots \\ J_i^{kT} \omega_i r_i \end{bmatrix} \tag{23}$$

B. Solving for the translation

The generalization of the translation equation system keeps the same structure as in the case of a pair of cameras

$$\underset{[t_1 t_2 \cdots t_M]}{\operatorname{argmin}} \sum_{j=1}^{M} \sum_{k=j+1}^{M} \sum_{i=1}^{N} \lambda_i(j,k)\omega_i(j,k)(d_i^j - d_i^k - \mathbf{t}_j \mathrm{R}_j \mathbf{n}_i^j + \mathbf{t}_k \mathrm{R}_k \mathbf{n}_i^k)^2 \tag{24}$$

and it is solved in the same way (note that in this case the rotation must be solved first).

Fig. 6. Omnidirectional RGB and depth images acquired by our RGB-D camera rig.

VI. CALIBRATION OF A RIG FOR OMNIDIRECTIONAL IMAGE ACQUISITION

We have designed a camera rig for omnidirectional RGB-D acquisition which comprises 8 Asus Xtion Pro Live (Asus XPL) sensors mounted in a radial configuration (see figure 5). This device motivated at the origin the work described in this paper, since the parameters from the construction design were not accurate for our application. Existing calibration approaches like those based on SLAM [8] are very time consuming and impose important restrictions on the trajectory, since planar movement (as we have in our robot) is a degenerate case where calibration cannot be achieved. Thus, we employed the calibration method described in the previous section, which was applied on a sequence of images taken with the robot (planar movement is not a degenerate case in our approach).

The relative positions between the RGB cameras is the same as these between their corresponding depth cameras for our sensor configuration. Therefore, both RGB and depth omnidirectional images can be built, see figure 6. The 3D point cloud can be also built from such images, figure 7.

The precision of calibration is tested with an experiment where the robot moves in a small circular trajectory ($\varnothing \sim$ 0.5 m) in the center of a room, taking 200 images. In table II, the average residual in orientation and translation for the plane correspondences of these images is presented for different extrinsic calibrations: design parameters (no extrinsic calibration) with and without intrinsic correction, and the extrinsic calibration also with and without intrinsic correction. The ICP residual of the alignment of the spherical point clouds from these images is also shown. For all the cases, the combination of intrinsic and extrinsic calibration offers the best results.

TABLE II

RESIDUAL ERRORS FOR DIFFERENT COMBINATIONS OF INTRINSIC AND EXTRINSIC CALIBRATIONS.

*C*alib / Error type	*R*es. rot (deg)	*R*es. trans (cm)	*R*es ICP (cm)
Design Specs	3.17	3.0	0.49
Design S.+Intrinsic	2.95	3.1	0.45
Extrinsic calib	1.78	2.9	0.34
Extrinsic+Intrinsic	1.60	2.5	0.29

VII. CONCLUSIONS

A new methodology for calibrating the extrinsic parameters of range camera rigs has been presented in this paper.

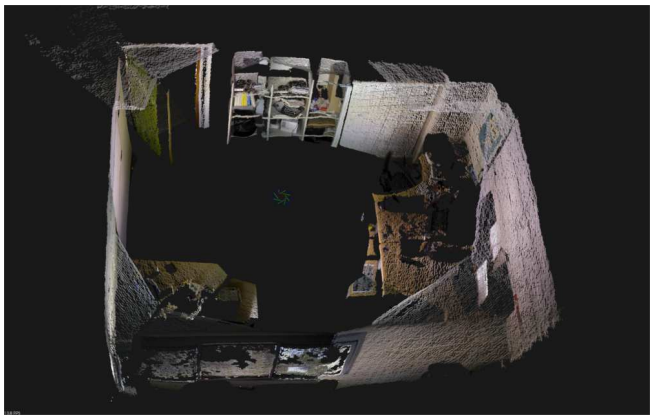

Fig. 7. Point cloud obtained from the reconstructed omnidirectional RGB-D image.

The method relies on the matching of plane observations from the different sensors. No constraints are put on the position of the cameras, where the only requirement for the system is that there is a planar surface that can be observed simultaneously. The observability conditions are analyzed, and a solution is presented based on MLE. With our method, performing calibration becomes very fast and easy for the user, avoiding problems of previous solutions which rely either on calibration patterns or trajectory estimation methods. The method has been tested for different configurations of cameras, including a camera rig designed for omnidirectional image acquisition. All the experiments have validated the claimed features of our proposal.

REFERENCES

[1] 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 (IROS 2004)*, vol. 3. IEEE/RSJ, 2004, pp. 2301–2306.

[2] J.-E. Ha, "Extrinsic calibration of a camera and laser range finder using a new calibration structure of a plane with a triangular hole," *International Journal of Control, Automation and Systems*, vol. 10, no. 6, pp. 1240–1244, 2012. [Online]. Available: http://dx.doi.org/10.1007/s12555-012-0619-7

[3] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *International Conference on Intelligent Robots and Systems (IROS 2013)*. IEEE/RSJ, 2013, pp. 1793–1800.

[4] S. Schneider, T. Luettel, and H.-J. Wuensche, "Odometry-based online extrinsic sensor calibration," in *International Conference onIntelligent Robots and Systems (IROS 2013)*. IEEE/RSJ, 2013, pp. 1287–1292.

[5] T. Gokhool, M. Meilland, P. Rives, and E. Fernández-Moral, "A Dense Map Building Approach from Spherical RGBD Images," in *International Conference on Computer Vision Theory and Applications (VISAPP 2014)*, Lisbon, Portugal, January 2014.

[6] R. Szeliski and H.-Y. Shum, "Creating full view panoramic image mosaics and environment maps," in *International Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 251–258.

[7] F.-A. Moreno, J. Gonzalez-Jimenez, J.-L. Blanco, and A. Esteban, "An instrumented vehicle for efficient and accurate 3d mapping of roads," *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 6, pp. 403–419, 7 2013.

[8] J. Brookshire and S. J. Teller, "Extrinsic calibration from per-sensor egomotion." in *Robotics: Science and Systems*, 2012.

[9] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.

[10] A. Teichman, S. Miller, and S. Thrun, "Unsupervised intrinsic calibration of depth sensors via slam," in *Robotics: Science and Systems*, Berlin, Germany, June 2013.

[11] C. Kerl, J. Sturm, and D. Cremers, "Dense visual slam for rgb-d cameras," in *International Conference on Intelligent Robots and Systems (IROS 2013)*. IEEE/RSJ, 2013.

[12] D. Holz and S. Behnke, "Fast range image segmentation and smoothing using approximate surface reconstruction and region growing," in *Intelligent Autonomous Systems 12*. Springer, 2013, pp. 61–73.

[13] M. Zuliani, C. Kenney, and B. Manjunath, "The multiransac algorithm and its application to detect planar homographies," *IEEE International Conference on Image Processing*, 2005.

[14] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter, "The 3d hough transform for plane detection in point clouds: A review and a new accumulator design," *3D Research*, vol. 2, no. 2, pp. 1–13, 2011.

[15] J. Poppinga, N. Vaskevicius, A. Birk, and K. Pathak, "Fast plane detection and polygonalization in noisy 3d range images," in *International Conference on Intelligent Robots and Systems (IROS 2008)*. IEEE/RSJ, 2008, pp. 3378–3383.

[16] K. Pathak, N. Vaskevicius, and A. Birk, "Uncertainty analysis for optimum plane extraction from noisy 3d range-sensor point-clouds," *Intelligent Service Robotics*, vol. 3, no. 1, pp. 37–48, 2010.

[17] K. Pathak, A. Birk, N. Vaskevicius, and J. Poppinga, "Fast registration based on noisy planes with unknown correspondences for 3-d mapping," *IEEE Transactions on Robotics*, vol. 26, no. 3, pp. 424–441, 2010.

[18] E. Fernández-Moral, W. Mayol-Cuevas, V. Arévalo, and J. González-Jiménez, "Fast place recognition with plane-based maps," in *International Conference on Robotics and Automation (ICRA 2013)*. IEEE, 2013.

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[20] J.-A. Fernández-Madrigal and J. L. B. Claraco, *Simultaneous Localization and Mapping for Mobile Robots: Introduction and Methods*. Information Science Reference, 2013.

[21] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 5, pp. 698–700, 1987.

[22] O. Sorkine, "Least-squares rigid motion using svd," *Technical notes*, vol. 120, 2009.

[23] J.-L. Blanco, "A tutorial on se(3) transformation parameterizations and on-manifold optimization," *University of Malaga, Tech. Rep*, 2010.