

# Jornadas de Automática

## A plane-based method for extrinsic calibration of RGB and depth cameras

Aigner, D. X.<sup>a,\*</sup>, Gonzalez-Jimenez, J.<sup>b</sup>

<sup>a</sup>Faculty of Mathematics and Economics, Ulm University, Germany. Machine Perception and Intelligent Robotics (MAPIR), Universidad de Málaga, Spain.

<sup>b</sup>Machine Perception and Intelligent Robotics (MAPIR). Dpto. de Ingeniería de Sistemas y Automática. Instituto Universitario de Investigación en Ingeniería Mecatrónica y Sistemas Ciberfísicos (IMECH.UMA), Universidad de Málaga, Spain.

**To cite this article:** Aigner, D. X., Gonzalez-Jimenez, J., 2024. A plane-based method for extrinsic calibration of RGB and depth cameras. *Jornadas de Automática*, 45. <https://doi.org/10.17979/ja-cea.2024.45.10928>

### Resumen

Este artículo presenta un método, implementado en ROS2, para la calibración extrínseca de un conjunto heterogeneo de cámaras que incluyen tanto RGB como de profundidad. El método propuesto estima las poses relativas entre dichas cámaras a partir de la observación de planos. Para ello, en primer lugar, se extraen y emparejan los vectores normales de las superficies planas de las imágenes (RGB y de profundidad). En segundo lugar, se plantea un problema de optimización que estima las rotaciones y traslaciones que minimizan los errores entre los pares de vectores normales en correspondencia. La aplicación utiliza algoritmos disponibles en librerías estándar para la extracción de planos (OpenCV, PCL) y optimización (Eigen). La eficacia y precisión del método se ilustran en una configuración con dos cámaras RGB y una cámara de profundidad.

**Palabras clave:** detección, integración de sensores y percepción, tecnología robótica, percepción y detección, información y fusión sensorial

### Abstract

This article presents a method implemented in ROS2 for the extrinsic calibration of camera rigs that combine heterogeneous sensors, including any combination of RGB and depth cameras. The proposed method estimates the relative poses between sensors using a two-step process. First, normal vectors of planar surfaces are extracted from RGB or depth images and matched across sensors. Second, an optimization problem determines the relative rotations and translations between all of them by minimizing errors between corresponding normal vector pairs. The implementation utilizes off-the-shelf algorithms for plane extraction (OpenCV, PCL) and optimization (Eigen). Experimental results on a setup with two RGB cameras and one depth camera demonstrate the effectiveness of the proposed approach.

**Keywords:** sensing, sensor integration and perception, robotics technology, perception and sensing, information and sensor fusion

## 1. Introduction

Mixing up different sensor types has turned out to be very advantageous in a variety of applications. This way, the downsides of a specific technology can be compensated by setting up heterogeneous systems. Accurate extrinsic calibration, which yields the relative pose between the sensors' coordinate systems, is essential for achieving precise and reliable measurements with these configurations. A widely used setup are depth-color sensor combinations. In particular,

we will discuss arbitrary combinations comprising structured-light-based RGB-Depth (RGB-D) cameras alongside RGB cameras, found in applications ranging from robotics to autonomous driving and 3D mapping. However, the variety of possible sensor combinations and the different data types present challenges in the extrinsic calibration of such systems. Specifically, this complicates the development of a general solution that can be widely applied without requiring a new approach for each unique sensor setup. Therefore, the primary

challenge addressed in this study is the calibration of heterogeneous sensor configurations that include depth and RGB cameras, as most existing calibration methods are limited to a specific setup. For this reason, we propose a method, implemented within a ROS2 environment, that allows for flexible calibration of sensor combinations, addressing both homogeneous and mixed sensor setups handling image and point cloud data.

The methodology applied to the described problem consists of splitting the problem into two parts. The first ROS2 node is responsible for:

1. time-synchronizing the incoming messages in pairs.
2. detection of planes in both the depth and RGB images.
3. calculation of plane correspondences based on a rough initial estimation of the sensor positions.
4. publishing the plane correspondences in a custom message format.

The second node is configured to receive the aforementioned messages, accumulating them and executing the calibration based on user-specified parameters and the plane data.

The remainder of this paper is organized as follows: Section 2 reviews related work discussing similar methods. Section 3 details the methodology, including the data processing and the calibration. Section 4 presents the experimental setup. Finally, Section 5 analyzes and discusses the findings, followed by the conclusions of the research in Section 6.

## 2. Related work

The extrinsic calibration related homogeneous and heterogeneous systems applied to depth and/or RGB cameras is a well-explored area with a variety of different approaches. In particular, we overview methods regarding homogeneous RGB-D and range camera systems as well as mixed depth-color sensor systems.

### 2.1. Homogeneous systems

Methods about the extrinsic calibration of homogeneous sensor systems are a well explored field. Hence, a big variety of methods exists for different sensor types. For multiple RGB cameras, Zhang (2000) is widely known and used, especially when it comes to 2D-object-based techniques. Compared to this type of method, self-calibration techniques don't require specific calibration environments (Lv et al., 2015). Later, new results have been presented mainly discussing refinements of Zhang (2000), e. g. applying RANSAC (Zhou et al., 2013), (Zhang et al., 2017).

Taking a look at extrinsic depth sensor calibration, existing methods are mostly target-based. The first option in this context are plane-based techniques. They utilize plane detection and comparison to find correspondences among different sensors' data sets. An approach of this kind was for example implemented by Fernandez-Moral et al. (2014). On the other hand, it is possible to calibrate from 3D objects such as spheres (Su et al., 2018). These methods usually outperform plane based methods, but require a partially overlapping field of view concerning the sphere.

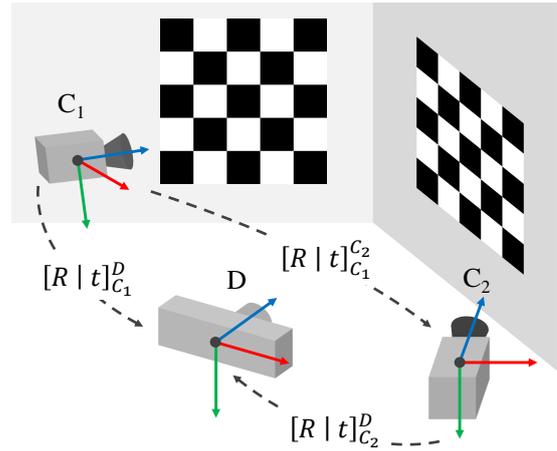


Figure 1: Schematic representation of the experimental system using a depth camera ( $D$ ) and two RGB cameras ( $C_1, C_2$ )

### 2.2. Heterogeneous systems

In the context of this paper, it is important to clarify the terms "intrinsic" and "extrinsic calibration". Traditionally, intrinsic calibration refers to determining a camera's internal parameters, such as focal length, principal point, and lens distortion coefficients. Extrinsic calibration, on the other hand, involves determining the translation and rotation (i.e., the pose  $[R|t]$ ) between different sensors or cameras. Most RGB-D devices combine a standard RGB sensor and a depth sensor based on structured (infrared) light. In relevant literature, the calibration of these components *within* the device is often referred to as extrinsic calibration because they are two separate sensors. However, this differs from our focus, which is on the extrinsic calibration of several independent devices.

Device-internal methods (Basso et al., 2018), (Chen et al., 2018) can rely on small displacements, small rotation angles, and a fixed setup that consists only of the color-depth pair. This makes the calibration process more straightforward compared to the extrinsic calibration of a system with separated devices featuring more complex setups.

Besides structured light technology, most depth sensors are based on time-of-flight or laser scanning (LiDAR). The latter, in particular, has several references in the literature regarding sensor fusion with RGB cameras.

Initial approaches work on 2D LiDAR systems, such as Zhang and Pless (2004). Methods involving a 3D LiDAR can be classified into *target-based* and *targetless*. Target-based methods, like those using planar surfaces to calibrate a laser-camera system (Park et al., 2014), often demonstrate higher accuracy. These methods frequently employ checkerboards (Geiger et al., 2012; Pandey et al., 2010) or 3D objects like boxes (Pusztai and Hajder, 2017). However, they imply more restrictive demands for the calibration environment (Pandey et al., 2012).

## 3. Methodology

The presented procedure addresses the challenge of determining the relative poses of mixed and rigid sensor systems. This extrinsic calibration is achieved without the necessity of *visual* overlap, provided that a *planar* overlap exists. This implies that sufficiently large planes, for example walls or the

floor, are visible to the sensors simultaneously from different viewpoints. The sensors employed in this setup may vary and can provide either depth data or image data. If RGB sensors are part of the configuration, the planes must feature checkerboards for detection purposes. Given these constraints, the specific combination of sensors becomes inconsequential.

By leveraging the simultaneously observed planes, we can identify plane parameters within each sensor's system. Subsequently, correspondences between planes from different sensors can be established by initially estimating the sensor positions roughly. Once identified, the transformation can be used as an argument of a cost function which is to be minimized in the calibration. The initial value is obtained from the user's guess. Assuming the presence of a principal sensor representing the world coordinate system, calibration can be performed either for each sensor with respect to this reference system or for every possible pair.

### 3.1. Plane extraction

Using reliable and exact methods for plane detection in the different data formats is crucial for a successful calibration process. Due to the flexibility to use either point cloud or image data, the detection methods differ in this aspect.

However, the used plane format is the same in both cases, namely the Hesse normal form. Each point  $p \in \mathbb{R}^3$  in the plane satisfies the equation

$$n^\top p + d = 0 \quad (1)$$

where  $n \in \mathbb{R}^3$  represents the planes normal vector with  $\|n\|_2 = 1$  and  $d \in \mathbb{R}$  is the plane's distance to the origin.

To extract planes from RGB images, we utilize checkerboards of a known size. Given the intrinsic and distortion parameters of the camera(s) (Heikkila and Silvén, 1997), traditional methods for solving the Perspective-n-Point (PnP) problem (OpenCV) are applied to calculate the plane's parameterization in 3D space (Lepetit et al., 2009). To maximize the information extracted from RGB images, the implementation can also handle multiple checkerboards. Thus, if the calibration environment provides more than one suitable plane, several of them can be equipped with (necessarily equal) checkerboards and detected accordingly.

Apparently, the plane recognition in point cloud data obtained from depth sensors requires a different approach. To address this, a region growing technique is applied (Holz and Behnke, 2013), implemented in the PCL (Trevor et al., 2013). However, alternatives such as RANSAC-based techniques (Honti et al., 2018) or graph-based methods (Nguyen and Le, 2013) hold similar promise.

Once the regions are segmented, a threshold ensures that only sufficiently large regions are processed further. Should a region fail to meet this threshold relative to the total points in the cloud, it is excluded from further consideration.

For each of these regions, the objective is to determine the optimal fitting plane estimating its parameters, given a set of 3D points  $r_i = (x_i, y_i, z_i)^\top$ , where  $i = 1, \dots, N$ . The aim is to minimize the sum of squared distances from each point to the plane. This objective is addressed via an eigenvector problem assumption under the assumption of isotropic Gaussian noise, following Poppinga et al. (2008). To be more exact,

the optimal normal vector  $n^*$  is the eigenvector to the smallest eigenvalue of the matrix

$$M = \sum_{i=1}^N (r_i - r_G)(r_i - r_G)^\top, \quad (2)$$

where  $r_G = \frac{1}{N} \sum_{i=1}^N r_i$  is the region's centroid. Once obtained, the optimal  $d^*$  can be calculated using (1):

$$d^* = -r_G^\top n^*. \quad (3)$$

### 3.2. Obtaining corresponding planes

After segmenting and selecting the principal regions as well as estimating their plane parameters for different images or point clouds, it is possible to proceed with searching for correspondences among the planes. At this point, the initial estimation of the sensors' position given by the user comes into consideration. Depending on the already mentioned calibration strategy, the following procedure is to be carried out either for all possible sensor pairs in the setup (see Figure 1) or only for those including the principal sensor.

Let  $A$  and  $B$  be a pair of sensors.  $\{(n_i^A, d_i^A)\}$ ,  $i = 1, \dots, K$  is considered the set of  $K$  planes seen by sensor  $A$  where  $n_i^A \in \mathbb{R}^3$  is the  $i$ -th normal vector and  $d_i^A \in \mathbb{R}$  holds the corresponding distance to the origin, both in the system of  $A$ . Analogously, this holds for the planes of  $B$ , given by the set of tuples  $\{(n_j^B, d_j^B)\}$ ,  $j = 1, \dots, L$ . Here  $n_j^B \in \mathbb{R}^3$  is the normal vector of the  $j$ -th plane and  $d_j^B \in \mathbb{R}$  its distance to the origin. A tilde ( $\sim$ ) above a letter indicates an estimated value.

---

#### Algorithm 1 Establish plane correspondences (simplified)

---

```

1: Input1:  $\{(n_i^A, d_i^A)\}$ ,  $i = 1, \dots, K$ 
2: Input2:  $\{(n_j^B, d_j^B)\}$ ,  $j = 1, \dots, L$ 
3: for  $i = 1, \dots, K$  do
4:   for  $j = 1, \dots, L$  do
5:      $\tilde{p}_j^A \leftarrow \text{transform}_B^A(n_j^B)$ 
6:      $\tilde{x}_j^A \leftarrow \text{transform}_B^A(n_j^B \cdot d_j^B)$ 
7:     New distance:  $\tilde{z}_j^A \leftarrow -\tilde{p}_j^A \cdot \tilde{x}_j^A$ 
8:      $\theta = n_i^{A\top} n_j^A$ 
9:      $s = \|d_i^A - \tilde{z}_j^A\|$ 
10:    if  $\theta < \theta_{\text{th}}$  and  $s < s_{\text{th}}$  then
11:      New correspondence:  $n_i/d_i$  with  $n_j'/d_j'$ 
12:    end if
13:  end for
14: end for

```

---

For readability, we assume that sensor  $A$  itself is the reference system. Therefore, its plane data does not need to be transformed. Algorithm 1, oriented to the structure presented in Fernandez-Moral et al. (2014), shows how correspondences between planes are detected. It iterates over all possible plane pairs. Then, the plane parameters of  $B$  are estimated in  $A$  using the initial guess chosen in advance. For the estimated transformed normal vector  $\tilde{p}_j^A$  only the coordinate transformation  $A \rightarrow B$  is required. To calculate the distance to the origin  $\tilde{z}_j^A$  an auxiliary point on the transformed plane  $\tilde{x}_j^A$  is calculated first.

If the planes are sufficiently similar, i. e. neither the angle between the normal vectors  $\theta$  nor the difference between the distances  $s$  exceed the thresholds  $\theta_{\text{th}}$  and  $s_{\text{th}}$ , they are considered the same plane seen from different viewpoints. The

choice of threshold values is closely linked to the accuracy of the initial estimation. For example, assuming that the initial estimations are too rough and the threshold values too small, fewer plane matches would be detected.

The output of this step is a list of plane matches or correspondences. In case of the intention to calibrate all sensors with respect to a principal one, this results in  $M - 1$  plane match lists,  $M$  being the total number of sensors in the system. If the calibration is to be executed for all possible pairs, the number of match lists grows to  $M \cdot (M - 1)/2$ .

### 3.3. Calibration problem formulation

Once the plane correspondences are obtained, the process comes down to sensor pair-wise calibrations based on this data. As the calibration of each sensor pair can be seen as an independent problem, we restrict the following discussion to one of them, keeping in mind that there might be additional pairs to calibrate.

Having processed the data up to this point, we can take advantage of existing methods to perform the optimization. Thus, the following relates to Arun et al. (1987) and Sorkine-Hornung and Rabinovich (2017), also used in a previous research by Fernandez-Moral et al. (2014).

We start with the previously introduced sensor pair where  $A$  represents the system of reference and  $B$  is located with a relative transformation  $[R|t] \in \mathbb{SE}(3)$  with respect to  $A$ . The rotation  $R \in \mathbb{SO}(3)$  is represented by a  $3 \times 3$  matrix, and the translation  $t \in \mathbb{R}^3$ . Furthermore, let now be  $\{(n_i, d_i)\}_A$  and  $\{(n'_i, d'_i)\}_B$ ,  $i = 1, \dots, N$ , the sets of  $N$  ordered plane correspondences between  $A$  and  $B$ , respectively. I. e., the planes with the same index  $i$  are associated with the same plane. We now aim to estimate the optimal  $[R|t]$ , considering that the measurements are affected by unbiased Gaussian noise. This problem can be divided into two separate ones since the rotation and translation constraints are decoupled.

The rotation matrix  $R$  can be found by solving the least square problem

$$R = \operatorname{argmin}_R \sum_{i=1}^N \omega_i \|n_i - R n'_i\|_2^2. \quad (4)$$

where  $\omega_i$  denotes the weight of the  $i$ -th correspondence. Applying Arun et al. (1987), it can be reformulated and solved non-iteratively by applying a singular value decomposition. Similarly, the translation can be obtained analytically, as well by considering a least square problem:

$$\operatorname{argmin}_t \sum_{i=1}^N \omega_i (d_i - d'_i + t^\top n_i)^2. \quad (5)$$

However, Zúñiga Noël et al. (2017) improved this solution, outlining that in the analytic solution in Fernandez-Moral et al. (2014) the presence of very noisy observations or outliers causes difficulties. Instead, iterative, more robust solutions using Gauss-Newton or Levenberg-Marquardt algorithms are proposed. To set up the problem correctly, (4) and (5) are combined as

$$\operatorname{argmin}_\xi \left( \begin{bmatrix} n_i \\ d_i - d'_i \end{bmatrix} - \begin{bmatrix} R & 0_{3 \times 1} \\ -t^\top R & 1 \end{bmatrix} \begin{bmatrix} n'_i \\ 1 \end{bmatrix} \right)^2. \quad (6)$$

Here,  $\xi \in \mathbb{R}^6$  holds the three angles and three translation components to reconstruct

$$T(\xi) = \begin{bmatrix} R(\xi) & t(\xi) \\ 0_{1 \times 3} & 1 \end{bmatrix} \in \mathbb{SE}(3). \quad (7)$$

It can be shown quickly that the second part of (6) is related to  $T(\xi)$ :

$$T^{-T}(\xi) = \begin{bmatrix} R & 0_{3 \times 1} \\ -t^\top R & 1 \end{bmatrix}. \quad (8)$$

Therefore, (6) is equivalent to

$$\operatorname{argmin}_\xi \left( \underbrace{\begin{bmatrix} n_i \\ d_i - d'_i \end{bmatrix} - T^{-T}(\xi) \begin{bmatrix} n'_i \\ 1 \end{bmatrix}}_{=: e_i} \right)^2, \quad (9)$$

such that  $e_i$  can be used as argument of

$$\xi^* = \operatorname{argmin}_\xi \left( \|e_i(\xi)\|_2^2 \right), \quad (10)$$

a function that is then to be minimized. Here  $\rho(s) = \log(1 + s)$  is used for robustness purposes.

As the second, iterative approach provides a better performance facing noise and/or outliers (Zúñiga Noël et al., 2017), this method will be used for the calibration approach in Section 4.



Figure 2: Test setup used for the experiment: RGB-D device (Orbbec Astra) in the middle, two RGB sensors (Logitech C922 Pro) left and right

## 4. Experimental setup

For the experiment, we chose a mixed setup featuring an RGB-D sensor, hereafter labeled as  $D$ , and two RGB sensors ( $C1$  and  $C2$ ), all operating at a frame rate of 30 Hz and a resolution of  $640 \times 360$ . The cameras were mounted on an aluminum frame as shown in Figure 2. Camera  $D$  was placed between and above  $C1$  and  $C2$ , with the two latter slightly tilted around their Y-axes, pointing inward.

To leverage the capability of detecting multiple planes in both sensor types, the test environment featured two detectable planes. These planes were equipped with checkerboards to enable detection by the RGB sensors, as described

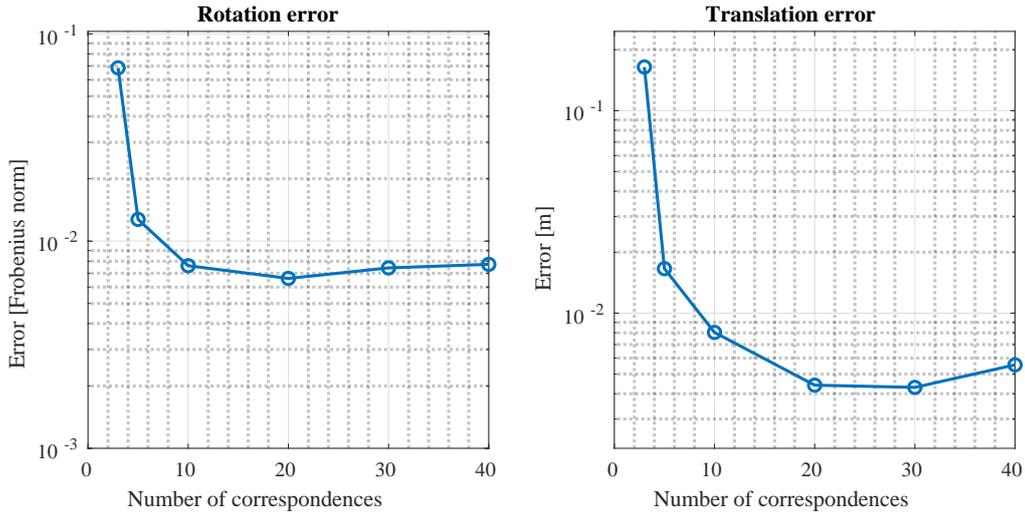


Figure 3: Mean rotation (left) and translation error (right) for different sample sizes of plane correspondences (lower means better), the calibration was executed 20 times for each sample size with a random set of plane matches

in Section 3.1. Also, there is no need to label the checkerboards. Since an initial estimation is provided, the assignment of the planes is possible by estimating the coordinate transformation to the other system.

Using this setup, data containing about 60 plane correspondences for every sensor pair was recorded simultaneously by moving the camera rig by hand and detecting the two planes from different points of view (see Figure 4). This means that from one synchronized image set featuring three sensor pairs with two plane observations each, a maximum of six plane correspondences can be extracted.

Due to the quadratic increasing error and bias of the depth camera with distance to the observed scene, the frame was kept close to the observed surfaces (0.5 - 1.5 m) during the recording. This approach should keep the total error (uncertainty and bias) of the used data within the range of  $\pm 10$  mm (Giancola et al., 2018).



Figure 4: Camera rig recording plane data from the floor and a wall equipped with one checkerboard each

## 5. Results

To properly check the implemented method, we utilize a technique that does not rely on knowing the exact relative poses of the sensors (*ground truth*). This was achieved by calibrating the three cameras with respect to each other, resulting in three transformations (see Figure 1). In the exact case, this *closed-loop* transformation  $T_{loop} = [R_{loop} | t_{loop}]$  (e.g.  $D \rightarrow C1 \rightarrow C2 \rightarrow D$ ),  $R_{loop}$  should be a  $3 \times 3$  identity matrix for rotation and a zero-vector for translation. Any deviation from this ideal indicates an error.

To quantify this, we define the rotation error ( $\epsilon_R$ ) as the Frobenius norm of the difference between the computed rotation matrix ( $R_{loop}$ ) and the identity matrix ( $I_{3 \times 3}$ )

$$\epsilon_R = \|R_{loop} - I_{3 \times 3}\|_F,$$

while the translation error ( $\epsilon_t$ ) is defined as the Euclidean distance of the translation vector ( $t_{loop}$ ) from the origin:

$$\epsilon_t = \|t_{loop}\|_2$$

Using the recorded dataset mentioned in Section 4, we evaluated how well the method performed with different numbers of plane matches: 3, 5, 10, 20, 30, and 40 correspondences. Each evaluation was repeated 20 times using random sets of correspondences from the dataset.

The results shown in Figure 3 indicate that both, the rotation and translation error drop rapidly when working with 10 plane correspondences. More plane matches do not lead to significantly better results. It can be observed that the translation error reaches the single-digit millimeter range, which is within the previously mentioned error bounds of  $\pm 10$  mm (Giancola et al., 2018). This indicates that the method converges successfully, providing a high level of accuracy. Furthermore, the rotation error also shows a strong decrease with an increasing number of correspondences, stabilizing after 10 correspondences. This rapid convergence suggests that the method is both efficient and reliable.

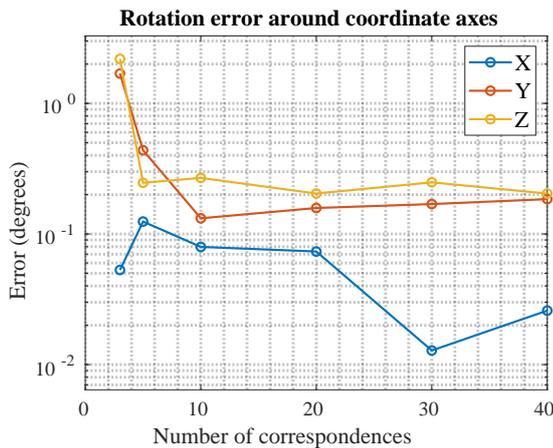


Figure 5: Errors of the three rotation angles for different sample sizes of correspondences

Figure 5 presents a detailed analysis of the rotation error, breaking it down into angles. The observed behavior aligns with expectations. Since the rotation of the setup around the X-axis (which is perpendicular to the aluminum profile, as depicted in Figure 1) is negligible in all transformations, the initial estimation of  $0^\circ$  is highly accurate. Consequently, the error is small even for samples involving 3 or 5 plane matches. The behavior of the other angles mirrors that of the overall rotation error depicted in Figure 3.

Our study enables a partial comparison with the method outlined in Zúñiga Noël et al. (2017). Like our approach, theirs employed a setup with three sensors. However, their experiment was restricted to RGB-D sensors, what might explain that larger errors were encountered and therefore causes a tenfold difference in error magnitudes between our results and theirs.

## 6. Conclusion

As seen in the experimental results, our method demonstrates the level of accuracy achievable with the employed sensor setup. Reliable results are obtained using a minimum of 10 plane correspondences, showcasing its practicality and effectiveness for extrinsic calibration tasks.

The versatility in data input, accommodating both image and point cloud data, enhances its applicability for future scenarios. For instance, this method could find utility in camera-LiDAR systems and other similar applications.

## Acknowledgements

This research has been carried out as part of the ARPEGGIO project (PID2020-117057GB-I00) funded by the Spanish Ministry of Science and Innovation.

## References

Arun, K. S., Huang, T. S., Blostein, S. D., 1987. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9 (5), 698–700.

- Basso, F., Menegatti, E., Pretto, A., 2018. Robust intrinsic and extrinsic calibration of rgb-d cameras. *IEEE Transactions on Robotics* 34 (5), 1315–1332.
- Chen, G., Cui, G., Jin, Z., Wu, F., Chen, X., 2018. Accurate intrinsic and extrinsic calibration of rgb-d cameras with gp-based depth correction. *IEEE Sensors Journal* 19 (7), 2685–2694.
- Fernandez-Moral, E., González-Jiménez, J., Rives, P., Arévalo, V., 2014. Extrinsic calibration of a set of range cameras in 5 seconds without pattern. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 429–435.
- Geiger, A., Moosmann, F., Car, Ö., Schuster, B., 2012. Automatic camera and range sensor calibration using a single shot. In: *2012 IEEE international conference on robotics and automation*. IEEE, pp. 3936–3943.
- Giancola, S., Valenti, M., Sala, R., 2018. A survey on 3d cameras: Metrological comparison of time-of-flight, structured-light and active stereoscopy technologies.
- Heikkilä, J., Silvén, O., 1997. A four-step camera calibration procedure with implicit image correction. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*. IEEE, pp. 1106–1112.
- Holz, D., Behne, S., 2013. Fast range image segmentation and smoothing using approximate surface reconstruction and region growing. In: *Intelligent Autonomous Systems 12: Volume 2 Proceedings of the 12th International Conference IAS-12, held June 26-29, 2012, Jeju Island, Korea*. Springer, pp. 61–73.
- Hontí, R., Erdélyi, J., Kopáčík, A., 2018. Plane segmentation from point clouds. *Pollack Periodica* 13 (2), 159–171.
- Leptit, V., Moreno-Noguer, F., Fua, P., 2009. Ep n p: An accurate o (n) solution to the p n p problem. *International journal of computer vision* 81, 155–166.
- Lv, Y., Feng, J., Li, Z., Liu, W., Cao, J., 2015. A new robust 2d camera calibration method using ransac. *Optik* 126 (24), 4910–4915.
- Nguyen, A., Le, B., 2013. 3d point cloud segmentation: A survey. In: *2013 6th IEEE conference on robotics, automation and mechatronics (RAM)*. IEEE, pp. 225–230.
- Pandey, G., McBride, J., Savarese, S., Eustice, R., 2010. Extrinsic calibration of a 3d laser scanner and an omnidirectional camera. *IFAC Proceedings Volumes* 43 (16), 336–341.
- Pandey, G., McBride, J., Savarese, S., Eustice, R., 2012. Automatic targetless extrinsic calibration of a 3d lidar and camera by maximizing mutual information. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 26. pp. 2053–2059.
- Park, Y., Yun, S., Won, C. S., Cho, K., Um, K., Sim, S., 2014. Calibration between color camera and 3d lidar instruments with a polygonal planar board. *Sensors* 14 (3), 5333–5353.
- Poppinga, J., Vaskevicius, N., Birk, A., Pathak, K., 2008. Fast plane detection and polygonalization in noisy 3d range images. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 3378–3383.
- Pusztai, Z., Hajder, L., 2017. Accurate calibration of lidar-camera systems using ordinary boxes. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. pp. 394–402.
- Sorkine-Hornung, O., Rabinovich, M., 2017. Least-squares rigid motion using svd. *Computing* 1 (1), 1–5.
- Su, P.-C., Shen, J., Xu, W., Cheung, S.-C. S., Luo, Y., 2018. A fast and robust extrinsic calibration for rgb-d camera networks. *Sensors* 18 (1), 235.
- Trevor, A., Gedikli, S., Rusu, R. B., Christensen, H. I., 2013. Efficient organized point cloud segmentation with connected components. *Semantic Perception Mapping and Exploration (SPME)* 1.
- Zhang, Q., Pless, R., 2004. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*(IEEE Cat. No. 04CH37566). Vol. 3. IEEE, pp. 2301–2306.
- Zhang, Q., Wu, S., Wang, W., Fang, Z., 2017. Improving 2d camera calibration by lo-ransac. *Int. J. Inf. Electron. Eng* 7, 93–98.
- Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22 (11), 1330–1334.
- Zhou, F., Cui, Y., Wang, Y., Liu, L., Gao, H., 2013. Accurate and robust estimation of camera parameters using ransac. *Optics and Lasers in Engineering* 51 (3), 197–212.
- Zúñiga Noël, D., Gómez Ojeda, R., Moreno, F. Á., González Jiménez, J., et al., 2017. Calibración extrínseca de un conjunto de cámaras rgb-d sobre un robot móvil. *Actas de las XXXVIII Jornadas de Automática*.