A Catadioptric Extension for RGB-D Cameras

Felix Endres Christoph Sprunk Rainer Kümmerle Wolfram Burgard

***Abstract*— The typically restricted field of view of visual sensors often imposes limitations on the performance of localization and simultaneous localization and mapping (SLAM) approaches. In this paper, we propose and analyze the combination of an RGB-D camera with two planar mirrors to split the field of view such that it covers both front and rear view of a mobile robot. We describe how to estimate the extrinsic calibration parameters of the modified sensor using a standard parametrization and a reduced one that exploits the properties of the setup. Our experimental evaluation on real-world data demonstrates the robustness of the calibration procedure. Additionally, we show that our proposed sensor modification substantially improves the accuracy and the robustness in a simultaneous localization and mapping task.**

I. Introduction

Performing localization or SLAM with consumer grade RGB-D cameras such as the Microsoft Kinect has been a topic of intensive research in recent years [1], [2], [3], [4]. Many approaches are extensions of algorithms commonly used with stereo cameras or lidars (laser range scanners), which typically have a horizontal field of view of 180° to 270° and provide accurate depth measurements in a range between a few centimeters and several dozen meters. For RGB-D sensors the distance in which the sensor provides depth measurements is generally limited to a range of about 0.6 m to 8 m with a restricted field of view of 43° vertically by 57° horizontally. In contrast to lidars, for which one can usually assume some meaningful geometric structure to be in the sensor's field of view (e.g., more than half of the room for a field of view of 180°), in the context of RGB-D cameras we often need to deal with visually and geometrically ambiguous structure, e.g., when perceiving only a flat part of a wall or one corner of a room. Accordingly, an extension of the perception capabilities of RGB-D sensors would be highly beneficial. Obviously, an improvement can be achieved by the use of multiple sensors. This, however, comes with the inherent increase of computational and financial costs.

In this paper, we propose a novel catadioptric setup for Kinect-style RGB-D cameras that is of extremely low cost (less than 20 EUR), and requires neither significant computational resources nor higher power consumption and substantially relaxes the above-mentioned limitations.

Catadioptric sensors have been extensively used in the robotics and computer vision community for localization, visual odometry and SLAM [5]. Various shapes of mirrors have been used to increase the field of view, including parabolic, hyperbolic and elliptic. Also mirrors have been used to capture stereo images with a single camera [6]. To the best of our knowledge, however, we are the first to propose to combine an RGB-D camera with mirrors.

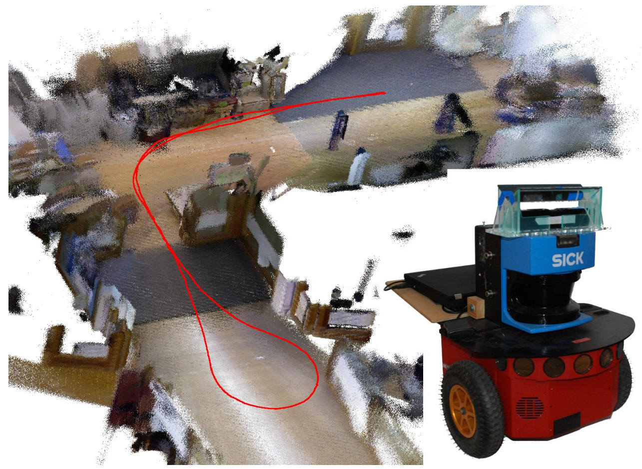


Fig. 1. A resulting map and the trajectory estimate from our evaluation experiment, where we show that the proposed catadioptric sensor substantially improves the accuracy in a robot SLAM task. Note that only the RGB-D data was used to generate the map.

In this work, we use two planar mirrors to split the field of view of the camera, such that the robot has effectively two fields of view of about 20° vertically by 57° horizontally in opposite directions. This setup is particularly beneficial for robots moving in the plane, as the sensor data gained by the horizontally extended perception yields much more information relevant to the planar motion than the sacrificed perception in the vertical direction. However, the proposed sensor extension can be also beneficial for robots moving with more degrees of freedom as, e.g., the ambiguity between translation and rotation is alleviated. Figure 1 depicts a Pioneer robot with the proposed catadioptric sensor and a map created only from the obtained RGB-D data.

To readily offer the developed device to the research community, we publish the used CAD model and our implementation of the calibration software[1].

II. Related Work

A. Calibration

Calibration is a crucial factor for the accuracy of sensor measurements. The calibration of an individual camera is called intrinsic calibration and concerns parameters that govern the image creation, i.e., focal length, principal point,

All authors are with the Department of Computer Science at the University of Freiburg, Germany.

The authors thank Janosch Dobler, Robert Grönsfeld and Pratik Agarwal for their support.

[1] see `http://ais.informatik.uni-freiburg.de/projects/datasets/catadioptric-rgbd`

and distortion model parameters. For RGB cameras, there are well-known techniques from the field of computer vision for estimating these parameters [7].

For an RGB-D camera we need to perform the intrinsic calibration for the color camera, the infrared camera, and the infrared projector. To this end, Herrera *et al.* [8] use a checkerboard to calibrate the intrinsics of such a camera. Recently, Teichman *et al.* [9] presented an approach to calibrate the depth-measurements of an RGB-D camera given a SLAM estimate.

There also exist several approaches for extrinsic calibration of cameras, i.e., the relative offset between two sensors. A known extrinsic calibration, for example, allows us to transform the motion of one sensor into the motion of the other sensor. Most notably a device can be calibrated by precise manufacturing according to a model. For example, if the proposed device is produced using the CAD model and a "perfect" laser cutter, the relative geometry of the mirrors is precisely known and can be used to compute the angular offset between the two virtual camera viewpoints. In practice, however, the extrinsic calibration needs to be estimated. For example, Carrera *et al.* [10] perform an extrinsic calibration of multiple cameras on a mobile robot, where they perform SLAM for each camera and additionally exploit that the cameras observe the same parts of the environment to recover the parameters. Zienkiewicz *et al.* [11] automatically calibrate the position of the camera from visual odometry. Given the trajectory of a vehicle, Maddern *et al.* [12] perform an extrinsic calibration of LIDAR sensors by optimizing the Rényi Quadratic Entropy of the point cloud as the robot traverses the environment.

Brookshire and Teller present an unsupervised approach that only requires ego-motion estimates for calibrating the offset between range sensors either in 2D [13] or 3D [14]. To determine the extrinsic calibration of two (or more) cameras with an unconstrained rigid motion, we need to constrain all six degrees of freedom by appropriate motions. In their analysis, Brookshire and Teller [14] find that rotational motion around at least two different axes is required to obtain the required constraints.

The calibration procedure described in this paper shares ideas with the related approaches described above but combines both relative motion estimates and loop closures to estimate the calibration parameters. Additionally, the variants of the calibration that are specific to our proposed catadioptric sensor require the estimation of a smaller set of parameters.

B. RGB-D SLAM

SLAM using RGB-D sensors has been intensively studied in the recent years. The first scientifically published RGB-D SLAM systems [1], [2] use visual features to estimate frame to frame transformations. They employ these transformations to construct a pose graph and estimate the global sensor trajectory using non-linear least squares optimization. They create the map by reprojecting the measurements in a common coordinate frame using the estimated maximum likelihood poses.

Fig. 2. The assembled catadioptric sensor using an Asus XtionPRO Live.

Alternative approaches integrate the sensor data into a voxel grid and directly fuse every measurement into the map representation [3], [15]. These approaches achieve a highly accurate visual odometry. It is more difficult, though, to update the voxel grid representation to reflect new information, e.g., loop closures or calibration data, that often affects already integrated measurements.

The RGB-D SLAM system used in this work is an extension of our previous technique [2] so that it works with images from two viewpoints.

III. CATADIOPTRIC EXTENSION

In this section, we will present the idea of using a catadioptric setup for an RGB-D camera. We will outline our goals, motivate our design choices, and discuss our implementation of the proposed modification.

A. Goal

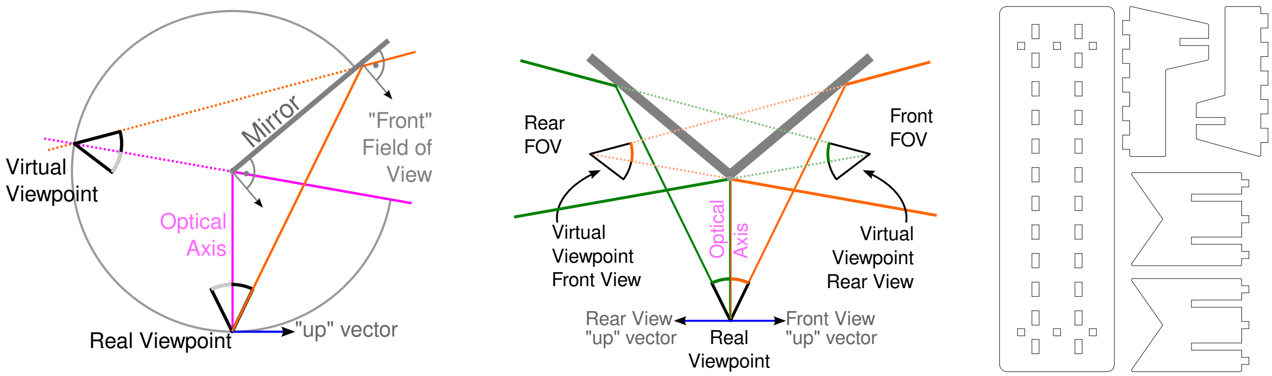
For the design of the catadioptric device proposed in this paper, we focus on the following goals:

- Improved performance of robotic applications.
- Quick and easy to build and integrate on the robot.
- Cheap (cheaper than a second camera).
- Easy handling of resulting data.

While many catadioptric configurations are conceivable, our goal is a setup that is generally applicable. We therefore do not address, e.g., the possible reduction of the minimum sensing distance with respect to the center of the robot. While this could easily be achieved by increasing the distance between mirrors and camera, it comes at the cost of equally reduced maximum range and a bigger size of the device.

B. Design

The most influential design choice is probably the shape of the mirror(s). Possibilities include the use of three paraboloidal or hyperboloidal mirrors, or a single mirror that is parabolically curved along the camera's up vector, or planar mirrors. We propose the use of planar mirrors, as this neither requires expensive optical components nor intricate changes in the computation of the registered RGB-D image (or the respective point cloud). This allows to use the built-in processing of the camera. Our design splits the field of view with two planar mirrors, creating two *virtual viewpoints* roughly opposite to each other. Figure 2 shows the assembled device. Figures 3a and 3b illustrate the formation of the virtual viewpoints. Further splits might be beneficial in



(a) Illustration of the concept of a virtual viewpoint. For clarity, only one mirror and the respective projection of the viewpoint is shown.

(b) For two planar mirrors, the focal point is projected to two virtual viewpoints. The "up" vector needs to be flipped for the lower part of the original image.

(c) The CAD model of the mirror mount for an "ASUS XtionPRO Live".

Fig. 3. Illustrations of the concept of a virtual viewpoint and the CAD model of the mirror mount.

specific scenarios to allow the robot to measure other aspects of its environment, e.g., floor or ceiling.

We align the intersection of the mirrors with the axis through the infrared camera, the projector, and the color camera. In this way, the projection and the two images are about equally split, which maximizes the overlap in the field of view. A rotation of the two mirrors around the (original) optical axis would lead to gaps in the registered RGB-D image, as the mirrors would reflect different proportions of the projector pattern and the field of view of the infrared camera. We choose the tilt of the two mirrors so that the centers of the two half fields of view are roughly horizontal. Note that ideal planar mirrors do not introduce any distortion to the depth perception principle, regardless of their pose.

The choice of the distance between camera and mirrors is a trade-off between loss of image area and the size of the device. The closer we mount the mirrors to the camera, the larger will be the unusable projection of the gap between the mirrors be in the images. If they are further away, the mirrors need to be larger, and the size and weight of the overall device grows. In our implementation the mirrors are about 2 cm in front of the camera and we crop out a rectangle with a height of 50 pixels from the center of the image.

To hold the mirrors and the sensor, we designed a multi-component structure, which we can easily assemble by manually plugging the individual components together. Figure 3c shows the CAD model of the structure, which we used to cut the pieces from a 30 cm $\times$ 30 cm $\times$ 0.5 cm sheet of polymethylmethacrylat (PMMA) in a laser cutter.

While it would be possible to use coated PMMA for the mirrors as well, we recommend to use glass mirrors. Due to the great planarity and rigidity of glass no noticeable distortions are introduced to the images.

IV. Calibration of the RGB-D Sensor

For calibrating our catadioptric setup we need to estimate the position of the two virtual cameras. Due to the small size of the used mirrors and the rigidity of glass, we assume no additional distortion from the mirrors. In our experiments we found that this assumption seems to hold in practice. Therefore, the intrinsic calibration procedure does not differ from that of a regular RGB-D camera. In this section, we therefore only discuss the extrinsic calibration of the virtual viewpoints created by the mirrors.

A. Calibration and SLAM

To estimate the motion of the cameras in one time step, we extended our existing RGB-D SLAM system [2], which estimates transformation between frames based on the correspondence of visual features.

This system employs the so-called graph-based formulation, which performs non-linear least-squares to estimate the maximum likelihood configuration of the sensor poses. Such a least-squares problem has the following form:

$$\mathbf{F}(\mathbf{x}) = \sum_{i \in \mathcal{G}} \mathbf{e}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij})^\top \, \boldsymbol{\Omega}_{ij} \, \mathbf{e}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij}), \tag{1}$$

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x}} \mathbf{F}(\mathbf{x}). \tag{2}$$

Here, $\mathbf{x}$ is the state vector, e.g., the poses of the sensor in case of SLAM. Furthermore, the error function $\mathbf{e}(\cdot)$ computes the difference between an expected measurement given the current state vector and a real measurement $\mathbf{z}_{ij}$. The error is $\mathbf{0}$ when $\mathbf{x}$ perfectly explains the measurement. Finally, the information matrix $\boldsymbol{\Omega}_{ij}$ models the uncertainty in the error. We employ g^{2}o [16] to obtain the solution of Eq. 2.

In case of a SLAM problem, the error function describes relative transformations between the camera poses at certain points in time. The transformation may relate subsequent poses, i.e., visual odometry, or span over large time intervals (loop closures). Furthermore, if a camera observes a part of the environment that was previously seen in another camera, the error function involves two different cameras.

If we also include the unknown extrinsic calibration, we obtain a second set of error functions that deals with estimating the calibration parameters corresponding to the setup of our mirrors. Without any further assumptions the calibration consists of estimating the offset $\mathbf{c} \in \mathrm{SE}(3)$ between the two cameras. Without loss of generality let us assume that the state vector contains ordered pairs of poses for the camera in each time step, i.e., $\mathbf{x} = \langle \mathbf{x}_1^{[1]}, \mathbf{x}_1^{[2]}, \ldots, \mathbf{x}_t^{[1]}, \mathbf{x}_t^{[2]} \rangle$, where $\mathbf{x}^{[j]}$ represents the poses of the camera j. The error for the

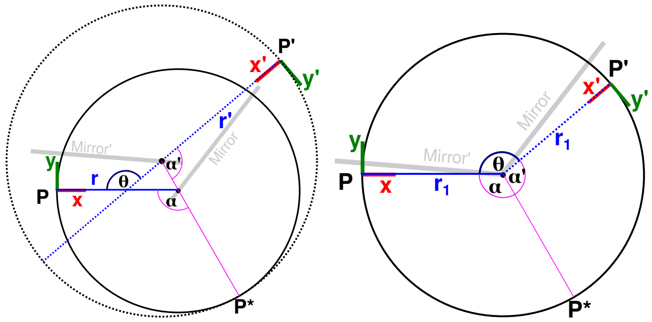


Fig. 4. Cross sections of the proposed catadioptric extension. P^* depicts the position of the real focal point, while P and P' depict the virtual focal points as projected by the two mirrors. (a) Assuming the intersection of the mirror planes to be parallel to the image plane (both orthogonal to the illustration) results in a displacement with three degrees of freedom (x', y', θ') between the virtual viewpoints. (b) Additionally assuming the edge between the mirrors to intersect the optical axis of the camera reduces the transformation between the virtual viewpoints to the two deg. of freedom r and θ.

offset parameter $\mathbf{c}$ is given by

$$\mathbf{e}_i(\mathbf{x}, \mathbf{c}) = \mathbf{x}_i^{[2]} \ominus \left(\mathbf{x}_i^{[1]} \oplus \mathbf{c} \right), \tag{3}$$

where $\oplus$ corresponds to the motion composition and $\ominus$ to its inverse.

Given these two types of error functions, our joint estimation problem is given by solving

$$\operatorname*{argmin}_{\mathbf{x}, \mathbf{c}} \sum_{ij \in \mathcal{G}} \|\mathbf{e}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij})\|_{\mathbf{\Omega}_{ij}} + \sum_i \|\mathbf{e}_i(\mathbf{x}, \mathbf{c})\|_{\mathbf{\Omega}_i}, \tag{4}$$

where $\|\mathbf{e}\|_{\mathbf{\Omega}} = \mathbf{e}^\top \mathbf{\Omega} \mathbf{e}$. Note that while we only consider two cameras, the approach itself is able to handle multiple cameras.

In the following, we will refine the error function $\mathbf{e}_i(\mathbf{x}, \mathbf{c})$ and the dimensionality of our calibration parameter $\mathbf{c}$ to correspond — under certain assumptions — to the specific catadioptric setup at hand.

B. Additional Constraints Imposed by the Device Structure

Due to the arrangement of the proposed catadioptric device, the offset between the virtual camera viewpoints is constrained to a manifold of less than six degrees of freedom.

If we assume that the intersection of the mirror planes is parallel to the transversal axis (the axis orthogonal to optical axis and "up" vector of the camera) and intersects with the optical axis, we obtain only two degrees of freedom for the virtual viewpoints, namely r and θ. See Figure 4b for an illustration of the corresponding configuration.

Let ${}^{\text{2DOF}}\mathbf{c} = \langle r, \theta \rangle$ be the two dimensional parameter. We can convert this to a 4×4 transformation matrix $\in \mathrm{SE}(3)$ as follows:

$${}^{\mathrm{SE}(3)}f_{\text{2DOF}}({}^{\text{2DOF}}\mathbf{c}) = \begin{pmatrix} R_x(\theta) & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}, \tag{5}$$

where $R_x(\theta)$ corresponds to a 3×3 rotation matrix around the x-axis and $\mathbf{t} = (r - r\cos(\theta), r\sin(\theta), 0)^\top$. The result of this conversion recovers the parameter $\mathbf{c}$ in its original space and can directly be plugged into Eq. 4 to estimate the calibration.

Under the assumptions stated above, these two parameters can be determined from planar motion only. The introduced constraints allow for the computation of the unobserved degrees of freedom.

If we relax the assumption that the edge of the mirrors intersects with the optical axis and we only assume that the edge between the mirrors is parallel to the transversal axis of the camera, we obtain three degrees of freedom for our setup. Figure 4a illustrates this configuration. Under these assumptions the roll and yaw angles of the virtual viewpoints are identical. Further, the offset in the transversal axis must be zero. The remaining degrees of freedom are the relative pitch θ and the translational offset $\mathbf{b} \in \mathbb{R}^2$. Thus, we obtain ${}^{\text{3DOF}}\mathbf{c} = \langle \mathbf{b}, \theta \rangle$ which we can transform to SE(3) as follows:

$${}^{\mathrm{SE}(3)}f_{\text{3DOF}}({}^{\text{3DOF}}\mathbf{c}) = \begin{pmatrix} R_x(\theta) & (0, \mathbf{b})^\top \\ \mathbf{0} & 1 \end{pmatrix}. \tag{6}$$

Again, the result of Eq. 6 can be directly considered in Eq. 4 to perform the calibration.

This calibration method has the advantage that it is more robust to small variations in the placement of the camera with respect to the mirrors.

It is again possible to fully calibrate the virtual viewpoints with planar motions under the stated assumption, if we use the rotation to constrain the translational degrees of freedom and a translation to constrain the rotational offset. This means, we now need to rotate about the axis along which the translation is known, which is the transversal axis - and therefore need to mount the device such that the wide aperture angle is vertical. Unfortunately, in a planar SLAM application using this configuration for online self-calibration would reduce the horizontal field of view by at least 62.5 %. Therefore the overlap of consecutive measurements would be substantially reduced when the robot rotates.

V. Experiments

In the following, we will present real-world experiments to evaluate the novel catadioptric sensor setup. First, we examine the calibration of the device and afterwards we investigate the benefits of the setup for performing SLAM.

A. Calibration

There are three interesting special cases in the calibration of the proposed sensor. First, we will examine the convergence using egomotion alone. Second, we will evaluate the impact of a loop closure between the two virtual viewpoints. A third case, interesting in particular for robots moving in the plane, is to look at the convergence when restricting the sensor to a planar motion. For the evaluation of the convergence, we recorded data while rotating the sensor about the optical axis, the transversal axis and the "up" vector of the real camera. Figure 5 shows the convergence of the individual components of the displacement. The initial guess for the optimization is the identity, i.e., both viewpoints are at the same place. For the methods with restricted degrees of freedom, we compute the shown displacement components as described in Section IV-B. As expected, on the left column we can see that only the two degrees of freedom calibration method is capable to compute a stable estimate for the displacement for a planar motion without loop closure. The three degrees of freedom calibration method does not converge, because the camera would need to rotate around the transversal axis. However, as discussed in Section IV-B, this would be an unfavorable configuration in a planar SLAM setting.

The loop closure between the virtual viewpoints at frame 50 then abruptly introduces sufficient constraints on all degrees of freedom. For rotational motion around more than one axis we observe quick convergence for all methods. While the calibration of three and six degrees of freedom behaves very similar, except for the improved roll estimate for planar motion, the behavior of the two degrees of freedom method deviates notably. Particularly interesting is the convergence to almost zero for the translation along the Y and Z axis, even though the roll converges correctly. The estimated calibration of the rotation between the virtual cameras corresponds well to the angle between the mirrors, which we measured manually. Since the location of the focal point of the RGB-D camera is affected by the geometry of the lens and also inside the housing of the camera, we are only able to obtain a rough estimate of the translational components manually. The estimated translation of the virtual cameras is in line with what we expect given the approximate manual measurement. Furthermore, a visual inspection of the point cloud data of the catadioptric extension revealed the accuracy of the range data, for example, both virtual cameras observe the ground plane at the expected location and orientation.

B. SLAM Error Measure

To quantify the performance gain from using the proposed sensor extension, we compare the trajectories of a mobile robot, which is driving through an indoor environment, reconstructed during SLAM with ground truth from 2D Monte-Carlo localization using a SICK LMS-200 laser scanner (see Figure 1). To compute the deviation from the ground truth, we use the root-mean-square of the absolute trajectory error (RMS-ATE), as described in [17]. For a trajectory estimate $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_1 \dots \hat{\mathbf{x}}_n\}$ and the corresponding ground truth $\mathbf{X}$ the RMS-ATE is defined as

$$\mathrm{ATE}_{\mathrm{RMS}}(\hat{\mathbf{X}}, \mathbf{X}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \| \mathit{trans}(\hat{\mathbf{x}}_i) - \mathit{trans}(\mathbf{x}_i) \|^2},$$

i.e., the root-mean-square of the Euclidean distances between the corresponding ground truth and the estimated poses. To make the error metric independent of the coordinate system in which the trajectories are expressed, the trajectories are aligned such that the above error is minimal. The correspondences of poses are established using the timestamps from the sensor image and the range scan.

C. Improved Accuracy for SLAM

Ideally, we would compare the accuracy of the SLAM trajectory estimate from the same motion for the baseline and the proposed setup. However, since RGB-D cameras actively project an infrared pattern, we refrain from running

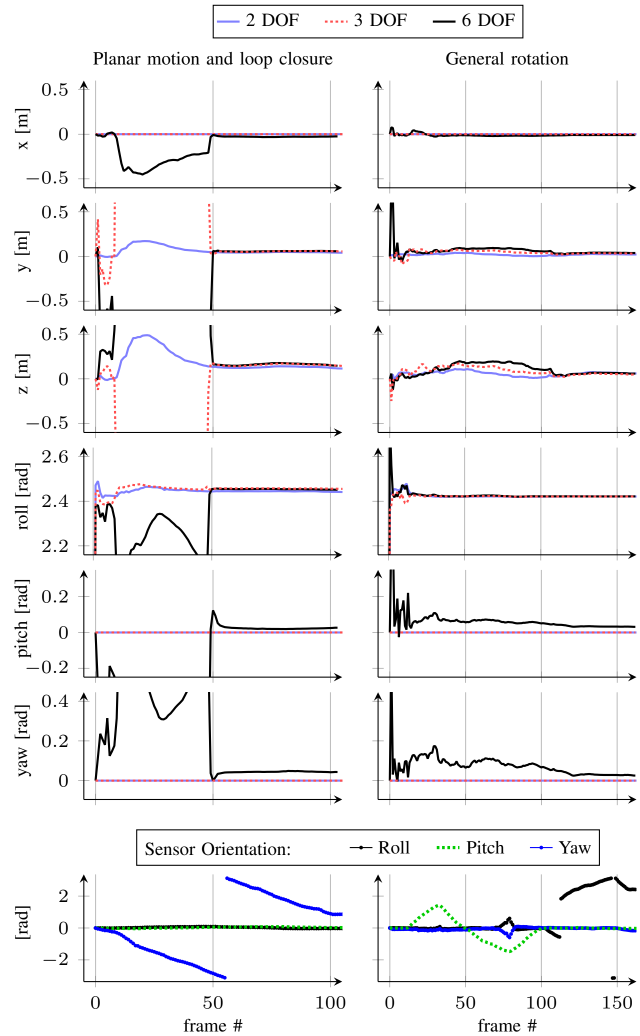


Fig. 5. Convergence of the calibration parameters for the proposed methods. The bottom row depicts the orientation of the virtual viewpoints of the camera facing to the front during the recorded motion. The rows above show the components of the computed relative transformation between the virtual viewpoints. *Left*: Rotation only about the vertical axis, with a loop closure between the viewpoints around frame 50. *Right*: Rotation of the device around all axes, without loop closure.

the catadioptric setup simultaneously with a regular sensor to avoid crosstalk effects between the sensors. To guarantee comparability, we use a highly accurate laser-based trajectory following approach [18], for which we teach a trajectory and let the robot repeat it, once with a regular RGB-D camera and once with the same camera embedded in the catadioptric device. The regular camera is mounted analogous to the front view of the catadioptric device. For ground truth, we use the laser scanner to create a 2D map in the "teach" run and use the laser-based localization result for comparison in each repetition run. In our experiments, the deviation of the mobile robot from the taught trajectory was always below 0.02 m and 2.5 deg. Therefore, the input to both sensor setups is suited for an unbiased comparison. The trajectory and the mapped environment is depicted in Figure 1. The RMS-ATE for the catadioptric sensor is 0.076 m. The corresponding error for mapping with the regular sensor, is 0.209 m. Thus, the error is reduced by 64 %. In an experiment with two regular RGB-D

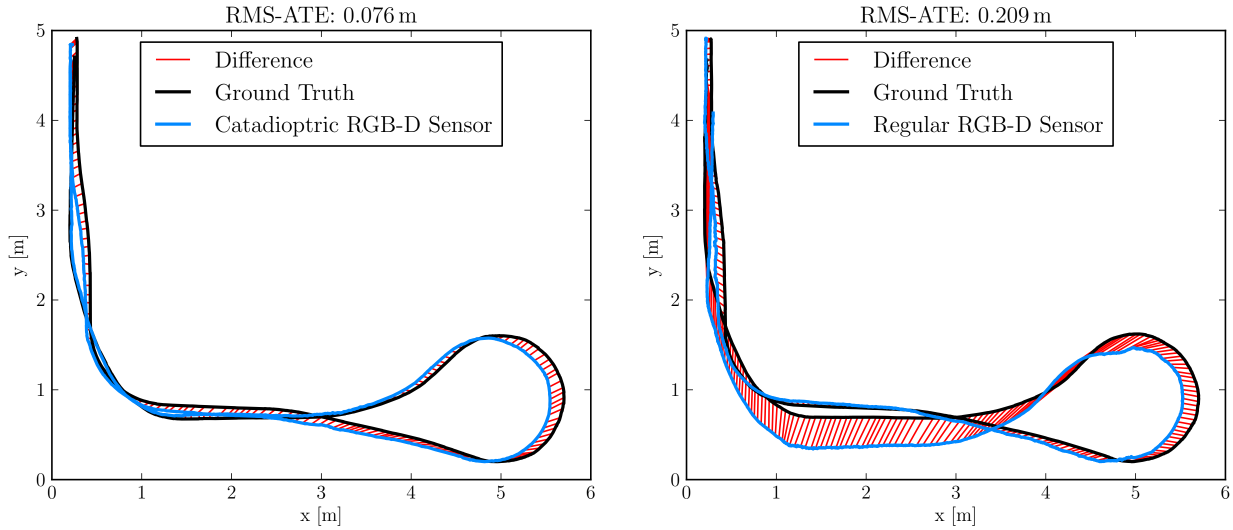


Fig. 6. For a trajectory of about 18 m the translational root mean squared error (RMSE) of our approach (left panel) with respect to the ground truth is only 0.076 m. The vertical RMSE is 0.021 m, which corresponds well to our robot driving on a planar indoor surface. The baseline experiment (right panel) uses a forward-facing, unmodified Xtion Pro Live for SLAM, the RMSE is 0.209 m. The vertical RMSE is 0.094 m.

cameras, attached back to back, we obtained an RMS-ATE of 0.070 m—a further reduction of only three percentage points. Figure 6 shows the trajectory from the respective RGB-D SLAM results in comparison with the ground truth.

The calibration used for the catadioptric sensor was determined beforehand using the full six degree of freedom procedure with the recorded data from Section V-A. Given the known calibration data and exploiting the common time stamp of the two splits, we project the visual features of both viewpoints into a common coordinate system and perform matching of frames as for a regular RGB-D camera. Experiments with the calibration result of the two and three degree of freedom parametrization resulted in comparable error values.

VI. Conclusions

In this paper, we proposed a catadioptric extension to RGB-D cameras by using two planar mirrors to split the field of view such that it covers both front and rear view. We furthermore described how to estimate the extrinsic calibration parameters of the modified sensor and developed methods to exploit the devices structure, e.g., to achieve full calibration from planar motion only. In experiments carried out with real-world data, we demonstrated that RGB-D SLAM applications can greatly benefit from the extended field of view.

References

[1] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using kinect-style depth cameras for dense 3D modeling of indoor environments," *International Journal of Robotics Research (IJRR)*, vol. 31, no. 5, pp. 647–663, April 2012.

[2] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3D mapping with an RGB-D camera," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 177–187, Feb 2014.

[3] T. Whelan, M. Kaess, J. Leonard, and J. McDonald, "Deformation-based loop closure for large scale dense RGB-D SLAM," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[4] J. Cunha, E. Pedrosa, C. Cruz, A. J. Neves, and N. Lau, "Using a depth camera for indoor robot localization and navigation," *DETI/IEETA-University of Aveiro, Portugal*, 2011.

[5] C. Plagemann, C. Stachniss, J. Hess, F. Endres, and N. Franklin, "A nonparametric learning approach to range sensing from omnidirectional vision," *Robotics and Autonomous Systems*, vol. 58, no. 6, pp. 762 – 772, 2010.

[6] J. Gluckman and S. Nayar, "Rectified Catadioptric Stereo Sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 224–236, Feb 2002.

[7] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[8] D. Herrera C., J. Kannala, and J. Heikkilä, "Joint depth and color camera calibration with distortion correction," *IEEE Journal on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 34, no. 10, Oct 2012.

[9] A. Teichman, S. Miller, and S. Thrun, "Unsupervised intrinsic calibration of depth sensors via SLAM," in *Proceedings of Robotics: Science and Systems*, 2013.

[10] G. Carrera, A. Angeli, and A. J. Davison, "SLAM-based automatic extrinsic calibration of a multi-camera rig," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011.

[11] J. Zienkiewicz, R. Lukierski, and A. J. Davison, "Dense, auto-calibrating visual odometry from a downward-looking camera," in *Proc. of the British Machine Vision Conference (BMVC)*, 2013.

[12] W. Maddern, A. Harrison, and P. Newman, "Lost in translation (and rotation): Rapid extrinsic calibration for 2D and 3D lidars," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2012.

[13] J. Brookshire and S. Teller, "Automatic calibration of multiple coplanar sensors," in *Proc. of Robotics: Science and Systems (RSS)*, 2011.

[14] ——, "Extrinsic calibration from per-sensor egomotion," in *Proc. of Robotics: Science and Systems (RSS)*, 2012.

[15] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. W. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2011, pp. 127–136.

[16] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011.

[17] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2012.

[18] C. Sprunk, G. D. Tipaldi, A. Cherubini, and W. Burgard, "Lidar-based teach-and-repeat of mobile robot trajectories," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013.