

Relative Pose Improvement of Sphere based RGB-D Calibration

David J. T. Boas^{1,2}, Sergii Poltaretskyi², Jean-Yves Ramel¹, Jean Chaoui², Julien Berhouet^{1,3}
and Mohamed Slimane¹

¹Laboratoire d'Informatique Fondamentale Appliquée de Tours (LIFAT), Université François Rabelais, 37200 Tours, France

²IMASCAP, 29280, Plouzané, France

³Service Orthopédie 1, Centre Hospitalier Universitaire de Tours, Avenue de la République, 37044 Tours Cedex 09, France

Keywords: Computer Vision, RGB-Depth Camera, Camera Calibration, Spherical Object.

Abstract: RGB-Depth calibration refers to the estimation of both RGB and Depth camera parameters, as well as their relative pose. This step is critical to align streams correctly. However, in the literature there is still no general method for accurate RGB-D calibration. Recently, promising methods proposed to use spheres to perform the calibration, the centers of these objects being well distinguishable by both cameras. This paper proposes a new minimization function which constrains spheres centers positions by requiring the knowledge of sphere radius, and a previously calibrated RGB camera. We show the limits of previous approaches and their correction with the proposed method. Results demonstrate an improvement in relative pose estimation compared to the original method on the selected datasets.

1 INTRODUCTION

The recognition of a real scene to allow digital interaction is one major problem in computer vision applications. A key component of many scene understanding approaches is the determination of 3D coordinates of objects of interest. The recent increase in the availability of low cost Depth cameras offers new approaches to this problem. The combination of a Depth sensor with an RGB camera provides broadened scene information, and is referred to as an RGB-D device.

To operate accurately, it is critical to know the calibration parameters of the sensor pair, i.e. the intrinsic parameters of both cameras, and the rigid transformation between the optic centers of the two cameras. However, the quality of manufacturer RGB-D calibration differs for each camera model, and is inadequate for high precision applications. The calibration of an RGB-D camera presents new challenges against the calibration of an RGB camera:

1. Popular color feature points, such as the corners of a checkerboard, are not visible by a Depth camera.
2. Depth maps given by Depth cameras are usually noisy, with noise increasing quadratically with the distance.

3. Depth measurements are unreliable on the border of objects.

Recent RGB-D calibration approaches still suffer from a few pixels shift. Moreover, the relative pose estimation between the two cameras may be highly imprecise. These approximations come from inaccuracies between RGB and Depth matching points. The origin of these errors can be explained by the observation that Depth cameras are subject to several types of noise. The noise originates from the shapes, positions or materials of the objects present in the scene. Their effect typically increases quadratically with the distance. Some recent publications try to address this issue (Herrera C. et al., 2012; Basso et al., 2018). To the best of our knowledge, no sphere based calibration method proposed to handle it.

This paper is built upon a previous work which proposed to use spheres for close distance RGB-D calibration (Staranowicz et al., 2014). The approach requires both a calibrated RGB camera and an uncalibrated Depth camera to observe a spherical object. This approach relies heavily on 2D data to perform calibration. However, the loss of 3D information eventually leads to relative pose estimation errors.

Alternatively, several sphere based calibration methods are based on the principle of precisely determining the center of a sphere (Schnieders and Wong, 2013; Sun et al., 2016). It has been proposed to

use 3D sphere's centers for camera network relative pose estimation for either Depth cameras (Minghao Ruan and Huber, 2014), or RGB cameras (Guan et al., 2015). This approach has yet to be applied to RGB-D calibration.

Based on these ideas, we propose to perform the calibration in 3D, using the estimated spheres centers of both RGB and the Depth cameras to perform the calibration. We introduce an alternative non-linear optimization function, which only requires supplementary knowledge of the sphere radius. An evaluation of our approach upon synthetically generated data, and real data (Boas et al., 2018) shows this information can be used to improve relative pose estimation in RGB-D calibration.

In this paper, we focus on the problem of determining the Depth intrinsic parameters, and the extrinsic parameters between the RGB and the Depth camera. We intentionally omit any determination and refinement of the RGB intrinsic parameters, as we consider this step previously performed.

This paper is structured as follows. Section 2 surveys recent works related to RGB-D calibration. Section 3 summarizes the initial approach with our contribution. In Section 4, experimental data are presented. Finally, Section 5 and 6 outline the significance of our contribution.

1.1 Problem Formulation

To align both RGB and Depth channels, it is necessary to determine the intrinsic parameters of both cameras. These parameters are commonly represented with a pinhole camera model, which models the image formation process by expressing the focal distance and the center of the image (also known as the principal point) into a 3×3 matrix.

The intrinsic parameters \mathbf{K} of the two cameras are noted as ${}^{\mathcal{R}}\mathbf{K}$ and ${}^{\mathcal{D}}\mathbf{K}$, where $\{\mathcal{R}\}$ refers to the RGB camera and $\{\mathcal{D}\}$ to the Depth camera; with (f_u, f_v) the focal length and (u_0, v_0) the principal point. The two channels alignment is then carried out by a rigid transformation $(\mathbf{R}|\mathbf{t})$, known as the extrinsic parameters. This transformation is composed of a 3×3 rotation matrix \mathbf{R} and a 3×1 translation vector \mathbf{t} .

2 RELATED WORKS

RGB-D calibration methods are mainly supervised, i.e. use objects with known geometric properties to perform the calibration. The idea is to find matching points between both cameras to determine the calibration parameters.

Checkerboard Based Calibration. Similarly to RGB calibration approaches, a commonly encountered object in RGB-D calibration is a checkerboard. The matching points are the checkerboard's corners. However, on a Depth map, the corners of a checkerboard are not visible, and its border is inaccurate, which negatively influences the results. Some remarkable approaches were proposed (Zhang and Zhang, 2014; Mikhelson et al., 2014). They, however, required either the user intervention to select points, or previous intrinsic camera calibration, making these approaches unpractical.

Approaches using checkerboard can rely on Infrared images, usually available from a Depth camera, to perform the RGB-D calibration (Herrera C. et al., 2012; Darwish et al., 2017). They propose to use disparity images, built from the RGB and IR images. However, these methods are often restricted to Structured Light Depth cameras, and the distortion observed between the IR and the Depth images has to be modeled. An alternative method (Basso et al., 2018) proposes the use of two Depth correction maps. However, this approach requires significant setup, as the checkerboard must be placed on a plane at various distances, making it less practical.

Sphere Based Calibration. Another object commonly used in camera calibration is the sphere. Its advantage in RGB-D calibration lies in its distinguishability by both cameras, and its full visibility regardless of the point of view. A method for both intrinsic and extrinsic calibration (Staranowicz et al., 2014) proposes to use a sphere with unknown geometric properties, with their centers as matching points. They correct the error introduced by shifted ellipses centers, and minimize the distance between the ellipse observed by the RGB camera and the projected sphere fitted by the Depth camera. Others approaches (Shen et al., 2014; Su et al., 2018), focus on camera network extrinsic parameters estimation, and mainly use data from the Depth camera.

3 PROPOSED METHOD

3.1 Original Method Description

Our work is based on the Staranowicz et al. method (Staranowicz et al., 2014), where N spheres are presented in front of the RGB-D camera, as shown on Figure 1. A sphere is observed as a points cloud for the Depth camera, and as an ellipse for the RGB camera. The ellipse originates from the sphere projection onto the camera plane, which can be assimilated to a conic

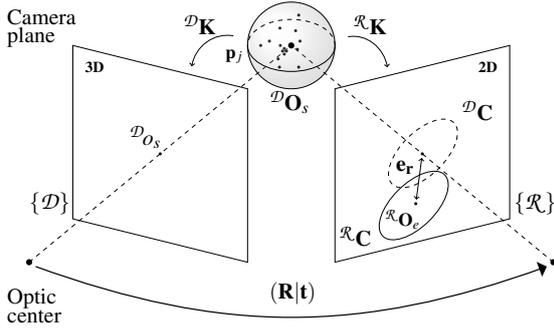


Figure 1: Overview of the RGB-D calibration scene. The sphere is seen as an ellipse ${}^{\mathcal{R}}\mathbf{C}$ with center ${}^{\mathcal{R}}\mathbf{O}_e$ for the RGB camera $\{\mathcal{R}\}$. The sphere is seen as a points cloud \mathbf{p}_j , with center ${}^{\mathcal{D}}\mathbf{O}_s$ for the Depth camera $\{\mathcal{D}\}$. This center is projected onto the Depth camera plane as ${}^{\mathcal{D}}o_s$. The two optic centers are linked by the rigid transform $(\mathbf{R}|\mathbf{t})$. The reprojection error e_r is the Euclidean distance between the ellipse center ${}^{\mathcal{R}}\mathbf{O}_e$ and the projected ellipse center of ${}^{\mathcal{D}}\mathbf{C}$.

section. The method is composed of two phases: an initialization and a non-linear optimization.

The initialization phase uses a variant of the Direct Linear Transformation (DLT) (Hartley and Zisserman, 2004) to determine a first estimate of the Depth camera intrinsic parameters ${}^{\mathcal{D}}\mathbf{K}$, as well as the extrinsic parameters $(\mathbf{R}|\mathbf{t})$. Given a set of at least six 2D to 3D points matches (projected on the plane \mathbf{Z} defined by $z = 1$), the DLT algorithm defines a homogeneous system whose resolution gives the calibration parameters. The matching points are the RGB ellipses centers ${}^{\mathcal{R}}\mathbf{O}_{e_i}$, and the Depth spheres centers ${}^{\mathcal{D}}\mathbf{O}_{s_i} := (x_i, y_i, z_i)^T$ projected on \mathbf{Z} as ${}^{\mathcal{D}}o_{s_i} := {}^{\mathcal{D}}(u_i, v_i, 1)^T$ (Equation 1).

$${}^{\mathcal{D}}\mathbf{O}_{s_i} = z_i {}^{\mathcal{D}}\mathbf{K}^{-1} {}^{\mathcal{D}}o_{s_i} \quad (1)$$

These parameters are then refined by minimizing a combination of the objective functions \mathcal{L}_1 (Equation 3) and \mathcal{L}_2 (Equation 4). Function \mathcal{L}_1 expresses the squared projection error e_r . The reprojection error is defined as the Euclidean distance between RGB ellipses centers ${}^{\mathcal{R}}\mathbf{O}_{e_i}$ and the estimated ones from the projection of the Depth spheres centers ${}^{\mathcal{D}}\mathbf{O}_{s_i}$ onto the RGB camera plane. It is represented by Equation 2 where ${}^{\mathcal{R}}z_i$ is the (z) parameter of the sphere center expressed in the RGB camera coordinate system.

$$e_{r_i} = \left\| {}^{\mathcal{R}}\mathbf{O}_{e_i} - \frac{1}{{}^{\mathcal{R}}z_i} {}^{\mathcal{R}}\mathbf{K}(\mathbf{R} {}^{\mathcal{D}}\mathbf{O}_{s_i} + \mathbf{t}) \right\| \quad (2)$$

$$\mathcal{L}_1 = \frac{1}{2N} \sum_{i=1}^N \|e_{r_i}\|^2 \quad (3)$$

Function \mathcal{L}_2 expresses the Frobenius distance between the RGB and the Depth conic sections, i.e. the

ellipses. These conic sections are represented by the 3×3 matrices ${}^{\mathcal{R}}\mathbf{C}$ and ${}^{\mathcal{D}}\mathbf{C}$; with \mathbf{C}^{-1} the dual conic of \mathbf{C} (in the case of an ellipse). The Depth conic section ${}^{\mathcal{D}}\mathbf{C}$ is estimated from the projection of the fitted Depth sphere ${}^{\mathcal{D}}\mathbf{O}_{s_i}$ onto the RGB camera plane (Staranowicz et al., 2014). $f({}^{\mathcal{D}}\mathbf{O}_{s_i})^2$ is a quadratic function of the Depth sphere center to represent the loss of accuracy of Depth cameras increasing with distance.

$$\mathcal{L}_2 = \frac{1}{2N} \sum_{i=1}^N \frac{1}{(f({}^{\mathcal{D}}\mathbf{O}_{s_i}))^2} \|{}^{\mathcal{R}}\mathbf{C}_i^{-1} - {}^{\mathcal{D}}\mathbf{C}_i^{-1}\|_F^2 \quad (4)$$

The combination of both function gives the objective function to minimize (Equation 5) :

$$\min_{{}^{\mathcal{D}}\mathbf{K}, \mathbf{R}, \mathbf{t}} \rho_1 \mathcal{L}_1 + \rho_2 \mathcal{L}_2 \quad (5)$$

Specific Aspects. Unlike the original method, the ellipses centers are not corrected, as the displacement effect is minimal at close distance. Moreover, their ellipse center correction proposition is valid only if the ellipse is oriented towards the principal point of the image, which is not the case of a standard ellipse fitting algorithm. As for Equation 4, we defined f^2 as the squared Depth sphere center distance with respect to the camera. The weight ρ_1 and ρ_2 are chosen similarly.

3.2 Proposed Objective Function

In order to reduce noise impact observed on Equation 5, we propose a new minimization function \mathcal{L}_3 (Equation 6) to better constrain the calibration parameters. Function \mathcal{L}_3 represents the mean of the squared Euclidean distances between the spheres centers ${}^{\mathcal{R}}\mathbf{O}_{s_i}$ estimated from the RGB camera, and the spheres centers ${}^{\mathcal{D}}\mathbf{O}_{s_i}$ estimated from the Depth camera. The RGB spheres positions ${}^{\mathcal{R}}\mathbf{O}_{s_i}$ are retrieved from the RGB ellipses ${}^{\mathcal{R}}\mathbf{C}_i$, the sphere radius R , as shown in Section 3.3.

$$\mathcal{L}_3 = \frac{1}{2N} \sum_{i=1}^N \|{}^{\mathcal{R}}\mathbf{O}_{s_i} - (\mathbf{R} {}^{\mathcal{D}}\mathbf{O}_{s_i} + \mathbf{t})\|^2 \quad (6)$$

The function \mathcal{L}_3 can be seen as the 3D equivalent of the function \mathcal{L}_2 in the case of fixed sphere radius. However, it does not express well the reprojection error observed on the RGB camera. We thus define a new objective function as the linear combination of \mathcal{L}_1 and \mathcal{L}_3 (Equation 7).

$$\min_{{}^{\mathcal{D}}\mathbf{K}, \mathbf{R}, \mathbf{t}} \rho_1 \mathcal{L}_1 + \rho_3 \mathcal{L}_3 \quad (7)$$

Figure 2 proposes an overview of this method, with our contribution.

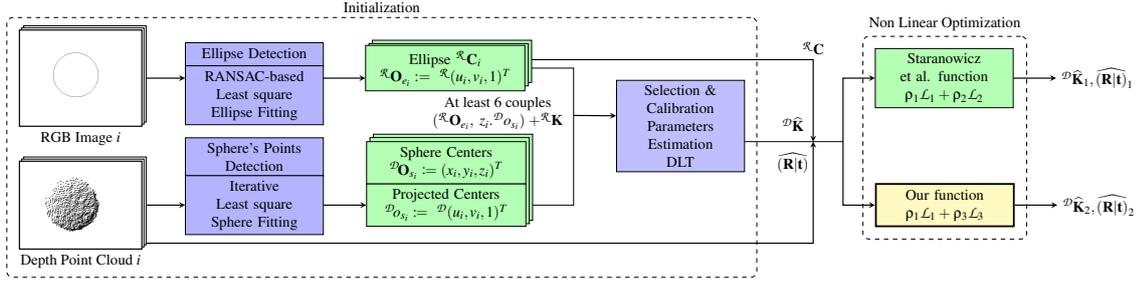


Figure 2: Simplified pipeline of the Staranowicz et al. (Staranowicz et al., 2014) method, with our proposed objective function (in yellow).

3.3 Retrieving a Sphere Center From an RGB Ellipse

To define Equation 7, it is necessary to determine the 3D Spheres centers ${}^{\mathcal{R}}\mathbf{O}_{s_i}$ with respect to the RGB camera. To retrieve the 3D sphere center ${}^{\mathcal{R}}\mathbf{O}_s$ from a 2D view, the knowledge of the conic shape ${}^{\mathcal{R}}\mathbf{C}$, the intrinsic parameters ${}^{\mathcal{R}}\mathbf{K}$ and the 3D sphere radius R are necessary. We propose to use a linear approach (Schnieders and Wong, 2013), which gives a perfect estimate given noise free data. This method is composed of four main steps:

Step 1 (3a) : The effect of ${}^{\mathcal{R}}\mathbf{K}$ on the conic ${}^{\mathcal{R}}\mathbf{C}$ is withdrawn, by normalizing it with ${}^{\mathcal{R}}\mathbf{K}^T$. The conic ${}^{\mathcal{R}}\mathbf{C}$ is transformed as the normalized conic ${}^{\mathcal{R}}\bar{\mathbf{C}}$:

$${}^{\mathcal{R}}\bar{\mathbf{C}} = {}^{\mathcal{R}}\mathbf{K}^T {}^{\mathcal{R}}\mathbf{C} {}^{\mathcal{R}}\mathbf{K} \quad (8)$$

Step 2 (3b) : A rotation ${}^{\mathcal{R}}\mathbf{U}^T$ is performed to transform the ellipse ${}^{\mathcal{R}}\bar{\mathbf{C}}$ as a circle ${}^{\mathcal{R}}\mathbf{D}$. This circle is centered at the origin $(0,0)$ of the camera plane. The rotation ${}^{\mathcal{R}}\mathbf{U}^T$ is found by applying a Singular Value Decomposition (SVD) on ${}^{\mathcal{R}}\bar{\mathbf{C}}$. ${}^{\mathcal{R}}\mathbf{U}$ is the orthogonal matrix whose columns are the eigenvectors of ${}^{\mathcal{R}}\bar{\mathbf{C}}$. Thus, the circle ${}^{\mathcal{R}}\mathbf{D}$ is defined as following:

$${}^{\mathcal{R}}\mathbf{D} = {}^{\mathcal{R}}\mathbf{U}^T {}^{\mathcal{R}}\bar{\mathbf{C}} {}^{\mathcal{R}}\mathbf{U} \quad (9)$$

The circle matrix ${}^{\mathcal{R}}\mathbf{D}$ is normalized by its first element as ${}^{\mathcal{R}}\bar{\mathbf{D}} = {}^{\mathcal{R}}\mathbf{D}/{}^{\mathcal{R}}\mathbf{D}_{00}$. The radius r of the circle can be computed as $r = \sqrt{-{}^{\mathcal{R}}\bar{\mathbf{D}}_{22}}$.

Step 3 (3c) : The sphere center \mathbf{O}_s is retrieved by using the relationship between r and the radius R of the sphere, as expressed in Equation 10.

Step 4 (3d) : The rotation ${}^{\mathcal{R}}\mathbf{U}$ is applied to reverse the rotation ${}^{\mathcal{R}}\mathbf{U}^T$, and express the sphere with respect to the ellipse as ${}^{\mathcal{R}}\mathbf{O}_s$.

$${}^{\mathcal{R}}\mathbf{O}_s = {}^{\mathcal{R}}\mathbf{U} \underbrace{\begin{bmatrix} 0 & 0 & R(\frac{\sqrt{1+r^2}}{r}) \end{bmatrix}^T}_{\mathbf{O}_s} \quad (10)$$

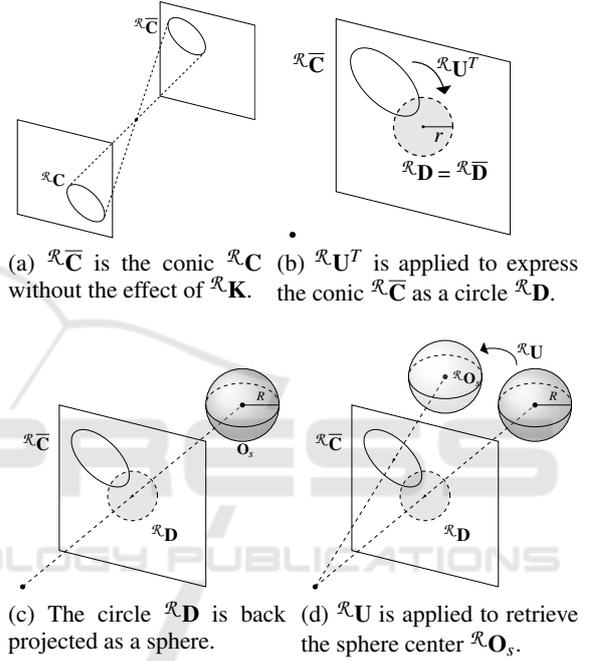


Figure 3: Overview of the steps in order to back project an ellipse \mathbf{C} into a sphere (Schnieders and Wong, 2013).

4 EXPERIMENTS

We evaluate the accuracy of the ellipse back projection approach exposed in Section 3.3 with synthetically generated data, and assess its suitability for RGB-D calibration. We then evaluate our proposition, derived from Equation 7 to Staranowicz et al. method. on several synthetic and real datasets (Boas et al., 2018). It is important to note that in order to have a fairer comparison, we use the sphere radius knowledge for both approaches.

4.1 Ellipse Back Projection Accuracy

The ellipse back projection method (Schnieders and Wong, 2013) provides the real sphere 3D position

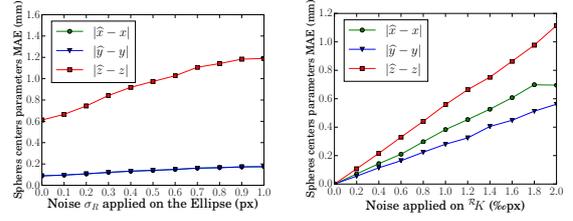
with noise free data. The method accuracy depends on the estimation of the ellipse ${}^{\mathcal{R}}\mathbf{C}$, the estimation of the intrinsic parameters of the RGB camera ${}^{\mathcal{R}}\mathbf{K}$, on the sphere position \mathbf{O}_s , and on the estimated sphere radius R . In a calibration setup, the sphere radius can be considered known with high accuracy. However, noise is typically observed on the estimations of ${}^{\mathcal{R}}\mathbf{C}$ and ${}^{\mathcal{R}}\mathbf{K}$. To further evaluate the method validity in a calibration context, synthetic scenes with known parameters were created.

Scene Description. A reference sphere of radius $R = 40 \text{ mm}$ is placed at $\mathbf{O}_s = (100, 100, 700)^T$. The sphere is projected onto the camera plane with known intrinsic parameters ${}^{\mathcal{R}}\mathbf{K}$ as $f_u = f_v = 1000$, and $(u_0, v_0) = (640, 480)$. The ellipse back projection method is then applied to estimate the sphere's center. During the experiments, a zero-mean Gaussian noise \mathcal{N} is applied on either the ellipse's points or the intrinsic parameters. To obtain a reliable metric, each experiment was performed 1000 times, and use the Mean Absolute Error (MAE) of the sphere's center parameters.

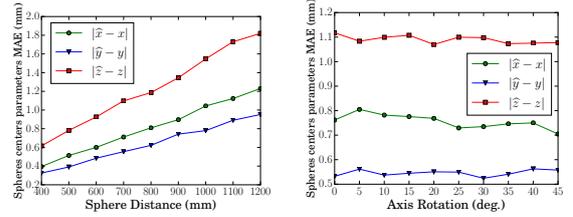
Analysis. The Figure 4a shows the results of the ellipse back projection method against noise applied on the RGB frame. This kind of noise originates from blur in the image and inaccuracies in ellipse fitting. To replicate this behavior, a zero-mean Gaussian noise $\mathcal{N}(0, \sigma_{\mathcal{R}})$ is applied on every pixel belonging to the ellipse. A Least-Square ellipse fitting algorithm is used to fit the new ellipse. On Figure 4b, the noise applied on the intrinsic parameters is observed. The noise observed on ${}^{\mathcal{R}}\mathbf{K}$ originates from inaccuracies during the RGB calibration method. A zero-mean Gaussian noise $\mathcal{N}(0, \sigma_K)$ is applied on all calibration parameters (f_u, f_v, u_0, v_0) . A study on current RGB calibration methods (Wong et al., 2011; Liu et al., 2017) allows us to estimate σ_K between 1‰ and 5‰ of the real calibration parameters values.

On Figure 4c, the sphere is placed along the vector between the origin and the reference sphere center \mathbf{O}_s , but with varying distance. The intrinsic Gaussian noise standard deviation is fixed at $\sigma_K = 2‰$. On Figure 4d, the effect of rotation is studied. A rotation is performed along the normal of the camera plane (i.e. the (z) axis) to modify the sphere's center position. The center position goes from $(\sqrt{2} * 100, 0, 700)^T$ (0 deg.) to the initial reference sphere center $(100, 100, 700)^T$ (45 deg.). The intrinsic Gaussian noise σ_K is also applied.

Figure 4 shows that the back projection method is sensible to noisy ellipse detection and inaccurate intrinsic parameter estimation. In all cases, the error is mostly on the (z) parameter estimation, (x, y) parameters estimation being more accurate. This error



(a) MAE against Gaussian noise $\mathcal{N}(0, \sigma_{\mathcal{R}})$ applied on the ellipse points. (b) MAE against Gaussian noise $\mathcal{N}(0, \sigma_K)$ applied on ${}^{\mathcal{R}}\mathbf{K}$.



(c) MAE with increasing distance and Gaussian noise $\mathcal{N}(0, \sigma_K = 2‰)$. (d) MAE with increasing angle and Gaussian noise $\mathcal{N}(0, \sigma_K = 2‰)$.

Figure 4: Ellipse back projection ${}^{\mathcal{R}}\mathbf{C}$ into a sphere \mathbf{O}_s mean absolute error (MAE) for several parameters.

is amplified by the distance between the sphere and the camera, but is robust to angle variations. At most, the back projection error is at 2 mm . Nonetheless, the suitability of this approach for RGB-D calibration will be demonstrated below.

It is important to note that this approach requires the sphere radius to be known with high accuracy (no more than a few ‰ of error). However, this is not a problem in the case of a 3D printed sphere (with sub-millimeter printing accuracy).

4.2 Synthetic Scene Setup

The synthetically generated sequences are designed to evaluate the robustness of both methods against various measurements noise (see Figure 5).

Data Generation. An RGB-D camera couple with known extrinsic parameters is placed in a 3D scene. The spheres positions and radius, the transformation $(\mathbf{R}|\mathbf{t})$ and the intrinsic parameters ${}^{\mathcal{R}}\mathbf{K}$ and ${}^{\mathcal{D}}\mathbf{K}$ are known and fixed.

The RGB camera is placed 30 mm above the Depth camera, with no rotation. The rotation vector is defined as $\theta = (\theta_x, \theta_y, \theta_z)^T$. θ is the representation of \mathbf{R} as Euler angles in degrees. The translation vector is defined as $\mathbf{t} = (t_x, t_y, t_z)^T$ in mm .

As proposed by Staranowicz et al. (Staranowicz et al., 2014), twenty 3D points ${}^{\mathcal{D}}\mathbf{O}_{s_i}, i \in [1, 20]$ are randomly generated at the intersection of the field of view of the two cameras. For each sphere center ${}^{\mathcal{D}}\mathbf{O}_{s_i}$,

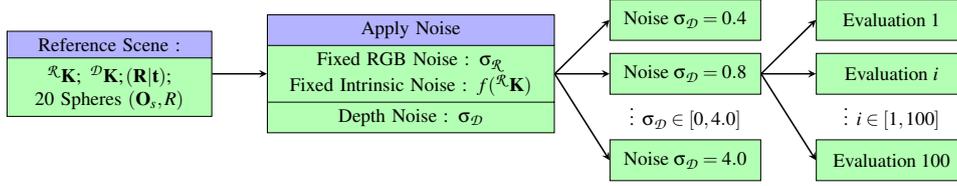


Figure 5: Overview of synthetic data generation.

100 3D points $\mathbf{p}_{ij}, j \in [1, 100]$ are randomly generated on the periphery of half of the sphere by using the parametric equation of a sphere. The input data of the algorithm is generated from these twenty sets of a hundred points. These spheres are projected onto the RGB camera plane to obtain the associated ellipses. Associated Depth maps are generated by projecting the points \mathbf{p}_{ij} onto the Depth camera plane.

Noise Application. The noise on the RGB camera is modeled with a zero-mean Gaussian noise on the RGB frames similarly to Section 4.1, with $\sigma_{\mathcal{R}} = 0.6$. A fixed Intrinsic noise on $\mathcal{R}\mathbf{K}$ is applied by multiplying each calibration parameters by the value $\eta = 1 + 2\%$ in order to have a similar noise on each sequence. Its value is determined similarly as $\sigma_{\mathcal{K}}$ in Section 3.3. Depth camera noise is modeled by a shift on the sphere points \mathbf{p}_{ij} , and thus $\mathcal{D}\mathbf{O}_{s_i}$, to characterize any kind of error. This noise is labeled $\sigma_{\mathcal{D}}$, and is our variable of interest. All tests were evaluated a hundred times.

4.3 Synthetic Data Comparison

Figure 6 summarizes the results after evaluating both methods on synthetically generated data. Increasing Depth noise $\sigma_{\mathcal{D}}$ is applied on Depth data. The Mean Absolute Error (MAE) is used to compare every parameters involved in the RGB-D calibration, as well as the 2D (e_r) and 3D reprojection error (E_r). The 3D reprojection error (Equation 11) is the Euclidean distance between the real spheres centers and the ones estimated from Section 3.3 approach. This metric more accurately accounts for Depth evaluation errors, as the 2D reprojection error tends to hide depth displacements (in the case both cameras are close).

$$E_{r_i} = \|\mathcal{R}\mathbf{O}_{s_i} - (\mathbf{R}^{\mathcal{D}}\mathbf{O}_{s_i} + \mathbf{t})\| \quad (11)$$

On Figure 6, better results are observed for our proposition (Equation 7), especially for parameters involving distances, as this approach allows to recover depth information from the RGB camera. This is particularly visible on Figures 6a and 6e, with the focal estimation f_u, f_v and the translation along the depth axis t_z . Both reprojection errors are higher on error free Depth data as our proposition is more sensible to RGB noise than the Staranowicz et al. method.

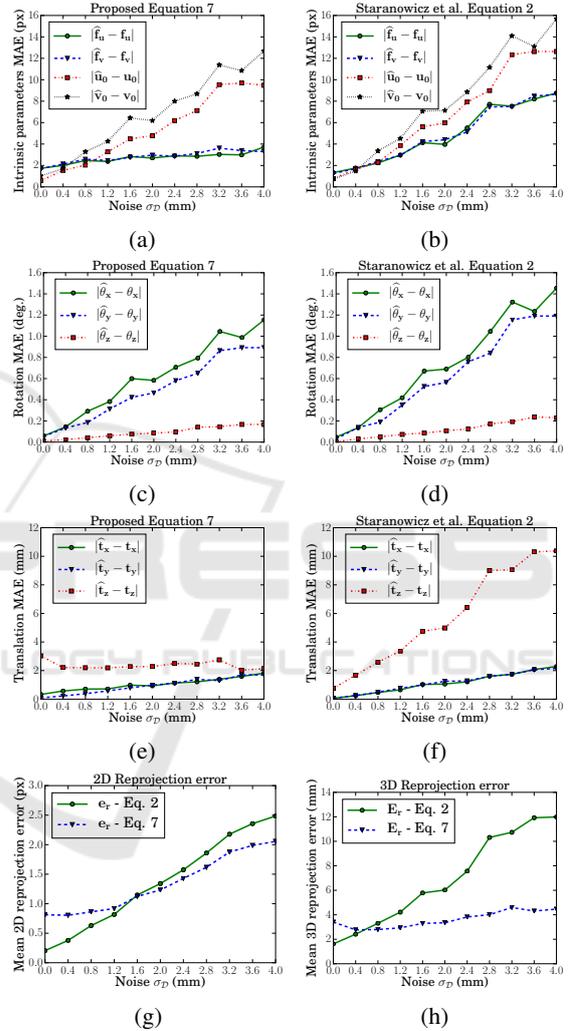


Figure 6: Simulation results of both approaches : (a) & (b) - Mean absolute Intrinsic parameters errors with increasing noise on Depth data. (c) & (d) - Mean absolute Translation error. (e) & (f) - Mean absolute Rotation error. (g) & (h) - 2D and 3D Reprojection error.

The main improvement is particularly visible on Figure 6h, with 3D reprojection error E_r results improving by many millimeters, something that is due to the better estimated Depth intrinsic parameters and relative pose between both cameras. The constraint on sphere centers introduced by the ellipse back pro-

jection attenuates calibration parameters divergence as σ_D increases. It is interesting to note that E_r closely follows the mean absolute translation errors on \mathbf{t}_z . Obviously, the more accurate the RGB intrinsic parameters estimation, as well as the ellipse detection, the better our results. Considering the ellipse back projection method introduced on average an error of 2 mm, an improvement of this approach should further improve Equation 7 results.

4.4 Real Data Comparison

Both approaches are evaluated on a real dataset, comprising three sequences, each representing a different scenario. For each of these sequences, RGB frames, Depth points clouds, and manufacturer intrinsic parameters are available. The Random and Equidistant sequences have been taken with a separated RGB-D camera couple, where the Depth camera is placed above the RGB camera. The Random sequence consists of 72 RGB-D frames couples of a randomly distributed sphere’s position. The Equidistant sequence consists of 27 RGB-D frames couples of a sphere positioned to equidistantly cover the field of view of both cameras. Both datasets use a Hololens (Microsoft, Redmond, USA) front facing RGB camera and Picoflexx (PMD Technologies, Siegen, Germany) Depth camera couple. The Double sequence has been taken with a RealSense SR300 (Intel, Santa Clara, USA), which allows us to compare against the manufacturer given extrinsic parameters ($\mathbf{R}(\mathbf{t})$). This sequence consists of 20 RGB-D frames couples of a support with two spheres. The spheres are painted in blue to ease the ellipse detection step.

Data Extraction. Ellipse detection follows the next approach. An initial estimate is obtained by color thresholding. A Canny-edge algorithm is then performed around the ellipse’s outline. The final ellipse is determined by a RANSAC based Least-Square (in the algebraic distance sense) ellipse-fitting method to have a robust estimate. Points cloud segmentation is performed by an internal algorithm. Finally, an Iterative Least-Square sphere fitting using the Radius R knowledge is performed to obtain ${}^D\mathbf{O}_s$. This algorithm minimize the difference between the radius R and the Euclidean distance of the points with respect to ${}^D\mathbf{O}_s$ using the centroid as the initial guess. The sphere center is then projected back onto the Depth camera plane using the manufacturer intrinsic para-

eters. All RGB-D frames couples are selected, as long as their sphere fitting error (Root Mean Square error of the points distance to the center) is below a fixed threshold. These couples allows to build the homogeneous system according to the DLT algorithm. The linear estimate is then refined by optimization, using the Levenberg-Marquardt algorithm.

Calibration results are shown on Table 1. A re-projection error improvement is visible for our proposition. The resulting calibration parameters are, however, quite different. The synthetic results suggest that the estimated parameters associated with the new approach are more accurate. Figure 7 gives a visualization of the RGB-D alignment at different step of the calibration. As expected by the synthetic results, the visualization is similar for the Staranowicz et al. and our method. However, the improvement is clearly visible against the manufacturer provided calibration, and the initialization phase.

5 DISCUSSION AND FUTURE WORK

The results demonstrate that the initial method (Staranowicz et al., 2014) is quite sensitive to noise, especially on depth sphere center ${}^D\hat{\mathbf{O}}_{s_i}$ estimate. This does not impact to a great extent the 2D reprojection error, but ultimately leads to inaccurate translation $\hat{\mathbf{t}}_z$ and focal (f_u, f_v) estimates. We identify this translation error as one of the main drawbacks of the Staranowicz et al. approach. Using the back projection of RGB ellipses into spheres allows for a more precise relative pose estimation. We believe our approach allows to use the advantages of both sensors.

In practice, it is possible to accurately detect an ellipse on an RGB frame in a controlled calibration scene. However, the sphere fitting on a points cloud offers more challenges, because of the presence of outliers and the depth distance error of the Depth camera, leading to a difference of several millimeters. Our contribution allows to mitigate this aspect, by making use of the RGB ellipses which can be estimated with the knowledge of the sphere radius. However, our approach requires a highly accurate calibration and ellipse detection.

Constraining the (z) axis is particularly important, considering the noise increasing with the distance ty-

Table 1: 2D Reprojection error e_r by applying both approaches on real datasets.

Metric \ Dataset	Random		Equidistant		Double		
	Stara.	Our	Stara.	Our	Stara.	Our	Manufacturer
e_r (px)	2.25 ± 1.28	2.17 ± 1.10	2.49 ± 1.29	2.33 ± 1.05	7.03 ± 2.76	6.48 ± 2.82	28.8 ± 3.61

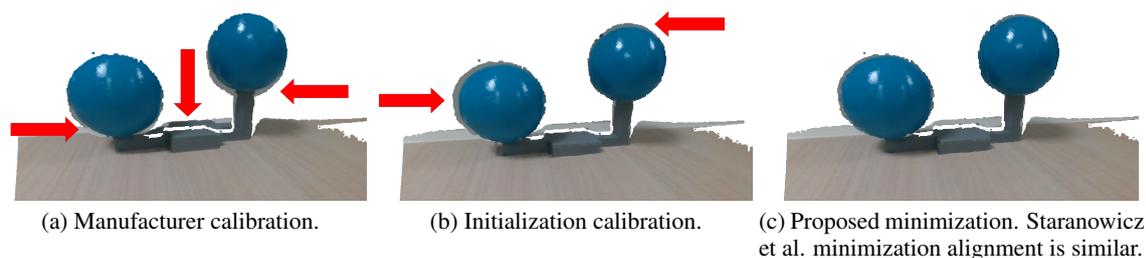


Figure 7: Alignment results on the Double dataset. Misalignments are highlighted in red.

pically observed by Depth cameras. These methods are restricted to low range RGB-D calibration, as no modelization of the Depth noise is proposed. In future work, we seek to deal with the noise increasing with the distance observed by Depth camera, and improve the ellipse back projection accuracy.

6 CONCLUSION

This work demonstrates the sensitivity of the Staranowicz et al. method (Staranowicz et al., 2014) to noisy measurements. We proposed a new minimization objective function to better constrain the relative translation and the intrinsic parameters of the Depth camera. Multiple simulations with both synthetic data that reproduces real conditions and real data were performed to determine and quantify the evolution of calibration parameters. We showed that using the 3D centers of spheres, instead of their 2D projection allows to improve the calibration parameters estimation, especially with high accuracy cameras and noisy Depth data.

REFERENCES

- Basso, F., Menegatti, E., and Pretto, A. (2018). Robust Intrinsic and Extrinsic Calibration of RGB-D Cameras. *IEEE Transactions on Robotics*, pages 01–18.
- Boas, D. J. T., Ramel, J.-Y., Slimane, M., and Poltaretskyii, S. (2018). Rgb-d sphere-based calibration dataset. <http://www.rfai.li.univ-tours.fr/tools-and-demos/>.
- Darwish, W., Tang, S., Li, W., and Chen, W. (2017). A New Calibration Method for Commercial RGB-D Sensors. *Sensors*, 17(6):1204.
- Guan, J., Deboeverie, F., Slembrouck, M., van Haerenborgh, D., van Cauwelaert, D., Veelaert, P., and Philips, W. (2015). Extrinsic Calibration of Camera Networks Using a Sphere. *Sensors*, 15(8):18985–19005.
- Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.
- Herrera C., D., Kannala, J., and Heikkila, J. (2012). Joint Depth and Color Camera Calibration with Distortion Correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):2058–2064.
- Liu, Z., Wu, Q., Wu, S., and Pan, X. (2017). Flexible and accurate camera calibration using grid spherical images. *Optics Express*, 25(13):15269–15285.
- Mikhelson, I. V., Lee, P. G., Sahakian, A. V., Wu, Y., and Katsagelos, A. K. (2014). Automatic, fast, online calibration between depth and color cameras. *Journal of Visual Communication and Image Representation*, 25(1):218–226.
- Minghao Ruan and Huber, D. (2014). Calibration of 3D Sensors Using a Spherical Target. *Proceedings of the International Conference on 3D Vision*, 01:187–193.
- Schnieders, D. and Wong, K.-Y. K. (2013). Camera and light calibration from reflections on a sphere. *Computer Vision and Image Understanding*, 117:1536–1547.
- Shen, J., Xu, W., Luo, Y., Su, P.-C., and Cheung, S.-c. S. (2014). Extrinsic calibration for wide-baseline RGB-D camera network. *IEEE 16th International Workshop on Multimedia Signal Processing*.
- Staranowicz, A., Brown, G. R., Morbidi, F., and Mariottini, G. L. (2014). Easy-to-Use and Accurate Calibration of RGB-D Cameras from Spheres. In *Image and Video Technology*, volume 8333, pages 265–278. Springer Berlin Heidelberg.
- Su, P.-C., Shen, J., Xu, W., Cheung, S.-C., and Luo, Y. (2018). A Fast and Robust Extrinsic Calibration of RGB-D Camera Networks. *Sensors*, 18(1):235.
- Sun, J., He, H., and Zeng, D. (2016). Global Calibration of Multiple Cameras Based on Sphere Targets. *Sensors*, 16(1):77.
- Wong, K.-Y. K., Zhang, G., and Chen, Z. (2011). A Stratified Approach for Camera Calibration Using Spheres. *IEEE Transactions on Image Processing*, 20(2):305–316.
- Zhang, C. and Zhang, Z. (2014). Calibration Between Depth and Color Sensors for Commodity Depth Cameras. In *Computer Vision and Machine Learning with RGB-D Sensors*, pages 47–64. Springer International Publishing.