Fast and Robust Registration and Calibration of Depth-Only Sensors

Andre Mühlenbrock, Roland Fischer, René Weller and Gabriel Zachmann

University of Bremen, Computer Graphic and Virtual Reality, Germany

Abstract

The precise registration between multiple depth cameras is a crucial prerequisite for many applications. Previous techniques frequently rely on RGB or IR images and checkerboard targets for feature detection, partly due to the depth data being inherently noisy. This limitation prohibits the usage for use-cases where neither is available. We present a novel registration approach that solely uses depth data for feature detection, making it more universally applicable while still achieving robust and precise results. We propose a combination of a custom 3D registration target - a lattice with regularly-spaced holes - and a feature detection algorithm that is able to reliably extract the lattice and its features from noisy depth images.

CCS Concepts

• ***Computing methodologies*** → ***Interest point and salient region detections;*** *Camera calibration;*

1. Introduction

RGB-D and range cameras are widely used throughout the research community, e.g. for 3D reconstruction, SLAM, or object recognition. In case of large spaces or to avoid occlusions, multiple cameras are often used. In these cases, the sensors have to be calibrated extrinsically (registered to each other). The classical approach to calibrate and registrate multiple cameras is based on feature detection on planar checkerboards as they provide a good target in color and IR images. For example, Macknojia et al. [MCAPL13] synchronously captured a checkerboard in a Kinect's color and IR images for extrinsic calibration between the RGB and depth sensor and between multiple Kinect's. Detection algorithms for checkerboards were continuously improved (e.g. [RSS08, DF18]) and generally yield robust and precise results. However, it is not always applicable, e.g., if no IR or color images are available.

Performing the registration directly on depth images or point clouds (PC), on the other hand, leads to the issue of inherently noisy depth data, which makes accurate feature detection difficult [SZZQ18]. The registration process proposed by Song et al. [SZZQ18] relies on a checkerboard with regularly-spaced hollow squares. Deviations in the noisy depth image are handled by a model-based approach that considers the hole centers to find hole corners for registration. [RAFSA20] proposed to use a 3D checkerboard as target and a process, involving normal estimation, edge detection, and thresholding, to detect it in the depth image.

We present a target detection and registration procedure that is exclusively based on depth images, therefore avoiding the need for supporting color or IR images. Our approach leads to accurate registration results despite the noisy input. We achieve this by using a special 3D checkerboard-like registration target and a pipeline designed for the precise recognition of the target's reference points.

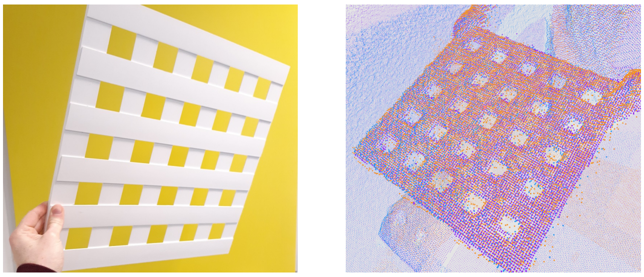

Figure 1: *Our 3D registration target: a lattice-like board with 25 holes. Left: photo; right: registered point cloud.*

2. Our Approach

Instead of detecting a checkerboard pattern in the IR image for calibration as is often done, we have designed a lattice that can be detected directly in the depth image. This lattice consists of 25 evenly spaced rectangular holes (4cm x 4cm each) (see Fig. 1).

The main challenge of the classical camera registration procedure is the correct recognition of the board's features in the image. Since we use only depth data instead of an RGB image, we cannot simply reuse the image-based detection algorithms. Instead, we present a novel approach that is fast and easy to implement yet achieves robust results in all our test cases.

In the first step, we search for gaps along the scanlines in the

DOI: 10.2312/egp.20211033

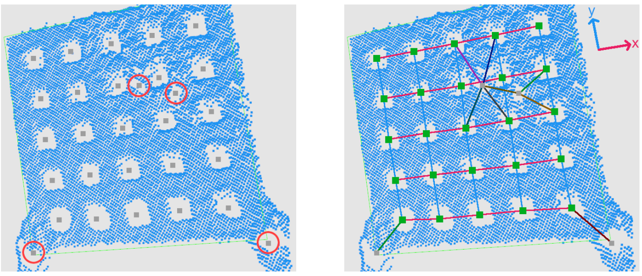

Figure 2: *Estimation of axes based on noisy grid nodes (left, gray) using our heuristic. The color of the edges (right) represents the cluster id. Using the two largest clusters (blue and pink), the x- and y-direction of the grid can be stably determined even with very noisy data and likely valid grid nodes (green) can be identified.*

depth image which correspond to the regular geometry of the lattice. In a second step, we try to determine the region of the physical lattice object by clustering all gap segments. Clusters that can not represent the lattice are filtered out by calculating eigenvectors over all gap segment center points of each and every cluster; these must meet certain thresholds in order to qualify for representation of the lattice. Next we determine all 3D points that could belong to the lattice based on their proximity to the gap segments. To filter out noise, e.g. due to the flying pixel effect, we use RANSAC to fit a plane into those points and classify the points into inliers and outliers. Only now do we determine the precise holes in 3D, which are characterized by outliers that are completely enclosed by inliers. To identify these holes, we iterate over all points in scanline order: all points, that are not inliers and are circumscribed by inliers are considered holes. This will yield a number of segments belonging to the same hole which can be merged efficiently by using the union find structure. We finally estimate the centers of each hole by averaging the inliers directly in the neighbourhood of the hole segments.

In order to determine the orientation of the lattice, we have developed a heuristic that can handle even noisy hole centers, which will be called grid nodes in the following. In order to do so, we define the following set of vectors:

$$V = \{n - m \mid l < dist(m,n) < h, \quad m \in H, \quad n \in H\},$$

where $l = d - \delta$, $h = d + \delta$, d = *lattice spacing* and δ is a tolerance (in our case $d = 8$cm and $\delta = 2$cm which represents the geometry of our lattice).

This set is clustered by the angle the vectors subtend with the x-axis. The two largest clusters represent the prevalent directions; the median of each is considered the direction of the x- and y-axis of the lattice (see Fig. 2). We now have the two very stable directions of the calibration lattice in space, but we don't now which one is which axis and their signs. To resolve this ambiguity, we consider the hands holding the lattice to estimate the orientation and align the x-axis such that it points towards the hands. For the third z-axis we use the normal of the plane determined by RANSAC earlier, and align it towards the camera.

Finally, we register the depth cameras using the SVD-based transformation estimation of the Point Cloud Library [RC11].

3. Results

We have evaluated our algorithm in a real-world application: three Kinect V2 cameras were mounted to the ceiling of a surgical room to record open, abdominal surgeries. We have recorded a test scene of ≈5.7k frames for each camera (≈17.2k frames in total) and removed those frames where the lattice was not fully visible (due to occlusion or the lattice was moved out of the view frustum) for each individual camera which results in 11,478 valid frames. Our results show that the lattice was detected in 9,815 frames (85.5%). On average, 22.2 of the 25 hole centers were used in each lattice. To get a first indication of the accuracy and stability of hole center detection, we constructed an ideal regular grid with origin in the central hole along the measured axes and calculated the deviation between detected grid nodes and hole centers of this ideal grid. The average deviation was 4.1mm with a standard deviation of 3mm. We performed our evaluation on an Intel Core i5-4570. Our lattice detection algorithm required an average of 76.67ms per frame, with a standard deviation of 41.94ms.

4. Conclusions & Future Work

We have presented a novel approach for the registration of depth sensors based exclusively on depth data. For this, we adapted the classical checkerboard-pattern: instead of using a colored plane, we designed a lattice-like board with regularly-spaced holes which are visible in the depth image. Moreover, we have developed an algorithm to detect the board reliably in depth images. A special focus was given to the noise that is inherent to depth data. Our evaluation with three Kinect V2 cameras and extensive recordings shows that our lattice detection is very robust and we achieve accurate registration results despite the noisy depth data within less than 80 ms per frame on average. As our procedure is not relying on RGB or IR images, it is highly versatile and applicable to a broad range of use cases. We are currently conducting a more extensive evaluation with ground truth data points. Our code is publicly available as part of an Unreal Engine 4 plugin: https://gitlab.informatik.uni-bremen.de/cgvr_public/lattice_based_registration.

References

[DF18] Duda A., Frese U.: Accurate detection and localization of checkerboard corners for calibration. In *29th British Machine Vision Conference. British Machine Vision Conference (BMVC-29), September 3-6, Newcastle, United Kingdom* (2018). 1

[MCAPL13] Macknojia R., Chavez-Aragon A., Payeur P., Laganiere R.: Calibration of a network of kinect sensors for robotic inspection over a large workspace. pp. 184–190. 1

[RAFSA20] Reyes-Aviles F., Fleck P., Schmalstieg D., Arth C.: Improving rgb image consistency for depth-camera. *Journal of WSCG 28* (01 2020), 105–113. 1

[RC11] Rusu R. B., Cousins S.: 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)* (Shanghai, China, May 9-13 2011). 2

[RSS08] Rufli M., Scaramuzza D., Siegwart R.: Automatic detection of checkerboards on blurred and distorted images. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2008), pp. 3121–3126. 1

[SZZQ18] Song X., Zheng J., Zhong F., Qin X.: Modeling deviations of rgb-d cameras for accurate depth map and color image registration. *Multimedia Tools and Applications 77* (06 2018). 1