Vis Comput (2012) 28:1139–1151
DOI 10.1007/s00371-011-0664-x

ORIGINAL ARTICLE

Time-of-flight sensor and color camera calibration for multi-view acquisition

Hyunjung Shim · Rolf Adelsberger · James Dokyoon Kim · Seon-Min Rhee · Taehyun Rhee · Jae-Young Sim · Markus Gross · Changyeong Kim

Published online: 28 December 2011

Abstract This paper presents a multi-view acquisition system using multi-modal sensors, composed of time-of-flight (ToF) range sensors and color cameras. Our system captures the multiple pairs of color images and depth maps at multiple viewing directions. In order to ensure the acceptable accuracy of measurements, we compensate errors in sensor measurement and calibrate multi-modal devices. Upon manifold experiments and extensive analysis, we identify the major sources of systematic error in sensor measurement and construct an error model for compensation. As a result, we provide a practical solution for the real-time error compensation of depth measurement. Moreover, we implement the calibration scheme for multi-modal devices, unifying the spatial coordinate for multi-modal sensors.

The main contribution of this work is to present the thorough analysis of systematic error in sensor measurement and therefore provide a reliable methodology for robust error compensation. The proposed system offers a real-time multi-modal sensor calibration method and thereby is applicable for the 3D reconstruction of dynamic scenes.

Keywords Depth sensing · Multi-modal sensor fusion · Multi-view acquisition · 3D video processing

H. Shim (✉) · J.D. Kim · S.-M. Rhee · T. Rhee · C. Kim
Samsung Advanced Institute of Technology, Giheung, South Korea
e-mail: kateshim@gmail.com

R. Adelsberger · M. Gross
ETH, Zurich, Swiss

J.-Y. Sim
Ulsan National Institute of Science and Technology, Ulsan, South Korea

1 Introduction

Upon the advancement in 3D display technologies, we envision that the next move in the display industries gears toward autostereoscopy for multiple users. Such a new shift demands the framework for modeling and visualizing a real 3D scene at arbitrary viewing directions. As one of candidates for future 3D sensing technology, time-of-flight (ToF) depth camera receives a great attention from various researchers and has been adopted in several topics such as novel view synthesis, 3D scene modeling, gesture recognition, human-computer interaction, etc. In this paper, we present a multi-view acquisition system for the 3D visualization of a dynamic scene, using multiple depth and color cameras.

A variety of multi-view acquisition and processing techniques has been proposed in the field of computer vision and computer graphics. In general, existing techniques can be classified into either a single-modal sensor-based approach (using color cameras only) or multi-modal sensor-based approach. Single-modal sensor-based approaches have adopted for light field imaging [1], stereo vision [2], photometric stereo [3], shape from X [4] and many others. Possible by the rapid progress in camera manufacturing technology, color cameras are suitable for capturing and processing a dynamic scene. However, they either require an excessive number of cameras (from tens to hundreds), hardly reproducible in practice [1], or may suffer from ambiguity in shape reconstruction [2–4]. As alternatives, multi-modal sensor-based approaches shall serve a compact system configuration, requiring much less devices than single-modal sensor-based approaches, and free from ambiguity in 3D reconstruction. Multi-modal sensors are typically composed of color cameras to capture the texture of scene and additional depth sensors; range scanner [5], structured light [6], time-of-flight (ToF) [8], etc.; to acquire the geometry of

scene. Yet, many of multi-modal sensor-based approaches are not suitable for handling the dynamic scene [5, 6]. In fact, the applicability for capturing the dynamic scene is determined by the principle of depth sensor.

Among existing depth sensors, ToF depth sensor is suitable for recording the dynamic 3D scene. Thus, we employ three pairs of ToF sensors and color cameras for composing our multi-view acquisition system. Lately, several techniques have been proposed to integrate ToF sensors into the multi-view acquisition system. However, their results presented the shape misalignment inherent by the error in sensor measurement. Although the ToF depth camera achieved the quality improvement in depth measurement during past years, it still produces the substantial errors, yielding lack in precision. Addressed by the previous work, the accuracy of sensors exhibits a significant dependency on environmental conditions as well as internal properties. These sources of error ask a complex non-linear compensation so to ensure the acceptable accuracy in 3D reconstruction. Although issues in measurement error are reported in previous work, the thorough analysis and study have not been conducted in sufficient depth so to be applicable for a practical system.

In this paper, we investigate and develop the practical pipeline for compensating errors in sensor measurement and calibrating multi-modal devices. For the error compensation, we first identify the sources of error. Upon the experimental validation, we conclude that the scene depth and the intensity of reflected infrared image (referred to as an amplitude image) are the major sources of error. Then, we construct the error model parameterized by two sources of error and use it to correct the depth measurement. More specifically, we generate a look-up-table (LUT) to perform the real-time error compensation for sensor measurement, which is essential for processing video inputs.

Because the ToF sensor only outputs the geometric information of scene, we use additional color cameras to capture the radiometric information of scene. We implement a new calibration scheme between ToF sensors and color cameras for better fitting onto the proposed system. As a result, we present a real-time 3D scene acquisition system, applicable for the 3D reconstruction of dynamic scene.

2 Related work

A number of techniques for ToF sensor calibration has been discussed in recent years. Previous work considered a pinhole imaging model for the ToF sensor and solved for intrinsic and extrinsic parameters for calibration. Zhang [7] has suggested the intrinsic calibration method for the ToF sensor. Given by the measurement points and their ground truth positions, the author computed intrinsic parameters by applying conventional algorithms such as Calde/Callab, Matlab-Toolbox, or OpenCV. To enhance the accuracy in estimation, Kahlmann et al. [8, 9] proposed to use a calibration pattern consisting of filled white circles on a black background. For error compensation in sensor measurement, they introduced a look-up-table (LUT) to correct the depth measurements. Their LUT considers the depth dependency and the exposure time for model construction. Recently, Fuchs et al. in [10, 11] computed intrinsic parameters using a conventional checkerboard pattern and reported the reasonable accuracy for estimating intrinsic parameters. Their extrinsic calibration of range sensors used a specialized hardware, an industrial robot. Such a robot-based system shall not be suitable for most practical applications.

A general system of calibrating multi-view ToF sensors has been presented by Kim et al. [12]. The authors applied a three-step calibration algorithm in order to remove the systematic bias in depth measurement: a rigid transformation, polar angular correction and a constant depth bias. Our calibration method is different from them in that we exhaustively analyze an error distribution and derive an empirical model for compensating the systematic error. Guan et al. [13] have presented a method to calibrate a network of camcorders and ToF cameras using a spherical calibration object. They assume predetermined intrinsic parameters, which require additional calibration objects. Moreover, they do not consider the error compensation, yielding a lack in precision. In [14], same authors combined multi-modal data, color images and depth maps. They used standard calibration techniques [15] based on a checkerboard pattern. Also, they compared the estimated position of plane with depth measurements to derive the error compensation model. As a result, their model accounted a constant bias, ray discrepancy and depth dependent bias into their compensation model.

This paper presents an effective calibration technique for the error compensation of sensor measurement as well as the extrinsic calibration of multi-modal sensors. Upon the extensive experiments and analysis, we derive the practical solution for error compensation, the LUT-based real-time error compensation. We identify the measured depth and the amplitude image as the major sources of error by empirical analysis. In addition, we suggest a modified technique for the extrinsic calibration method for ToF sensors. Instead of computing the intrinsic parameters of ToF sensor, we directly use depth measurements for calculating extrinsic parameters. In this way, we avoid the lack in precision for intrinsic parameter estimation.

3 System specification

As shown in Fig. 1, the proposed system consists of three device pairs which are rigidly mounted onto a rig. Each

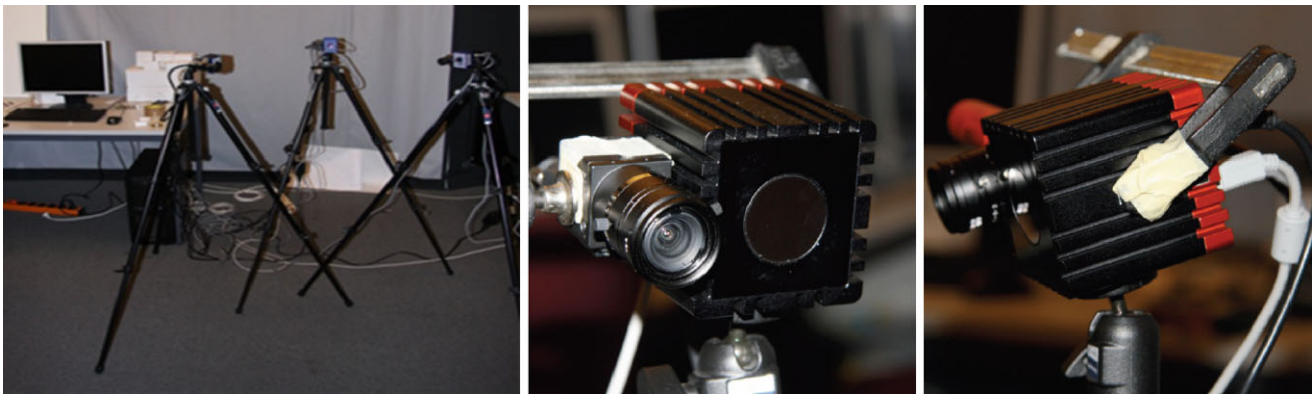


Fig. 1 Experimental setup. *Left*: setup with three camera pairs, *middle*: frontal view of color camera attached to the ToF sensor, *right*: side view of the pair

pair combines a time-of-flight (ToF) depth sensor from the *MESA*^{TM} *Swissranger 4000* (SR4000) with a color camera from the *PointGrey Research Flea2*. These pairs of sensors stay 1 m (or more) apart and have an overlapping field of view. We merge the combined depth and color data from all sensors into a single, consistent 3D scene for free viewpoint rendering.

The ToF sensor produces a depth map and an amplitude image at every frame and their resolutions are 176×144. The depth map provides the 3D position of scene points with a floating-point precision and the amplitude image contains the intensity values of the reflected infrared light corresponding to the depth map. SR4000 offers three modulation frequencies for infrared light; 29 MHz with a range of 5.17 m, 30 MHz with a range of 5.0 m and 31 MHz with a range of 4.84 m. Notice that higher modulation frequencies are of high interest for practitioners as it provides a dense depth resolution. Besides, the color camera runs 30 fps at 1032×776 and its field of view is around $43.6° \times 34.6°$.

4 Time-of-flight sensor calibration

As noted in the literature, the ToF sensor has the systematic bias, yielding distorted depth maps. In the following sections, we first identify the sources of error (Sect. 4.1) and then carry out the empirical analysis to derive an error model that accounts for all sources of error simultaneously (Sect. 4.2).

Based on the information from the manufacturer, the previous work and our own experiences, we have selected the potential sources of error for the SR4000 ToF sensor. They are; scene depth, amplitude image, temperature, lens distortion and spatial interference. For each component, we analyze its characteristics and evaluate whether the error is present, relevant, and possible to be integrated into our model for error compensation.

It is important to note that, ideally, we should analyze each source of error from depth measurements only affected by the corresponding source of error. However, actual measurements are the result of multiple error sources and it is difficult to separate each source of error in practice. Hence, it is impossible to measure the error corresponding to a single source alone.

Therefore, instead of modeling each source of error, we first identify the major sources of error by proving or disproving a hypothesis of each source being the significant source of error. Then, we build the error compensation model that accounts for all the major sources of error simultaneously.

4.1 Sources of errors

In this section, we will list the potential sources of error and eliminate the term if it does not possess the systematic nature. In this way, we will arrive at two major sources of error; the scene depth and the amplitude image.

4.1.1 Scene depth

Upon our experiences, we found that the error in depth measurements depends on the actual depth of scene. Our hypothesis is that the error can be modeled by a non-linear warp function, which is parameterized by the measured depth. To prove or disprove the presence of depth warp, we carry out the following experiment.

As depicted in the left of Fig. 2, we design a calibration object using two planar bodies attached orthogonally to a connection. Then, we capture amplitude images and depth maps of this object at multiple scene depths, covering the entire operating range of ToF sensor. From each captured amplitude image, we manually select the region at each plane (blue and red rectangle in Fig. 2). Then, we apply RANSAC [17] for fitting the corresponding 3D points of the frontal

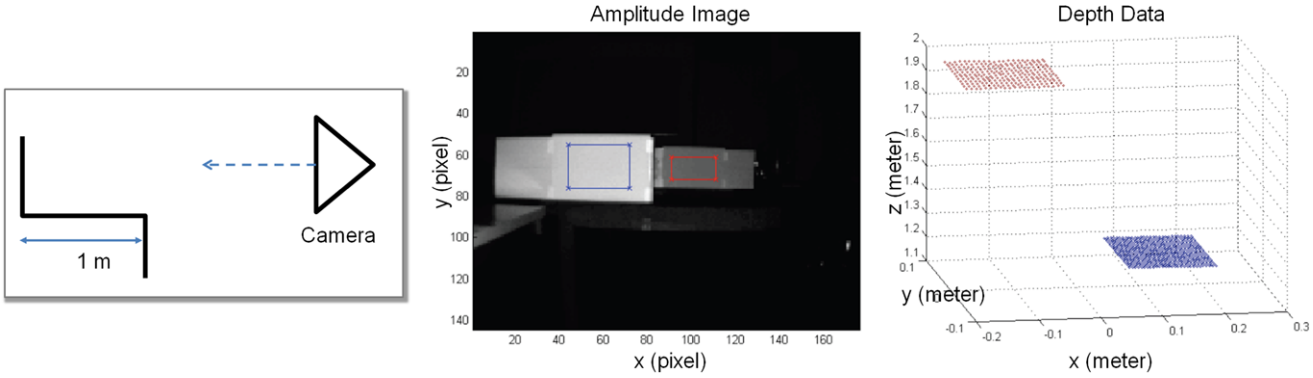


Fig. 2 *Left*: illustration of our calibration object (two planar objects are attached orthogonally to a connection and have a 1 m distance to each other), *middle*: *blue rectangle* includes the foreground part in the calibration object and *red rectangle* includes a region in the farther away plane of the calibration object, *right*: 3D points for the selected regions in the *middle image*

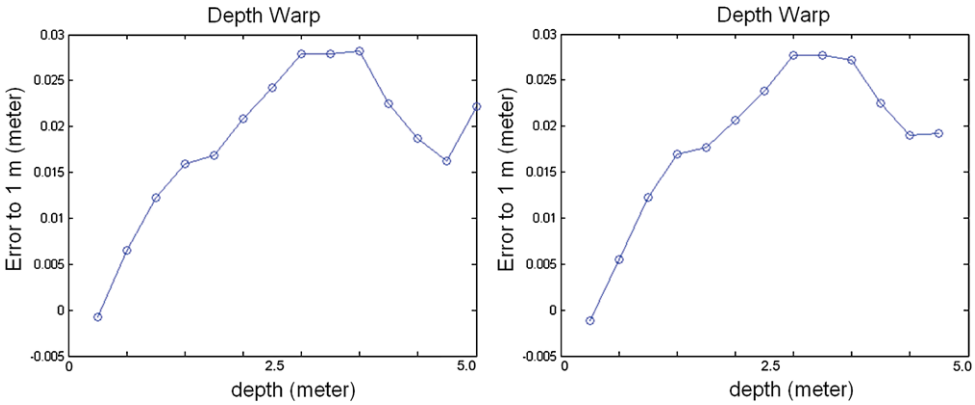


Fig. 3 *Left*: deviation from ground truth (1 m) over multiple shots using the calibration object in Fig. 2, *right*: results from another camera

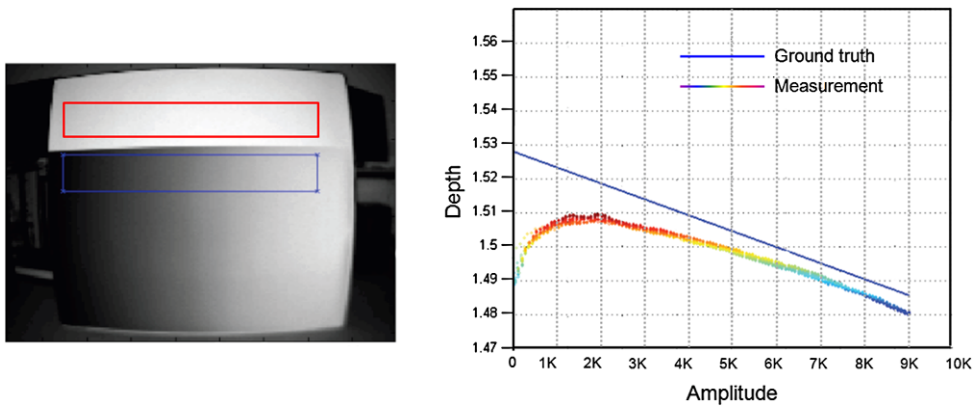


Fig. 4 *Left*: amplitude image of a calibration object for amplitude variation. (*Red rectangle*: regions for acquiring the ground truth, *blue rectangle*: regions for the measurement) *right*: depth deviation within the selected gradient pattern

plane into the plane equation. We perform the same operation on the back plane and then calculate the orthogonal distance between two planes. Figure 2 illustrates these steps.

The depth warp does not exist if both conditions satisfy: (1) the measured distance between two fitted planes is very close to the ground truth, (2) this value stays consistent at the different scene depth. Based on our experiments, neither of them were true and therefore we conclude that the depth warp is present (see Fig. 3). We observe the maximum amount of error up to 3 cm within the operating range.

4.1.2 Amplitude image

The amplitude image is the amount of infrared light that has been reflected back from the scene. We observe that there is also a depth error introduced by varying amplitude values. The amount of reflected light is influenced by various factors in the scene, such as material properties and the surface orientation. Also, the attenuation factor of light would be inversely proportional to the square of the traveling distance of light.

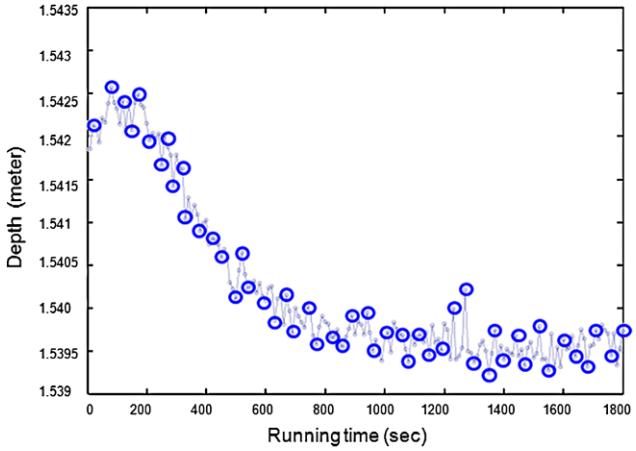


Fig. 5 Depth measurement over the running time (second). Ground truth depth is located at 1.5 m

To see if the error in depth measurements is sufficiently correlated with the amplitude image, we observe the depth values at varying amplitude values. To simulate the variation in amplitude image, we use a gray scale gradient pattern attached on top of the planar board as a calibration object (see the left of Fig. 4). Then, we record the amplitude image and the depth map of this calibration object. From the amplitude image, we select the plain-colored region (red rectangle) of the object and compute the plane position from the selected points. This serves the ground truth for the plane position. After that, we measure the depth over the gradient area (blue rectangle) and compare the measured depth with the ground truth in the right side of Fig. 4.

As shown in Fig. 4, we observe that the depth measurements deviate from the real depth values. Upon the experiment, the error introduced by the amplitude reaches up to 4 cm.

4.1.3 Temperature

The temperature of sensor is another source of error in the depth measurements. [16] and [8] explain the underlying physical principle of this phenomenon in detail.

To simulate the changes in temperature, we power up the device and measure the depth within the first 30 minutes. By turning on the power, we expect the temperature of device is increased gradually over the time. By fixing all other conditions, we record the depth value of the same object, a white plane, over the time and plot them accordingly (see Fig. 5).

As a result, the maximum amount of error by the temperature changes is roughly 0.3 cm within 20 minutes, relatively small compared to other sources, and its distribution is stabilized after 10 minutes. Hence, we decide to ignore the temperature from the sources of error. Instead, we ensure, for every experiment, the device having been powered up for at least 20 minutes before the usage.

4.1.4 Lens distortion

Similar to the conventional color camera, the ToF sensors also present the optical distortion. The ToF sensor consists of a set of LEDs emitting the modulated infrared light, and the lens receiving the reflected infrared light from the scene. To cover the wide range of the scene, the lens of ToF sensor has a large field of view, approximately $43.6° \times 34.6°$. On the other hand, the lens diameter is small in order to keep the sensor size being reasonably small. This sensor configuration yields the severe lens distortion in the measurements. One common approach to model the lens distortion is suggested by [18], combining radial distortion and tangential distortion. That is,

$$\tilde{x} = x + x\left[k_1 r^2 + k_2 r^4\right] + \left[2p_1 xy + p_2\left(r^2 + 2x^2\right)\right],$$
$$\tilde{y} = y + y\left[k_1 r^2 + k_2 r^4\right] + \left[2p_2 xy + p_2\left(r^2 + 2y^2\right)\right].$$

$\{\tilde{x}, \tilde{y}\}$ are the distorted coordinates, $\{k_1, k_2\}$ represent the radial distortion parameters, $\{p_1, p_2\}$ stand for the tangential distortion parameters and r^2 equals to $x^2 + y^2$.

In fact, the manufacturer (SwissRanger) provides universal distortion parameters using the above distortion model, meaning the same parameters for every device. Hence, we attempt to derive optimal distortion parameters per device and see if the measurement accuracy is enhanced. Our estimates on distortion parameters and the universal distortion parameters by the manufacturer are listed in Table 1. From these experimental results, we find that the default parameters are the valid approximation to optimal parameters per device. That is, our estimates for each device are sufficiently close to manufacturer' parameters. Differences for $\{k_1, k_2\}$ are in the order of 10^{-11}.

Finally, we conclude that the radial distortion exists in the raw data (amplitude image and the depth map). However, the default parameters are sufficiently accurate and thereby the additional distortion model is not necessary. As a result, the additional factor for the lens distortion can be discarded from the sources of error.

4.1.5 Spatial interference

The spatial interference indicates the error influenced by the position (2D coordinates) of a pixel. It is correlated to the propagation delay when the CCD is read out [11, 19]. To validate the presence of error, we measure the depth of a planar object and compute a depth deviation from the ground truth.

As a result, the computed deviation is ignorable; the standard deviation is about 2 mm. We compute the statistics of the pixel dependent error and present it in Table 2. Since the overall deviation is sufficiently small, we decide to discard it from the final error model.

Table 1 Lens distortion parameters

Camera	Set	k_1	k_2	p_1	p_2	c_x	c_y
Camera 1	Manufacturer	$7.82 \times e^{-10}$	$2.01 \times e^{-9}$	0	0	87.5	71.5
	Our estimates	$7.82 \times e^{-10}$	$2.01 \times e^{-9}$	0	0	91.88	71.5
Camera 2	Manufacturer	$7.82 \times e^{-10}$	$2.01 \times e^{-9}$	0	0	87.5	71.5
	Our estimates	$7.89 \times e^{-10}$	$2.03 \times e^{-9}$	$9.64 \times e^{-5}$	$-6.77 \times e^{-5}$	88.27	72.13
Camera 3	Manufacturer	$7.82 \times e^{-10}$	$2.01 \times e^{-9}$	0	0	87.5	71.5
	Our estimates	$7.82 \times e^{-10}$	$2.01 \times e^{-9}$	$1.49 \times e^{-4}$	$-1.24 \times e^{-5}$	89.5	71.64

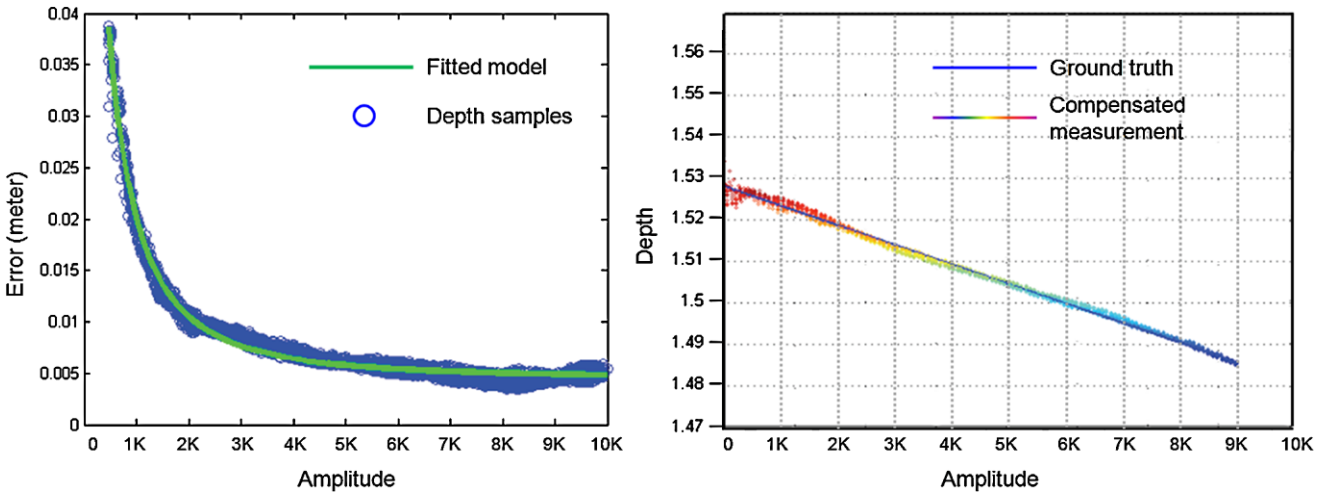


Fig. 6 Amplitude dependent error model. *Right*: plots for depth samples (*blue*) and the fitted model (*green*) of error, *left*: plots for amplitude corrected samples and the ground truth

Table 2 Statistics of pixel dependent error

Values in meter	Pixel dependent error statistics
Mean	2.62×10^{-19} (m)
STD	1.9×10^{-4} (m)
VAR	3.98×10^{-6} (m)

4.2 Error compensation model

Based on the faithful experiments and analysis in Sect. 4.1, we identify two major sources of error; the scene depth and the amplitude image. Knowing two major components, we construct a global error model for compensating the depth error dependent on amplitude image and depth measurement.

On the left side of Fig. 6, we visualize the amplitude dependent error distribution. By observing these measurements, we find that the measured data points fit well into an inverse quadratic function. Thus, we propose a parametric model to represent the error distribution upon the changes in amplitude. That is,

$$E(r) = \frac{\alpha}{(\frac{r}{\beta} + \gamma)^2} + \delta. \quad (1)$$

The input value r is the amplitude value, the parameter α is the scaling factor, β is constant and set to 16000, γ shifts the function along the abscissa and δ is the offset of the function. We have measured the error distribution upon the amplitude changes at several different distances. By observing multiple error distributions, we empirically find that the parametric model shown in (1) matches well with measurements.

This formula is capable of representing the characteristics of error distribution at a fixed depth value. Key to our model is that we constitute the global error model by expanding this amplitude dependent error model upon the changes in depth values.

At the fixed scene depth, we fit the model into the data by non-linear optimization and compute the model parameters $\{\alpha, \gamma, \delta\}$. Considering the depth dependency in error, we calculate the model parameters at every depth value within the operating depth range. In practice, it is impossible to calculate the model parameters for all possible depth values.

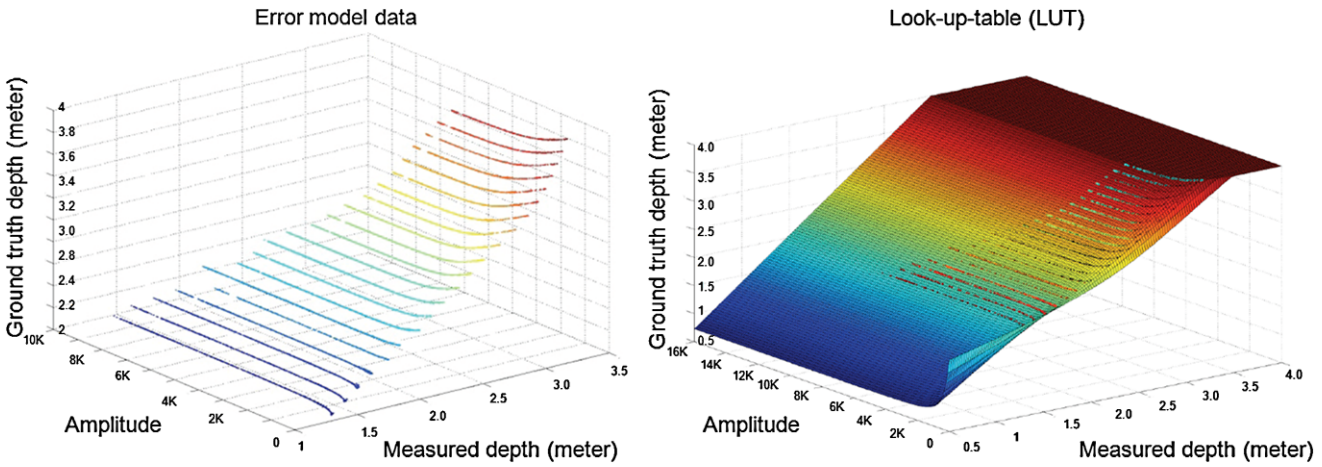


Fig. 7 *Left*: plots for the measurements, *Right*: LUT. Z-axis: ground truth depth, Y-axis: amplitude values at a measured depth, X-axis: measured depth

Hence, we compute the model parameters at discrete depth samples and use them from the adjacent depth to interpolate the model parameters for in between depth.

Precisely, we sample the depth by positioning the sensor at every 10 cm within the operating range. Then, at every depth sample, we record the gradient pattern attached to the white wall for acquiring the error distribution upon the amplitude changes. Finally, we compute the model parameters $\{\alpha, \gamma, \delta\}$ at every depth sample.

4.2.1 LUT calculation

Given the measurements (Fig. 7) and the parametric model (1), we have optimized the model parameters $\{\alpha, \gamma, \delta\}$ at the corresponding depth. When the measurements are not available, we interpolate $\{\alpha, \gamma, \delta\}$ using those of adjacent depth values. Then, we can calculate a look-up-table (LUT) by reproducing the measurements based on the estimated model parameters. Note that we can use this LUT to directly compensate the depth error at runtime of acquisition. That is, having the amplitude value and the measured depth of a given pixel, the LUT returns the corrected depth value.

The LUT is computed as follows. First, we are given model parameters at the known depth values and a grid resolution of the LUT. Then, we compute the truth depth value for each grid in the LUT and store it accordingly. We present the measurements through the operating range on the left side of Fig. 7 and the resultant LUT on the right side of Fig. 7. For the actual implementation shown in Fig. 7, we use 150 entries in each input dimension, resulting a depth resolution of approximately 2.35 cm. In practice, we perform the bilinear interpolation using four adjacent values along two dimensions; amplitude and measured depth; to increase the precision. The execution time for error compensation depends on the implementation scheme. Currently, on our implementation, it performs roughly at 120 fps in average, which is much faster than the frame rate for acquisition. This capability for real-time acquisition enables to model the dynamic scene.

5 Multi-modal sensor fusion

In previous section, we have discussed our error compensation approach for a single ToF sensor. In this section, we present our calibration scheme for the multi-modal devices. For that, we should identify the relative pose of all devices with respect to the reference device, equivalent to solving for extrinsic parameters in camera calibration. Standard approaches use feature points (e.g. checkerboard pattern or similar structured pattern) and estimate the poses of color cameras reasonably well. We employ a standard camera calibration approach, the MATLAB camera calibration toolbox by Bouguet, for calibrating the poses of color cameras.

Finally, we need to position all devices, the ToF sensor and color cameras, onto the same coordinate system. Hence, we first solve for the relative poses of ToF sensors (Sect. 5.1) and then link the poses of color cameras with respect to the ToF sensor (Sec. 5.2). As a result, all cameras shall share the same coordinate system.

5.1 Multiple ToF sensors

Unlike the situation of color cameras, standard camera calibration approaches are infeasible for the ToF sensor due to its poor resolution. These errors in extrinsic parameter estimation cause the serious collapse in 3D scene reconstruction. Hence, we suggest a different approach for calibrating the pose of ToF sensor.

We capture the checkerboard pattern and select more than five points on the same plane in the amplitude image. Then,

we calculate the plane equation using the selected feature points. During the plane calculation, we use the corresponding 3D positions of selected feature points, retrieved from the depth map. This is distinguished from existing methods in that they match the correspondences in 2D image domain so to estimate extrinsic parameters; minimizing the back-projection error from 3D to 2D. We instead handle the correspondences in 3D space and hence gain the precision from the extra dimension. This effectively alleviates the shortcoming from the poor resolution of the ToF sensor.

In the next step, we set a reference camera, typically the center camera, and compute the rotation and translation from other ToF sensors i to the reference camera j. By fitting the selected feature points into (2), we can compute the rotation matrix $\mathbf{R}_{i,j}$ and the translation vector $\mathbf{t}_{i,j}$, transforming all pixels from the camera i to the camera j.

$$\begin{bmatrix} x^j \\ y^j \\ z^j \end{bmatrix} = \mathbf{R}_{i,j} \begin{bmatrix} x^i \\ y^i \\ z^i \end{bmatrix} + \mathbf{t}_{i,j}. \tag{2}$$

Our method of extrinsic calibration for the ToF sensor aim to remove the computational error sources as many as possible. Hence, we do not estimate the intrinsic parameters of low-resolution range sensors. In this way, we prevent from the inaccurate estimate of intrinsic parameters; focal length, the center of projection and the distortion parameters; yielding to the failure in the pose estimation. Instead, we utilize the measurement, the 3D position of features, for directly computing extrinsic parameters.

5.2 ToF sensors and color cameras

To combine a set of ToF sensors and color cameras, we relate the pose of color camera with respect to the ToF sensor. To achieve the correspondences between the ToF and color cameras, we select the same set of features from both the color image and the amplitude image. Since we know the pose of color camera from the standard approach, we can calculate additional rotation and translation parameters that transform the coordinate systems of color camera to the ToF sensor or vice versa.

6 Experimental results and evaluation

We evaluate our calibration approach in two stages. First, we evaluate the performance of our error compensation model for the ToF sensor in Sect. 6.1. Then, we present the accuracy of the pose estimation in Sect. 6.2.

6.1 Error compensation

In this section, we evaluate our error compensation model for the ToF sensor. To justify the reliability, we perform the evaluation for various camera settings. For that, we conduct the experiments under two different camera settings; a modulation frequency of 30 MHz with an integration time of 30 (Setting 1) and that of 40 MHz with the integration time of 70 (Setting 2). Notice that the integration time introduced in this paper is consistent to the integration time parameter of SR4000. In reality, each of these numbers $\{30, 70\}$ translates into $\{0.3 + 30 \times 0.1 = 0.6, 0.3 + 70 \times 0.1 = 1\}$ millisecond. Note that we constitute our LUT for the modulation frequency of 30 MHz with the integration time of 70. Then, we use the same LUT for all other conditions. The error compensation results are visualized in Fig. 8.

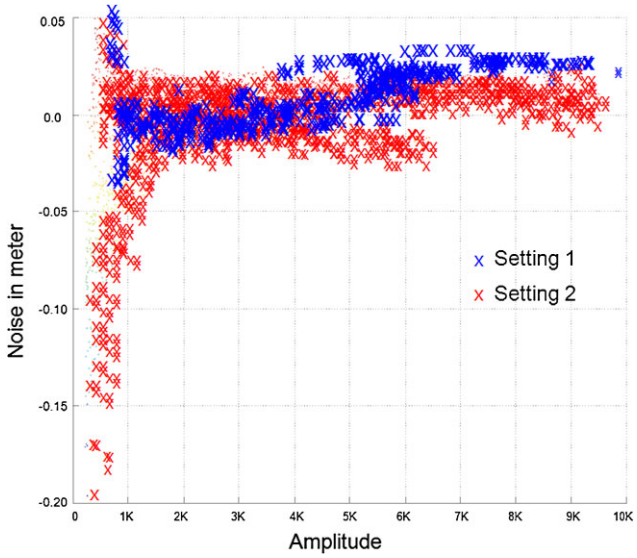


Fig. 8 Error distribution after the error compensation. Setting 1 (*blue cross*): modulation frequency of 30 MHz and an integration time of 30, setting 2 (*red cross*): modulation frequency of 40 MHz and an integration time of 70

Notice that the experiment is conducted with an identical scene and setup to Fig. 4. From Fig. 8, we find that our error compensation method performs reasonably well for both camera settings. Based on this experiment, we also compute the statistics of error distribution and present it in Table 3. These results show that we successfully reduce the error by an order of magnitude.

Finally, we present rendering results upon the error compensation. Figure 9 demonstrates the significant improvement in rendering, achieved by our error compensation procedure. We acquire an image of a checkerboard. The black patches have a low amplitude and thereby they suffer from severe error in depth, including many bumps and other artifacts. After the compensation, the mean deviation from the plane is significantly subsided (right image).

6.2 Multi-modal sensor calibration

For the extrinsic calibration, existing work either rely on the optical estimation of the pose of all the sensors or use the

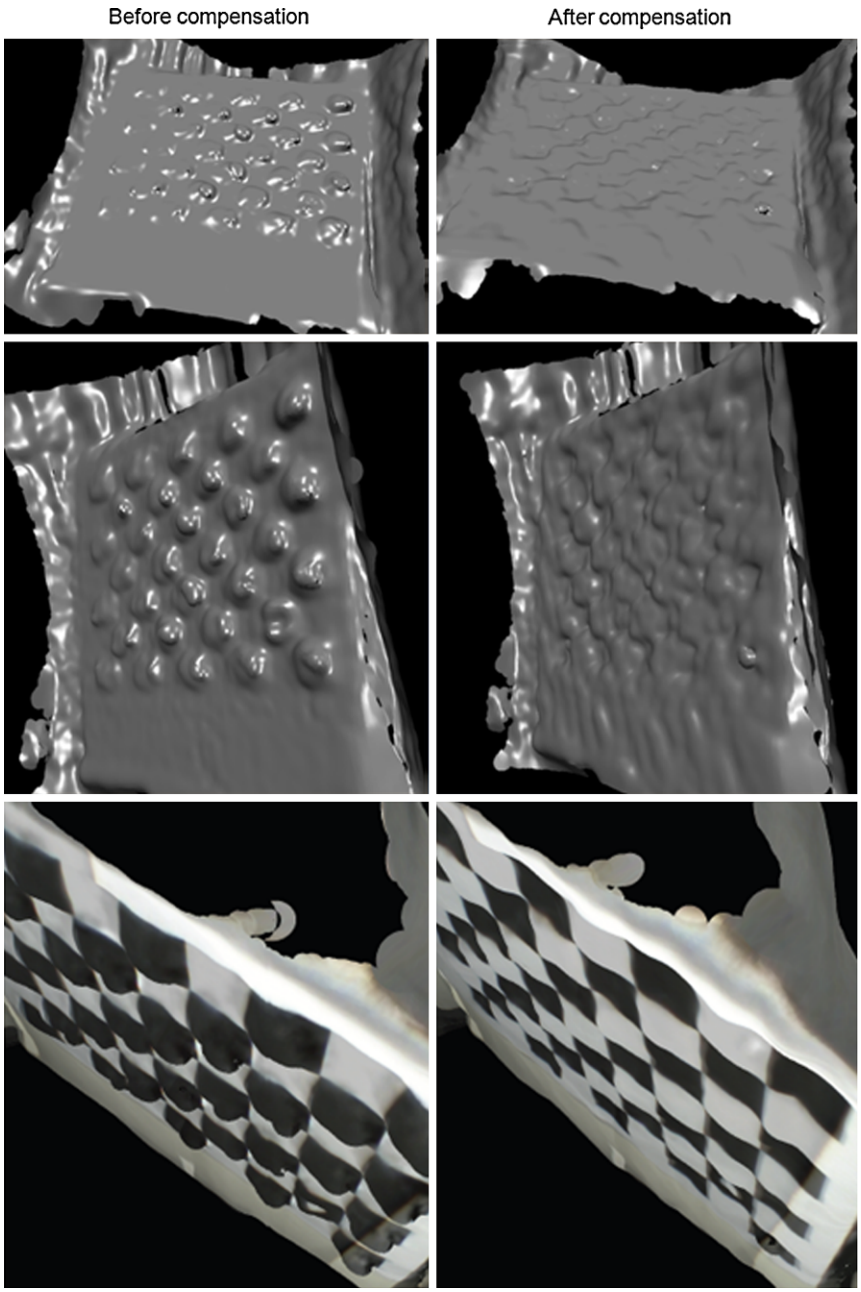


Fig. 9 Visual impact of the error compensation. *Left*: before the error compensation, *right*: after our error compensation

Table 3 Results of our error compensation approach. (Modulation frequency: 30 MHz, Integration time: 30)

Values in meter	Error without compensation	Error with our compensation
Mean	1.70×10^{-2} (m)	1.36×10^{-3} (m)
STD	1.53×10^{-2} (m)	1.24×10^{-2} (m)
VAR	2.33×10^{-4} (m)	1.54×10^{-4} (m)

specialized equipment to measure their position. Often, using special hardware to position the sensors is impractical for most practitioners. Hence, as suggested by the previous work [12], we use the optical features for extrinsic calibration.

After positioning all sensors in the same coordinate system, we render the complete 3D scene and present it in Fig. 10. This figure illustrates three example scenes, (1) checkerboard plane, (2) cushions with doll on the sofa and (3) human sitting on the chair. Each scene is chosen to

Fig. 10 Rendering results using the combined data (three pairs of depth map and color image). *Left*: checkerboard, *right*: multiple objects with textures, *top*: rendering results at novel views, *middle*: three captured color images, *bottom*: three captured depth maps

represent the different type of scene, (1) simple geometry with challenging texture, (2) complex geometry with detailed texture, (3) complex geometry with complex texture. To visualize the synthesized view, we choose the novel view being different from the originally captured views and still show the acceptable quality in rendering. For checkerboard plane, we could produce the synthesized view with a moderate quality, well-aligned checkerboard pattern. For the sofa scene, we could reproduce the sharpness in characters and this reflects some success in sensor calibration. Typically, the characters will clearly be blurred out even with a small calibration error. The example of human body includes complex facial structure and crease on clothes, which is challenging both in terms of geometry and texture. Even with the complex scene, we could show the reasonable quality in rendering, possible by our effective calibration method. All of them demonstrate the effectiveness of the proposed system.

During the experiments, we set the integration time to a constant value of 70, the modulation frequency to 30 MHz (suggested by the manufacturer). The integration time is carefully chosen to ensure the followings; the sensor can operate without saturation even at the position close enough to the wall and it has low noise levels at farther distances.

7 Conclusion

We present a multi-view acquisition system using a set of ToF sensors and color cameras. The main contributions of the proposed work is to provide a practical solution of compensating the systematic error in ToF sensor measurement and an effective method for the extrinsic calibration of multi-modal devices.

The goal of systematic error compensation is to derive an empirical model for ease use. We have identified possible error sources, stated hypotheses of their impact, evaluated our hypotheses, and provided a model to compensate the resultant error. Based on our analysis, the amplitude dependent error, the depth dependent error and lens distortion are the sources of errors. The conclusion of our analysis makes sense as follows. Since ToF sensors accumulate the photons traveling back to the detector after each photon hits the surface point, they inherently possess the systematic bias dependent on the strength of reflected IR signal (amplitude image), the traveling distance (actual depth value) and the optical distortion at detector (lens distortion).

We have employed the optical features in estimating the pose of ToF sensors and color cameras. Unlike related work, our approach does not suffer from erroneous estimation in intrinsic parameters for the ToF sensors. In case of color cameras, we can estimate the intrinsic parameters reliably and accurately because the resolution is sufficiently high.

Finally, we could successfully combine all sensors into the single coordinate system.

We presented a hardware and software system that acquires and processes a scene for 3D visualization. The hardware consists of three sensor pairs, each including one ToF sensor and one color camera. The sensors are arranged with a wide baseline ($\geq$1 m) to capture scenes from different perspectives, expanding the entire view of the scene. The depth data are then combined and processed to generate smooth surfaces. Finally, we successfully generate the synthetic views of the combined scene. Extensive evaluation and analysis show the effectiveness of our system, acquiring and visualizing the multi-view dynamic scene.

8 Limitation and future work

Currently, we investigate the characteristics of depth sensor alone for compensating the systematic error in measurement. For higher accuracy, it is possible to employ the high quality/resolution color image so to account for the error compensation framework. Yet, it is important to note that, depending on the application scenario, the color image may not be a reliable cue for 3D reconstruction if the extreme illumination and shadows appear in the scene.

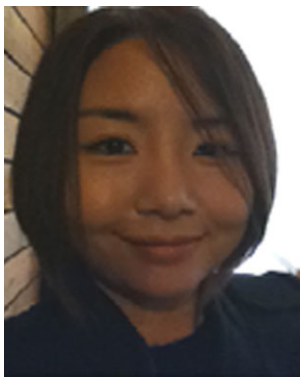
From Table 3, although the proposed method successfully subsides the systematic bias in depth measurement, the standard deviation of measurement error did not show the significant improvement. Based on our analysis, the standard deviation is closely related to the measurement noise, considered as the random noise. For now, in order to greatly reduce the measurement noise, we need to increase the integration time. In the future, we will study an effective technique to reduce the measurement noise by exploring the spatial and temporal consistency in depth map.

For the simple scene structure, similar to the checkerboard plane, ones might consider a model fitting by assuming that a certain structure appears in the scene; planes, sphere, lines, etc. However, unless we have the knowledge about the scene structure, it is hard to know whether it is a plane, the bumpy surface or another complex structure. In this paper, we focus on calibration technique, which is independent on the scene context. Still, adopting the scene context for improving the quality is attractive for the future work.

Also, in addition to the calibration of multi-modal devices, we would like to investigate the effective representation and rendering scheme, suitable for the proposed acquisition system.

References

1. Levoy, M., Hanrahan, P.: Light field rendering. In: Proc. ACM SIGGRAPH, pp. 31–42 (1996)
2. Scharstein, D.: Stereo vision for view synthesis. In: Proc. Computer Vision and Pattern Recognition, pp. 852–858 (1996)
3. Woodham, D.R.J.: Photometric method for determining surface orientation from multiple images. Opt. Eng. **19**(1), 139–144 (1980)
4. Zhang, R., Tsai, P.-S., Cryer, J.E., Shah, M.: Shape from shading: a survey. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 690–706 (1999)
5. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The digital Michelangelo project. In: Proc. ACM SIGGRAPH, pp. 131–144 (2000)
6. Batlle, J., Mouaddib, E.M., Salvi, J.: Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. Pattern Recogn. 1–63 (1998). doi:10.1109/ROBOT.1197.620027
7. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 1330–1334 (2000). doi:10.1109/34.888718
8. Kahlmann, T., Ingesand, H.: Calibration and improvements of the high-resolution range-imaging camera SwissRanger TM. In: Conference on Videometrics VIII, Part of the IS&T/SPIE Symposium on Electronic Imaging (2005)
9. Kahlmann, T., Ingensand, H.: High-precision investigations of the fast range imaging sensor swissranger. In: Optics East (2007)
10. Fuchs, S., May, S.: Calibration and registration for precise surface reconstruction. In: Proceedings of the DAGM Dyn3D Workshop (2007)
11. Fuchs, S., Hirzinger, G.: Extrinsic and depth calibration of ToF-cameras. In: Proc. Computer Vision and Pattern Recognition, pp. 1–6 (2008)
12. Kim, Y.M., Chan, D., Theobalt, C., Thrun, S.: Design and calibration of a multi-view ToF sensor fusion system. In: Proc. Computer Vision and Pattern Recognition (2008)
13. Guan, L., Pollefeys, M.: A unified approach to calibrate a network of camcorders and ToF cameras. In: Proc. European Conference on Computer Vision (2008)
14. Guan, L., Franco, J.-S., Pollefeys, M.: 3D object reconstruction with heterogeneous sensor data. In: 3DPVT (2008)
15. Bouguet, Y.J.: Camera calibration toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/
16. Gut, O.: Untersuchungen des 3D-sensors swissranger, Diploma thesis
17. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
18. Fryer, J.G., Brown, D.C.: Lens distortion for close-range photogrammetry. Photogramm. Eng. Remote Sens. **52**, 51–58 (1986)
19. Payne, A., Dorrington, A., Cree, M.: Optimization of 3D range-imaging sensors. SPIE Newsroom (2008)



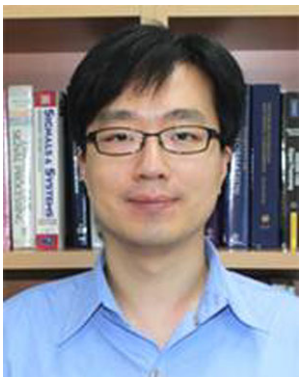
Hyunjung Shim received her PhD and MS degree in Electrical and Computer Engineering from Carnegie Mellon University at 2008. She is currently a research scientist at the Samsung Advanced Institute of Technology, Samsung Electronics. Her research interests include 3D modeling and reconstruction, inverse lighting and reflectometry, face modeling, image-based relighting and rendering, light field capturing and processing algorithms and color enhancement algorithms.

Rolf Adelsberger holds a MSc in Computer Science from the Federal Institute of Technology Zürich (ETHZ). He has written his Master Thesis at Cambridge, MA, where he was involved in a joint Rolf research project between MIT, Mitsubishi Electric Research Lab (MERL) and ETHZ. The motion capture system for "everyday surroundings" was published at the 2007 SIGGRAPH conference. After that Mr. Adelsberger conducted research on infrared (IR) time-of-flight (ToF) cameras at the computer graphics lab (CGL) of ETHZ. Before he focused on the calibration of IR ToF cameras he was involved in the development of a Spatially Adaptive Flash Unit (patented) for DSLR cameras he moved into the area of wireless sensor networks (WSN). Small, unobtrusive computer systems fascinated him since the time of his Master Thesis. Until the end of 2011 he worked at the IBM Research Lab in Ruschlikon (Switzerland). There, he developed wireless inertial measurement sensor boards for IBM's WSNs. Mr. Adelsberger is currently pursuing a PhD in Electrical Engineering while working at the Wearable Computer Lab at ETHZ. His research focuses on medical and sport-related motion and gait analysis of humans.

James Dokyoon Kim received his BS and MS degrees from Yonsei University in 1993 and 1995 respectively. He is the director of 3D Mixed Reality Group of Advance Media Lab. in Samsung Advanced Institute of Technology since 2007. He is the Samsung Research Master of 3D Graphics field and received honorary doctorate degree in 2010. His research interests include 3D image processing, graphics and augmented reality.

Seon-Min Rhee received her PhD and MS degree in computer science and engineering from Ewha Womans University at Seoul, Korea in 2007. Currently she is a research scientist in Samsung Advanced Institute of Technology. Her research interests include Mixed Reality, Computer Graphics, and 3D Vision.

Taehyun Rhee is a Senior Researcher and Research Staff Member of Samsung Advanced Institute of Technology (SAIT) in Samsung Electronics. He received his PhD in Computer Science from University of Southern California in 2008. His research concern is to solve scientific problems related to Computer Graphics, Computer Animation, Computer Vision, and Medical Image. His current research activity is focused on realistic human body modeling and animation, soft-tissue deformation, surface/volume reconstruction from scans, medical image, and realistic rendering algorithms. He was a Senior Engineer of Research Innovation Center at Samsung Electronics from 1996 to 2003 while developing photorealistic rendering algorithms, 3D user interfaces, and VR applications.



Jae-Young Sim received the BS degree in electrical engineering and the MS and PhD degrees in electrical engineering and computer science from Seoul National University, Seoul, Korea, in 1999, 2001, and 2005, respectively. From 2005 to 2009, he was a Research Staff Member, Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd. In 2009, he joined the School of Electrical and Computer Engineering, Ulsan National Institute of Science and Technology (UNIST) as an Assistant Professor. His research topics include image and 3-D visual signal processing, multimedia data compression, and computer vision.

Markus Gross is a Professor of Computer Science at the Swiss Federal Institute of Technology Zürich (ETH), head of the Computer Graphics Laboratory, and the Director of Disney Research, Zürich. He joint the ETH Computer Science faculty in 1994. His research interests include physically based modeling, computer animation, immersive displays, and video technology. Before joining Disney, Markus was director of the Institute of Computational Sciences at ETH. He received a master of science in electrical and computer engineering and a PhD in computer graphics and image analysis, both from Saarland University in Germany in 1986 and 1989. Markus serves on the boards of numerous international research institutes, societies, and governmental organizations. He received the Technical Achievement Award from EUROGRAPHICS in 2010 and the Swiss ICT Champions Award in 2011. He is a fellow of the EUROGRAPHICS Association and a member of the German Academy of Sciences Leopoldina. Prior to his involvement in Disney Research he cofounded Cyfex AG, Novodex AG, LiberoVision AG, and Dybuster AG.

Changyeong Kim received his MS and PhD degrees from Korea Advanced Institute of Science Technology in 1987 and 1996 respectively. He is the head of Advance Media Lab. in Samsung Advanced Institute of Technology since 2010. He is the Samsung Electronics Fellow since 2006 and IS&T Honorary Member. His research interests include 3D image processing, color processing and medical imaging.