To appear in an IEEE VGTC sponsored conference proceedings

Volumetric Calibration and Registration of RGBD-Sensors

Stephan Beck * Bernd Froehlich †

Virtual Reality and Visualization Research Group at Faculty of Media, Bauhaus-Universität Weimar

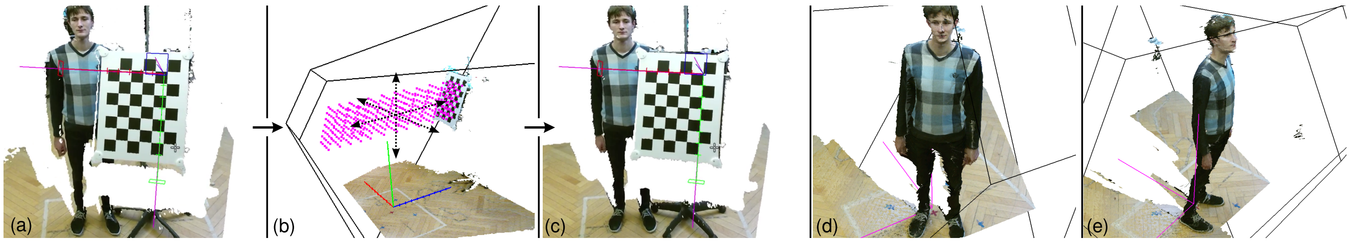

Figure 1: (a) Initial basic calibration. (b) Sampling of references at checkerboard crossing points. (c) Our calibration applied. (d) and (e) Results for two overlapping Kinects V2 positioned about 45 degrees left and right of the user visualized with the real-time 3D reconstruction from [2].

Abstract

We present an integrated approach for the calibration and registration of color and depth (RGBD) sensors into a joint coordinate system without explicitly identifying intrinsic or extrinsic camera parameters. Our method employs a tracked checkerboard to establish a number of correspondences between positions in color and depth camera space and in world space. These correspondences are used to construct a single calibration and registration volume per RGBD sensor which directly maps raw depth sensor values into a joint coordinate system and to their associated color values. Our evaluation demonstrates an accuracy with an average 3D error below 3 mm and an average texture deviation smaller than 0.5 pixels for a space of about 1.5 m x 1.8 m x 1.5 m.

Keywords: Telepresence, 3D capturing, camera calibration, registration, depth camera, Kinect.

1 Introduction

The accurate calibration of RGBD sensors is a challenging task. In 3D telepresence systems, multiple sensors have to be precisely registered to an application's world coordinate system and they have to match in 3D space [5, 2].

We developed an approach which directly maps raw depth sensor values to 3D positions in world space and to their corresponding texture coordinates of the associated color camera image. A tracked checkerboard is placed at various positions in our capturing volume to establish correspondences between raw depth values of an RGBD sensor and the associated positions in world space. A depth camera's infrared image of the checkerboard is used to also establish correspondences between the raw depth values and the texture coordinates of the associated color camera. These correspondences are entered into a 3D lookup table whereby empty cells are filled by scattered data interpolation. This process is performed once for each RGBD sensor. During runtime, this 3D lookup table can be used on the CPU or GPU to map the raw depth values of a sensor to 3D positions in world space and to the corresponding coordinates of the color image.

The main properties of our novel approach are

- a low-latency single-step mapping of raw depth sensor values to positions in world space and to texture coordinates in an associated color image,
- no reliance on any specific lens or camera model and
- an accuracy with an average 3D error below 3 mm and an average texture deviation smaller than 0.5 pixels.

2 Related Work

The specific challenges of calibrating 3D capturing systems were investigated by several researchers [4, 5, 3]. In order to obtain an accurate metric depth measuring Beck et al. [2] and, more recently, Avetisyan et al. [1] explicitly calibrate the depth sensors. Kainz et al. [4] present a multi-Kinect registration that fits a three-dimensional polynomial function to 3D correspondences which are captured from a calibration sphere. Deng et al. [3] suggest a smooth field of rigid transformations and achieve better extrinsic registration compared to systems that only use a single rigid body transform.

The limitations of inter-camera based approaches as presented by [4, 3] are that at least two sensors have to overlap and that the registration results in small geometrical distortions of the captured scene. A more general limitation of depth correction approaches based on sweeping [2, 1] is the missing synchronization between the depth sensor and the tracking system. This can lead to interferences between the different sampling processes and therefore inaccuracies in the depth calibration. We therefore prefer a calibration method that operates with a static target. Furthermore, most of the aforementioned methods depend on the accuracy of a large set of interdependent and error-prone parameters. In contrast, our integrated process makes the resulting calibration independent of any specific lens or camera model. It is independent of the real type of involved distortions and it does not require the sensors to overlap.

3 Calibration Method

The central idea of our method is to correct an initial calibration by sampling a set of references R at the crossing points of a tracked checkerboard for various locations in our area of interest (see also Figure 1 (b)). The references are filled into a 3D lookup table – our calibration volume – whereby empty cells are filled by scattered data interpolation. The calbration is performed using the following steps (cf. Figure 2):

*e-mail:stephan.beck@uni-weimar.de
†e-mail:bernd.froehlich@uni-weimar.de

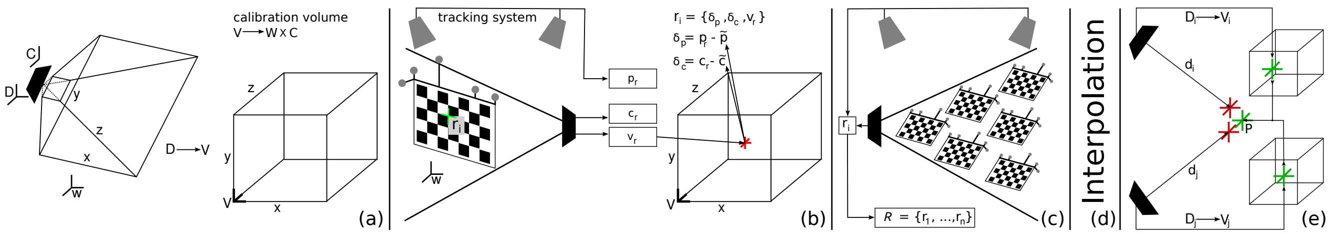

Figure 2: Overview of our proposed method: (a) The depth sensor space D is transformed into a normalized calibration volume space V. (b) The calibration volume can be locally corrected based on a set of references R which we sample at the crossing points of a tracked checkerboard. (c) Reference sampling at various checkerboard locations in our area of interest. (d) Correction of calibration volume at each voxel based on the offsets at its neighbors in R using scattered data interpolation. (e) During runtime, a point $\mathbf{p} \in W$ is reported by, e.g., two sensors to be at local positions $d_i \in D_i$ and $d_j \in D_j$. The correct location of $\mathbf{p}$ can be reconstructed from both sensors by look-ups in their calibration volumes.

1. Initial calibration: We start from an initial calibration Υ of an RGBD-sensor. Υ can be any approximation of the intrinsic and extrinsic parameters of the RGBD sensor. We suppose that Υ is calibrated such that the captured values from depth space D are registered to a joint world coordinate system W. We transform the depth sensor space D into a normalized calibration volume space V and initialize each voxel $\mathbf{v} \in V$ with a point $\mathbf{p} \in W$ and a coordinate $\mathbf{c}$ in the sensor's color texture space C. At this point the calibration volume defines a valid mapping with the accuracy of the initial calibration.

2. Reference sampling: For each reference $r_i \in R$, its texture coordinate $\mathbf{c_r} \in C$ in the sensor's color image is detected and its normalized volume coordinate $\mathbf{v_r} \in V$ is computed using z from the depth and x, y from infrared image of the sensor. The tracking system monitors the checkerboard and reports the measured 3D world position $\mathbf{p_r} \in W$ of r_i. For each r_i, the offsets δ_p and δ_c are computed and stored (Figure 2 (b) and (c)).

3. Interpolation: In a final interpolation step, the volume is locally corrected at each voxel based on the offsets to its neighbors in R. We investigate two interpolation schemes: inverse distance weighting (*IDW*) and natural neighbor interpolation (*NNI*). As a result, the calibration volume contains the corrected calibration with a higher accuracy compared to Υ.

4 Results

We measured the accuracy and the effects of sparse and dense reference sampling and calibrated the Kinect V2 into our joint coordinate system to cover a capturing volume of about 1.5 m x 1.8 m x 1.5 m. The initial calibration was performed using OpenCV. We took approximately 2000 reference samples and divided these into three sets: R_{eval} of size 1000 for evaluation, R_{dense} of size 1000 and R_{sparse} of size 500, a subset of R_{dense}. The sets R_{dense} and R_{sparse} were then used as input for interpolation. We were interested in the errors in terms of the absolute distance to the ground truth. For the 3D error, the ground truth is our tracking system. For the 2D error, the ground truth is the detected crossing point in image space at the reference sample $r_i \in R_{eval}$. The results for *NNI*, as well as *IDW* for different neighborhoods k (5, 10 and 20) and for different resolutions of the calibration volume, are listed in Table 1. Our evaluation clearly shows that our method is able to achieve a very high accuracy. It scales with the density of the reference set (R_{dense} vs. R_{sparse}) and with the resolution of the calibration volume. Figure 1 illustrates the result of our proposed method for a capturing system in the context of 3D telepresence.

Table 1: Average absolute errors in 3D world space and 2D texture space compared to a basic initial calibration; measured for R_{eval} based on the input sets R_{dense} and R_{sparse} for different interpolation schemes and volume resolutions (upper row $64 \times 64 \times 128$, lower row $128 \times 128 \times 256$) per method and row. 3D errors are given in mm, 2D errors in texels, standard deviations in parentheses, and maximum errors in brackets.

Method	3D dense	3D sparse	2D dense	2D sparse
Initial	35.7 (13.0)[75.0]	-	24.2 (1.0)[28.0]	-
IDW5	3.2 (2.1)[14.0]	4.2 (2.9)[18.5]	0.3 (0.2)[1.1]	0.3 (0.2)[1.2]
	3.2 (2.1)[13.6]	4.1 (2.9)[17.0]	0.3 (0.2)[1.9]	0.3 (0.2)[1.2]
IDW10	3.1 (2.0)[14.7]	4.3 (2.8)[16.0]	0.3 (0.2)[1.2]	0.3 (0.2)[1.2]
	3.1 (2.1)[13.9]	4.2 (2.8)[15.8]	0.3 (0.2)[1.2]	0.3 (0.2)[1.2]
IDW20	3.0 (2.0)[17.7]	4.5 (2.7)[15.8]	0.3 (0.3)[5.0]	0.4 (0.2)[1.1]
	3.0 (2.1)[16.5]	4.4 (2.8)[16.5]	0.3 (0.2)[1.0]	0.4 (0.2)[1.1]
NNI	1.7 (1.0)[5.0]	2.0 (1.2)[5.7]	0.2 (0.2)[1.5]	0.3 (0.2)[1.5]
	1.7 (1.1)[5.8]	2.0 (1.3)[6.9]	0.2 (0.2)[1.3]	0.3 (0.2)[1.9]

5 Conclusion

We presented an integrated method for the accurate calibration and registration of multiple RGBD-sensors into a joint coordinate system without explicitly identifying intrinsic or extrinsic camera parameters. Our evaluation shows that we are able to register the sensors with an average accuracy of about 2-3 mm for the Kinect V2 into our joint coordinate system. We also achieved a texture coordinate deviation smaller than 0.5 pixels in the color camera. Our calibration and registration process could be accelerated by sweeping the checkerboard through the capturing space. However, this would require that the tracking system and the RGBD-sensors be in sync, which might become possible with next generation hardware.

References

[1] R. Avetisyan, M. Willert, S. Ohl, and O. Staadt. Calibration of depth camera arrays. In *Proceedings of the 13th SIGRAD 2014 Conference of the Swedish Eurographics Chapter, Eurographics Association*, 2014.

[2] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. In *IEEE Transactions on Visualization and Computer Graphics*, volume 19, pages 616–625, Los Alamitos, CA, USA, April 2013. IEEE Computer Society.

[3] T. Deng, J. C. Bazin, T. Martin, C. Kuster, J. Cai, T. Popa, and M. H. Gross. Registration of multiple rgbd cameras via local rigid transformations. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2014.

[4] B. Kainz, S. Hauswiesner, G. Reitmayr, M. Steinberger, R. Grasset, L. Gruber, E. Veas, D. Kalkofen, H. Seichter, and D. Schmalstieg. Omnikinect: real-time dense volumetric data acquisition and applications. In *Proc. of VRST 2012*, VRST '12, pages 25–32, New York, NY, USA, 2012. ACM Press.

[5] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *Proc. of ISMAR 2011*, pages 137–146, Washington, DC, USA, 2011. IEEE Computer Society.