Learning Depth Calibration of Time-of-Flight Cameras

David Ferstl, Christian Reinbacher, Gernot Riegler, Matthias Rüther, Horst Bischof {ferstl,reinbacher,riegler,ruether,bischof}@icg.tugraz.at



Figure 1: System overview. Using a planar target with known dimensions the system automatically detects feature points in both the RGB and the depth image. These points are used for a full stereo calibration and serve as ground truth 3D points for depth calibration.

Time-of-Flight (ToF) sensors are widely used in many applications such as autonomous navigation, 3D reconstruction and human-computer interaction. Thanks to an adoption by the gaming industry, ToF cameras have become reliable and affordable and are present in many living rooms as an accessory to gaming consoles. These modern devices combine a depth sensor with a traditional RGB sensor to form a so-called RGB-D camera.

Unlike passive 3D sensors based on stereo triangulation, ToF cameras emit light and therefore deliver acquisitions mostly independent from lighting conditions resulting in dense, real-time depth-maps of nearly arbitrary surfaces. The quality of the depth-maps depends on the scene geometry and surface properties in the scene. The two main error sources are 1) intrinsic calibration errors and 2) a spatially varying, context sensitive error in measured depth. This paper tackles both error sources of a RGB-D camera system using a novel calibration framework. The main goal of this work is to provide a highly accurate calibration and error compensation with a minimum of user interaction and without relying on a complex, specifically machined 2.5D calibration target.

According to the two error sources, our method is divided into two main parts: 1) The geometric camera calibration, to accurately estimate the intrinsic and extrinsic parameters of the RGB-D system. Features are automatically detected on a calibration target with sub-pixel accuracy which are used in both single-camera and stereo calibration. 2) The compensation of the measured depth errors learned from ground-truth depth data acquired during geometric camera calibration. Both parts of our method require a minimum of user interaction besides providing the calibration images. An overview of our RGB-D calibration framework is shown in Fig. 1.

We treat the intensity image of the depth camera as a regular intensity camera which allows us to parameterize the depth camera using a standard pinhole model with non-linear lens distortion. After estimating the extrinsic and intrinsic parameters of both depth and RGB camera we aim at minimizing the spatially varying depth bias which is inherently present in ToF-based depth measurements [1]. It has been shown that this error bias depends on scene geometry, reflection properties of the scene and position of the pixel in image space [2, 3]. Instead of directly modeling each error distribution with its own heuristics. Unfortunately, the different error sources highly correlate, which makes it hard to calibrate each error on its own. In our work we solve the problem of depth-bias compensation directly by using machine learning. We use a Random Regression Forest (RRF) to directly infer the mapping from depth and intensity features to a depth offset, which eliminates all error sources. Graz University of Technology Institute for Computer Graphics and Vision Inffeldgasse 16, 8010 Graz, AUSTRIA



Figure 2: Evaluation of the influence of depth camera calibration to dense 3D reconstruction. The first row shows the color-coded error to GT(d) for one depth acquisition. It compares the uncalibrated camera output *D* (a), the camera output corrected with depth dependent polynomial fitting *D*-*Fit* (b) to our method *OURS* (c). Regions where no depth data is available are marked with zero. The second row shows the rendered *KinectFusion* output of an image series of the *Desk Scene* with magnified regions of interest.

		Desk Scene				
	D	S3D-calib	I-Fit	D-Fit	OURS	
RMSE(E)[mm]	168.48	158.23	47.45	46.09	39.83	
Ē[mm]	-141.19	-139.33	-23.40	-18.27	2.65	
$\sigma(E)[mm]$	91.92	74.99	41.67	42.32	39.74	

Table 1: Accuracy evaluation. The depth error is measured as RMSE(E), mean error \overline{E} and standard deviation $\sigma(E)$ to GT in mm. The accuracy is evaluated for the recorded *Desk Scene*. The error of our method is compared to the uncalibrated camera output (*D*), the camera output calibrated with the values provided by the manufacturer (*S3D-calib*), to error fitting to ToF intensity (*I-Fit*) and to error fitting to depth (*D-Fit*) according to [1, 2].

The quality of our method is shown for a *Desk Scene* which is selected to incorporate a high texture and structure variation. Since the utilized depth camera has a depth range of ≈ 300 to 1000 mm the scene size is chosen appropriately. The GT measurement is generated using a Structured Light (SL) scanner which consists of two 4MP intensity cameras and one projector. The depth uncertainty of the SL system at the given baseline is ≤ 1.2 mm. We compare the different methods on single ToF acquisitions in Tab. 1 and show an improvement over separate depth or intensity fitting methods currently used in practice ([1, 2]). Further, we show that our method outperforms existing correction techniques and improves the quality of dense 3D reconstruction in Fig. 2. A more detailed evaluation can be found in the paper.

The MATLAB code for the automatic feature detection and calibration is available on our website¹ since we believe that it can not only benefit RGB-D camera calibration but also a variety of other camera calibration tasks.

- [1] S. Fuchs and G. Hirzinger. Extrinsic and depth calibration of tofcameras. In *CVPR*, 2008.
- [2] J. Jung, J.-Y. Lee, Y. Jeong, and I.S. Kweon. Time-of-flight sensor calibration for a color and depth camera pair. *TPAMI*, PP(99):1–1, 2014.
- [3] Y. M. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view tof sensor fusion system. In *CVPR Workshops*, 2008.