Machine Vision and Applications manuscript No.
(will be inserted by the editor)

Extrinsic Calibration of Multi-Modal Sensor Arrangements with Non-Overlapping Field-of-View

Carolina Raposo · João Barreto · Urbano Nunes

Received: date / Accepted: date

Abstract Several applications in robotics require complex sensor arrangements that must be carefully calibrated, both intrinsically and extrinsically, to allow information fusion and enable the system to function as a whole. These arrangements can combine different sensing modalities - such as color cameras, Laser-RangeFinders, and depth cameras - in an attempt to obtain richer descriptions of the environment. Finding the location of multi-modal sensors in a common world reference frame is a difficult problem that is largely unsolved whenever sensors observe distinct, disjoint parts of the scene. This article builds on recent results in object pose estimation using mirror reflections to provide an accurate and practical solution for the extrinsic calibration of mixtures of color cameras, LRFs, and depth cameras with non-overlapping Field-of-View. The method is able to calibrate any possible sensor combination as far as the setup includes at least one color camera. The technique is tested in challenging situations not covered by the current state-of-the-art, proving to be practical and effective. The calibration software is made available to be freely used by the research community.

Keywords Extrinsic Calibration · Laser-Rangefinder · Depth Camera · Non-Overlapping FOV · Mirror

Carolina Raposo
E-mail: carolinaraposo@isr.uc.pt

João Barreto
E-mail: jpbar@isr.uc.pt

Urbano Nunes
E-mail: urbano@isr.uc.pt

Carolina Raposo · João Barreto · Urbano Nunes
Institute of Systems and Robotics, University of Coimbra, Portugal

1 Introduction

Many applications in robotics and intelligent transportation systems (ITS) require the use of multiple sensors that can be of the same modality (homogeneous sensor networks) [14,13,19] or of different modalities (heterogeneous or hybrid sensor setups) [4,11,2]. These setups must be calibrated both intrinsically and extrinsically. The intrinsic calibration consists in determining the sensor parameters that enable to convert measurement units into metric units (e.g. pixels into mm) [6]. The extrinsic calibration consists in locating the different sensors in a common coordinate frame, allowing the platform to function as a whole. The literature in extrinsic calibration is vast and includes methods for finding the relative pose between sensors of different modalities [21,7]. However, most of these solutions require the fields-of-view (FOVs) to overlap and cannot cope with situations in which sensors are observing different, disjoint parts of the scene (see Figure 1). This is not the case of the recent work by Bok *et al.* [2] where the first solution to the calibration of a color camera and a LRF with non-overlapping FOVs is proposed.

This article revisits the problem of the extrinsic calibration of multi-sensor arrangements that can comprise color cameras[1], Laser-Rangefinders (LRF), and/or depth cameras. It builds on recent results for estimating the pose of an object observed by a color camera through planar mirror reflections [18] and proposes a systematic, practical approach for calibrating mixtures of color cameras, LRFs, and depth cameras with non-overlapping FOV. The method uses a checkerboard pattern as calibration object and handles situa-

[1] The term *color camera* is used when referring to regular cameras, either RGB or grayscale, in order to better distinguish between depth cameras.

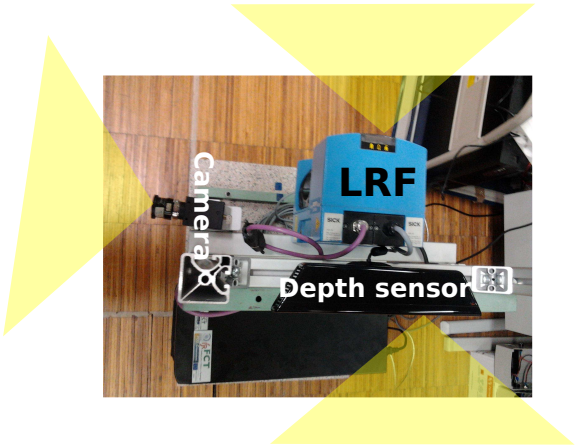


Fig. 1: Example of a heterogeneous sensor network with non-overlapping field-of-view placed in a moving platform.

tions for which there is no simple, effective solution in the state-of-the-art.

Since the estimation of the relative pose of an object from mirror views requires $N \geq 3$ reflections [18,12], the calibration in the case of non-overlapping FOV must be based in conventional extrinsic calibration methods that use a small number M of object images, otherwise the total number $N \times M$ of frames to be acquired becomes prohibitive. While the extrinsic calibration of a color camera and a LRF can be accomplished from $M = 3$ images of a moving checkerboard pattern [21], the existing approaches for calibrating a color and depth camera require 20 to 30 calibration frames to deliver accurate estimation results [7]. In order to overcome this difficulty we also describe a new method for the joint calibration of color and depth cameras that accomplishes the accuracies reported in [7] using one sixth of the input frames. The method has been first introduced in a prior conference publication [17] and is herein extended to handle the situation of the color and depth cameras having disjoint FOVs.

In summary, the article combines recent results in estimating the pose of an object that is observed through mirror reflections [18] with explicit methods for calibrating color camera / LRF and color camera / depth camera arrangements [21,7], and presents for the first time a solution for the extrinsic calibration of mixtures of heterogeneous sensors with non-overlapping FOVs. The contributions are as follows:

1. An accurate, easy to use method that combines different modules for accomplishing the extrinsic calibration of multi-modal sensor setups with non-overlapping FOV such as the one depicted in Figure 1.
2. A new calibration method for mixtures of color and depth cameras that outperforms the state-of-the-art approach by Herrera *et al.* [7]. This new method performs accurately when using a small number of input frames [17], which is of paramount importance for accomplishing calibration in the case of non-overlapping FOVs.
3. A thorough experimental assessment of the solution that enables to decide about the number of mirror reflections M and object views N that are needed to reach a certain accuracy level.
4. An experiment of the calibration of a sensor platform with an application in Structure-from-Motion that evinces the usefulness of heterogeneous sensor systems.
5. A complete MATLAB toolbox[2] that implements the described calibration methods and that will be made publicly available to the robotics and ITS communities. It includes a calibration dataset to be used as benchmark.

1.1 Related Work

There are many applications in robotics and ITS that combine several sensing devices, either from the same or different modalities. Stationary camera networks are used in surveillance and object tracking [14], while multi-camera rigs allow for the coverage in vehicles of the whole surrounding environment [13]. In [1], Auvinet *et al.* describe a system that uses multiple active cameras for reconstructing the volumes of bodies in motion from the acquired depth maps. Its main application is in gait analysis which has become an increasingly interesting area of research. LightSpace [22] combines depth cameras and projectors to provide interactivity on and between surfaces in everyday environments, allowing a convincing simulation of the manipulation of physical objects. The literature also reports heterogeneous sensor setups comprising both combinations of color cameras and LRF, and combinations of color and depth cameras. Color camera and LRF networks have recently been used in object classification for the construction of maps of outdoor environments [4], by integrating visual and shape features. In [15], these features are combined for pedestrian detection in urban environments. In [11], multiple color and depth cameras are used for generating high-quality multi-view depth, allowing for the construction of 3D video.

For most of these applications the relative pose between sensor nodes must be known in advance in order

[2] The toolbox that accompanies this article can be accessed at `http://arthronav.isr.uc.pt/~carolina/toolbox_multimodal/`.

Table 1: Methods for calibrating multi-sensor systems.

		Color Cam.	LRF	Depth Cam.
Overlap	Bouguet [24]	X		
	Vasconcelos [21]	X	X	
	Herrera [7]	X		X
NonOverlap	Rodrigues [18]	X		
	Bok [2]	X	X	
	Our Contribution	**X** **X**	**X**	**X**

for the acquired multi-modal information to be fused and the platform to work as a whole. There are several methods for performing both intrinsic and extrinsic calibration of color cameras, LRFs, and depth cameras. A brief overview of current approaches to calibrate either sensors of the same modality, or mixtures of two modalities is now provided. Table 1 summarizes the results of this overview clearly showing that the majority of existing solutions are unable to handle the problem of generic extrinsic calibration between sensors of different modalities with non-overlapping Field-of-View (FOV).

1. *Color Camera Calibration:* The literature is vast but explicit methods using a known checkerboard pattern are specially popular because they are stable, accurate, and the calibration rig is easy to build. Bouguet's camera calibration toolbox [3] implements Zhang's method [24] that, given 3 or more images, estimates the intrinsic parameters, as well as the poses of the checkerboard with respect to the camera. These poses can be used to find the relative rigid displacement between different camera nodes (extrinsic calibration) by simply assuring that some planes are simultaneously observed across different nodes. However, there are camera networks for which the FOVs of the different nodes do not overlap. This happens either in surveillance, where many times a broad region must be covered with a small number of cameras [14,10], or in robotics, whenever cameras are placed to obtain an omnidirectional view of the scene around the vehicle [13]. A possible solution in these cases is to use mirrors for computing the pose with respect to an object that is outside the FOV [18,12,20,8]. The idea has been first used in [12] to calibrate a camera network with the pose of the object being estimated from a minimum of 5 mirror reflections. Sturm *et al.* [20] proved that such relative pose could be determined from a minimum of 3 images and Rodrigues *et al.* [18] introduced a minimal, closed-form solution for the problem that outperforms the methods suggested in [12,20]. An exhaustive experimental evaluation showed that in practice 5 to 6 reflections are more than enough to obtain very stable and accurate results.
2. *Color Camera / LRF Calibration*: Zhang and Pless [23] proposed a practical method for the extrinsic calibration of a color camera and a LRF that uses at least 5 images of a known checkerboard pattern. Later on, Vasconcelos *et al.* [21] described a minimal solution for the problem leading to a robust algorithm that clearly outperforms the method in [23]. These solutions only deal with the overlapping case. More recently, an algorithm for calibrating a color camera and a LRF whenever their FOVs do not intersect was proposed [2]. The method makes assumptions about the relative pose between the checkerboard and the environment's structure that may be difficult to satisfy in small or cluttered spaces. Moreover, due to these assumptions, the sensor platform must move in order to acquire calibration data. In case of large platforms, such as ground vehicles, this method is not appropriate since it would be extremely difficult to acquire the required calibration data in different positions and orientations. In this article the method [21] is extended to the non-overlapping case, providing a simple solution that works for any color camera / LRF configuration that can be attached to either a small or a large platform.
3. *Color Camera / Depth Camera Calibration*: Scene reconstruction from a color-depth camera pair measurements requires the system to be calibrated, both intrinsically and extrinsically. Kinect cameras have a standard calibration from factory that is not accurate enough for many situations. Herrera *et al.* [7] have recently modeled the Kinect's depth camera distortion and proposed a method for calibrating a depth camera and additional color cameras whose FOV overlap. Its main strength is in an explicit depth distortion term. Unfortunately it requires many images (over 20) and it is not prepared for handling non-overlapping situations. To tackle the first issue, we propose a new calibration method for color camera / depth camera pairs that has proven to perform accurately with only 6 to 10 calibration images. Moreover, this new approach is extended to the non-overlapping case, solving the calibration problem for any possible sensor configuration.

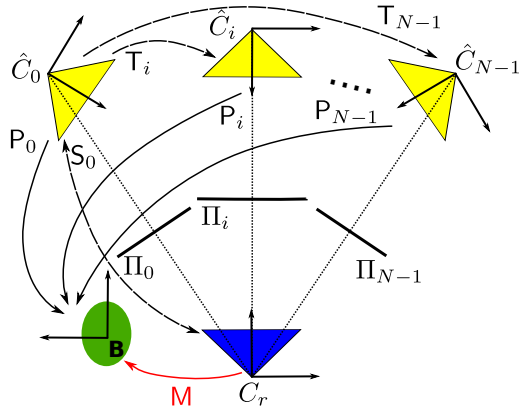


Fig. 2: Object **B** is seen by camera C_r through N planar mirror reflections $\mathbf{\Pi}_i$, originating N virtual cameras $\hat{C}_i$. Our goal is to find the pose M of the real camera C_r with respect to object **B**.

1.2 Overview of the Article and Used Notation

The structure of the paper is as follows. Section 2 revisits the problem of determining the pose of an object from planar mirror reflections and introduces the estimation algorithm by Rodrigues *et al.* [18] that will be a cornerstone of the proposed extrinsic calibration framework. Section 3 starts by quickly reviewing the minimal solution for calibrating a color camera and a LRF [21] and shows how the method can be used in conjunction with mirror reflections to handle the case of camera and LRF pointing towards different directions. The section ends with a real experiment that validates the approach and assesses how accuracy varies as a function of the number of mirror reflections N and object poses M. Section 4 starts by quickly overviewing the method of Herrera *et al.* [7] for calibrating mixtures of color and depth cameras. It then proposes several modifications to [7] that dramatically improve the robustness, the computational time, and the number of input frames [17]. The new method is used to accomplish extrinsic calibration in the case of non-overlapping FOV, and the approach is validated through real experiments with ground truth. Finally, Section 5 shows how these contributions can be used to calibrate an heterogeneous combination of color camera, LRF, and depth camera facing different directions around a moving platform.

The notation used in this paper is as follows. Scalars are represented by plain letters, e.g. t, vectors are indicated by bold symbols, e.g. $\mathbf{t}$, and vectors with unit norm by $\overrightarrow{\mathbf{t}}$. Matrices are denoted by letters in sans serif font, e.g. R. Planes are represented by a 4D homogeneous vector that is indicated by an uppercase Greek letter, e.g. $\mathbf{\Pi}$. Sets of intrinsic parameters are defined by uppercase calligraphic letters, e.g. $\mathcal{L}$. Entities in LRF reference frame are represented using $'$ and in depth camera reference frame using $\mathring{}$, e.g. $\mathbf{\Phi}'$ and $\mathring{\mathbf{\Phi}}$.

Throughout the article most geometric entities, such as points, lines, or planes will be represented in projective coordinates. The symbol $=$ will be used to denote strict equality and $\sim$ to represent equality up to scale between projective representations. The rigid transformations between reference frames are represented by 4×4 matrices and, in the schemes, the direction of the arrow indicates the mapping of coordinates, e.g. in Figure 2 the coordinates in C_r are mapped to **B** by M.

2 Camera Pose Estimation from Mirror Reflections

It can be shown that the image acquired by a camera looking at a planar mirror is equivalent to the image that would be acquired by a virtual camera placed behind the mirror plane. In this case, the virtual and real cameras have the exact same intrinsic parameters and their local reference frames are related by a symmetry transformation S with respect to the mirror plane. Rodrigues *et al.* [18] propose to freely move a planar mirror in front of a camera in order to obtain images of an object that lies outside the FOV. It was shown that, given $N \geq 3$ images, it is possible to estimate the rigid displacement M between camera and object (Figure 2), as well as the plane coordinates of the N mirrors. Since their algorithm will be extensively used to accomplish the extrinsic calibration of sensors with non-overlapping FOV, this section overviews its steps.

2.1 Review of the Algorithm presented in [18]

Figure 2 shows an object **B** being observed by camera C_r through N planar mirror reflections. Each virtual camera $\hat{C}_i$ is originated by the mirror plane $\mathbf{\Pi}_i$, which is uniquely defined by its unitary normal vector $\overrightarrow{\mathbf{n}}_i$, and the scalar euclidean distance d_i, with $i = 0, \ldots, N-1$. The pose of the object P_i in each virtual camera reference frame is determined by either applying the *PnP* algorithm [5], in case **B** is a known 3D object, or by estimating and factorizing a planar homography [6], in case **B** is a plane surface. For the sake of convenience the article always uses a planar checkerboard pattern as calibration object. With abuse of notation, the pose and the homography will be denoted by the same symbol, whenever it is convenient.

Given the $N \geq 3$ object poses P_i, Rodrigues *et al.* [18] choose a reference virtual view and determine the position of the corresponding mirror plane. Let $\hat{C}_0$ be the reference camera. The first step is to compute the

relative pose T_i of the remaining virtual views, which can be easily accomplished by applying the following formula:

$$\mathsf{T}_i = \mathsf{P}_i^{-1}\mathsf{P}_0, i = 1, 2, \ldots, N-1, \tag{1}$$

with T_i being a 4×4 matrix with format

$$\mathsf{T}_i = \begin{bmatrix} \mathsf{R}_i & \mathbf{t}_i \\ \mathbf{0} & 1 \end{bmatrix}, \tag{2}$$

where R_i is a rotation matrix with rotation axis direction $\overrightarrow{\omega}_i$ and rotation angle θ_i, and $\mathbf{t}_i$ is the translation vector. It can be shown that each rigid motion T_i gives rise to two independent linear constraints on the parameters of the mirror plane $\mathbf{\Pi}_0$ that can be stacked for the $N-1$ motions, originating a system of linear equations. The least squares solution can be found by applying SVD, and $\mathbf{\Pi}_0$ is computed. The symmetry transformation S_0, that relates the reference frames of virtual camera $\hat{C}_0$ and real camera C_r, is given by:

$$\mathsf{S}_0 = \begin{bmatrix} \mathsf{I} - 2\overrightarrow{\mathbf{n}}_0\overrightarrow{\mathbf{n}}_0^{\mathsf{T}} & 2d_0\overrightarrow{\mathbf{n}}_0 \\ \mathbf{0} & 1 \end{bmatrix}. \tag{3}$$

This symmetry matrix is involutory, meaning that $\mathsf{S}_0 = \mathsf{S}_0^{-1}$. From P_0 and S_0, the pose M of the object $\mathbf{B}$ comes in a straightforward manner as:

$$\mathsf{M} = \mathsf{S}_0\mathsf{P}_0. \tag{4}$$

Note that due to the mirror reflection, if the reference frame C_r is right-handed, then the reference frame $\hat{C}_0$ is left-handed, and vice-versa, making the multiplication in Equation (4) to be defined this way. To improve robustness, this algorithm is performed N times independently, each time considering a different virtual camera as the reference frame. Then, the average of all estimations of M is considered.

A singular configuration occurs whenever all the mirror planes intersect into a single line in 3D. This can be caused by either rotating the mirror around a fixed-axis, or when the reflection planes are all parallel (the intersection line is at infinity).

2.2 Calibration of Cameras with Non-Overlapping FOV

As explained in [18], the fact that it is now possible to determine the pose of an object M that is outside the camera's FOV enables the extrinsic calibration of cameras that do not observe overlapping regions of the scene.

Consider the situation of Figure 3 where cameras C_F and C_B, mounted on a platform, observe object $\mathbf{B}$ directly and through mirror reflections, respectively. The extrinsic calibration D can be carried by computing M_F and M_B, and then finding $\mathsf{D} = \mathsf{M}_F^{-1}\mathsf{M}_B$. M_F can be computed as a planar homography using a standard approach [24], while estimating M_B is performed using the algorithm from [18].

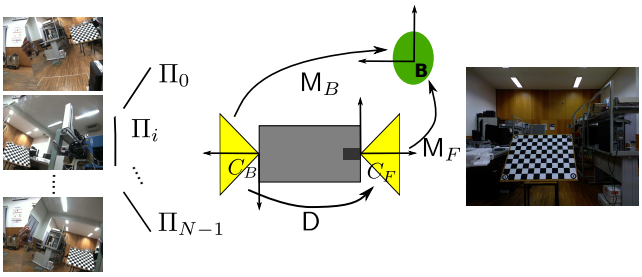


Fig. 3: The extrinsic calibration D is accomplished by showing the object directly to camera C_F (checkerboard image on the right) and showing it through mirror reflections to camera C_B (checkerboard images on the left).

In practical terms, according to [18], for $N = 6$ mirror views it is possible to achieve subpixel accuracy, corresponding to a rotation error slightly below 1° and a translation error of approximately 3.5%.

3 Extrinsic Calibration of a Color Camera and a Laser-Rangefinder

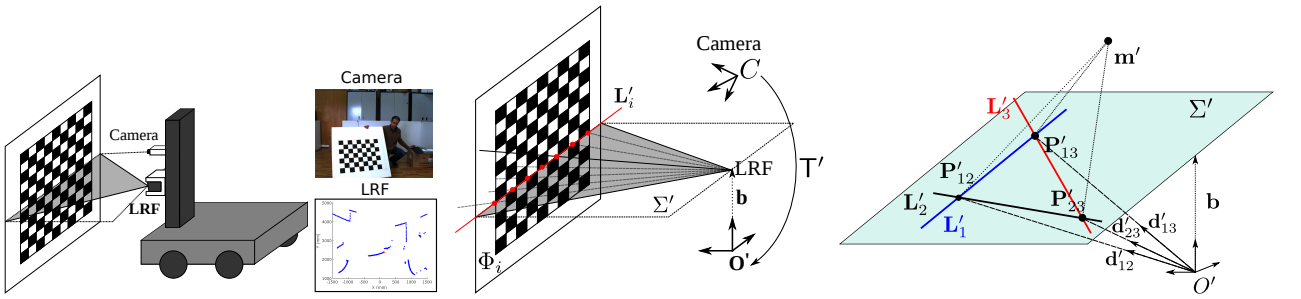
Let us now consider the problem of finding the extrinsic calibration T' between a color camera C and a LRF O' as illustrated in Figure 4.

Vasconcelos *et al.* [21] have recently shown that a color camera and a LRF can be calibrated from a minimum of $M = 3$ images of a planar grid. The problem of finding T' is cast as the problem of registering a set of planes $\mathbf{\Phi}_i, i = 1, 2, 3$, expressed in color camera coordinates, with a set of 3D lines $\mathbf{L}'_i, i = 1, 2, 3$ in the LRF coordinate system. They show that there are 8 solutions, with the correct one being selected by an additional plane-line correspondence.

We start by reviewing the algorithm [21] and then extend the method to the case of non-overlapping FOV by using mirror reflections.

3.1 Review of the Algorithm presented in [21]

Consider M planes $\mathbf{\Phi}_i \sim [\mathbf{n}_i^{\mathsf{T}} \quad 1]^{\mathsf{T}}, i = 1, 2, \ldots, M$ expressed in color camera coordinates and the corresponding lines $\mathbf{L}'_i$, expressed in LRF coordinates, where $\mathbf{L}'_i$ is the locus of intersection between $\mathbf{\Phi}_i$ and the scan plane $\mathbf{\Sigma}'$, as shown in Figure 4b.



(a) Camera and LRF mounted on a platform (b) Relative pose T' between camera and LRF (c) Line intersections in the scan plane Σ'

Fig. 4: (a) The extrinsic calibration in [21] is carried by moving a checkerboard pattern in front of both color camera and LRF. (b) The color camera C that observes the plane $\mathbf{\Phi}_i$ and LRF O' that sees line $\mathbf{L}'_i$ are related by a transformation T'. (c) Lines $\mathbf{L}'_i$ lie in the scan plane $\mathbf{\Sigma}'$ and intersect in points $\mathbf{P}'_{ij}$, which define the directions $\mathbf{d}'_{ij}$ with the origin of the LRF reference frame O'.

Vasconcelos *et al.* [21] show that the relative rotation R' can be determined by solving the system of nonlinear equations

$$\begin{cases} \alpha_{12}\mathbf{d}_{12} = \mathsf{R}'^{\mathsf{T}}(\mathbf{P}'_{12} + \mathbf{m}') \\ \alpha_{13}\mathbf{d}_{13} = \mathsf{R}'^{\mathsf{T}}(\mathbf{P}'_{13} + \mathbf{m}') \\ \alpha_{23}\mathbf{d}_{23} = \mathsf{R}'^{\mathsf{T}}(\mathbf{P}'_{23} + \mathbf{m}') \end{cases}, \tag{5}$$

where $\mathbf{d}_{ij}$ is the direction of the line where planes $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$ intersect, $\mathbf{P}'_{ij}$ is the point in plane $\mathbf{\Sigma}'$ where lines $\mathbf{L}'_i$ and $\mathbf{L}'_j$ meet, $\mathbf{m}'$ is an unknown vector and α_{ij} are unknown scalars that assure algebraic equality. The authors observed that the system of equations (5) corresponds to solving the *P3P* problem [5] for determining the relative pose between a color camera and an object from 3 object-image point correspondences. Figure 4c shows the nature of this *P3P* problem, where the virtual perspective camera is centered in point $\mathbf{m}'$ where planes $\mathbf{\Phi}'_i$ intersect, $\mathbf{d}_{ij}$ play the role of image points and $\mathbf{P}'_{ij}$ of object points. *P3P* enables to find the relative orientations R'^{T} and the position $\mathbf{m}'$ of the intersection of the 3 planes in LRF coordinates.

Finally, to find the translation $\mathbf{t}'$, it is shown in [21] that it suffices to compute $\mathbf{t}' = \mathsf{A}^{-1}\mathbf{c}$ with

$$\mathsf{A} = \begin{bmatrix} \mathbf{n}'^{\mathsf{T}}_1\mathbf{n}'_1\mathbf{n}^{\mathsf{T}}_1 \\ \mathbf{n}'^{\mathsf{T}}_2\mathbf{n}'_2\mathbf{n}^{\mathsf{T}}_2 \\ \mathbf{n}'^{\mathsf{T}}_3\mathbf{n}'_3\mathbf{n}^{\mathsf{T}}_3 \end{bmatrix} \mathsf{R}'^{\mathsf{T}} \text{ and } \mathbf{c} = \begin{bmatrix} \mathbf{n}'^{\mathsf{T}}_1\mathbf{n}'_1 - \mathbf{n}'^{\mathsf{T}}_1\mathsf{R}'\mathbf{n}_1 \\ \mathbf{n}'^{\mathsf{T}}_2\mathbf{n}'_2 - \mathbf{n}'^{\mathsf{T}}_2\mathsf{R}'\mathbf{n}_2 \\ \mathbf{n}'^{\mathsf{T}}_3\mathbf{n}'_3 - \mathbf{n}'^{\mathsf{T}}_3\mathsf{R}'\mathbf{n}_3 \end{bmatrix}, \tag{6}$$

where $\mathbf{n}'_i$ refers to the normal to planes $\mathbf{\Phi}'_i$, expressed in LRF coordinates ($\mathbf{n}'_i = \mathsf{R}'\mathbf{n}_i$).

A discussion about the singular configurations of this method is given in [21]. In general terms, whenever the lines where the checkerboard planes intersect are parallel, or the checkerboard planes intersect in a point that lies in the danger cylinder (refer to [21]) a singular configuration occurs.

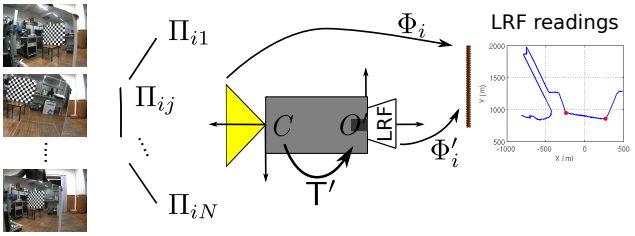


Fig. 5: Calibration of a color camera / LRF setup in case of non-overlapping FOV. It is shown how to determine the pose $\mathbf{\Phi}_i$ of the checkerboard in color camera coordinates from N mirror reflections $\mathbf{\Pi}_{ij}$. The line segment between the red dots in the LRF readings plot corresponds to the calibration plane.

The registration procedure originates a total of $R \leq 8$ rigid transformations T'_i that align 3 planes with 3 coplanar lines. For sets with $M > 3$ plane-line correspondences, the best solution is chosen in a RANSAC framework to select inliers, which are then used in a Bundle Adjustment step to refine the solution. This is done by simultaneously minimizing reprojection error and distance to LRF depth measurements.

According to [21], a total of $M = 5$ checkerboard planes are sufficient for achieving good accuracy.

3.2 Calibration of a Color Camera / LRF Pair with Non-Overlapping FOV

As shown in the scheme of Figure 5, the checkerboard pattern is placed in front of the LRF and, for each pose $\mathbf{\Phi}'_i, i = 1, \ldots, M$, the color camera observes the pattern through N mirror reflections $\mathbf{\Pi}_{ij}, j = 1, \ldots, N$. The coordinates of the checkerboard plane $\mathbf{\Phi}_i$ in the color

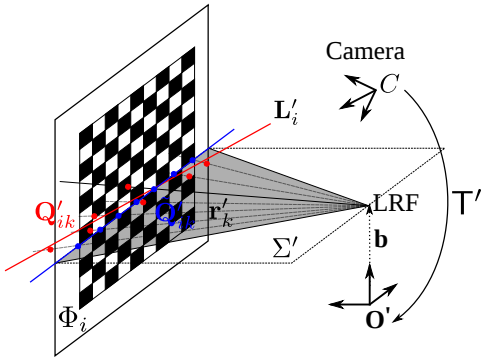


Fig. 6: Points $\mathbf{Q}'_{ik}$ (red) are constructed by measuring depth along the directions $\mathbf{r}'_k$ in the scan plane $\mathbf{\Sigma}'$. Points $\tilde{\mathbf{Q}}'_{ik}$ (blue) are obtained by mapping planes $\mathbf{\Phi}_i$ into the LRF reference frame and intersecting with $\mathbf{r}'_k$.

camera reference frame can be found using the method described in section 2. By applying this algorithm for retrieving each plane pose, the extrinsic calibration between the color camera and the LRF can be directly obtained by using the method from [21]. Note that, in order to be possible to perform such a calibration, a minimum of 9 images are necessary, since the algorithm presented in [18] (reviewed in Section 2) requires $N \geq 3$ mirror reflections and the algorithm from [21] (reviewed in Section 3.1) requires $P \geq 3$ plane poses. However, and as discussed before, in practice more images are required to obtain robust, accurate results.

The first step for performing the extrinsic calibration is to find an initial estimation. In this case, it is done similarly to the method from [21], having the difference that planes $\mathbf{\Phi}_i$ are determined from the algorithm described in Section 2 and not directly from plane-to-image homographies. A refinement step is then applied to minimize the reprojection errors in the color camera and LRF. The minimization is performed over the parameters T', inlier planes $\mathbf{\Phi}_i$ and corresponding mirror poses $\mathbf{\Pi}_{ij}$, and the color camera's intrinsic parameters K,

$$\min_{\mathsf{T}',\mathbf{\Phi}_i,\mathbf{\Pi}_{ij},\mathsf{K}} e = e_{LRF} + k e_{CAM}, \tag{7}$$

where k is a weighting parameter. The LRF residue e_{LRF} is the sum of the squared distances between the points $\tilde{\mathbf{Q}}'_{ik}$ and the points $\mathbf{Q}'_{ik}$ that are reconstructed from the depth readings (refer to Figure 6):

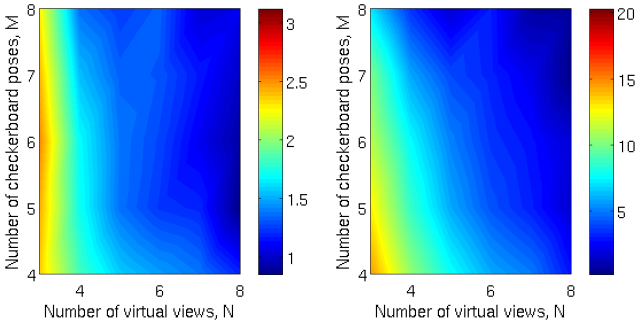
$$e_{LRF} = \sum_i \sum_j ||\mathbf{Q}'_{ik} - \tilde{\mathbf{Q}}'_{ik}||^2. \tag{8}$$

In the present case of a non-overlapping configuration, since the checkerboard images result from mirror reflections, the reprojection error of the color camera is computed differently, when compared to the algorithm proposed in [21]. The virtual camera relative to the mirror pose $\mathbf{\Pi}_{ij}$ is computed by using transformation S_0 in Figure 2, which is a reflection about a plane. Since the virtual cameras "observe" the checkerboard directly, they are used for computing the reprojection errors of the plane-to-image homographies, yielding e_{CAM}. Note that the computation of the color camera residue depends on the intrinsic parameters K, so that this formulation allows to refine both the intrinsic and the extrinsic calibrations. The main steps of the proposed method are outlined in Algorithm 1.

Algorithm 1: New Method for the Calibration of a Color Camera / LRF Pair with Non-Overlapping FOV

Inputs: Scan plane $\mathbf{\Sigma}'$, lines $\mathbf{L}'_i$ and object poses in relation to the virtual cameras $\mathsf{P}_{ij}, i = 1, \ldots, M$, $j = 0, \ldots, N-1$
Output: Extrinsic calibration T', mirror poses $\mathbf{\Pi}_{ij}$, and planes $\mathbf{\Phi}_i$

1. Compute each plane pose $\mathbf{\Phi}_i$ using the method from [18] (reviewed in Section 2).
2. Use the algorithm presented in [21] (reviewed in Section 3.1) for initializing the extrinsic calibration T' from the obtained planes $\mathbf{\Phi}_i$ and the input lines $\mathbf{L}'_i$.
3. Refine the estimated transformation T', planes $\mathbf{\Phi}_i$, mirror poses $\mathbf{\Pi}_{ij}$, and color camera intrinsics K in a bundle adjustment step as defined in Equation (7).



(a) Rotation error (in degrees) (b) Translation error (in percentage)

Fig. 7: Extrinsic calibration errors obtained with the LRF-color camera pair in a non-overlapping configuration, for varying number of checkerboard poses and virtual views.

3.3 Experimental Results for Non-Overlapping FOV

The method proposed by Vasconcelos *et al.* [21] is able to achieve robust, accurate results for the LRF-color camera pair calibration in an overlapping configura-

tion from $M \geq 6$ checkerboard images. In the case of non-overlapping FOV, besides the rigid displacement T' between the sensors, the pose of the checkerboard $\mathbf{\Phi}$, which is observed through mirror reflections, in color camera coordinates, must be estimated. Rodrigues *et al.* [18] show that in practice $N \geq 6$ virtual views are required for the estimation to be reasonably accurate.

In this section a real experiment with ground truth is carried, not only to validate the approach presented in section 3.2, but also to assess the number of $M \times N$ images that are necessary in practice to obtain a certain degree of accuracy.

A setup that has overlapping FOV was used in order to enable calibration with the algorithm of section 3.1 that works as ground truth. A dataset of $M = 12$ direct image-LRF cuts was collected for calibrating the sensor pair. Without moving the color camera with respect to the LRF, we collected a second dataset where the checkerboard is observed by the color camera through mirror reflections in order to mimic the non-overlapping situation. A total of $M = 12$ checkerboard poses with each pose being observed by $N = 12$ mirror reflections was acquired.

In order to find the number of checkerboard poses M and mirror reflections N required to obtain accurate results, $M = 4, \ldots, 8$ and $N = 3, \ldots, 8$ were considered, and for each of the 30 possible combinations of M and N, 50 calibrations sets of M checkerboard poses and N corresponding image reflections were randomly selected. The average rotation and translation errors with respect to the ground truth obtained with these calibration sets are shown in Figure 7. It can be seen that using $M < 6$ checkerboard poses or $N < 6$ mirror reflections frequently leads to translation errors over 4%. This can be explained by the fact that, in average, not even the original methods perform more accurately with less images. Note that, as observed in experiments performed with the original methods, the rotation error is always relatively low (below 2.4°). Moreover, using the minimum number of mirror reflections ($N = 3$) provides poor results, as reported in [18]. However, the method was able to achieve very accurate results, with translation errors of approximately 1% and rotation errors up to 1° with datasets of 7 or more checkerboard poses and mirror reflections. A good compromise would be to use a dataset of 36 images, consisting of $M = 6$ checkerboard poses and $N = 6$ mirror reflections, as the obtained average errors are of about 3.5% in translation and 1.26° in rotation. Thus, in overall terms, the results are satisfactory and prove that, for carefully chosen checkerboard and mirror orientations, the extrinsic color camera / LRF calibration can be accurately achieved using small datasets.

4 Extrinsic Calibration of a Color Camera and a Depth Camera

In this section the calibration of a color camera / depth camera pair when their FOVs do not overlap is considered. A similar approach to the one proposed in section 3 is considered: using mirror reflections for enabling the color camera to observe a calibration target placed in the FOV of the depth camera.

Recently, Herrera *et al.* [7] proposed a method for solving this problem in the case of overlapping FOV, using as depth camera the Kinect. They showed that jointly calibrating both sensors improved accuracy, as opposed to a separate calibration, and proposed a new depth distortion model for the depth camera. Unfortunately, their method requires $K > 20$ images of a checkerboard to provide good results. The straightforward extension to the non-overlapping case would lead to the need of collecting at least 120 images because of the mirror reflections. Thus, a modified version of Herrera's method is proposed, which is able to achieve comparably accurate calibration results using only about 6 images. This is accomplished by both initializing the relative pose using a new minimal, optimal solution and including a metric constraint during the iterative refinement to avoid a drift in the disparity to depth conversion. This contribution has been briefly presented in a prior conference paper [17] and favors calibration both in overlapping and non-overlapping situations. Finally, the extension of our method [17] to the non-overlapping case is presented.

4.1 Depth Camera Model

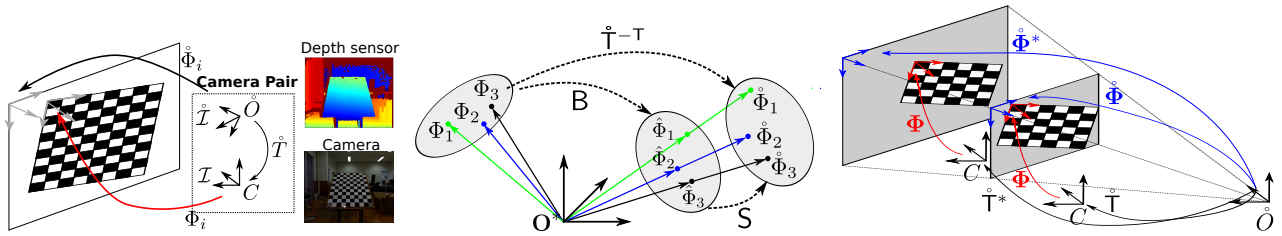
In this article, the Kinect's depth sensor will be used as the depth camera. It provides an image of disparity values computed between the observed image and a prerecorded one. In order to obtain a depth value z_d, a model defined by

$$z_d = \frac{1}{c_1 d_u + c_0} \tag{9}$$

is used, where c_1 and c_0 are two intrinsic parameters of the depth camera, and d_u is the disparity obtained after applying distortion correction. The Kinect's depth sensor presents distortion which has been modeled by Herrera *et al.* [7]:

$$d_u = d + \mathcal{D}_\delta(u, v) \cdot e^{\alpha_0 - \alpha_1 d}, \tag{10}$$

where d is the disparity returned by the depth sensor in pixel (x_d, y_d), $\mathcal{D}_\delta$ contains the spacial distortion pattern, and $\alpha = [\alpha_0, \alpha_1]$ models the decay of the distortion effect. Although this distortion model might be



(a) Calibration of the color-depth camera pair (b) Registration in the dual projective space $\mathcal{P}^{3*}$ (c) Scale drift in the relative pose estimation

Fig. 8: (a) The color camera C, with intrinsic parameters $\mathcal{I}$, and the depth camera $\mathring{O}$, with intrinsic parameters $\mathring{\mathcal{I}}$, are related by a rigid transformation $\mathring{\mathsf{T}}$, which is initialized by registering planes $\mathbf{\Phi}_i$ with planes $\mathring{\mathbf{\Phi}}_i$, $i = 1, 2, 3$. (b) This registration can be interpreted as a projective transformation in $\mathcal{P}^{3*}$ that maps points $\mathbf{\Phi}_i$ into points $\mathring{\mathbf{\Phi}}_i$, which can be factorized as $\mathring{\mathsf{T}}^{-\mathsf{T}} \sim \mathsf{SB}$. (c) The relative pose estimation may be affected by a drift in scale: the pose of the checkerboard in the color camera reference frame C is fixed, while the depth camera may observe the calibration plane at different depths.

considered for better accuracy, it will not be used in the experiments, and d_u is replaced by d in Equation (9). The set of intrinsic parameters of the depth camera is referred to as $\mathring{\mathcal{I}} = \{\mathbf{f}_d, \mathbf{c}_d, \mathbf{k}_d, c_0, c_1\}$, where $\mathbf{f}_d$ is the focal length, $\mathbf{c}_d$ the principal point and $\mathbf{k}_d$ the distortion coefficients. These are initialized using preset values, which are publicly available for the Kinect. Remark that although Kinect's depth sensor is being considered, the method can be applied to any other depth camera.

4.2 Fast, Robust Algorithm for Calibration in Overlapping Configurations

Figure 8a depicts the calibration process that consists in moving a checkerboard pattern in front of both color and depth cameras for collecting K image-disparity map pairs. The color camera's intrinsic parameters and initial plane poses $\mathbf{\Phi}_i$ in color camera coordinates are obtained using a standard calibration algorithm [24]. While Herrera *et al.* use this initialization step for both the color and the depth cameras, we perform differently for the latter.

In our case, the depth camera's intrinsic parameters are preset. Then, for each input disparity map i, the plane is segmented and each point (x_d, y_d), with disparity d, on the plane is reconstructed:

$$X = (x_d - x_{0d})\frac{z_d}{f_{dx}}, \quad Y = (y_d - y_{0d})\frac{z_d}{f_{dy}}, \quad Z = z_d, \tag{11}$$

where the depth value z_d is computed using (9), and $\mathbf{c}_d = [x_{0d}, y_{0d}]$ and $\mathbf{f}_d = [f_{dx}, f_{dy}]$ are the intrinsic parameters. To each 3D point cloud, a plane $\mathring{\mathbf{\Phi}}_i$ is fitted using a standard total least squares algorithm.

4.2.1 Initialization of extrinsic calibration

The extrinsic calibration is the problem of finding rotation $\mathring{\mathsf{R}}$ and translation $\mathring{\mathbf{t}}$ such that

$$\mathring{\mathbf{\Phi}}_i \sim \underbrace{\begin{bmatrix} \mathring{\mathsf{R}} & \mathbf{0} \\ -\mathring{\mathbf{t}}^{\mathsf{T}}\mathring{\mathsf{R}} & 1 \end{bmatrix}}_{\mathring{\mathsf{T}}^{-\mathsf{T}}} \mathbf{\Phi}_i, i = 1, 2, 3 \tag{12}$$

verifies. While Herrera *et al.* [7] use a linear sub-optimal algorithm to carry this estimation, a new solution is herein proposed.

The idea is to find the relative pose $\mathring{\mathsf{T}}$ between the color and depth cameras through $3D$ registration of two sets of planes $\mathring{\mathbf{\Phi}}_i = [\mathring{\mathbf{n}}_i^{\mathsf{T}} \quad 1]^{\mathsf{T}}$ and $\mathbf{\Phi}_i = [\mathbf{n}_i^{\mathsf{T}} \quad 1]^{\mathsf{T}}$. Knowing that points and planes are dual entities in $3D$ - a plane in the projective space $\mathcal{P}^3$ is represented as a point in the dual space $\mathcal{P}^{3*}$, and vice-versa - Equation (12) can be seen as a projective transformation in $\mathcal{P}^{3*}$ that maps points $\mathbf{\Phi}_i$ into points $\mathring{\mathbf{\Phi}}_i$. As shown in Figure 8b, the transformation $\mathring{\mathsf{T}}^{-\mathsf{T}}$ can be factorized into a rotation transformation B, mapping points $\mathbf{\Phi}_i$ into points $\widehat{\mathbf{\Phi}}_i$, and a projective scaling S that maps points $\widehat{\mathbf{\Phi}}_i$ into points $\mathring{\mathbf{\Phi}}_i$:

$$\mathsf{B} = \begin{bmatrix} \mathring{\mathsf{R}} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}, \quad \mathsf{S} = \begin{bmatrix} \mathsf{I} & \mathbf{0} \\ -\mathring{\mathbf{t}}^{\mathsf{T}} & 1 \end{bmatrix}, \tag{13}$$

where I is the 3×3 identity matrix. B and S can be estimated from $K = 3$ point-point correspondences by firstly computing the rotation $\mathring{\mathsf{R}}$ and afterwards the translation $\mathring{\mathbf{t}}$. This provides an easy two-step process to perform the registration:

1. Since the length of a vector is not changed by rotation, $\mathbf{n}_i$ and $\mathring{\mathbf{n}}_i$ are normalized, and an algorithm derived from [9] for computing a transformation between two sets of unitary vectors is applied. From

Equation (12), it is known that this transformation is a pure rotation, and thus the translation component is not computed.

2. The computation of the projective scaling S that maps points $\widehat{\mathring{\mathbf{\Phi}}}_i$ into points $\mathring{\mathbf{\Phi}}_i$, with $\widehat{\mathring{\mathbf{\Phi}}}_i = \mathsf{B}\mathbf{\Phi}_i$, $i = 1, 2, 3$, is done similarly to [21]. We can write $\mathring{\mathbf{\Phi}}_i \sim \mathsf{S}\widehat{\mathring{\mathbf{\Phi}}}_i$, yielding

$$\lambda_i \begin{bmatrix} \mathring{\mathbf{n}}_i \\ 1 \end{bmatrix} = \begin{bmatrix} \mathsf{I} & \mathbf{0} \\ -\mathring{\mathbf{t}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathring{\mathsf{R}}\mathbf{n}_i \\ 1 \end{bmatrix}, \tag{14}$$

where λ_i is an unknown scale factor. After some algebraic manipulation, it comes that

$$\mathring{\mathbf{n}}_i^\mathsf{T}\mathring{\mathbf{n}}_i\mathbf{n}_i^\mathsf{T}\mathring{\mathsf{R}}^\mathsf{T}\mathring{\mathbf{t}} - \mathring{\mathbf{n}}_i^\mathsf{T}\mathring{\mathbf{n}}_i + \mathring{\mathbf{n}}_i^\mathsf{T}\mathring{\mathsf{R}}\mathbf{n}_i = 0. \tag{15}$$

Each pair $\mathbf{\Phi}_i$, $\mathring{\mathbf{\Phi}}_i$ gives rise to a linear constraint in the entries of the translation vector $\mathring{\mathbf{t}}$, which can be computed by $\mathring{\mathbf{t}} = \mathsf{A}^{-1}\mathbf{c}$, with A and $\mathbf{c}$ defined as in Equation (6), by substituting $\mathbf{n}_i'$ with $\mathring{\mathbf{n}}_i$, $i = 1, 2, 3$.

This algorithm has a singular configuration if the three normals do not span the entire 3D space, meaning that planes have a configuration such that either two of all of their normals are co-planar.

When the number of acquired image-disparity map pairs is $K > 3$, the best solution is determined in a RANSAC framework, similar to [21]:

1. For each possible triplet of pairs of planes $\mathbf{\Phi}_i$, $\mathring{\mathbf{\Phi}}_i$, a transformation $\mathring{\mathsf{T}}$ is estimated.
2. For each solution $\mathring{\mathsf{T}}$, the depth camera coordinates $\mathring{\mathbf{\Phi}}_i^*$ for all $\mathbf{\Phi}_i$ are computed using (12), and the euclidean distance d_i in the dual space between the computed $\mathring{\mathbf{\Phi}}_i^*$ and $\mathring{\mathbf{\Phi}}_i$ is determined.
3. Each solution is ranked by $rank(\mathring{\mathsf{T}}) = \sum_j \max(t, d_i)$, where t is a predefined threshold. The correspondences for which $d_i < t$ are considered as inliers.
4. Find $\mathring{\mathsf{T}}$ for which $rank(\mathring{\mathsf{T}})$ is minimum.

After obtaining an initial estimation for the transformation $\mathring{\mathsf{T}}$, and a set of inlier correspondences, a bundle adjustment procedure is performed.

4.2.2 Bundle Adjustment

It was experimentally observed that under poor initialization and a small number of images, Herrera's method tends to drift in depth. After careful analysis, we came up with an hypothesis for this observation. From Equation (16), it can be seen that if c_0 and c_1 are affected by a scale component, an error in the extrinsic calibration will occur, while the reprojection error does not change. Figure 8c depicts the problem, where it can be seen that a scale drift will cause the pose of the calibration plane w.r.t. the depth camera $\mathring{\mathbf{\Phi}}^*$ to be incorrectly estimated, while its pose relative to the color camera $\mathbf{\Phi}_i$ is not affected. This automatically originates an erroneous estimation of the relative pose, represented by $\mathring{\mathsf{T}}^*$. In order to tackle this problem, a new cost function was presented that, unlike Herrera's, not only takes into account the reprojection error in the color camera and the difference between measured $\widehat{d}$ and estimated d disparities, but also the difference between Euclidean distances between known points of an object λ and the measured distances between those points $\widehat{\lambda}$:

$$\min_{\mathcal{I},\mathring{\mathcal{I}},\mathring{\mathsf{T}},\mathbf{\Phi}_i} e = \sum ||\mathbf{x} - \widehat{\mathbf{x}}||^2 + k_1 \sum (d - \widehat{d})^2 + k_2(\lambda - \widehat{\lambda})^2, \tag{16}$$

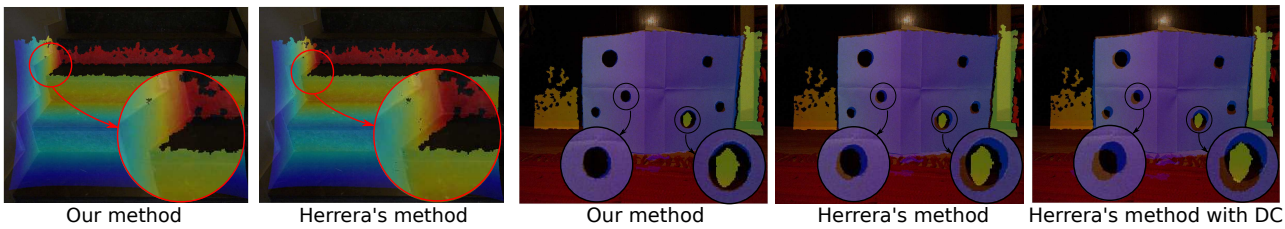
where k_1 and k_2 are weighting parameters, $\mathbf{x}$ and $\widehat{\mathbf{x}}$ are the reprojected and measured pixel positions, and $\mathcal{I}$ are the intrinsic parameters of the color camera.

As a final step of our approach, the whole depth distortion function of Equation (10) is recovered using the optimized intrinsic and extrinsic calibrations. Since it is done in open-loop, significantly lower run times were obtained, when compared to [7] that runs multiple consecutive optimization steps. However, this distortion correction will not be performed in the experiments.

4.3 Experimental Results for an Overlapping Configuration

In order to assess the accuracy of our proposed calibration approach, and to compare it with the state-of-the-art method [7], a small dataset of 8 image-disparity-map pairs was acquired for estimating the relative pose between color and depth cameras, and the corresponding intrinsic parameters, using both methods.

Using the same sensor setup, images of a flight of perpendicular stairs were acquired and used for assessing the depth camera's intrinsic calibration. This was done by computing the angles between all possible pairs of reconstructed planes, and comparing them either with 90° or 0° depending whether they are orthogonal or parallel, respectively. Results are shown in Figure 9a, that report an average angular error of 0.495° with our method and 0.743° with Herrera's method. The figure also shows the overlaid depth map with the RGB images, showing that there is a misalignment for Herrera's method. This suggests that its estimation of the extrinsic calibration is also worse. Figure 9b evinces these observations. An image of an object with holes was acquired for assessing the alignment between the depth map and the intensity image when they are overlaid. Results with our method show that the misalignment is very slight, while for Herrera's method it is significant, and is larger when using distortion correction



(a) Reconstruction of a flight of stairs (b) Overlaid depth maps with the RGB images

Fig. 9: (a) A flight of stairs was reconstructed and the angles between all possible pairs of planes were computed, originating an average angular error of 0.495° with our method and 0.743° with Herrera's method. Note that a slight misalignment is observed when the depth map is overlaid with the RGB image using Herrera's solution, confirming it is less accurate than ours. This inaccuracy is evident in (b), where the depth maps are overlaid with the RGB image of an object with holes. While our method provides a good alignment, Herrera's method does not, and performs worse when distortion correction (DC) is applied.

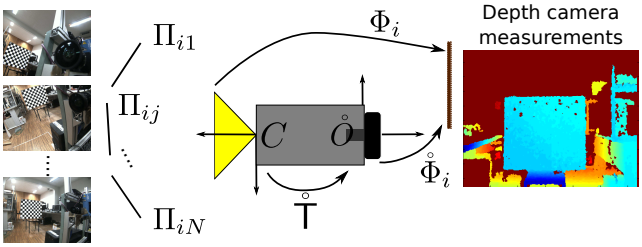


Fig. 10: Calibration of a color-depth camera pair in a non-overlapping configuration, similar to the color camera / LRF situation of Figure 5.

(DC). This indicates that Herrera's method is not able to properly model the depth distortion with small data sets. Note that none of the experiments was performed with distortion correction, except for the last one in Figure 9b.

The general conclusion is that Herrera's method is not able to provide accurate results with small datasets. However, our method performs well under these conditions, being favorable in non-overlapping configurations where mirror reflections must be used.

4.4 Calibration in the case of Non-Overlapping FOV

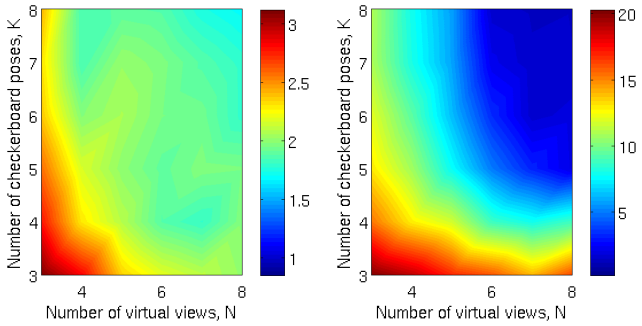
Calibrating a depth and a color camera with non-overlapping FOVs is done in an analogous manner as for the LRF case (section 3.2). Figure 10 shows that the checkerboard pattern is moved in front of the depth camera and, for each plane pose $\mathbf{\Phi}_i$, the color camera observes the pattern through N mirror reflections $\mathbf{\Pi}_{ij}$. Again, the checkerboard poses with respect to the color camera are computed using the algorithm reviewed in Section 2, and the extrinsic calibration between the color camera and the depth camera is carried using the new method described in the present section. However, and as previously explained in section 3.2, since mirror reflections are being used for estimating the checkerboard poses $\mathbf{\Phi}_i$, the first term of the error function in Equation (16), which corresponds to the color camera reprojection error, must be changed. The virtual camera corresponding to mirror pose $\mathbf{\Pi}_{ij}$ is computed using transformation S_0 in Figure 2, and used for finding the reprojected pixel positions $\mathbf{x}$ that appear in Equation (16). Moreover, since the mirror positions $\mathbf{\Pi}_{ij}$ are being taken into account, these are also refined, along with $\mathcal{I}, \mathring{\mathcal{I}}, \mathring{\mathsf{T}}$ and $\mathbf{\Phi}_i$.

For a non-overlapping configuration, the extrinsic calibration of a color camera and a depth camera requires a minimum of 9 images since the algorithm in Section 2 needs $N \geq 3$ mirror reflections and the present method requires $K \geq 3$ checkerboard poses.

4.5 Experimental Results for Non-Overlapping FOV

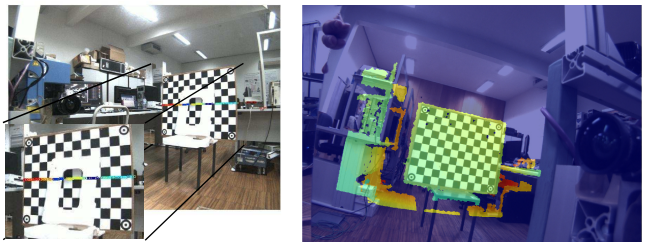
The method presented in section 4.2, that works with sensors whose FOVs overlap, reported accurate results for datasets of 6 image-disparity maps pairs. As in section 3.3, the purpose of this experiment is both to validate the approach and to assess the number of checkerboard poses and mirror reflections required to produce a certain accuracy, when working with a non-overlapping configuration. We performed similarly as in section 3.3, concerning the sensor setup and dataset acquisition. Our ground truth is the result of the calibration performed with the method of section 4.2.

A total of 50 calibration sets were randomly selected for each combination of $K = 3, \ldots, 8$ checkerboard poses and $N = 3, \ldots, 8$ mirror reflections. These sets were used as input to the method from section 4.4



(a) Rotation error (in degrees). (b) Translation error (in percentage).

Fig. 11: Extrinsic calibration errors obtained with the color camera / depth camera pair in a non-overlapping configuration, for varying number of checkerboard poses and virtual views.



(a) Overlaid LRF readings (b) Overlaid depth map

Fig. 12: LRF readings (a) and depth map (b) overlaid with the mirror reflections using the known mirror pose. Different colors identify different measured depths.

and the average rotation and translation errors are presented in Figure 11. The overall conclusions are the same as in section 3.3. For datasets using less than 6 checkerboard and mirror poses, the translation errors tend to be higher than 4% and the rotation errors slightly over 2°. Increasing the number of acquired images to 36 ($K = 6$ and $N = 6$) leads to average errors of 3.6% and 1.9° in translation and rotation, respectively. This is an acceptable result for many applications. However, when higher accuracy is required, increasing the number of checkerboard poses and the number of mirror reflections up to 8 originates results with errors smaller than 1.5% in translation and 1.6° in rotation.

Remark that in general, the errors obtained with the color camera / depth camera pair are slightly larger than the ones obtained with the color camera / LRF pair. This can be explained by the fact that in the first there are more parameters to be optimized, requiring more images to achieve the same accuracy. However, this difference is not significant and good results can be achieved with calibration sets comprising the same number of images.

5 Application Scenario

In this section, the heterogeneous sensor setup shown in Figure 1 is calibrated, having the fields-of-view of all sensors non-overlapping. A first set of calibration data was acquired by moving the checkerboard pattern in front of the LRF, and moving a mirror in front of the color camera so that it observes reflections of the pattern. The checkerboard is placed in $M = 8$ positions and, for each position, $N = 5$ mirror reflections are acquired. An identical calibration set is acquired for the depth-color camera pair. The methods described in sections 3.2 and 4.4 are used for calibrating each sensor pair.

In order to assess the quality of the calibration, and due to the absence of ground truth, the LRF readings and the depth maps were overlaid with the RGB image of a mirror reflection for which the mirror pose was known. This is shown in Figure 12, where it can be seen that a good alignment is obtained, indicating that both the mirror poses and the extrinsic calibration are well estimated. For the LRF case (Figure 12a), an object with an hole was used so that it can be seen that the depth variations between consecutive readings are aligned with the locations that correspond to transitions between the object and the plane behind it.

For the second part of the experiment (Figure 13), the platform traveled a short path in a corridor with parallel walls and 180 cm of width, so that the LRF and the depth camera observed one of the walls. Using the estimated depth camera intrinsic parameters, the wall plane observed by the depth camera was reconstructed. Moreover, the relative pose between the LRF and the depth camera was used for representing the LRF measurements in the depth camera coordinates. The angular and distance errors between the reconstructed plane and line were then computed, in degrees and percentage, respectively, for each of the 5 acquisitions, showing the results in the first two rows of Table 2. Due to noise in the data, slightly different results were obtained for different views. However, the accuracy of the calibration can be confirmed by the good results achieved in all cases.

This configuration can be used for obtaining a textured piecewise planar reconstruction of the scene. The procedure was the following:

1. Using the color camera, the homography of the floor plane was estimated from points with known distances.

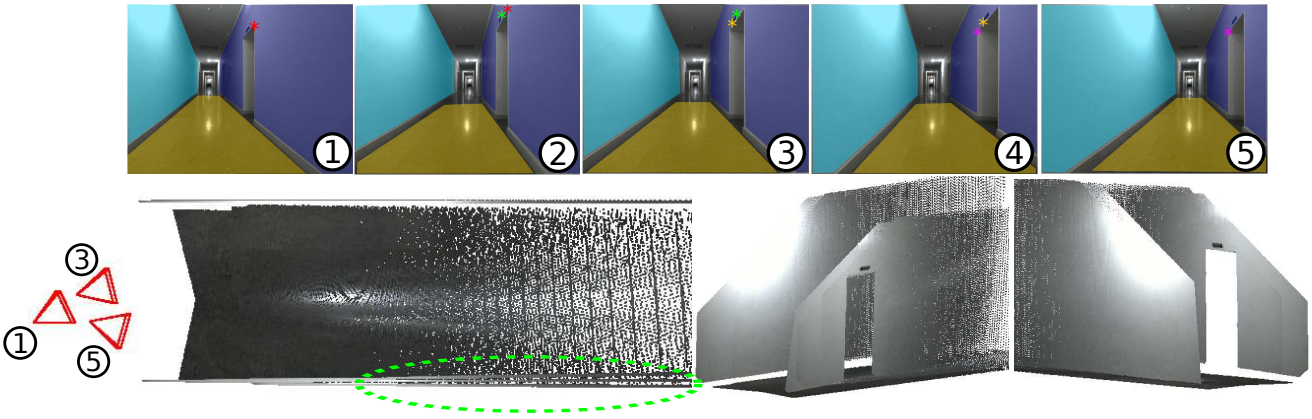


Fig. 13: The setup of Figure 1 moved through a corridor with parallel walls, with the LRF and depth camera pointing to each wall. 5 sets of measurements were acquired. By representing both planes in color camera coordinates, it is possible to estimate the platform motion and obtain textured reconstructions of all scene planes. The color camera positions corresponding to the first, third and fifth views are represented by red camera symbols.

Table 2: Angular (e_α) and distance (e_d) errors between reconstructed lines and planes observed from 5 different views. Angular errors between the floor plane and the wall planes reconstructed by the depth camera (e_{g1}) and the LRF (e_{g2}).

View	1	2	3	4	5
$e_\alpha(^\circ)$	0.11	0.33	0.34	0.16	0.47
$e_d(\%)$	0.97	1.72	1.24	1.33	1.50
$e_{g1}(^\circ)$	0.1967	0.4512	0.2116	0.3769	0.1615
$e_{g2}(^\circ)$	0.1951	0.4538	0.7642	0.4201	0.2053

2. Using the assumption that the two walls are parallel, the plane that contains the line reconstructed by the LRF and that is as parallel as possible to the plane reconstructed by the depth camera was estimated.
3. All the planes were represented in the camera reference frame, and the areas corresponding to each plane in the RGB image were manually segmented and reconstructed.

This procedure enables to have a set of planes represented in the same reference frame for each of the 5 platform positions shown in Figure 13. Since the plane correspondences between different views are known, the minimal solution proposed in [16] was applied for computing the platform motion. In this case there are only two correspondences of non-parallel planes (the floor and the walls), so it is necessary to extract one point correspondence for computing all 6 degrees-of-freedom. The extracted point correspondences are shown in Figure 13 with identifying colors. This example is particularly interesting because the complete lack of texture in the wall planes hampers the reconstruction using RGB cameras, and thus sensors that provide depth measurements are required. Figure 13 shows the segmented regions corresponding to each plane (top row), and the obtained 3D model after concatenating the individual reconstructions using the estimated platform motion (bottom row). Qualitatively, by observing the alignment between the individual 3D models, particularly in the area surrounding the door entrance, it can be seen that the registration was well performed. The green ellipse corresponds to the area of misalignment between planes, which, as can be seen, is very slight. This indicates that the platform motion estimation is accurate, which could not have been possible if the surface planes or the extrinsic calibration had been poorly recovered. In quantitative terms, this reconstruction was assessed by computing the angular errors between the floor plane and the reconstructed wall planes, shown in the last two rows of Table 2. Errors bellow 0.8° were always achieved, being another indicative that the reconstruction, and thus the extrinsic calibration, is accurate. Note that each plane was recovered using one of the sensors independently and only a good extrinsic calibration would provide small errors. The results confirm that this kind of sensor setup can indeed be used for obtaining accurate reconstructions of the scene, even in situations of lack of texture.

In general, the good results obtained prove that the proposed method is practical, effective and useful, solving a problem that so far did not have a simple solution.

6 Conclusion

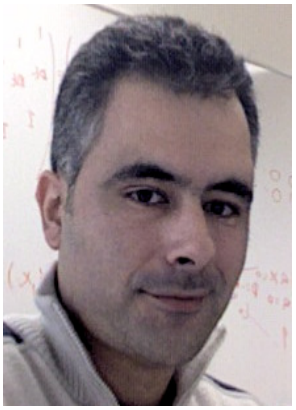
A new and systematic approach for the extrinsic calibration of setups with mixtures of color cameras, Laser-

Rangefinders (LRF) and depth cameras with non-overlapping FOV is presented. Our main contributions are the extension of existing methods for calibrating color camera / LRF and color camera / depth camera pairs to work with non-overlapping fields-of-view. Experimental validation for each case shows that the proposed method provides accurate results. The proposed method makes it possible to find the relative pose between LRFs and depth cameras in non-overlapping configurations, being the only requirement that a color camera is involved in the setup. A final application scenario for which there is no simple solution in the current state-of-the-art is also presented, evincing the practicality and effectiveness of the proposed method.

References

1. Auvinet, E., Meunier, J., Multon, F.: Multiple depth cameras calibration and body volume reconstruction for gait analysis. In: Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on (2012)
2. Bok, Y., Choi, D.G., Vasseur, P., Kweon, I.S.: Extrinsic calibration of non-overlapping camera-laser system using structured environment. In: Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, pp. 436–443 (2014)
3. Bouguet, J.Y.: Camera calibration toolbox for matlab. URL `www.vision.caltech.edu/bouguetj/calib_doc/index.html`
4. Douillard, B., Fox, D., Ramos, F., Durrant-Whyte, H.: Classification and semantic mapping of urban environments. Int. J. Rob. Res. (2011)
5. Haralick, B., Lee, C.N., Ottenberg, K., Nlle, M.: Review and analysis of solutions of the three point perspective pose estimation problem. International Journal of Computer Vision (1994)
6. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, second edn. Cambridge University Press, ISBN: 0521540518 (2004)
7. Herrera C., D., Kannala, J., Heikkil, J.: Joint depth and color camera calibration with distortion correction. Pattern Analysis and Machine Intelligence, IEEE Transactions on (2012)
8. Hesch, J., Mourikis, A., Roumeliotis, S.: Determining the camera to robot-body transformation from planar mirror reflections. In: Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on (2008)
9. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America A (1987)
10. Javed, O., Rasheed, Z., Alatas, O., Shah, M.: Knight trade;: a real time surveillance system for multiple and non-overlapping cameras. In: Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on (2003)
11. Kang, Y.S., Ho, Y.S.: High-quality multi-view depth generation using multiple color and depth cameras. In: Multimedia and Expo (ICME), 2010 IEEE International Conference on (2010)
12. Kumar, R., Ilie, A., Frahm, J.M., Pollefeys, M.: Simple calibration of non-overlapping cameras with a mirror. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (2008)
13. Pagel, F.: Calibration of non-overlapping cameras in vehicles. In: Intelligent Vehicles Symposium (IV), 2010 IEEE (2010)
14. Pflugfelder, R., Bischof, H.: Localization and trajectory reconstruction in surveillance cameras with nonoverlapping views. Pattern Analysis and Machine Intelligence, IEEE Transactions on (2010)
15. Premebida, C., Monteiro, G., Nunes, U., Peixoto, P.: A lidar and vision-based approach for pedestrian and vehicle detection and tracking. In: Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE (2007)
16. Raposo, C., Antunes, M., Barreto, J.: Piecewise-planar stereoscan: Structure and motion from plane primitives. In: Computer Vision ECCV 2014, Lecture Notes in Computer Science. Springer International Publishing (2014)
17. Raposo, C., Barreto, J., Nunes, U.: Fast and accurate calibration of a kinect sensor. In: 3D Vision - 3DV 2013, 2013 International Conference on (2013)
18. Rodrigues, R., Barreto, J.P., Nunes, U.: Camera pose estimation using images of planar mirror reflections. In: K. Daniilidis, P. Maragos, N. Paragios (eds.) Computer Vision ECCV 2010, Lecture Notes in Computer Science. Springer Berlin Heidelberg (2010)
19. Schenk, K., Kolarow, A., Eisenbach, M., Debes, K., Gross, H.: Automatic calibration of multiple stationary laser range finders using trajectories. In: Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on (2012)
20. Sturm, P., Bonfort, T.: How to compute the pose of an object without a direct view? In: Computer Vision ACCV 2006, Lecture Notes in Computer Science. Springer Berlin Heidelberg (2006)
21. Vasconcelos, F., Barreto, J.P., Nunes, U.: A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder. IEEE Transactions on Pattern Analysis and Machine Intelligence (2012)
22. Wilson, A.D., Benko, H.: Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In: Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology, UIST '10 (2010)
23. Zhang, Q., Pless, R.: Extrinsic calibration of a camera and laser range finder (improves camera calibration). In: Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on (2004)
24. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on (1999). DOI 10.1109/ICCV.1999.791289

Carolina Raposo received the Integrated Master's degree (BSc+MSc) in Electrical and Computer Engineering from the University of Coimbra, Portugal, in 2012. Since 2012 she is a computer vision researcher at the Institute for Systems and Robotics - Coimbra. She is currently a PhD student at the University of Coimbra, Portugal. Her main research interests lie on geometric computer vision, 3D reconstruction and Structure-from-Motion.



João P. Barreto (M'99) received the "Licenciatura" and Ph.D. degrees from the University of Coimbra, Coimbra, Portugal, in 1997 and 2004, respectively. From 2003 to 2004, he was a Postdoctoral Researcher with the University of Pennsylvania, Philadelphia. He has been an Assistant Professor with the University of Coimbra, since 2004, where he is also a Senior Researcher with the Institute for Systems and Robotics. His current research interests include different topics in computer vision, with a special emphasis in geometry problems and applications in robotics and medicine. He is the author of more than 70 peer-reviewed publications and recipient of several distinctions and awards including a Google Faculty Research award and 5 *Outstanding Reviewer Awards*. He is Associate Editor for Computer Vision and Image Understanding and Image and Vision Computing Journals.

Urbano Nunes (S'90-M'95-SM'09) received the Ph.D. degree in electrical engineering from the University of Coimbra, Portugal, in 1995. Prof. Nunes is a Full Professor with the Electrical and Computer Engineering Department of Coimbra University, and a researcher of the Institute for Systems and Robotics (ISR-UC) where he is the coordinator of the Automation and Robotics for Human Life Group (AR4LIFE-G). He has research interests in several areas in connection with intelligent vehicles and human-centered mobile robotics with more than 150 published papers in these areas. Dr. Nunes was Vice President for Technical Activities of the IEEE ITS Society (2011-2012), and a Cochair of the Technical Committee on Autonomous Ground Vehicles and ITS (2006-2011) of the IEEE Robotics and Automation Society (RAS). He serves as Associate Editor the journals: IEEE Transactions on Intelligent Transportation Systems and IEEE Intelligent Transportation Systems Magazine. He was the General Chair of the 2010 IEEE Intelligent Transportation Systems Conference, Funchal-Madeira, Portugal, and a General Chair of the 2012 IEEE Intelligent Robots and Systems, Vilamoura, Portugal.