

Time-of-flight sensor calibration for a color and depth camera pair

Jiyoung Jung, *Student Member, IEEE*, Joon-Young Lee, *Student Member, IEEE*,
Yekeun Jeong, and In So Kweon, *Member, IEEE*

Abstract—We present a calibration method of a time-of-flight (ToF) sensor and a color camera pair to align the 3D measurements with the color image correctly. We have designed a 2.5D pattern board with irregularly placed holes to be accurately detected from low resolution depth images of a ToF camera as well as from high resolution color images. In order to improve the accuracy of the 3D measurements of a ToF camera, we propose to perform ray correction and range bias correction. We reset the transformation of the ToF sensor which transforms the radial distance into the scene depth in Cartesian coordinate through ray correction. Then we capture a planar scene from different depths to correct the distance error that is shown to be dependent not only on the distance but also on the pixel location. The range error profiles along the calibrated distance are classified according to their wiggling shapes and each cluster of profiles with similar shape are separately estimated using a B-spline function. The standard deviation of the remaining random noise is recorded as an uncertainty information of distance measurements. We show the performance of our calibration method quantitatively and qualitatively on various datasets, and validate the impact of our method by demonstrating an RGB-D shape refinement application.

Index Terms—Time-of-flight sensor calibration, time-of-flight range error analysis, color-depth camera fusion, Kinect

1 INTRODUCTION

FOR the past few decades, depth sensors have been commonly used for navigation in robot application. Then many researchers have shown promising results to make use of a depth sensor as an active way for shape recovery of an object [1]. Unlike 3D reconstruction methods using images such as stereo, structure-from-motion, and shape-from-shading, a depth sensor provides the metric distance of the scene from the sensor, which can relieve ambiguity and scale problems that the image based reconstruction methods bear.

Recently, depth sensors have become popular as the sensors are getting cheaper and smaller. They are widely used in computer vision community, for any application that can take advantage of real-time 3D information such as gesture and action recognition for human-computer interaction [2], [3], augmented reality and handheld 3D scanning for mobile devices [4], and SLAM for autonomous vehicle and drone navigation [5]. Depth measurements are now often acquired with color images as a format of RGB-D input [6], [7], [8].

A 3D time-of-flight (ToF) camera is a type of depth sensors which modulates its illumination LEDs and measures the phase and the amplitude of the returned signal with its CCD/CMOS imaging sensor at each pixel. The new generation Kinect is installed with a ToF camera

which has greater accuracy compared to the previous Kinect with structured-light [9].

A 3D ToF camera provides an amplitude image which represents the amplitudes of the returned signals as well as a 3D point cloud of the scene. One might think that since a ToF camera provides amplitude images which look similar with traditional grayscale images, the existing calibration methods would work successfully on estimating the camera parameters of a ToF sensor as well. However, the methods dedicated to estimate intrinsic and extrinsic parameters of color cameras [10] or to extrinsic calibration of a camera with a 2D laser range finder [11], [12] are not adequate to calibrate a ToF camera or a sensor fusion system containing a ToF camera.

The main reason that the traditional calibration methods do not work well on calibrating a ToF-color camera pair comes from the characteristics of the ToF sensor. Unlike a 2D laser range finder, the range measurement that a typical 3D ToF camera provides is less accurate, and the amplitude images are blurry and in low resolution compared with general color images. Therefore the process of calibrating a sensor fusion system which includes a ToF camera has to be designed to overcome the weakness of the sensor.

In this paper we present an accurate and practical calibration method for a time-of-flight and a color camera pair. Since an accurate correspondence acquisition is the key in homography based calibration, we have designed a 2.5D pattern board with holes for feature detection on both color and ToF depth images. The holes are more robust to be detected in a low resolution image because the center of a circular pattern is preserved when the

-
- J. Jung, J.-Y. Lee, and I. S. Kweon are with the Department of Electrical Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea. E-mail: {jjjung, jylee}@rcv.kaist.ac.kr, iskweon@kaist.ac.kr
 - Y. Jeong is with Microsoft Corporation, One Microsoft Way, Redmond WA 98052-6399. E-mail: yejeong@microsoft.com

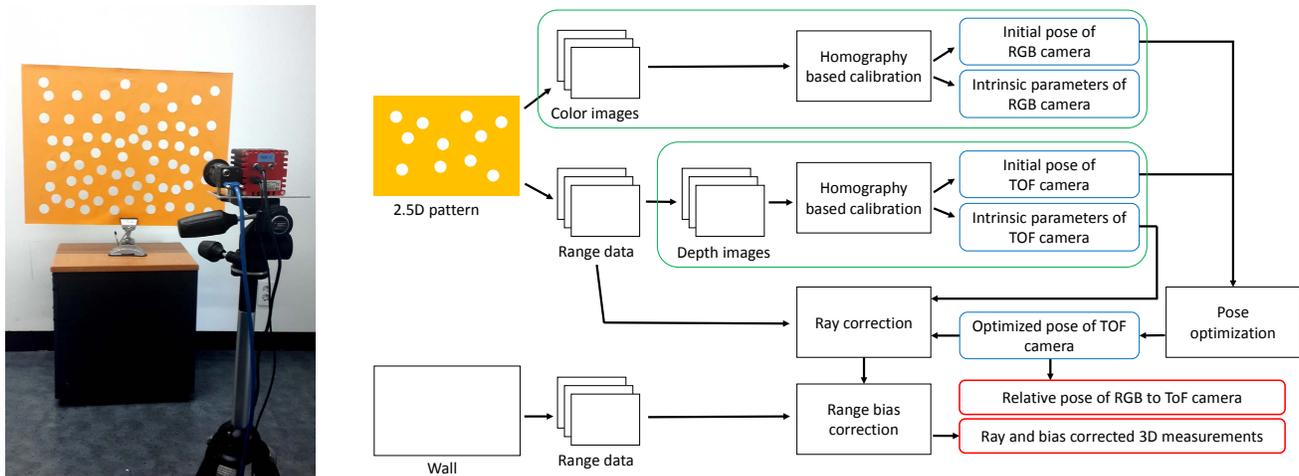


Fig. 1: System setup and an overview of the proposed calibration method.

image is isotropically blurred.

We calibrate the original per-pixel range measurement of the ToF camera, which is the radial distance of the scene, instead of the converted distance in Cartesian coordinate. We first correct the ray direction of each pixel because the manufacturer provided range-to-3D transformation may be incorrect to individual sensors due to mechanical differences in manufacturing process. After ray correction, the range error along each ray is measured by capturing a planar scene (e.g. a wall) from different depths. The range error for each pixel is wiggling along the range measurement. To the best of our knowledge, this is the first report that the wiggling shape or the amount of bias of the range error profile depends on its pixel location on the ToF sensor. We cluster the range error profiles with similar shape and estimate the range bias of the pixels in each cluster separately using a B-spline function.

The resulting 3D measurements are ray corrected and bias eliminated values and they are transferred onto the color image using the optimized pose to be rendered as an accurate 3D scene. The flow chart of the system is illustrated in Fig. 1 with our system setup. We use a MESA SwissRanger SR4000 and a PointGrey Flea3.

A preliminary version of this work appears in [13]. Ray correction and spatially varying range bias estimation are added to compensate the error of the range measurements of a ToF camera. Quantitative and qualitative analyses on various experiments are performed to show the effectiveness of the proposed method. Also, we apply our calibration result to an RGB-D shape refinement application and demonstrate a compelling result. The source code of our software is available online at our website [14].

1.1 Previous work

One of the most popular calibration methods is [10]. It is a homography based camera calibration process that uses 2D metric information of a checkerboard. Based on

this method, an extrinsic calibration of a camera and a 2D laser range finder with the constraint on depth measurements has been presented [11]. These two methods are often treated as a baseline of later calibration methods of various sensor systems which are usually modified to consider the specific setups and sensor characteristics.

The calibration of a 3D ToF camera has been studied in a comparatively recent time. Several works have been focused on calibrating a 3D ToF sensor fusion system. Fuchs and Hirzinger [15] suggest a distance error model which compensates distance, amplitude, and latency related errors. Kim *et al.* [16] present a depth denoising algorithm based on parametric noise modeling. Kim *et al.* [17] estimate the pose of the sensor by applying [18] to the ToF amplitude images and use it to model the systematic error of the sensor, which is to compare the transformed coordinate with the distance measurement.

The sensor characterizations and the calibration methods of a ToF camera are examined in metrology community as well. A number of approaches are presented to overcome the low quality of the ToF amplitude images. Lindner and Kolb [19] add a preprocessing step to the amplitude images to stabilize and speed up the pattern recognition process. Kahlmann *et al.* [20] propose a calibration pattern consisting of filled white circles on a black background. Kern [21] uses a plane with holes but their objective is to calibrate a laser scanner which provides much more accurate depth measurements than a 3D ToF camera. Since their holes are arranged in a grid, they have to go through another algorithm to identify the holes, whereas our pattern has holes spread uniquely so that identification process becomes very simple. Beder and Koch [22] incorporate the measured depth values in addition to the amplitude images of the checkerboard, which makes a single image suffice to estimate the focal length and the pose of the sensor. However, since the lens distortion and distance measurement errors are neglected, this method yields a lack of precision.

Several metrological literatures focus on the systematic

distance error of a ToF camera. As stated in [23], a periodic “wiggling” error which causes the calculated depth to oscillate around the actual depth arises due to the deviations of the modulated light signal from a perfect sine function. In order to calibrate the measured depth, Lindner and Kolb [19] combine a pixelwise linear calibration with a global B-splines fit, whereas Schiller *et al.* [24] use a polynomial to model the distance deviation. Schmidt [25] uses a model based approach to predict the behavior of the non-sinusoidal light signal.

There are studies on other sources of errors of the ToF camera. Lindner and Kolb [26] deal with intensity related distance error and Steiger *et al.* [27] report the change of distance measurements along with the camera temperature. For depth denoising as a post processing to the depth data, Swadzba *et al.* [28] propose a filtering pipeline and Reynolds *et al.* [29] propose to identify and discard the incorrect depth values using confidence measures. To reconstruct the depths for discarded pixels, super-resolution approaches that use depth information from the surrounding pixels are introduced [30], [31].

There are noticeable works presented in a couple of years which improve strengths in individual stages and arrange them into a high performance calibration solution. Schiller *et al.* [24] use an analysis-by-synthesis approach which synthesizes the depth and amplitude images from GPU and find the camera parameters that minimizes the error with the input images through non-linear optimization. Lindner *et al.* [32] combine intrinsic, distance, and reflectivity related errors in the similar framework. They initialize all the parameters with standard computer vision techniques from OpenCV [33] on checkerboard images and use averaging for temporal denoising and bilateral filtering for spatial denoising as suggested in [34]. The depth errors are corrected globally with a B-spline function. The performance is improved by synthesizing depth images with additional high resolution color cameras. However, they focus less on the spatial distribution of the depth error and the acquisition of precise location of features on the images.

Though targeted for a Kinect providing a disparity map, Herrera *et al.* [35] correct disparity distortion along the depth and spatial location, motivated by [36]. They focus on the fact that the disparity error shows radial distribution that decays exponentially as the Kinect disparity increases. The pose between the color and depth sensors is estimated using coplanar constraint rather than correspondence matching, which may result in inaccurate transfer of the depth measurements on the image corners with large radial distortion at the cost of simplicity of the calibration process.

2 ACCURATE CORRESPONDENCE ACQUISITION

A traditional calibration method of a camera such as [10], [18], [33] generally locates the correspondences between a model object with known geometry and its projection

on the image by feature detection. A non-planar 3D calibration object with very high geometric precision may be preferred in high-quality photogrammetric calibration. However, since the object should cover the complete 3D measurement range of the camera system, 3D calibration objects are difficult to manufacture and handle. In addition, when calibrating an optical camera together with a depth camera, the design of such 3D pattern is often not possible due to the different imaging modalities of depth and color [23]. Therefore, a planar 2D calibration pattern is preferred which allows a much easier calibration procedure.

The correspondences between the pattern and the images are used to model the relationship between the world and the image coordinates. The calibration includes pose estimation of the sensor, which computes the 3D transformation from the world coordinate to the camera coordinate, and intrinsic parameter estimation, which models the projection relationship between the camera coordinate and the image coordinate.

It is essential to acquire a sufficient number of accurate correspondences between the world and the camera system in sensor calibration. In case of calibrating a color camera which provides high quality color images, corner detection on the images of a black and white checkerboard yields very accurate correspondences of the known 2D geometry and the image. However, we cannot expect such an accurate set of correspondences through corner detection on the amplitude images of a ToF camera because they are blurry and in low resolution as shown in Fig. 2(f). A preprocessing to the amplitude images [19] or manual selection of the feature points may be possible alternatives, but instead, we propose to use the depth images of the ToF camera.

2.1 A 2.5D pattern board

The major product of a time-of-flight camera is the 3D measurement of the scene, not the amplitude images. The amplitude images are merely offering a hint of the texture of the scene due to difference in reflectivity of near-infrared light. Since our goal is to align the depth information from a ToF camera onto the color image, it is proper to estimate the necessary camera parameters using 3D measurements rather than amplitude images.

We have designed a 2.5D pattern board that consists of features to be easily and correctly detected by the range measurements of a ToF camera as well as by color images. The 2.5D pattern board consists of 64 holes irregularly placed on a plane as shown in Fig. 2(a). The board is 80×60cm in size and the diameter of a hole is 4cm, which is large enough for the near-infrared rays to pass through so that the circular pattern is clearly shown in a 176×144 sized depth image as shown in Fig. 2(b). The location of the centers of the holes are predefined. We detect the ellipses in the depth images, find their centers, and use them as feature points to estimate the homography using the calibration toolbox for a color camera in [37].

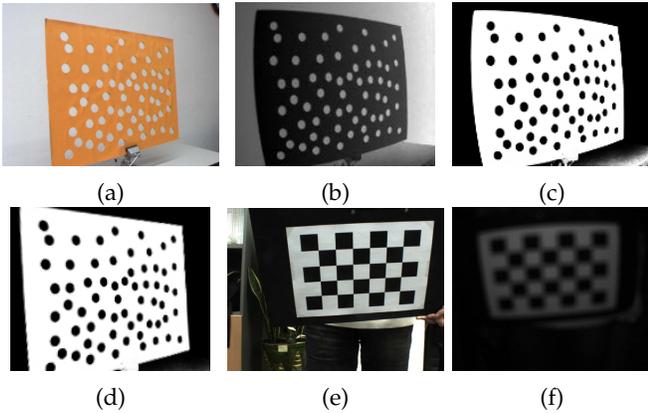


Fig. 2: (a) The 2.5D pattern board in a 640×480 color image and (b) in a 176×144 depth image. (c) The same depth image rescaled for better feature detection and (d) radial distortion removed for better homography estimation. (e) A black-and-white checkerboard in a color image and (f) in an amplitude image of a ToF camera.

2.2 Depth image processing

A depth image refers to a grayscale image of the z -components of the 3D measurements. Although it is still possible to use the depth image as it is, the accuracy of the estimated homography of each depth image improves when we rescale, i.e. redefine the intensity mapping of the depth image and remove the radial distortion as shown in Fig. 2(c) and (d).

The rescaling process improves the performance of ellipse detection. We manually locate four corners of the pattern board in each depth image and calculate the plane parameter using 3D measurements of those four corners by SVD. We rescale the depth image as,

$$I(z_t) = \exp(-d_{plane}^2/\tau^2) \tag{1}$$

where z_t is the z -component of a 3D measurement at a pixel and d_{plane} is the distance between the plane and the 3D measurement. τ is a weight to control the intensity decay, which we have set as $\tau = 200$.

In the process of homography estimation using the detected features, i.e. the centers of ellipses, we reject the feature points which do not satisfy the homography constraint. While this outlier rejection is necessary in order to disregard erroneous detection of ellipses in the scene, the correct features near the image boundary tend to be eliminated along with the outliers if we use radially distorted depth images. Therefore, we first apply a homography based calibration to rescaled depth images, remove radial distortion from the depth images using the estimated radial distortion parameters, and then apply a homography based calibration process again to the rescaled and radially undistorted depth images.

Table 1 shows the effect of the proposed 2.5D pattern board on correspondence acquisition for ToF camera calibration. To evaluate the accuracy of the feature detection, we estimate the homography between the model plane

TABLE 1: Average projection error of homography estimation [pixel]

Camera	Color	ToF		
		Amplitude	Depth	Radially undistorted depth
Type of images	Color			
Corner detection of a checkerboard [38]	0.479	0.495	-	-
Ellipse detection of the 2.5D pattern board (proposed)	0.832	-	0.226	0.154

and its images using the detected features and compute the pixel deviation of the projected 3D points by the estimated homography. The traditional corner detection of a black-and-white checkerboard is compared using the public calibration software [38], which is the implementation of the state-of-the-art calibration methods [24], [32]. The checkerboard we used is 80×64 cm in size that consists of squares of 80 millimeters on a side. 63 corner points are used as correspondences to be fair with 64 holes of the proposed 2.5D pattern board. The average projection error of the corner detection in Table 1 is the best result of several trials with some user assistance on feature selection.

In spite of the mixed depth pixels around the holes and perspective effects that cause circular patterns to appear as ellipses, our proposed calibration pattern shows clear strength on ToF camera calibration. Although the corner detection of a checkerboard is highly accurate for high resolution color images, the ellipse detection of the 2.5D pattern board on low resolution depth images with or without rescaling and radial distortion removal shows better performance on correspondence acquisition.

For the experiment in Table 1, we used a color camera with a resolution of 1600×1200 and a ToF camera with a resolution of 176×144 .

3 POSE OPTIMIZATION

We detect the centers of ellipses in the depth and color images and obtain the intrinsic parameters of the ToF and color cameras as well as the pose of the camera for each image using the homography based calibration method [37]. The initial estimates of parameters of the two cameras are obtained independently. However, since the ToF camera and the color camera are mounted on a sensor rig as shown in Fig. 1, we need to design a function to optimize the poses of both sensors together.

We use an intrinsic model as [10], which consists of a pinhole model with radial distortion correction. A 3D point $\mathbf{x}_t = [x_t, y_t, z_t]^T$ in the ToF camera coordinate is first normalized by $\mathbf{x}_n = [x_n, y_n]^T = [x_t/z_t, y_t/z_t]^T$. Radial distortion is modeled as:

$$\mathbf{x}_k = (1 + k_1 r^2 + k_2 r^4) \mathbf{x}_n \tag{2}$$

where $r^2 = x_n^2 + y_n^2$ and $\mathbf{k}_t = [k_1, k_2]$ is a vector containing the distortion coefficients. The image coordinate $\mathbf{p}_t =$

$[u_t, v_t]^T$ is obtained:

$$\begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} = \mathbf{K}_t \begin{bmatrix} \mathbf{x}_k \\ 1 \end{bmatrix} = \begin{bmatrix} f_{tx} & 0 & u_{0t} \\ 0 & f_{ty} & v_{0t} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ 1 \end{bmatrix} \quad (3)$$

where $\mathbf{f}_t = [f_{tx}, f_{ty}]$ is the focal length and $\mathbf{p}_{0t} = [u_{0t}, v_{0t}]$ is the principal point. The same model applies to the color camera.

In the following optimization, the intrinsic parameters of both cameras remain fixed as the initial estimates. The pose of the ToF camera of each image and the relative pose of the color camera with respect to the ToF camera are optimized.

Let the projection of a 3D point \mathbf{X} in the world coordinate onto the i -th ToF depth image be \mathbf{p}_{ti} and onto the i -th color image be \mathbf{p}_{ci} :

$$\mathbf{p}_{ti} = \mathbf{K}_t [\mathbf{R}_{ti} \quad \mathbf{t}_{ti}] \mathbf{X} \quad (4)$$

$$\mathbf{p}_{ci} = \mathbf{K}_c [\mathbf{R}_{t2c} \quad \mathbf{t}_{t2c}] \begin{bmatrix} \mathbf{R}_{ti} & \mathbf{t}_{ti} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{X} \quad (5)$$

where \mathbf{K}_t and \mathbf{K}_c are the intrinsic camera matrices defined in Eq. (3) of the ToF camera and the color camera respectively. Radial distortion of both cameras are modeled as Eq. (2). $[\mathbf{R}_{ti} \quad \mathbf{t}_{ti}]$ is the i -th ToF camera pose with respect to the world coordinate and $[\mathbf{R}_{t2c} \quad \mathbf{t}_{t2c}]$ is the color camera pose with respect to the ToF camera, which are optimized through the following error minimization:

$$\min \sum_i w_t \|\mathbf{p}_{ti} - \widetilde{\mathbf{p}}_{ti}\|^2 + w_c \|\mathbf{p}_{ci} - \widetilde{\mathbf{p}}_{ci}\|^2 \quad (6)$$

where $\widetilde{\mathbf{p}}_{ti}$ and $\widetilde{\mathbf{p}}_{ci}$ are the feature points extracted in the i -th ToF depth image and the i -th color image respectively.

We use Levenberg-Marquardt optimization to minimize Eq. (6). w_t and w_c are the weights to balance the projection errors in depth and color images. The weights are determined according to the ratio of the resolution of a color image Ω_c to that of a depth image Ω_t . We set $w_c = 1$ and $w_t = \sqrt{\Omega_c/\Omega_t}$, which means that 3.5-pixel error in a color image is considered same as 1-pixel error in a depth image when $\Omega_t = 176 \times 144$ and $\Omega_c = 640 \times 480$. Note that the weights should reflect the actual overlapping portion of the images when the field of views of two cameras are appreciably different.

4 DEPTH CORRECTION

There are a number of error sources in a ToF camera which influence the accuracy of its 3D measurements. A comprehensive depth correction has to be performed to obtain accurate depth data from the sensor. Since the ToF camera measures the time of flight along the light path, the error should be corrected with respect to the range measurement, not in Cartesian coordinates [23]. Moreover, the correct range-to-3D transformation is also important because the 3D measurements after the conversion is the depth data which we actually use.

In Section 4.1, we correct the range-to-3D transformation for each pixel so that its ray direction coincides with the estimated intrinsic parameters. The estimation of the range error and its complex behavior are described in Section 4.2. The analysis on the distribution of the range bias error along with the range measurement and the pixel location followed by an effective error correction are described in Section 4.3.

4.1 Ray correction

A ToF camera provides a 3D measurement $[x_t, y_t, z_t]^T$ of the scene but it is actually a transformed value of the range measurement along the certain ray direction for each pixel, represented in a 3D Cartesian coordinate. The transformation is predefined by the manufacturer and the ray direction $[x_t/z_t, y_t/z_t]^T$ at each pixel is fixed when the modulation frequency of the ToF camera remain unchanged during the capture. However, this range-to-3D transformation may be inaccurate to individual sensors due to mechanical differences arisen from the manufacturing process. Therefore we first reset the transformation to be in accord with the intrinsic parameters estimated by the homography based calibration, before correcting depth errors included in the range measurement.

Given an image coordinate $[u_t, v_t]^T$, the ray direction $\mathbf{x}_n = [x_n, y_n]^T$ at the pixel is defined by the intrinsic parameters of the camera $\Pi_t = \{\mathbf{f}_t, \mathbf{p}_{0t}, \mathbf{k}_t\}$ by Eq. (2) and Eq. (3). Since the 3D space defined by the manufacturer provided range-to-3D transformation tends to be inaccurate to individual sensors, we propose to reset the transformation to obtain the correct 3D measurements.

Let a range measurement be R and we model the range-to-3D transformation as,

$$\begin{aligned} z_t &= s(R + R_b) \\ x_t &= x_n z_t = x_n s(R + R_b) \\ y_t &= y_n z_t = y_n s(R + R_b) \end{aligned} \quad (7)$$

where R_b is a range bias, and s is a scalar to normalize each component, i.e. $s = 1/\sqrt{x_n^2 + y_n^2 + 1}$.

Since there are no unknowns left in Eq. (7) except R_b after the homography based calibration, the ray corrected 3D measurements are obtained by simply applying Eq. (7) to the raw range measurements R of a ToF camera by setting $R_b = 0$, assuming R_b is relatively small.

4.2 Range bias estimation

To estimate R_b for better accuracy, we use the range data of the 2.5D pattern board. We identify the range measurements fallen on the plane (not the holes) using the plane parameter estimated in Section 2.2. We select the pixels having manufacturer provided 3D measurements within a small distance threshold from the plane (e.g. 10mm). Since we know the optimized pose of the camera, we can calculate the optimized plane parameter \mathbf{N}_{ti} as,

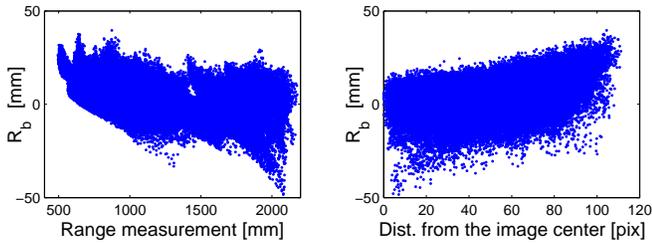


Fig. 3: The distribution of R_b according to range measurement (left) and distance of the pixel location from the image center (right).

$$\mathbf{N}_{ti} = \begin{bmatrix} \mathbf{R}_{ti} & \mathbf{t}_{ti} \\ \mathbf{0}^T & 1 \end{bmatrix}^{-T} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad (8)$$

Without loss of generality, we assume that the plane parameter of the 2.5D pattern board in the world coordinate is $[0, 0, 1, 0]^T$. Then, the ray corrected 3D measurement $[x_t, y_t, z_t]^T$ should satisfy the plane equation:

$$[x_t, y_t, z_t, 1]\mathbf{N}_{ti} = 0 \quad (9)$$

Using Eq. (7), Eq. (8), and Eq. (9), we estimate R_b for each range measurement of the plane. However, as shown in Fig. 3, it is difficult to find any connection of R_b with the range measurement or with the pixel location of the range measurement on the image plane. The range bias varies in both domains and it is impossible to represent the range bias using either one of those variables.

To identify the relationship of the range bias with both the range measurement and the pixel location of the measurement on the image more clearly, we capture a planar scene at different distances and estimate the depth error along the corrected ray direction in the next section.

4.3 Range bias correction

The range measurements that a time-of-flight sensor provides suffer from random noise and systematic bias. The most crucial error source of a ToF sensor is formed by a systematic wiggling error altering the measured distance by shifting the distance information significantly toward or away from the sensor [32]. The systematic wiggling error arises in the process of distance calculation from the phase difference of the reference signal and the returned signal of a ToF sensor. Due to hardware and cost limitations, the theoretical assumption of a sinusoidal signal shape is generally not suitable in reality. As a result, range errors appear as shown in Fig. 4. Instead of modeling the range bias of all the measurements using a single B-spline function, we focus on the spatial distribution of the range error on the image plane as well.

TABLE 2: RMS error before and after range bias correction [mm]

	Before correction	After correction
All pixels corrected together with a single B-spline function [32]	6.18	6.12
Each cluster corrected separately (proposed)	6.18	3.80
Cluster 1 (Fig. 4(b))	3.84	2.92
Cluster 2 (Fig. 4(c))	10.6	5.61
Cluster 3 (Fig. 4(d))	24.2	8.95
Cluster 4	4.87	3.72
Cluster 5	6.28	4.14

Range bias correction is important in enhancing the quality of the 3D measurements of a ToF camera because not only it reduces the error in range measurement but also it removes the offset bias from the error to make the error distribution more like a zero-mean Gaussian. The bias corrected depth becomes more suitable to depth related applications such as [3], [8].

To estimate the error of the range measurement, we have captured a wall at different distances, making sure that the wall is captured more than 30 times at each distance. We have averaged those range measurements to obtain a reliable representative frame of the wall at each distance. We model the plane using SVD because a large portion of the 3D measurement after ray correction is already highly accurate. We estimate the range error as the distance between the 3D measurement and the fitted plane along the ray.

Fig. 4 shows the range error along the measurement of different pixels on the image. The figure in the middle of the column (a) shows the range error profiles of all the pixels. The error is wiggling along the range measurement, showing a wide range of variance. Lindner *et al.* [32] model this error using a single B-spline function. Instead, we have applied k -means clustering to the error data to classify them according to shape of fluctuation. As a result, the wiggling errors are successfully divided into a number of clusters, each of which having a similar shape to be modeled by a single B-spline function. The measurements in each cluster are radially distributed on the image plane, as shown in the top row of Fig. 4. The columns (b-d) show the spatial distribution of the three out of five ($k = 5$) clusters of pixels and their corresponding range error profiles with the estimated B-spline functions.

The RMS errors before and after bias correction are shown in Table 2. It is shown to be much more effective when range error is corrected each cluster independently than when corrected as a whole. The average RMS error of spatially varying range bias correction is 3.80mm, which is much less than that of a single function correction. Note that the pixels on the image corners (Cluster 3) show large error compensation result due to range bias correction. The range bias clearly has a spatial distribution and it is more effective when corrected by

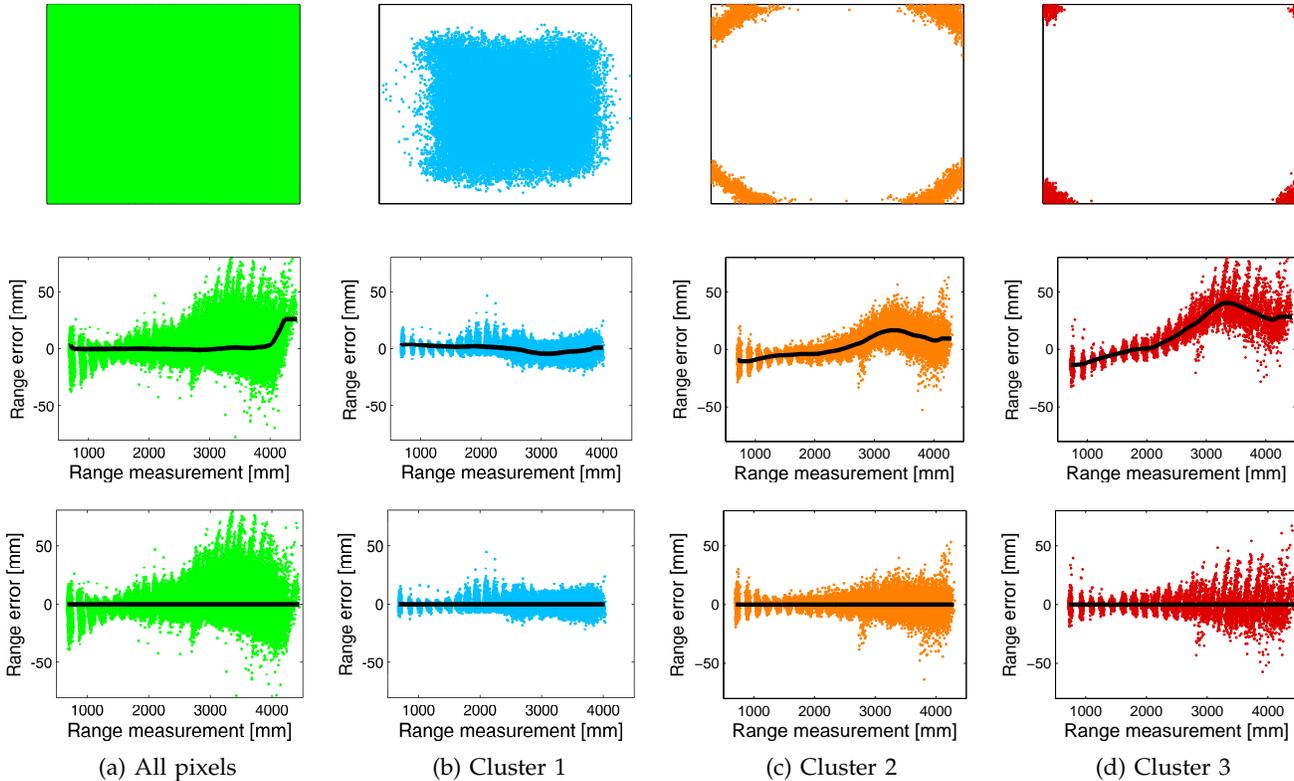


Fig. 4: (Top) Spatial distribution of the measurements. (Middle row) Range errors and the estimated B-spline functions. (Bottom) Range error after correction. (a) Range bias correction of all the measurement using a single B-spline function. (b-d) Three clusters (Cluster 1,2,3 in Table 2) of the measurements showing similar error profiles and their corrections.

using different B-spline functions.

5 EXPERIMENTS

We have performed various experiments to validate the proposed calibration method. In Section 5.1, quantitative analyses are presented to show that each stage of the proposed framework is effective. To provide a clue on the number of necessary images to obtain a certain level of performance, we show performance differences using different number of images of the pattern board and the planar scene in calibration. The angular error between the captured faces of a cuboid is analyzed to show the ray correction is effective on alleviating the depth distortion due to incorrect range-to-3D transformation. The planar scene comparison test show that the different range bias correction for different pixels is necessary, especially for the corner pixels.

The performance comparison with the previous method is described in Section 5.2 and the 3D rendering results of the real scenes are presented in Section 5.3.

We used two sets of color and depth cameras for the experiments. Each set of cameras is mounted together to have a fixed relative pose as shown in Fig. 1. One setting consists of a MESA SwissRanger SR4000 which provides range data of 176×144 in resolution and a PointGrey Flea3 which provides a color image of $640 \times$

480 in resolution. Another set consists of a different ToF camera of the same manufacturer and a color camera with a resolution of 1600×1200 .

The two different ToF cameras have presented the same problem of incorrect interpretation of its ray direction and the similar error distribution on its sensor plane, as a general characteristic of the sensor.

We implemented our algorithm in Matlab. The code has been released for the research community. We provide the exact coordinates of the holes on the 2.5D pattern board so that any one can make one of his own.

The calibration with 18 images of the pattern and 21 images of the planar scene takes 157 seconds (including 129 seconds of depth correction) on a 3.4 GHz computer, but we can expedite the process to 137 seconds (including 122 seconds of depth correction) by using 9 images of the pattern and 6 planar scenes without significant performance degradation.

5.1 Quantitative analysis

Table 3 and Table 4 show the actual effectiveness of the ray correction and range bias correction. When 18, 9 and 5 images are used in calibration, we use another set of 18 images to validate the calibration performance. For evaluation, we calculate the camera poses of 18 color and depth images of the validation set. Then, using

TABLE 3: Average projection errors on color and depth images [pixel] and average range measurement errors of the pattern plane [mm]

		Number of images used in calibration	18	9	5
Projection error [pixel]	Color images		0.307	0.398	0.781
	Depth images		0.291	0.308	0.313
Range error [mm]	Manufacturer provided		12.727	12.727	12.727
	Ray corrected		6.131	6.134	6.150
	Ray and bias corrected		5.731	5.757	5.771

TABLE 4: Percentage of the range measurements having errors smaller than 5 / 10 / 20mm [%]

No. of imgs used in calib.	18	9	5
Manufacturer provided	26.8/51.2/77.1	26.8/51.2/77.1	26.8/51.2/77.1
Ray corrected	48.9/81.1/98.9	49.7/80.7/98.7	49.6/80.6/98.7
Ray and bias corrected	52.2/83.3/99.1	52.8/82.8/98.9	52.6/82.7/98.9

the intrinsic parameters, the optimized pose between two cameras, the corrected range-to-3D transformation, and range error profiles obtained by images of the calibration set and the range data of the planar scene, we calculate the projection errors and the depth errors on validation set. The depth errors are calculated as explained in Section 4.2, as the distance between the range measurements of the pattern board (except for the holes) and the intersection between per-pixel ray and the estimated plane using the pose of each image (Eq. (8)).

When 18 images are used in calibration, the average projection errors on color and depth images after the pose optimization are very small (0.307 and 0.291 respectively). However, the manufacturer provided 3D measurements still show a large misalignment in the rendering result shown in Fig. 10 (b) because its ray direction does not coincide with the estimated intrinsic parameters. The average depth error of the manufacturer provided 3D measurements is not affected by the number of images in calibration set because the intrinsic parameters and the optimized pose between two cameras obtained from the calibration set do not affect the depth error. Given the range measurement, the depth error is determined by the range-to-3D transformation, the range bias elimination and the estimated pose of the image in the validation set.

The correction of ray direction reduces the average depth error in half. The additional correction of the range bias reduces average depth error by 0.4mm. The amount of reduction may seem small, but it is shown that the percentage of the range measurements having a very small depth error (less than 5mm) has increased from 48.9% to 52.2% in Table 4. The range bias correction eliminates a few large bias errors, which barely affects

TABLE 5: Performance difference of bias correction using different number of planar scene range data. Average error of range measurements [mm] and the percentage of the range measurements having errors smaller than 5 / 10 / 20mm [%]

Number of planar scenes	21	16	11	6
Average range error [mm]	5.757	5.775	5.783	5.786
Less than 5mm error [%]	52.8	52.6	52.6	52.6
Less than 10mm error [%]	82.8	82.7	82.6	82.6
Less than 20mm error [%]	98.9	98.9	98.9	98.9

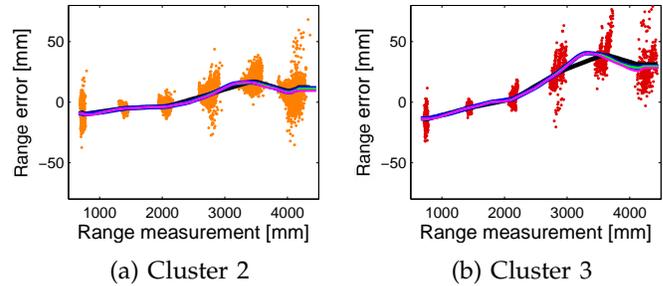


Fig. 5: Range error profiles obtained from 6 planar scenes. The lines show range bias estimation using a planar scene captured from 6 (black), 11 (blue), 16 (green), and 21 different distances (magenta).

the average error but makes the measurements on the image corners more reliable and eventually results in a larger field of view providing accurate range measurements. As a result, 99% of all the range measurements are guaranteed to be in 20mm error using the proposed calibration method.

5.1.1 Analysis on range bias correction

Table 5 shows the performance difference of the range bias correction using the range data of a planar scene at 21, 16, 11 and 6 different distances. For all the cases, the nearest range measurement is 0.7m, and the farthest is 4.4m. The performance is quite similar regardless of the number of the scenes, as long as the capturing positions are regularly placed along the range. Fig. 5 shows the depth error of the Cluster 2 and Cluster 3 in Fig. 4 when 6 planar scenes are used instead of 21. The estimated range bias (in black) is practically the same as the ones using 11 (in blue), 16 (in green) and 21 scenes (in magenta).

It is clear that the systematic range bias can be reduced if its spatial distribution is considered in estimation. The range bias correction can be left as optional, since most of the pixels near the image center are already highly accurate without bias elimination. However, the additional range bias correction using a planar scene at several distances can improve the accuracy of the pixels at the image corners substantially. It can be a strong advantage of the proposed calibration considering that the field of view of a ToF camera is very small.



Fig. 6: A color image of the cuboid and a manually constructed mask for its three faces.

TABLE 6: Angular error between reconstructed planes in degrees. \angle_{ab} represents the angle between the fitted planes of face a and b.

	$90^\circ - \angle_{12}$	$90^\circ - \angle_{23}$	$90^\circ - \angle_{31}$
Manufacturer provided	1.64°	5.98°	-0.78°
Ray corrected	1.49°	3.68°	-0.32°
Ray and bias corrected	1.16°	3.54°	-0.19°

5.1.2 Real cuboid analysis

We evaluated the performance of calibration by reconstructing a cuboid whose faces are known to be at 90 degrees from each other. Fig. 6 shows a color image of a cuboid and its mask. We manually divided the three faces of the cuboid based on its appearance in the color image.

We constructed 3D mesh from the manufacturer provided, ray corrected, ray and bias corrected 3D measurements and obtained depth maps by using OpenGL from the viewpoint of the color camera using the optimized pose between the ToF and the color camera. The 3D points that belong to each face are selected using the mask in Fig. 6. A plane is fitted to each face using RANSAC and SVD. Table 6 shows the angle between each pair of adjacent faces of three reconstruction. The results show that the stage of ray correction notably improves the metric depth accuracy by alleviating the depth distortion.

5.1.3 Planar scene evaluation

The ray correction makes the estimated intrinsic and extrinsic parameters meaningful to represent a ToF camera as well as to a color camera. It is shown in rendering results that ray correction reduces projection errors and aligns the range measurements with the color image correctly. The range bias correction mainly corrects the range measurements on the image corners. We have captured planar scenes that cover the entire image to visualize the accuracy enhancement in Fig. 7.

We have also captured an uneven gray wall as shown in Fig. 8(a). For qualitative comparison, the wall is reconstructed using a structured-light method [39], which is known to be highly accurate. The reconstructed mesh

TABLE 7: Calibration result of projection error (in pixel) and depth error (in millimeter)

	Color	ToF	Depth (All)	Depth (Corners)
Schiller <i>et al.</i> [24] (Automatic feature detection)	0.313	2.330	7.45	16.6
Schiller <i>et al.</i> [24] (Some manual feature selection)	0.230	0.237	7.51	16.7
Our method	0.379	0.207	7.18	13.7

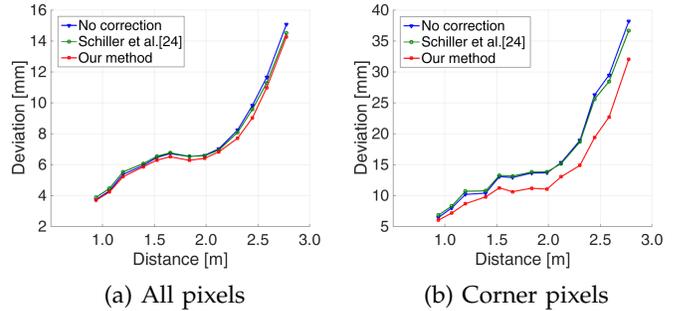


Fig. 9: Average depth error along the calibrated distance using different depth correction methods

using manufacturer provided transformation, ray correction, and ray and bias correction are compared with the reconstructed point cloud using structured light method by iterative closest point (ICP) [40]. Green regions indicate the distance between the point cloud and the mesh is less than 10mm. Yellow and blue regions indicate the absolute distance is between 10mm and 20mm.

5.2 Performance comparison

We compared the performance of our calibration method with Schiller *et al.* [24] using the public software [38]. Since Schiller *et al.* [24] require checkerboard images, we captured 41 images of the checkerboard described in Section 2.2 within the distance range from 0.9m to 2.9m. When corner detection was performed automatically, there were several incorrect localization of the corners in ToF amplitude images because the corners were too blurry as the cameras got further away from the board. As a result, the average projection error on the ToF amplitude images became very large, as shown in Table 7. With some user assistance on feature selection, we achieved the best result.

For comparison, we used the proposed 2.5D pattern board and depth images for correspondence acquisition for the ToF camera, as described in Section 2. 27 images captured in the same distance range were used. The projection error in ToF depth images using the proposed method is shown to be smaller.

Note that our method is effective in reducing the depth error of the corner pixels. The performance of the depth correction along the calibrated distance is compared in Fig. 9.

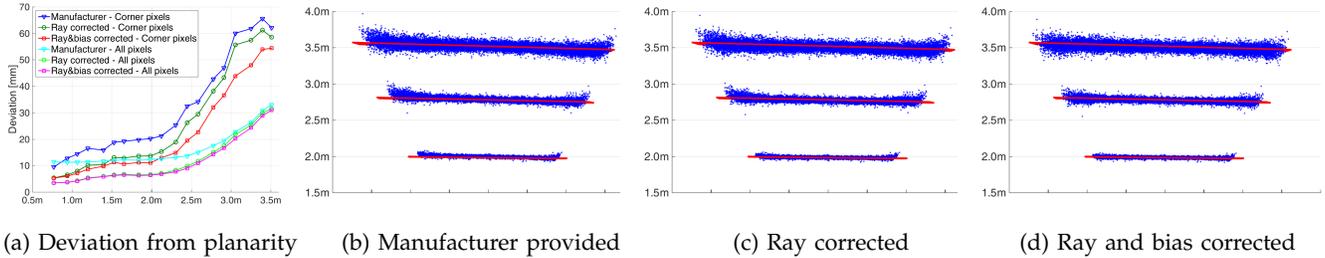


Fig. 7: Flat wall evaluation. The point clouds are the reconstructed walls captured from 3.5m, 2.8m, and 2.0m. The deviation from planarity is reduced using ray and bias correction. The curvature on the boundaries are flattened.

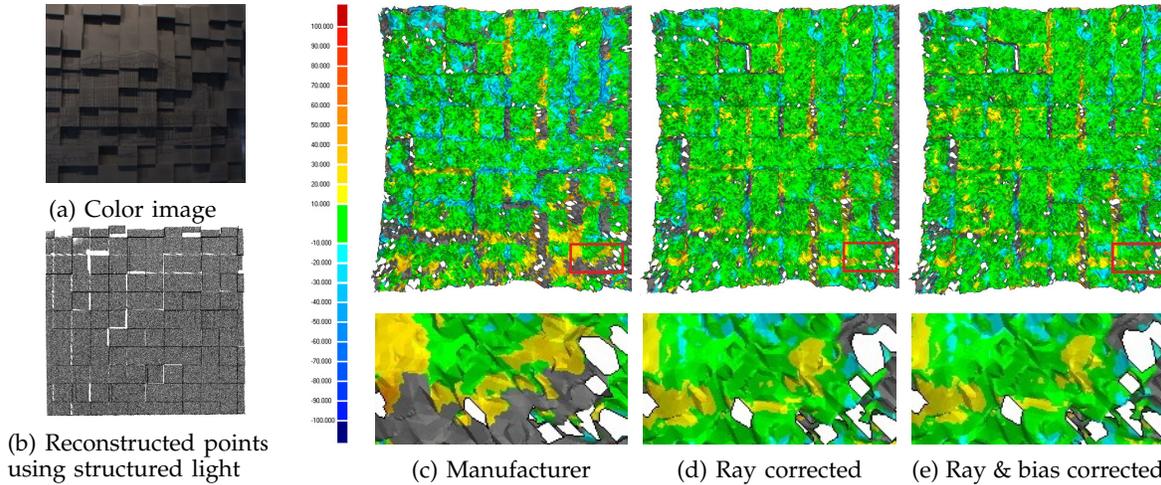


Fig. 8: Comparison with (b) structured light reconstruction method [39]. The reconstructed meshes using (c) manufacturer provided transformation, (d) ray correction, and (e) ray and bias correction are aligned using iterative closest point (ICP) [40]. Green regions indicate the distance between the point cloud and the mesh is less than 10mm.

5.3 Rendering results

We have captured various real indoor scenes to show that our method calibrates the color and depth cameras accurately to align the depth measurements with the high resolution color image.

Fig. 10 shows the 3D rendering results of different scenes. The top two rows are the front and side views of the same rendering of the 2.5D pattern plane using a color-depth data pair in the validation group. It is shown that ray correction improves the alignment of the color pixel with the transferred 3D measurements compared with the manufacturer provided transformation. Note that in the second row of Fig. 10, the rendered pattern plane is slightly curved in the result of ray correction, whereas it is rather straight in the result of ray and range bias correction. The third row shows the rendering results of a scene with large depth of field. The scanned lines at the bottom show the distortion of the depth measurement of the wall on the corners and their correction.

More complex indoor scene renderings are presented in Fig. 11 using different calibration methods. The magnified views show that our method greatly reduces the misalignment of the color and depth data. Note that our result wears the correct texture on the thin structure of the armrest and that the black antislip tape on the stairs is correctly located on the edge of the step.

6 DISCUSSION

6.1 Applications

The sensor calibration of a color and depth camera pair is a vital prerequisite for any applications using RGB-D input because color and depth information generally carry parallax from each other. Fig. 12 shows an example of shape refinement by combining noisy depth data from a ToF camera with a high resolution color information [8]. Since Han *et al.* [8] enhance the depth quality using the local shading information from the color image, the consistency between color and depth information is critical. The shading information generally changes drastically around the depth boundaries, which makes a small calibration error cause an apparent quality degradation as in Fig. 12 (c) and (d). It is shown in Fig. 12 (e) that the exact color and depth alignment by the proposed calibration method with depth correction improves the quality of the result significantly.

In addition, since the depth data provided by a ToF camera is generally noisy, the error distribution of a ToF sensor can also be used for further depth processing applications. From the remaining noise variance information with respect to the pixel location and the distance measurement after sensor calibration, we can build a depth noise variance map. If we assume a zero-mean

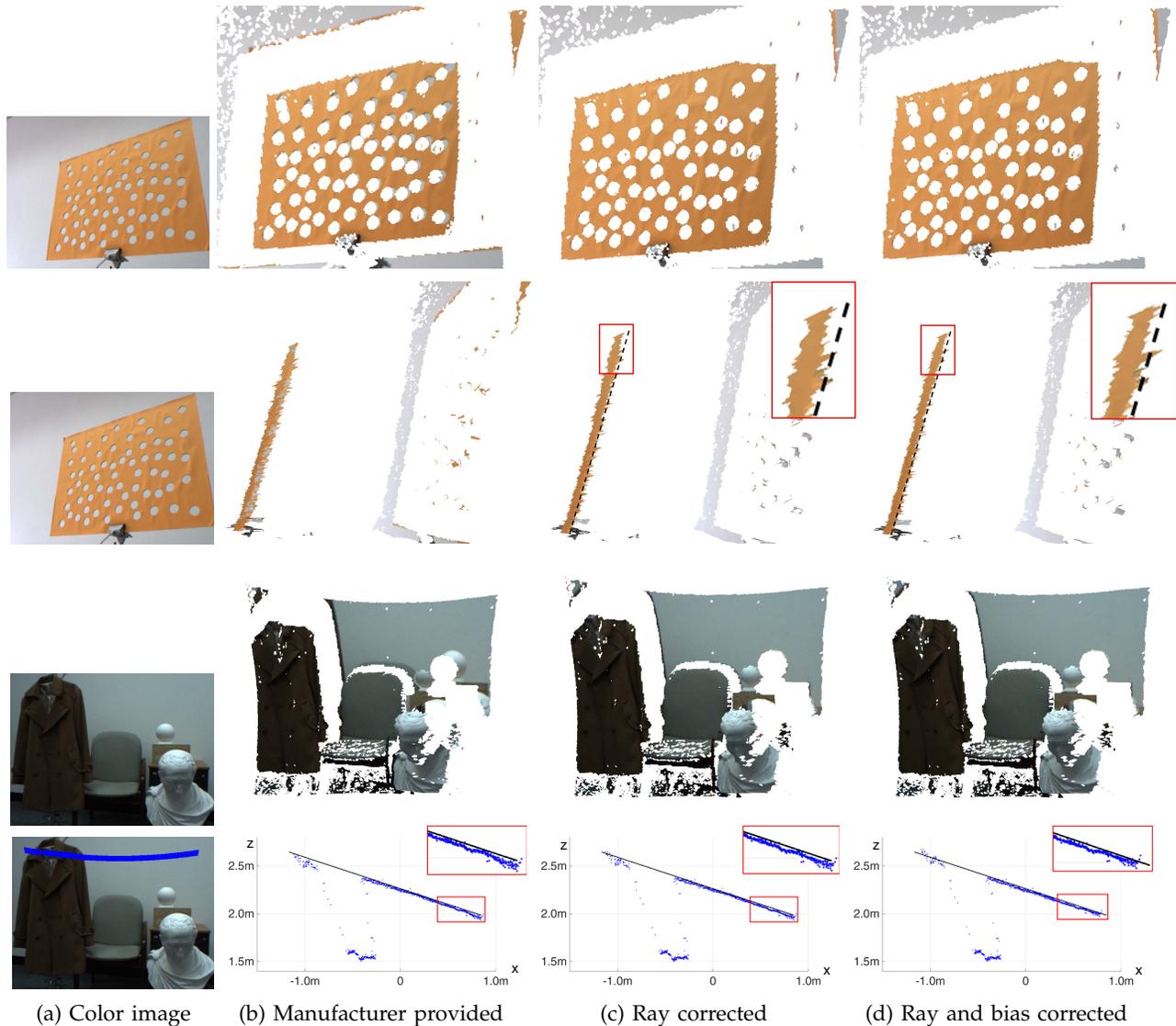


Fig. 10: 3D rendering results of different scenes. (b) Manufacturer provided 3D measurements, (c) ray corrected 3D measurements, and (d) ray and bias corrected 3D measurements are aligned with the corresponding color values using the optimized pose of the ToF camera with respect to the color camera. The magnified views in the second row show that the range bias correction straightens the corner of the rendered pattern plane. The scanned lines in the bottom row show the distortion of the depth measurement of the wall on the corners and their correction.

Gaussian noise, an adaptive depth denoising algorithm which uses the noise variance information such as [41] may be a strong application.

6.2 Limitations

Like other calibration methods using checkerboards, the proposed 2.5D pattern board cannot be captured from far away where the holes are too small to be detected in the ToF depth images. The pattern presented in this paper covers the calibrated range of 3 meters. To cover the further distance range, the pattern should be modified to have a proper scale, and the same calibration method can be applied.

Our method does not handle other sources of errors of a ToF camera such as reflectivity due to surface texture and surface direction, sensor temperature, and lighting

environment. We have focused on effective modeling of the systematic distance error of the sensor and have shown outperforming results compared to the previous method. The modeling of other sources of errors which are well studied in metrology community remains as our further work to enhance the performance.

6.3 Conclusion

We presented a calibration method for a color and depth camera pair. To acquire an accurate set of correspondences between the world and the camera coordinates of a time-of-flight camera as well as a color camera, we have designed a 2.5D pattern board with holes for near-IR rays of the ToF sensor to pass through. For depth correction of the ToF sensor, the range-to-3D transformation is corrected and the range error profile for each

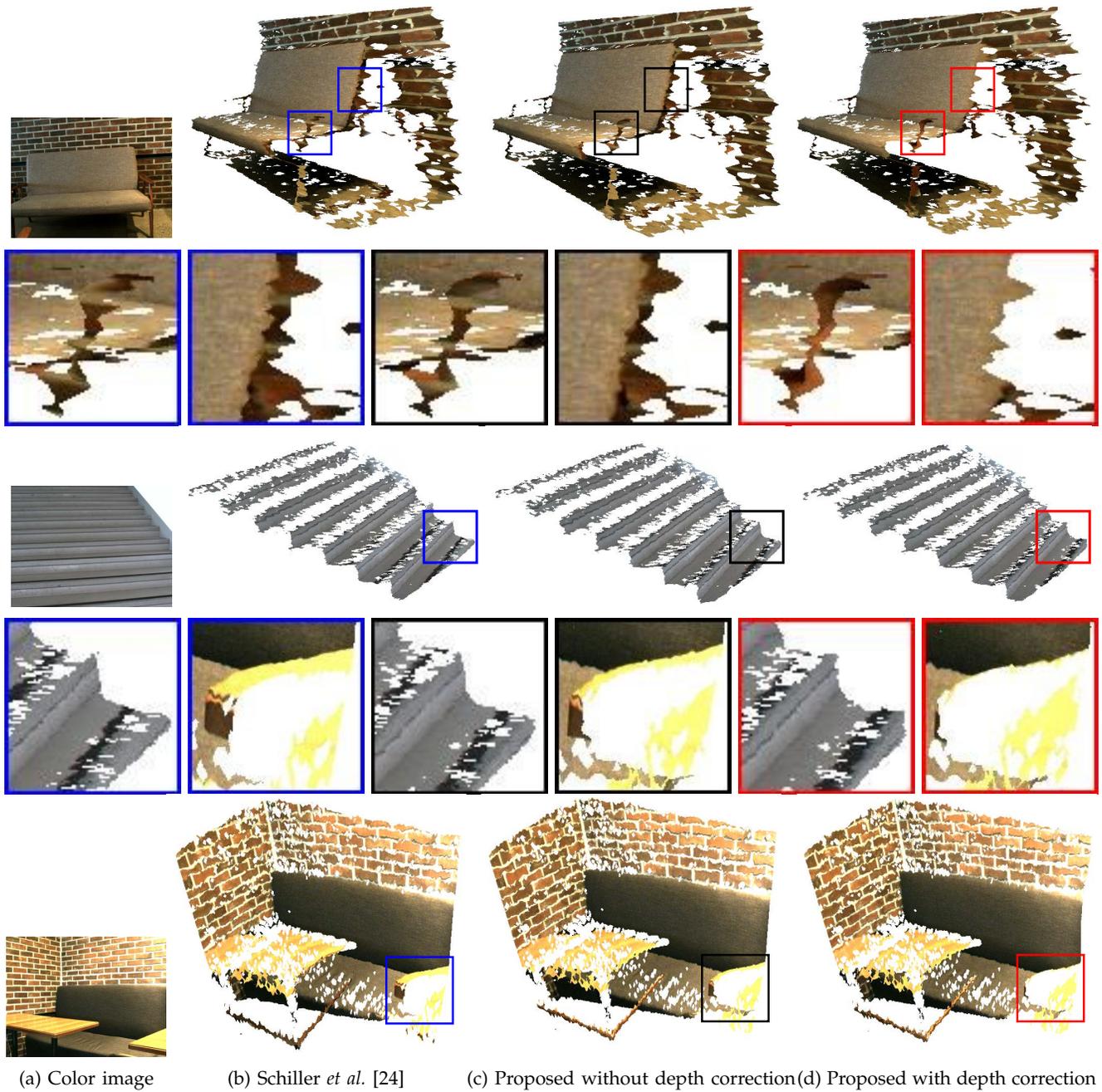


Fig. 11: 3D renderings of the various indoor scenes using different calibration methods. The magnified views in blue squares are the results of Schiller *et al.* [24], those in red and black squares are the results of the proposed method with and without depth correction respectively. The magnified views show that our method greatly reduces the misalignment of the color and depth data. Note that our result wears the correct texture on the thin structure of the armrest instead of the back wall, and that the black antislip tape on the stairs is correctly located on the edge of the step.

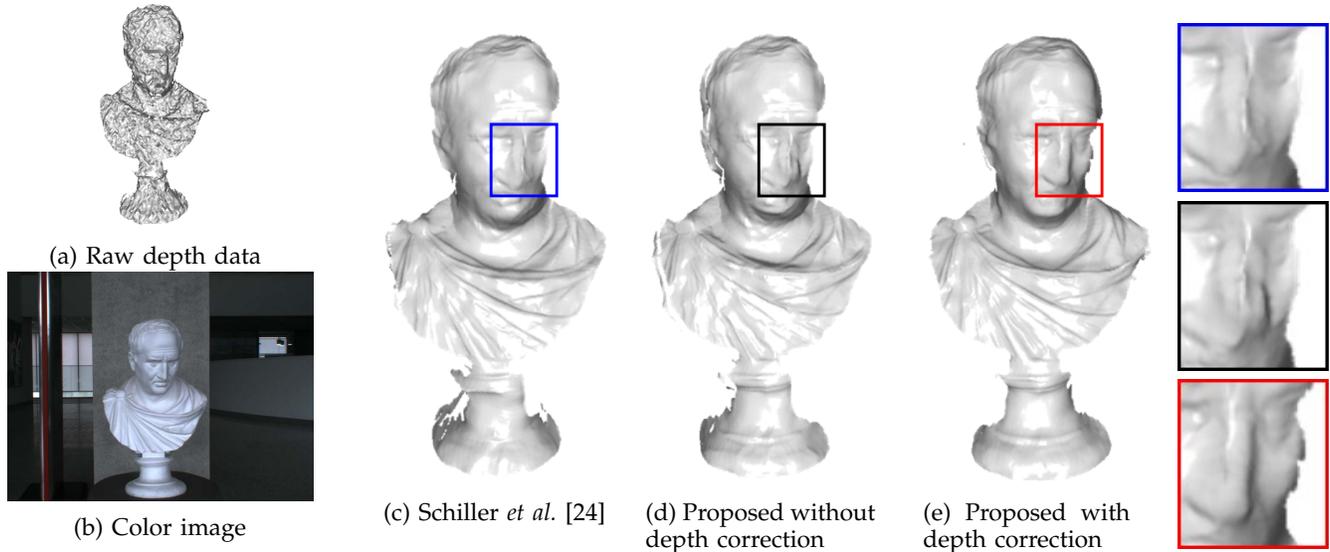


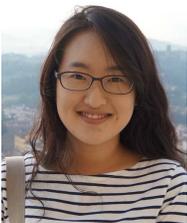
Fig. 12: An example of RGB-D shape refinement application using different calibration methods.

pixel is analyzed. We have presented that the amount of error depends on the pixel location as well as the distance measurement, which makes corner pixels with larger depth errors require more powerful error correction than center pixels. This improvement is important for ToF cameras since they usually suffer from very low resolution.

We closely evaluated the performance of each stage of the proposed framework and compared the performance of our method with the previous calibration method through various qualitative and quantitative experiments. We have shown that our improvement in sensor calibration plays an important role to enhance the quality of the related application. Moreover, we have released our code along with the calibration data for the research community. As a future work, we plan to utilize the error analysis of this sensor calibration for further enhancement of the depth information of the scene.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680). The authors thank Hyowon Ha for his help with the comparison evaluation using a structured light method.



Jiyoung Jung received the B.S and the M.S degrees in Electrical Engineering from KAIST, Korea in 2008 and 2010 respectively. She is currently a Ph.D. candidate in Robotics and Computer Vision Lab in the Department of Electrical Engineering in KAIST. Her research interests include photometric stereo, multiple sensor fusion, and 3D modeling of indoor/outdoor scenes. She is a recipient of Presidential Science Scholarship by Korea Student Aid Foundation. She is a student member of the IEEE.



member of the IEEE.

Joon-Young Lee received the B.S degree in Electrical and Electronic Engineering from Yonsei University, Korea in 2008, and the M.S degree in Electrical Engineering from KAIST, Korea in 2009. He is currently working toward the Ph.D. degree in Electrical Engineering at KAIST. His research interests include photometric methods in computer vision, image enhancement, and computational photography. He is a recipient of the Samsung HumanTech Paper Award and the Qualcomm Innovation Award. He is a student



Yekeun Jeong received the B.S. and Ph.D. degrees in Electrical Engineering from KAIST, Korea in 2006 and 2012 respectively. He is currently working in Augmented Reality group at Microsoft. His research interests include structure from motion, bundle adjustment, and their applications to large-scale 3D modeling with multiple sensors and mobile robots. He is a recipient of the Samsung Human-tech Thesis Award in 2010 and 2012 and the National Research Foundation of Korea scholarship.



In So Kweon received the B.S. and M.S. degrees in mechanical design and production engineering from Seoul National University, Korea, in 1981 and 1983, respectively, and the PhD degree in robotics from the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1990. He worked for Toshiba R&D Center, Japan, and joined the Department of Automation and Design Engineering, KAIST, Korea in 1992, where he is now a Professor with the Department of Electrical Engineering. He is a recipient

of the best student paper runner-up award at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 09). His research interests are in camera and 3-D sensor fusion, color modeling and analysis, visual tracking, and visual SLAM. He was the program co-chair for the Asian Conference on Computer Vision (ACCV 07) and was the general chair for the ACCV 12. He is also on the editorial board of the International Journal of Computer Vision. He is a member of the IEEE and the KROS.

REFERENCES

- [1] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk, "The digital Michelangelo project: 3D scanning of large statues," in *Proceedings of ACM SIGGRAPH*, 2000.
- [2] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from a single depth image," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [3] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera," in *ACM Symposium on User Interface Software and Technology*, 2011.
- [4] T. Simonite, "Depth-sensing cameras head to mobile devices," *MIT Technology Review*, <http://www.technologyreview.com/news/>, 2013.
- [5] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.
- [6] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. S. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011.
- [7] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, and B. Wilburn, "Edge-preserving photometric stereo via depth fusion," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [8] Y. Han, J.-Y. Lee, and I. S. Kweon, "High quality shape from a single RGB-D image under uncalibrated natural illumination," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013.
- [9] R. Knies, "Collaboration, expertise produce enhanced sensing in Xbox One," *The Official Microsoft Blog* http://blogs.technet.com/b/microsoft_blog/, 2013.
- [10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [11] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.
- [12] Y. Bok, Y. Jeong, D.-G. Choi, and I. S. Kweon, "Capturing village-level heritages with a hand-held camera-laser fusion sensor," *International Journal on Computer Vision (IJCV)*, vol. 94, no. 1, pp. 36–53, 2011.
- [13] J. Jung, Y. Jeong, J. Park, H. Ha, J. D. Kim, and I. S. Kweon, "A novel 2.5D pattern for extrinsic calibration of ToF and camera fusion system," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011.
- [14] J. Jung, "Website for public source code," <https://sites.google.com/site/jyjungcv/>.
- [15] S. Fuchs and G. Hirzinger, "Extrinsic and depth calibration of ToF cameras," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [16] Y. S. Kim, B. Kang, H. Lim, O. Choi, K. Lee, J. D. Kim, and C. Kim, "Parametric model-based noise reduction for ToF depth sensors," in *Three-Dimensional Image Processing (3DIP) and Applications II*, 2012.
- [17] Y. M. Kim, D. Chan, C. Theobalt, and S. Thrun, "Design and calibration of a multi-view TOF sensor fusion system," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008.
- [18] J.-Y. Bouguet, "Camera calibration toolbox for matlab," http://www.vision.caltech.edu/bouguetj/calib_doc, Last updated 2013.
- [19] M. Lindner and A. Kolb, "Lateral and depth calibration of PMD-distance sensors," in *Advances in Visual Computing*, vol. 4292. Springer, 2006, pp. 524–533.
- [20] T. Kahlmann, F. Remondino, and H. Ingensand, "Calibration for increased accuracy of the range imaging camera SwissrangerTM," in *Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*, 2006.
- [21] F. Kern, "Supplementing laser scanner geometric data with photogrammetric images for modeling," in *XVIII. International CIPA Symposium. Surveying and Documentation of Historic Buildings - Monuments - Sites Traditional and Modern Methods*, 2001, pp. 454–461.
- [22] C. Beder and R. Koch, "Calibration of focal length and 3D pose based on the reflectance and depth image of a planar object," *International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3, pp. 285–294, 2008.
- [23] D. Lefloch, R. Nair, F. Lenzen, H. Schäfer, L. Streeter, M. J. Cree, R. Koch, and A. Kolb, "Technical foundation and calibration methods for time-of-flight cameras," in *Time-of-Flight and Depth Imaging. LNCS*, M. Grzegorzec, C. Theobalt, R. Koch, and A. Kolb, Eds. Springer, 2013, vol. 8200, pp. 3–24.
- [24] I. Schiller, C. Beder, and R. Koch, "Calibration of a PMD-camera using a planar calibration pattern together with a multi-camera setup," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Part B3a, Beijing, China*, vol. XXXVII. ISPRS Congress, 2008, pp. 297–302.
- [25] M. Schmidt, "Analysis, modeling and dynamic optimization of 3D time-of-flight imaging systems," *Ph.D. thesis, University of Heidelberg, Germany*, 2011.
- [26] M. Lindner and A. Kolb, "Calibration of the intensity-related distance error of the PMD TOF-camera," in *SPIE: Intelligent Robots and Computer Vision XXV*, vol. 6764, 2007.
- [27] O. Steiger, J. Felder, and S. Weiss, "Calibration of time-of-flight range imaging cameras," in *Proceedings of International Conference on Image Processing (ICIP)*, 2008, pp. 1968–1971.
- [28] A. Swadzba, N. Beuter, J. Schmidt, and G. Sagerer, "Tracking objects in 6D for reconstructing static scenes," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–7.
- [29] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G. Brostow, "Capturing time-of-flight data with confidence," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 945–952.
- [30] K. Pathak, A. Birk, and J. Poppinga, "Sub-pixel depth accuracy with a time of flight sensor using multimodel gaussian analysis," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2008, pp. 3519–3524.
- [31] M. Lindner, M. Lambers, and A. Kolb, "Sub-pixel data fusion and edge-enhanced distance refinement for 2D/3D images," *International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3, pp. 344–354, 2008.
- [32] M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-Flight sensor calibration for accurate range sensing," *Computer Vision and Image Understanding (CVIU)*, vol. 114, no. 12, pp. 1318–1328, 2010.
- [33] OpenCV, <http://opencv.org>, Last updated 2014.
- [34] H. Rapp, "Experimental and theoretical investigation of correlating TOF-camera systems," *Master's thesis, University of Heidelberg, Germany*, 2007.
- [35] D. Herrera C., J. Kannala, and J. Heikkilä, "Joint depth and color camera calibration with distortion correction," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 34, no. 10, pp. 2058–2064, 2012.
- [36] J. Smisek, M. Jancosek, and T. Pajdla, "3D with kinect," in *Proceedings of IEEE International Conference on Computer Vision Workshops*, 2011.
- [37] G. Vogiatzis and C. Hernández, "Automatic camera pose estimation from dot pattern," <http://george-vogiatzis.org/calib/>, 2010.
- [38] I. Schiller, "MIP - MultiCameraCalibration," <http://www.mip.informatik.uni-kiel.de/tiki-index.php?page=Calibration>, Last updated 2011.
- [39] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.
- [40] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 14, no. 2, pp. 239–256, 1992.
- [41] D. Yeo, E. U. Haq, J. Kim, M. W. Baig, and H. Shin, "Adaptive bilateral filtering for noise removal in depth upsampling," in *International SoC Design Conference*, 2010.