Citation: A. Pérez-Yus, G. López-Nicolás, J. J. Guerrero. A novel hybrid camera system with depth and fisheye cameras. International Conference on Pattern Recognition (ICPR), pp. 2789-2794. Cancún, México, December 4-8, 2016.

A Novel Hybrid Camera System with Depth and Fisheye Cameras

A. Perez-Yus, G. Lopez-Nicolas and J.J. Guerrero
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Spain

***Abstract*—We introduce a novel hybrid camera configuration composed by a fisheye camera attached to a RGB-D system. Current RGB-D sensors provide the 3D information and scale of the scene, but they are limited by a small field of view. In contrast, wide field of view cameras capture a larger portion of the scene, but providing highly distorted images that require specific algorithms. By coupling a fisheye camera to a RGB-D system we take advantage of both type of cameras and reduce their drawbacks. The system provides a portion of the fisheye image with depth data and the capacity to use this seed information to perform scaled operations in the complete image. We also present a calibration procedure of the system in order to map depth information to the wide angle image. With this purpose, we have developed the first depth-fisheye calibration algorithm nurturing from state-of-the-art camera models and methods. Several experiments test the accuracy of the system with real images.**

I. Introduction

The development of cheap RGB-D cameras in the consumer market has been a breath of fresh air in the fields of robotics and computer vision. With them it was possible to retrieve the 3D of the scene simultaneously with the RGB image, with a single device and no extra computational cost. But the color camera specifications can fall a bit short depending on the application, which has led some researchers to calibrate the system with an external camera in order to get higher resolution images [1].

Here we propose to substitute the conventional RGB camera (Fig. 1(a)) with a fisheye camera to increase the field of view (FoV) of the system (Fig. 1(b)). In many robotic applications where the perception of the environment is relevant (e.g. navigation, SLAM, object recognition), a wider FoV can be useful to reduce the required number of images and the computational cost. We have attached a camera with a fisheye lens to the RGB-D sensor (Fig. 1 (c)), and both devices have been calibrated in order to map the depth information in the fisheye camera image (Fig. 1(d)). Although there is a great difference in FoV between both devices, the portion of the image with shared information can be highly valuable for the aforementioned applications.

Up to our knowledge this is the first work combining both types of cameras. As a result, currently there is no available tool to calibrate this hybrid camera system in order to be straightforwardly used in any further applications. In this work we present the procedure to perform the calibration of this new device, including the intrinsic and extrinsic calibration of both fisheye and depth cameras.

Fisheye cameras require more complex models than conventional camera models due to their extremely wide field of view and the distortion of the images. This is the main reason why existing approaches of extrinsic calibration between RGB and Depth cameras are not applicable to this problem. Using a explicitly designed camera model to calibrate the fisheye camera alone is first required in this method. We propose using the Scaramuzza's omnidirectional calibration method [2] to this end. To calibrate the whole system we propose and evaluate two alternative methods. The first consists in calibrating the intrinsics of both cameras separately to then compute the extrinsics and finally perform the calibration of the depth measurements and distortion. In the second only the fisheye is required to be calibrated on its own, and the depth camera is jointly calibrated with the relative pose of both devices by doing a non-linear minimization of the reprojection error in both fisheye and depth images. We present results comparing both methods and showing their accuracy on real images.

(a) View from RGB-D camera (b) View from fisheye camera

(c) Our proposed hybrid system (d) Depth mapped to the fisheye image

Fig. 1: (a) Scene view from a conventional RGB-D camera. (b) Same scene view from a fisheye camera. (c) Our proposal: hybrid camera system with Depth and Fisheye. (d) After the calibration we can map the depth to the whole scene image.

II. Related work

The interest in fusing visual information with range data has been approached for more than ten years now [3]. The most popular range sensors back then were the laser range finder (LRF). Not much later, works concerning the extrinsic calibration of cameras to LRF start appearing. For instance, the work of Zhang and Pless [4] presented a method of

camera - 2D LRF extrinsic calibration that presents some similarities with the algorithm proposed in this paper, like the utilization of a checkerboard as calibration pattern. Recent similar approaches such as [5], [6] propose minimal solutions for the same problem with outperforming results. Scaramuzza et al. [7] proposed another method of extrinsic calibration of a 3D laser scanner (a 2D LRF mounted on a rotating platform already calibrated) and any central camera (perspective or omnidirectional). Though it does not require to use any calibration pattern, a set of point correspondences between the two images must be selected manually. Gomez-Ojeda et al. [8] use structural corners to perform extrinsic calibration between a 2D LRF and a camera.

Usual range sensors are based on structured light or Time-of-Flight. Some of them also port a RGB camera next to the Depth camera, which makes them part of the so-called RGB-D camera type. Though most drivers for these cameras include precalibrated parameters to map the information between both sensors, it is often recommended to calibrate the RGB-D camera due to small differences that can be present due to the manufacturing process [9]. To improve the quality of the color images, some researchers use high resolution external cameras. In this vein Herrera et al. [1] proposed a method to calibrate both the intrinsics and extrinsics of a RGB-D plus external camera system. They also noted the absence in other methods of the correction of depth distortion not included in the default calibration and they proposed a method using the observed disparity instead of the metric coordinates as [10] did. In [11], a toolbox that calibrates both laser range sensor or a Kinect to a color camera in one single shot is presented. Mikhelson et al. [12] ease the process of calibrating the extrinsics of a depth-color camera pair by proposing an online method which removes the need to recalibrate the intrinsics once they are known.

Recently, other researches focused on the extension of the FoV of the depth sensor: [13] used two planar mirrors as a catadioptric extension of the RGB-D device and [14] a consumer set of wide angle lens. Fernandez-Moral et al. [15] proposed a method of extrinsic calibration without pattern to calibrate a RGB-D multi-camera rig. However, up our knowledge there are not previous works combining fisheye cameras and RGB-D devices.

III. System Description

We have created a hybrid camera system by rigidly coupling a high resolution camera with a fisheye lens to an Asus Xtion Pro Live RGB-D sensor (Fig. 1 (c)). The difference in field of view is large, as Fig. 1 shows. On the one hand, the FoV of the RGB-D camera may be too small for many applications, specially in close distances. Several works have shown the advantages of wide field of view (or omnidirectional) cameras in robotics and computer vision applications [16], [17]. On the other hand, the depth perception can help detecting obstacles, providing scale or enhancing the recognition at least in one portion of the scene. Next we describe the camera models used in this work before moving to the calibration itself.

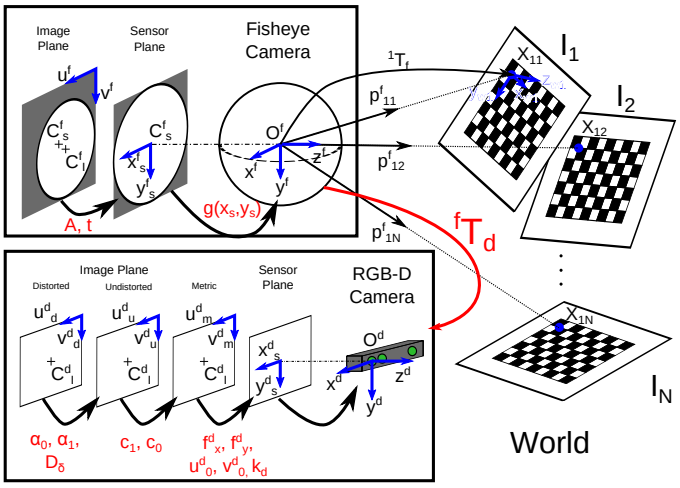


Fig. 2: Scheme containing the main coordinate systems and parameters of both cameras and its interaction with the world. In red the intrinsic and extrinsic parameters to calibrate.

A. Fisheye camera model

We chose the parametric camera model described by Scaramuzza et al. in [2], which considers the omnidirectional image as a highly distorted image and the calibration consists in retrieving the parameters of the polynomial which describes this distortion. With this model it is not necessary to provide a specific model of the sensor and works with all kind of projective, catadioptric or dioptric cameras. Although in this work we focus on fisheye cameras, the usage of this model makes our approach valid for the other types of cameras.

Using that model, the world points $\mathbf{X}^f = (X, Y, Z)$ have the origin of coordinates in the optical center of the camera O^f, where the coordinate system have the z^f component following the axis of the (cata)dioptric system (Fig. 2). Orthogonal to the z^f axis there is the Sensor Plane (x_s, y_s), a theoretical plane where the coordinates are still metrical. The images are represented in the Image Plane $\mathbf{u} = (u, v)$, where the position of the points is expressed in pixels. It is assumed that there is a misalignment and deformation between the Image Plane and the Sensor Plane, given by an affine transformation $\mathbf{x}_s = \mathbf{A}\mathbf{u} + \mathbf{t}$, where $\mathbf{t}$ is the image center, $\mathbf{t} = (u_0, v_0)$. The vector $\mathbf{p}$ pointing at the world point $\mathbf{X}$ from O^f follows the equation:

$$\lambda \cdot \mathbf{p} = \lambda \mathbf{g}(\mathbf{A}\mathbf{u} + \mathbf{t}) = P \cdot \mathbf{X}, \qquad \lambda > 0 \tag{1}$$

where P is the perspective projection matrix and the function $\mathbf{g}(x_s, y_s)$ is defined as follows:

$$\mathbf{g}(x_s, y_s) = (x_s, y_s, f(x_s, y_s))^T = (x_s, y_s, f(\rho_s))^T \tag{2a}$$

$$f(x_s, y_s) = a_0 + a_2\rho_s^2 + \ldots + a_N\rho_s^N \tag{2b}$$

where the function f is modeled as a Taylor expansion defined by the polynomial whose coefficients are $a_0, a_2, \ldots, a_N$, and where $\rho_s = \sqrt{x_s^2 + y_s^2}$. These coefficients along with the matrix $\mathbf{A}$ and vector $\mathbf{t}$ are the parameters needed to calibrate the intrinsics.

B. Depth camera model

The proposed intrinsic model of the depth camera is based in [1]. In essence is a standard projective camera model (as

the depth camera works with a conventional IR camera) with radial and tangential distortion correction. The images captured by the depth sensor have in every pixel a value of disparity (in disparity units, *du*) which increases with depth. Some drivers provide an automatic conversion of these values to metric distances given an internal calibration of the camera during the manufacturing process. However, to increase the accuracy of the measurements, we do not only calculate the parameters mapping these pixels to the real world, but also recover the metric distances per-pixel.

The IR camera model, based on [18], transforms coordinates from the Image Plane to the Sensor Plane via the focal distances f_x^d and f_y^d, the position of the center $C_I^d = (u_0^d, v_0^d)$ and some distortion coefficients $k_d = \{k_1, k_2, k_3, k_4, k_5\}$ (Fig. 2). If we have a metric value of depth per-pixel z_d, the metric 3D coordinates of the point $\mathbf{X}^d = (X^d, Y^d, Z^d)$ in the world with respect to the optical center O^d would be:

$$\mathbf{X}^d = (X^d, Y^d, Z^d) = (x_s z_d, y_s z_d, z_d) \tag{3}$$

However, we must transform the disparity values of the images as received from the camera to the metric values, which is done in two phases. First, in [1] it is mentioned the existence of a fixed error pattern that distorts the depth image with a per-pixel offset which is obtained from the intrinsic calibration. In the case of the sensor we use it follows the function:

$$d_u = d_d + \mathbf{D}_\delta(u_d, v_d) \cdot exp(\alpha_0 - \alpha_1 d_d) \tag{4}$$

where d_u is the resulting undistorted disparity and d_d the distorted disparity. Second, to convert the values of the pixels given in disparity units (*du*) to metric units, it is necessary to get the following coefficients c_1 and c_0 which forms the following equation:

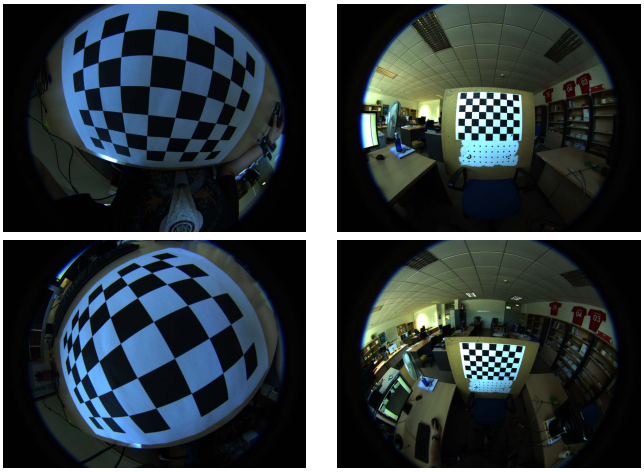
$$z_d = \frac{1}{c_1 \cdot d_u + c_0} \tag{5}$$

The set of parameters obtained from the intrinsic calibration of the depth camera are: the focal lengths (f_x^d, f_y^d), the image center (u_0^d, v_0^d), the distortion coefficients k_d, the disparity-depth transformation coefficients c_1 and c_0, the matrix $\mathbf{D}_\delta(u_d^d, v_d^d)$ and the exponential decay parameters α_0 and α_1.

IV. Depth-Fisheye camera calibration

Our algorithm use a checkerboard on a flat surface as main calibration pattern. The calibration requires the capture of several images of the planar checkerboard from different points of view, watching carefully that both the checkerboard in the external camera and the planar surface in the depth image are observed at the same time. The dimensions of the pattern are known and used as input for the algorithm. With this setup we have tackled this calibration problem from two alternate approaches:

1) Calibrating the intrinsic parameters of the fisheye and then performing a joint calibration of the rest of parameters involved in the system all at once.
2) Separating the process in stages, performing first the intrinsic calibration of the cameras separately, then retrieve the extrinsic parameters and finally compute the distortion correction and proper conversion from disparity to metric units.



(a) (b)

Fig. 3: Example of images from the sets of fisheye intrinsic calibration (a) and depth intrinsic/extrinsic calibration (b).

The specific nature of a the proposed camera system prevents the intrinsic calibration of the fisheye camera to be jointly calibrated properly with the extrinsic parameters. It is caused by the fact that to get an accurate intrinsic fisheye calibration it is necessary to fill the fisheye images with the checkerboard as much as possible (Fig. 3a). But it requires to pose the camera too close to the board that the depth sensor cannot retrieve information. Placing the camera pair at a reasonable distance makes the system prone to fail in the estimation of the intrinsic parameters of the fisheye camera because the distortion cannot be properly perceived (Fig. 3b). The fisheye camera can be easily calibrated by using the toolbox released by [2]. The two alternate methods to perform the calibration of our system are described in the following sections.

A. Joint calibration of fisheye and depth cameras

The first calibration methods for structured light-based RGB-D systems used images from the IR camera to perform traditional camera to camera extrinsic calibration [19]. However, as [1] pointed out, the depth images are not perfectly aligned with the IR images. The work [1] uses disparity images as mean to calibrate the intrinsics of the depth camera as well as the extrinsics of the system. Our method is inspired by this approach, introducing some modifications to make it suitable for our system.

We assume that the checkerboard is on a planar surface to relate in each image pair the pose of the fisheye camera from the checkerboard to its plane equation. As the checkerboard cannot be observed by the depth camera, we provide the points from its surface by manually selecting the vertices of the polygon which contains what is visible of the board in each image. That way we get the disparity values of a set of pixels belonging to the board. To relate the pose of the checkerboard as seen by the fisheye camera to the depth measurements, we need to have a rough estimation of the depth camera parameters. When the board has known

shape, e.g. rectangular, an initial estimate can be obtained selecting the corners from the board and using homographies as explained in [4]. The calibration of the final values of the depth intrinsics will be done jointly with the extrinsics solving the minimization problem, so the values set at this point just help the convergence. In the following sections the main parts of the method are detailed.

1) Retrieving camera pose from fisheye images: This section describes how for each calibration image I_i we obtain the pose of the fisheye camera with respect to the pattern ${}^{I_i}\mathbf{T}_f \in \mathbb{R}^{3\times4}$ (Fig. 2). It is formed by the rotation matrix ${}^{I_i}\mathbf{R}_f \in \mathbb{R}^{3\times3} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$ and the translation vector ${}^{I_i}\mathbf{t}_f \in \mathbb{R}^{3\times1}$. Calling $\mathbf{X}_{ij} = [X_{ij}, Y_{ij}, Z_{ij}]$ the 3D coordinate of every of the j points from the pattern in the pattern coordinate system and $\mathbf{u}_{ij} = [u_{ij}, v_{ij}]^T$ the correspondent pixel coordinates in the Image Plane, from (1) we get the following equation:

$$\lambda_{ij} \cdot \mathbf{u}_{ij} = \lambda_{ij} \cdot \begin{bmatrix} u_{ij} \\ v_{ij} \\ a_0 + ... + a_N \rho_{ij}^N \end{bmatrix} = [\mathbf{r}_1^i, \mathbf{r}_2^i, \mathbf{r}_3^i, \mathbf{t}^i] \cdot \begin{bmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \\ 1 \end{bmatrix} \quad (6)$$

We can assume without loss of generality that $Z_{ij} = 0$ for all points in the pattern, as they all belong to a planar surface in $X_{ij} - Y_{ij}$. If we multiply both sides vectorially by $\mathbf{u}_{ij}$:

$$\lambda_{ij} \cdot \mathbf{u}_{ij} \wedge \mathbf{u}_{ij} = \begin{bmatrix} u_{ij} \\ v_{ij} \\ a_0 + ... + a_N \rho_{ij}^N \end{bmatrix} \wedge [\mathbf{r}_1^i, \mathbf{r}_2^i, \mathbf{t}^i] \cdot \begin{bmatrix} X_{ij} \\ Y_{ij} \\ 1 \end{bmatrix} = 0 \quad (7)$$

Which gives the following system for every image I_i:

$$v_j \cdot (r_{31}X_j + r_{32}Y_j + t_3) - f(\rho_j) \cdot (r_{21}X_j + r_{22}Y_j + t_2) = 0 \quad (8a)$$

$$f(\rho_j) \cdot (r_{11}X_j + r_{12}Y_j + t_1) - u_j \cdot (r_{31}X_j + r_{32}Y_j + t_3) = 0 \quad (8b)$$

$$u_j \cdot (r_{21}X_j + r_{22}Y_j + t_2) - v_j \cdot (r_{11}X_j + r_{12}Y_j + t_1) = 0 \quad (8c)$$

In these equations we know the position of the points in both the world and the image (X_j, Y_j, u_j, v_j) and the function $f(\rho)$ from a previous calibration of the fisheye camera. Having these three equations for each of the n points $\mathbf{u}_j$ in the pattern we need at least $n = 3$ points to solve the system. The checkerboards always contain more than three points, which generates an overdetermined system. To solve this system we reformulate it as follows:

$$M \cdot H = \begin{bmatrix} M_1 \\ .. \\ M_j \\ .. \\ M_L \end{bmatrix} \cdot H = 0 \quad (9)$$

where

$$H = [r_{11}, r_{12}, r_{21}, r_{22}, r_{31}, r_{32}, t_1, t_2, t_3]^T \quad (10a)$$

$$M_j = \begin{bmatrix} 0 & f(\rho_j)X_j & -v_jX_j \\ 0 & f(\rho_j)Y_j & -v_jY_j \\ -f(\rho_j)X_j & 0 & u_jX_j \\ -f(\rho_j)Y_j & 0 & u_jY_j \\ v_jX_j & -u_jX_j & 0 \\ v_jY_j & -u_jY_j & 0 \\ 0 & f(\rho_j) & -v_j \\ -f(\rho_j) & 0 & u_j \\ v_j & -u_j & 0 \end{bmatrix}^T \quad (10b)$$

whose solution can be computed using SVD. The rest of the elements of the matrix $\mathbf{R}_i$, i.e. $\mathbf{r}_3 = [r_{13}, r_{23}, r_{33}]$ are obtained with the cross product of $\mathbf{r}_1$ and $\mathbf{r}_2$. This is repeated for each image I_i, computing all the transformations ${}^{I_i}\mathbf{T}_f$.

2) Non-linear minimization: The optimization is a minimization of the weighted sum of squares of the reprojection errors over all parameters, which we perform using cost function similar to the one from [1]. The cost function has two terms, one from each camera. The costs of the fisheye camera are the sum of the Euclidean distance between the pixel position of the corner $\mathbf{p}_{ij}$, and the position of its reprojection $\hat{\mathbf{p}}_{ij}$ given the pose of the checkerboard in every image ${}^{I_i}\mathbf{T}_f$ and the intrinsic parameters of the fisheye already calculated. Although the inner parameters of the camera are pre-computed, it is necessary to add this error because of the optimization of the poses ${}^{I_i}\mathbf{T}_f$. The costs of the depth camera are the sum of squared difference among the measured disparity from the image d_d and its predicted disparity $\hat{d}_d$. To compute the predicted disparity, we know from Section IV-A1 the pose of the camera with respect to a group of points lying on a plane for every image I_i. Therefore, we can obtain the plane where the checkerboard is on from these transformations. The normal $\mathbf{n}_i$ and the distance to the origin D_i are:

$$\mathbf{n}_i = \mathbf{r}_3^i, \qquad D_i = {\mathbf{r}_3^i}^T \mathbf{t}^i \quad (11)$$

The plane equation from the fisheye camera frame is then:

$${\mathbf{n_i}}^T \mathbf{X} - D_i = 0 \quad (12)$$

for each image. With the rough initial estimate of the parameters we can compute the predicted disparity and therefore the depth cost. To avoid including the whole matrix $\mathbf{D}_\delta(u_d, v_d)$ which adds many parameters to the optimization (in our case 640×480), it is optimized independently as done in [1]. Because of that, the depth cost is computed with the disparity values from the Undistorted Image Plane.

The main cost function is the following, weighted by the inverse of the measurement variance due to the difference in the units of the terms:

$$J = \beta \frac{\sum \|\hat{\mathbf{p}} - \mathbf{p}\|^2}{\sigma_f^2} + \frac{\sum \|\hat{d} - d\|^2}{\sigma_d^2} \quad (13)$$

where β is an additional weighting factor to give equal importance to both terms regardless of the number of points in each of them. The optimization has two phases: First, the main minimization, with the cost function from (13), where the parameters to minimize are ${}^{I_i}\mathbf{T}_f$, ${}^{f}\mathbf{T}_d$, c_1, c_0, f_x^d, f_y^d, $\mathbf{C}_I^d$ and k_d. Second, minimization of the distortion offset given by [1], where the parameters to minimize are α_0, α_1 and $\mathbf{D}_\delta(u, v)$. We have used an iterative non-linear minimization using the Levenberg−Marquardt algorithm for both processes.

B. Stepwise calibration of fisheye and depth cameras

The idea behind this approach is to separate the calibration of the whole system in stages instead of performing a single global optimization of the parameters. Herrera et al. [1] mention the improvement of the results in the calibration when both the intrinsic and extrinsic parameters are optimized simultaneously. However, the optimization may get stuck in a local minimum if the estimation of the seed values of some parameters is not good enough. We propose to perform following stages instead:

- Intrinsic calibration of the fisheye ($f(\rho)$, $\mathbf{A}$ and $\mathbf{t}$).
- Intrinsic calibration of the IR camera ($\mathbf{f}^d$, $\mathbf{C}_I^d$ and $\mathbf{k}_d$), which can be solved using traditional methods [19].

- Obtain the poses of the checkerboard with respect to the cameras (${}^{I_i}\mathbf{T}_f$, ${}^{I_i}\mathbf{T}_d$) and with them, the extrinsic calibration (${}^{f}\mathbf{T}_d$).
- Global optimization to compute disparity-depth correction parameters (c_1, c_0, $\mathbf{D}_\delta(u_d^d, v_d^d)$, α_0 and α_1) and refine the previously computed parameters.

The stages not included in the description of the first approach are detailed as follows.

1) Computation of the extrinsics of the system: Having computed the relative poses of the checkerboard from both cameras (${}^{I_i}\mathbf{T}_f$ from Section IV-A1 and ${}^{I_i}\mathbf{T}_d$ from [19]), for each image pair we have:

$$ {}^{d}\mathbf{R}_f^{(I_i)} = {}^{I_i}\mathbf{R}_f \cdot {}^{d}\mathbf{R}_{I_i} \tag{14a} $$

$$ {}^{d}\mathbf{t}_f^{(I_i)} = {}^{I_i}\mathbf{t}_f - {}^{d}\mathbf{R}_f \cdot {}^{I_i}\mathbf{t}_d \tag{14b} $$

Averaging the rotations (turned into rotation vectors) and the translations provides a good estimate of the extrinsics of the system. To refine the extrinsic parameters (6 DOF), we can minimize the reprojection error of the corner points from one reference frame to the other backprojected to the image ($\hat{\mathbf{p}}_{ij}$) with respect to the measured point ($\mathbf{p}_{ij}$), respectively:

$$ \arg\min \|\hat{\mathbf{p}}_{ij}^{f} - \mathbf{p}_{ij}^{f}\| + \|\hat{\mathbf{p}}_{ij}^{d} - \mathbf{p}_{ij}^{d}\| \tag{15} $$

where $\hat{\mathbf{p}}_{ij}^{c_1} = proj\left({}^{c_2}\mathbf{T}_{c_1} \cdot \mathbf{X}_{ij}^{c_2}\right)$, $proj$ the projection function (which changes depending on the type of camera) and c_1, c_2 means two different camera frames (in this case, f and d).

2) Global optimization: The global optimization proposed in this approach has the same formulation (Section IV-A2). In this case, we do not need to perform the simultaneous calibration of the intrinsics of the IR camera and the extrinsics of the system, as the values initially provided are already close to the final ones. These parameters can be fixed and ignored in the optimization process, and use them only to estimate the values of c_1, c_0, $\mathbf{D}_\delta(u_d^d, v_d^d)$, α_0 and α_1. However, if we include all the parameters as in Section IV-A2 we can use this optimization as a global refinement of the already estimated parameters, taking into account the depth image instead of the IR image.

V. Experiments

In the experiments, we use the proposed hybrid camera system shown in Fig. 1b: The RGB-D sensor is the Asus Xtion Live Pro, which provides a depth image resolution of 640×480 pixels at 30Hz. The fisheye camera is a camera uEye UI-3580CP of 2560×1920 pixels at 15Hz with a USB 3.0 interface, and a fisheye lens Lensagon CF5M1414, with a field of view of 182°. The images have been captured synchronized using Robot Operating System (ROS). The software used for the implementation of the algorithm is Matlab. The calibration pattern is a 9×7 checkerboard printed in a DIN-A2 sized paper attached to a rectangular piece of wood.

A set of fisheye images without depth information has been used to calibrate the intrinsics of the fisheye, resulting in a mean reprojection error of less than 0.2 pixels, which can be considered as an accurate calibration for such high resolution. For the depth intrinsics and camera pair extrinsic calibration we used two sets of images for comparison purposes. Set A has 25 fisheye and 25 depth images and set B has 28 of each. In Table I it is shown the mean reprojection error (MRE) for both sets of images using the results from four calibration procedures:

TABLE I: MRE in the fisheye and depth image for both sets of images A and B with four calibration results.

	Evaluated Set A		**Evaluated Set B**	
	Fisheye (px)	Depth (du)	Fisheye (px)	Depth (du)
1) Joint calib. A	0.185	0.839	0.187	1.477
2) Step calib. A	0.185	0.806	0.187	1.514
3) Joint calib. B	0.165	1.322	0.183	0.856
4) Step calib. B	0.165	1.321	0.182	0.835

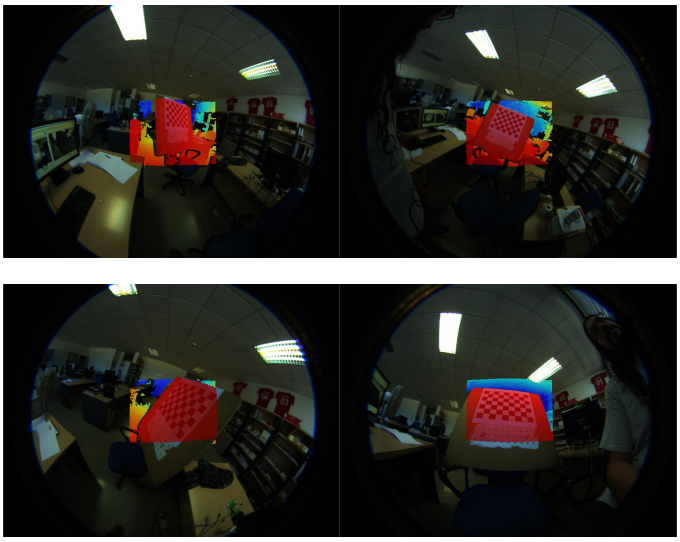


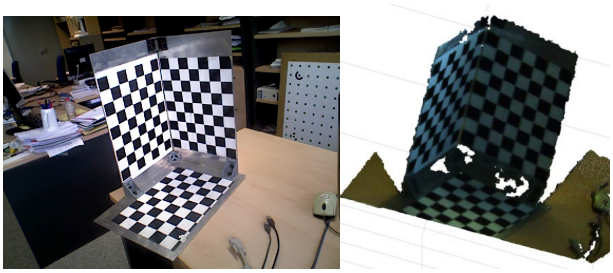
Fig. 4: Semi-transparent depthmap superimposed over the fisheye image to illustrate the quality of the calibration.

1) Joint Calibration of Set A (Section IV-A).
2) Stepwise Calibration of Set A (Section IV-B).
3) Joint Calibration of Set B (Section IV-A).
4) Stepwise Calibration of Set B (Section IV-B).

The MRE for each set from its own calibration results is in the fisheye camera less than 0.2 pixels, whereas in the depth it is less than 1du of error. Using the calibration values of extrinsics and depth intrinsics from the other set, the MRE of the depth increases to 1.5du. These values can be considered highly satisfactory considering the complexity of calibrating two cameras of such different kind. We can see how the difference among the methods is marginal, making both of them equally valid approaches for the task.

The quality of the mapping of the depth information can also be qualitatively analyzed superposing the depth maps in the fisheye image (Fig. 4). It can be observed how the borders in the depth coincide with the borders in the image, even at large distances. It is also possible to reconstruct the point cloud with this data, where the accuracy can also be appreciated (Fig. 5b).

To measure the quality of the calibration we have also used a 3D pattern. It consists of three metal plates screwed and



(a) (b)

Fig. 5: (a) 3D pattern consisting of three orthogonal planes. (b) Point cloud obtained using the calibration.

TABLE II: Angular difference between planes(in deg) with respect to the ground truth in the fisheye and depth image.

	Fisheye Image	Depth Image
α_{12}	-0.3559	1.3715
α_{23}	1.7045	-0.3339
α_{31}	1.2229	0.1908
Mean	1.0944	0.6321

secured at 90° with calibration patterns attached to them. We have accurate ground truth of the 3D position of the points of these patterns from a photogrammetric reconstruction by bundle adjustment. Obtaining the pose of the camera from the fisheye image and computing the planes from the depth image we can compare the angles between the metal plates to see how good the calibration is. Table II shows the results of the experiment, with an error of $\approx 1^\circ$ in the fisheye image and $\approx 0.6^\circ$ in the depth image.

VI. Conclusion

In this work we have presented a new hybrid camera system composed of a depth sensor and a fisheye camera. Many applications could be likely to benefit from such configuration, including navigation, SLAM or object detection. The fisheye provides wide field of view and the depth provides certainty and scale. However, as a consequence of its novelty, none of the existing methods of calibration can be directly applied to this set. We have proposed a method which combines some works from the state of the art for this purpose. The generality of the camera models used makes it exportable to other camera/range sensor configurations as well. The method have been tested with real images showing promising results in terms of accuracy.

Acknowledgment

This work was supported by Ministerio de Economía y Competitividad and European Union under FPI grant BES-2013-065834 and projects DPI2014-61792-EXP and DPI2015-65962-R.

References

[1] C. Herrera, J. Kannala, J. Heikkilä *et al.*, "Joint depth and color camera calibration with distortion correction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2058–2064, 2012.

[2] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 5695–5701.

[3] H. Baltzakis, A. Argyros, and P. Trahanias, "Fusion of laser and visual data for robot motion planning and collision avoidance," *Machine Vision and Applications*, vol. 15, no. 2, pp. 92–100, 2003.

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

[5] F. Vasconcelos, J. P. Barreto, and U. Nunes, "A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2097–2107, 2012.

[6] L. Zhou, "A new minimal solution for the extrinsic calibration of a 2d lidar and a camera using three plane-line correspondences," *Sensors Journal*, vol. 14, no. 2, pp. 442–454, 2014.

[7] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007, pp. 4164–4169.

[8] R. Gomez-Ojeda, J. Briales, E. Fernandez-Moral, and J. Gonzalez-Jimenez, "Extrinsic calibration of a 2d laser-rangefinder and a camera based on scene corners," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 3611–3616.

[9] C. Zhang and Z. Zhang, "Calibration between depth and color sensors for commodity depth cameras," in *International Workshop on Hot Topics in 3D, in conjunction with ICME*. IEEE, 2011.

[10] J. Smisek, M. Jancosek, and T. Pajdla, "3d with kinect," in *Consumer Depth Cameras for Computer Vision*. Springer, 2013, pp. 3–25.

[11] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 3936–3943.

[12] I. V. Mikhelson, P. G. Lee, A. V. Sahakian, Y. Wu, and A. K. Katsaggelos, "Automatic, fast, online calibration between depth and color cameras," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 218–226, 2014.

[13] F. Endres, C. Sprunk, R. Kummerle, and W. Burgard, "A catadioptric extension for rgb-d cameras," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2014, pp. 466–471.

[14] R. Tomari, Y. Kobayashi, and Y. Kuno, "Wide field of view kinect undistortion for social navigation implementation," in *Advances in Visual Computing*. Springer, 2012, pp. 526–535.

[15] E. Fernandez-Moral, J. González-Jiménez, P. Rives, and V. Arévalo, "Extrinsic calibration of a set of range cameras in 5 seconds without pattern," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 429–435.

[16] A. Rituerto, L. Puig, and J. Guerrero, "Visual slam with an omnidirectional camera," in *20th IEEE International Conference on Pattern Recognition (ICPR)*, 2010, pp. 348–351.

[17] J. Bermudez-Cameo, G. Lopez-Nicolas, and J. Guerrero, "Line-images in cone mirror catadioptric systems," in *22nd IEEE International Conference on Pattern Recognition (ICPR)*, 2014, pp. 2083–2088.

[18] J. Heikkilä, "Geometric camera calibration using circular control points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1066–1077, 2000.

[19] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 1999, pp. 666–673.