Design and Calibration of a Multi-view TOF Sensor Fusion System

Young Min Kim Derek Chan Christian Theobalt Sebastian Thrun

Stanford University, USA

[ymkim, ddc, theobalt, thrun]@stanford.edu

Abstract

This paper describes the design and calibration of a system that enables simultaneous recording of dynamic scenes with multiple high-resolution video and low-resolution Swissranger time-of-flight (TOF) depth cameras. The system shall serve as a testbed for the development of new algorithms for high-quality multi-view dynamic scene reconstruction and 3D video. The paper also provides a detailed analysis of random and systematic depth camera noise which is important for reliable fusion of video and depth data. Finally, the paper describes how to compensate systematic depth errors and calibrate all dynamic depth and video data into a common frame.

1. Introduction

The advancement of computing and camera technology in recent years has rendered it feasible to attack the problem of dynamic scene reconstruction with the goal of novel viewpoint rendering or 3D video. In order to provide high-quality 3D video renderings, several aspects of a dynamic scene have to be reliably reconstructed, including geometry and textural appearance. Simultaneous reconstruction of all these scene aspects is only feasible if the measurement principle of the sensors does not visually interfere with the real world. In the past, purely camera-based or simple active measurement algorithms have been proposed to meet these requirements. Shape-from-silhouette methods [9] and model-based approaches [2] are suitable for 3D videos of foreground objects, but either, like the former, suffer from mediocre shape reconstruction, or, like the latter, are limited to certain types of scenes. Stereo-based [16] or light-field-based methods [15] can reconstruct entire scenes, albeit at a more limited virtual viewpoint range. Despite their greater flexibility, the latter two approaches are also confronted with the difficult multi-view correspondence problem, which makes unambiguous depth reconstruction a challenging task.

Recently, the measurement quality of time-of-flight flash LIDARS, such as the MESA^{TM} Swissranger 3000 [10, 11], has dramatically improved. This new type of sensor measures scene depth at video rate by analyzing the phase shift between an emitted and a returned infrared light wavefront. In contrast to stereo, depth accuracy is independent of textural appearance. Sadly, the current camera models feature only sensors with a comparably low pixel resolution. Also, the depth readings are starkly influenced by random and systematic noise. We therefore built a multi-view recording system that simultaneously captures scenes from multiple depth and vision cameras. The system is intended to be a test-bed for developing new multi-view sensor fusion methods which join the forces of vision and depth sensors to recover high-quality geometry and texture. This paper addresses the two first challenges we are facing, namely the analysis of TOF camera noise and the development of a multi-view sensor fusion calibration procedure. The main contributions of this paper are:

- the design of a multi-view TOF sensor fusion system (Sect. 3);
- an analysis of random and systematic measurement errors of the SR-3000 (Sect. 4);
- a simple and efficient calibration to compensate systematic depth errors and accurately align dynamic depth and video data in a common frame (Sect. 5).

2. Related Work

Most previous multi-view 3D video recording systems employed synchronized video cameras only, such as [2, 9, 16, 15] to name a few. Waschbuesch et al. [14] proposed to use several DLP projectors throwing noise patterns into a scene to improve multi-view stereo. To our knowledge, we present here the first system to combine several video and time-of-flight depth cameras. It enables us to jointly acquire shape and multi-view texture at video rate, and requires neither visual interference with the scene, nor any form of segmentation or highly textured surfaces.

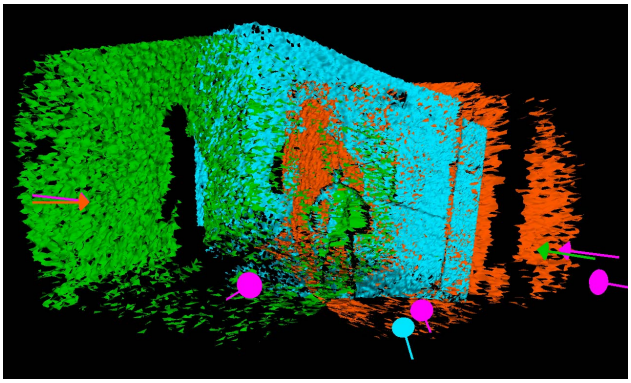


Figure 1. Calibration result: Data captured with 3 depth cameras (different colors) registered into a common world frame. The scene shows a person standing in front of the corner of the room. Overall, the depth maps almost perfectly align. The inter-twisted color pattern on the back wall is due to the random noise level in the data. The arrows show positions of the 3 depth and 5 additional video cameras (25 frames of static scene averaged to minimize random noise; depth maps rendered as triangle meshes with filtered elongated triangles at boundaries)

Due to the complex noise characteristics of the Swissranger^{TM} depth camera, calibration of all captured data into a common world frame is a non-trivial task. Some previous work has primarily looked at the random noise characteristics of earlier TOF camera models [1, 3]. In our research we capitalize on their insights and extend them into a generative noise model parametrized in actual sensor readings. The most problematic property of the Swissranger, however, is its systematic measurement bias that yields consistent measurement inaccuracies. Previous work has characterized and compensated systematic depth errors for a different camera, the PMD photonic mixer device [6, 7]. Kahlman et al. [5] computed look-up tables for the Swissranger to compensate integration time-dependent biases. Rapp [12] conducted theoretical and experimental analysis on the noise characteristics of single TOF sensors. In this work, we derive detailed models of both random and systematic bias under fixed measurement conditions and show how to calibrate multiple depth cameras into a common world frame.

3. System Architecture

The building block of our multi-view depth and video system is a so-called *fusion-unit*, Fig. 2. The computing power of a fusion unit is provided by a Dual Core Athlon 64 5600+ computer featuring 4 GB of memory. The computer is connected to up to two Point Grey Flea2 FireWire^{TM} B cameras that we typically run at 22 fps and 1024x768 pixels frame resolution. In addition, each unit can control one Swissranger SR 3000 time-of-flight camera that typically records at a 30 ms integration time which turned out to be

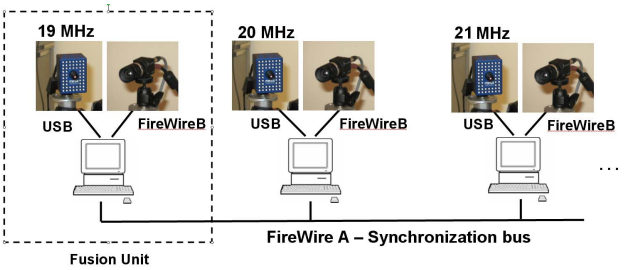


Figure 2. The architecture of our hybrid multi-view video and depth camera system. The building blocks of the system are connected fusion units, each of which controls up to two video and one depth cameras.

a good compromise between frame rate and data fidelity. The X/Y resolution of the TOF camera is 176×144 pixels. Each fusion unit runs our MVFusionRecorder software which enables sensor adjustment, on-the-fly-preview, synchronization of sensors and different recording modes.

A complete fusion recording setup connects several fusion units. It is of utmost importance that all employed sensors synchronously capture frames. Since the Swissranger cameras provide no means to feed in an external hardware synchronization pulse, we resort to the following hybrid hardware and software solution to synchronize all types of sensors across multiple fusion units.

All video cameras are hardware synchronized to an accuracy of 0.125 ms via a second FireWire^{TM} bus that links all fusion units in the system. To ensure that all Swissrangers capture a frame at the same time instant, and in order to be certain that this time instant coincides with the integration window of the video cameras, the depth camera integration is started simultaneously in software with the sync pulse that triggers video frame capture.

To prevent I/O streaming bottlenecks, data are captured to memory and stored to hard disk after recording is finished. Since infrared-based time-of-flight cameras show interference errors if multiple infrared emitters with the same modulation rate illuminate the scene, we set the modulation frequency of each camera to a different value (19 MHz, 20 MHz, 21 MHz or 30 MHz). As this range of frequencies is hardware-limited to four specific values, we can currently run at most four depth cameras simultaneously in one setup, Fig. 2. We are now flexible in placing our cameras throughout the scene. A typical converging camera arrangement with overlapping view frustra is illustrated in Fig. 1.

4. Depth Sensor Characteristics

The Swissranger is a so-called amplitude-modulated continuous wave (AMCW) measurement device that infers 3D scene structure by measuring scene depth along the rays through all sensor pixels (u, v). Depth along measurement rays is computed from the phase shift between a

sinusoidally-modulated wave-front emitted from IR diodes on the sensor, and the returning light wavefront reflected from the scene. One important feature of AMCW sensors is that, in contrast to triangulation sensors, measurement uncertainty only exists along the viewing ray from the depth camera to a point in 3D. This measurement uncertainty comprises of two components, a random component $d_r(u, v)$ and a systematic bias $d_s(u, v)$. While the random component accounts for the random per-frame deviation of the measured depth $d_m(u, v)$ from the ground truth depth $d_g(u, v)$, the systematic bias models discrepancies that are consistent over time. The measurement model for the distance along a ray through pixel (u, v) can therefore be represented as:

$$d_m(u, v) = d_g(u, v) + d_r(u, v) + d_s(u, v)\,. \qquad (1)$$

In the following, we provide models for random noise and, most importantly, systematic bias.

4.1. Random Noise

Motivated by earlier work of Hebert et al. [3], Anderson et al. [1] conclude that the standard deviation of the random distance variations along measurement rays σ_r can be approximated by:

$$\sigma_r \propto \frac{\lambda d_g^2}{\rho \cos \alpha}\,. \qquad (2)$$

Here, λ is the wavelength, ρ the reflectance of the target, and α the angle of incidence. While important from a functional point of view, this formulation is less suited as a generative model since it incorporates dependencies that are not immediately measurable. In our practical setting, the wavelength dependency can be ignored as the LEDs emit a fixed wavelength. Furthermore, it is fair to assume that the measured distance is reasonably close to the ground truth distance and hence d_g can be approximated by d_m. Finally, it can be assumed that reflectance and orientation dependency correlate with actual amplitude variations. Low amplitude typically leads to a low signal-to-noise ratio which results in a higher standard deviation. Experiments with gray cards of different albedo that were recorded under different angles showed that the effect of reflectance and orientation on *random noise* under typical lab conditions is insignificant and can in practice be ignored. Note that currently we also ignore dependencies on temperature and integration time as we hardly vary them in practice. We would also like to note that the above model does not account for the increased noise variance in mixed pixels, i.e. pixels that integrate over a depth discontinuity. In practice, we can discard these unreliable measurements from recorded data by enforcing a depth difference threshold around depth discontinuities.

4.2. Systematic Bias

The systematic measurement bias $d_s(u, v)$ leads to depth inaccuracies that are consistent over time. In order to understand and eventually correct the systematic measurement errors, we need to acquire ground truth 3D measurements that we compare against the sensor output. To this end, it is feasible to employ the off-the-shelf MATLAB calibration toolbox for normal vision cameras [13] since the depth sensor, in addition to 3D depth measurements, provides an amplitude image at each frame. Similar to [8], a depth camera can thus be treated as a normal optical camera for now. We can therefore record images of a checkerboard pattern to compute an intrinsic matrix $\mathbf{K}$, lens distortion coefficients, and extrinsic parameters $\mathbf{R}$ and $\mathbf{T}$ – henceforth we will refer to this model of a depth camera as the model in *space I*.

The Swissranger also provides for each pixel (u, v) a point in metric 3D space $\mathbf{p}(u, v) = (x, y, z)$ whose location is determined via time-of-flight. Henceforth, we refer to this time-of-flight measurement space as *space II*.

Due to the systematic bias, 3D depth point clouds of checkerboard corners according to space I do not exactly align with the checkerboard corners measured in space II. Physical origins of this bias are manifold, but include inaccuracies in the measurement model assumed by the manufacturer which may, for instance, not correctly cater for lens aberrations, amplitude effects, mixed pixels etc. Our experiments show that the systematic measurement bias can be modeled by the following factors: Rigid misalignment $\mathcal{R}$, ray-space misalignment $\mathcal{D}$, a distance bias along the measurement ray $\mathcal{B}$, and the subtle influence of orientation and translation that we can implicitly model as dependencies on ratios of normalized amplitudes. The first three types of misalignment are depicted in Fig. 3. Our calibration procedure described in Sect. 5 will eventually estimate and compensate exactly those effects in the range measurements.

Rigid misalignment, $\mathcal{R}$, means that the 3D point clouds in space I and space II are off by a rigid transformation. Ray-space misalignment, $\mathcal{D}$, means that, even after rigid correction, viewing rays (or measurement rays) towards corresponding 3D points in space I and space II do not point exactly in the same direction but are angularly misaligned. By far the strongest misalignments originate from the bias $\mathcal{B}$ in distances along measurement rays, Fig. 5(a). Finally, variations in surface orientation and reflectance are likely to have an influence on the systematic error as they lead to changes in measured amplitude.

Overall, our model of systematic measurement bias takes the following form:

$$d_s(u, v) = K e^{-br} d'_s(u, v)\,, \qquad (3)$$

where $d'_s(u, v)$ is the systematic measurement bias due to

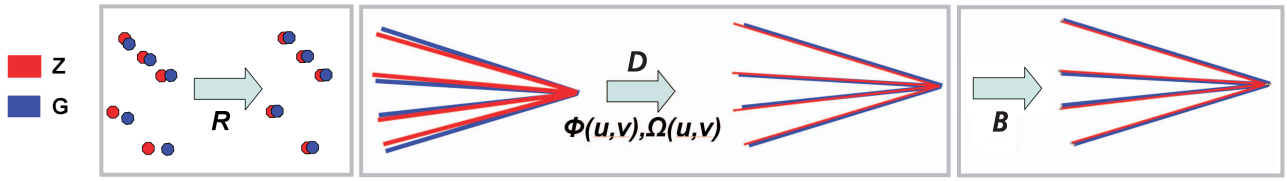


Figure 3. Steps to compensate the systematic bias in the depth data and align them with an optical camera model (space I) of the sensor. For clarity, steps are illustrated in 2D. Red shows the measured TOF data, blue the ground truth data in space I; circles represent points in 3D, lines represent measurement rays.

rigid, ray and distance misalignment. The exponential term preceding $d'_s(u,v)$ models the influence of the orientation and reflectance by reducing them to changes in measured normalized amplitude. This formulation originates from the fact that, even though the depth measurements along the rays that have smaller amplitude values are biased to be farther away compared to the true depth, we found no obvious direct relationship between the measured amplitude and the systematic error in practice (also reported in [7]). Instead, ratios of normalized amplitude are used based on two experimental observations: (1) The systematic bias is dependent on the relative amplitude value of the particular pixel compared to the other pixels within the same frame. (2) The relationship between the systematic bias and the amplitude value is affected by the image plane location of a pixel. For example, when one measures against a planar white wall, one will see a radial amplitude pattern, Fig. 4, but the depth measurements nonetheless correctly show the flat surface of the wall.

In Eq. (3) $\bar{A}(u,v) = \frac{A(u,v)}{\sum_{u',v'} A(u',v')}$ denotes the normalized amplitude at pixel location (u,v). The normalized reference amplitude $\bar{A}_r(u,v) = \frac{A_r(u,v)}{\sum_{u',v'} A_r(u',v')}$, is the radial amplitude pattern seen whenever the sensor is looking at a flat surface perpendicular to the viewing angle, Fig. 4. This pattern is almost independent of overall distance to the wall and thus serves as our baseline. Given this, $r(u,v) = \frac{\bar{A}(u,v)}{\bar{A}_r(u,v)}$ is the ratio of normalized amplitudes at pixel location (u,v). This way, we can implicitly model surface orientation and reflectance dependencies as they relate to measurable variations in normalized amplitude ratios.

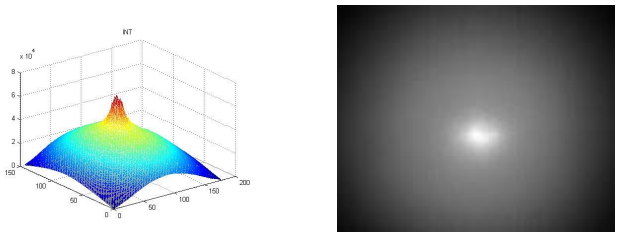


Figure 4. Normalized intensity amplitude $A_r(u,v)$ plotted as height field (left). Corresponding amplitude image recorded by the sensor (right).

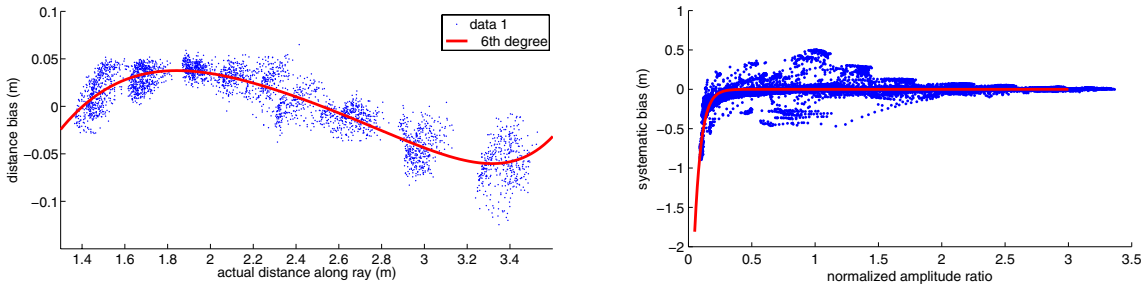
The exponential formulation in Eq. (3) was found by recording gray cards of 10%, 50% and 90% reflectance under inclination angles ranging from $10°$ to $70°$ in $10°$ increments. $K = -4.6428$ and $b = 18.8410$ are typical coefficients obtained by fitting the above exponential to our Swissranger data. Fig. 5(b) shows that the error was explained by Eq. (3). In practice, the amplitude ratio effect only plays a role for samples with $r \leq 0.3$, which are only very few samples with an extremely dim signal return, e.g. at points seen under grazing angles. This threshold thus enables us to filter out unreliable samples. If this is done, we can in practice simplify Eq. (3) and approximate $d_s(u,v)$ by $d'_s(u,v)$.

5. System Calibration

Based on the insights on systematic depth errors, we can derive a practical calibration procedure that aligns all depth maps in 3D without discontinuities and that calibrates the video cameras into the same world frame. Our method does not require special calibration objects for the depth cameras and works with the same checkerboard pattern used for video camera calibration.

The calibration procedure comprises of three stages. First, the intrinsic parameters of each video camera are computed using a checkerboard and the MATLAB calibration toolbox. The intrinsics of each depth camera $d \in 1,\ldots,N_d$ are found as well by analyzing amplitude images of the checkerboard. Secondly, the extrinsics of both depth and video cameras are found by placing the checkerboard in the center of the scene and computing rotation and translation matrices with respect to a common world frame.

The third step is the estimation and compensation of each depth camera's systematic measurement error, Sect. 4.2, which is performed as follows: To obtain ground truth measurements, the checkerboard is placed at $k = 25 - 35$ different locations in the viewing frustum of a Swissranger. At each position k, 50 frames of the static checkerboard are recorded. It is important that over the whole set of measurements the checkerboard positions cover the depth range of the camera. Moreover, the checkerboard corners should reproject to as many different sensor pixel locations as possible. At each position k, an extrinsic calibration



(a) (b)

Figure 5. (a) Error in distance along ray in dependence of ground truth distance only: the data is approximated by a 6-degree polynomial (red). (b) Full model of distance bias: a 10%, 50%, and 90% reflective grey cards were (at a constant distance) recorded under different inclinations (10° - 70°); the plot shows the ratio of normalized amplitude against systematic distance error; the red curve is the exponential factor in Eq. (3).

is performed yielding $k \cdot N_B$ ground truth 3D positions $\mathbf{G} = \{\mathbf{g}_i | \mathbf{g}_i = (x, y, z)_i,\ i \in 1, \ldots, k \cdot N_B\}$ in the Swissranger's eye space. Here, N_B denotes the number of corners in the checkerboard. Following the terminology from Sect. 4.2, we refer to these data as the measurements in space I.

Corresponding space II measurements $\mathbf{Z} = \{\mathbf{z}_i | \mathbf{z}_i = (x, y, z)_i, i = 1, \ldots, k \cdot N_B\}$ of the same data are obtained by averaging the set of 50 measurements taken at each k. By averaging all 50 depth measurements at each k, space II 3D measurements $\mathbf{Z} = \{\mathbf{z}_i | \mathbf{z}_i = (x, y, z)_i, i = 1, \ldots, k \cdot N_B\}$ of the same data set are obtained in which the influence of random noise is minimized.

The goal of depth calibration now is the computation of a warping function $\mathcal{W}$ that maps the elements of $\mathbf{Z}$ onto the elements of $\mathbf{G}$, $\mathbf{g}_i = \mathcal{W}(\mathbf{z}_i)$. Following the analysis of the systematic bias in Sect. 4, we propose that $\mathcal{W}$ should be of the form (see also Fig. 3):

$$\mathbf{g}_i = \mathcal{W}(\mathbf{z_i}) = \mathcal{B} \circ \mathcal{D} \circ \mathcal{R}(\mathbf{z}_i) \tag{4}$$

In the following, we take a closer look at each component of $\mathcal{W}$.

Rigid Compensation $\mathcal{R}(\mathbf{z}_i)$ applies a rigid body transformation that aligns $\mathbf{Z}$ and $\mathbf{G}$. Since correspondences are given, the optimal rigid body transformation can be estimated from the data in closed form using the method of Horn [4].

Directional Compensation The second component, $\mathcal{D}$, corrects directional misalignments between the viewing rays from the camera to the 3D points $\mathbf{G}$ and $\mathbf{Z}$. To estimate $\mathcal{D}$ from the data, we compute for each pair of $\mathbf{g}_i \in \mathcal{G}$ and $\mathbf{z}_i \in \mathcal{Z}$ the corresponding pair of viewing ray directions from the coordinate origin, $\delta(\mathbf{g}_i)$ and $\delta(\mathbf{z}_i)$. For each pair $\delta(\mathbf{g}_i)$ and $\delta(\mathbf{z}_i)$, we can derive polar angular corrections $\Omega(u_i, v_i)$ and $\Phi(u_i, v_i)$ to bring $\delta(\mathbf{z}_i)$ into alignment

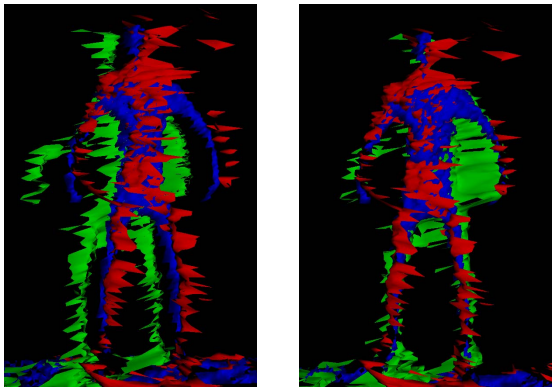


Figure 6. Impact of bias correction: (left) three synchronously recorded depth maps of a person that do not line up if only extrinsics are computed and no systematic bias is corrected. (right) correctly aligned depth maps after extrinsic calibration and bias correction (OpenGL rendering of each depth map as triangle mesh: triangles at occlusion boundaries were only partly filtered and are still visible in the green map captured from a lateral viewpoint).

with $\delta(\mathbf{g}_i)$, (u_i, v_i) being the reprojected pixel locations of points i. Unfortunately, in practice only for a subset of pixels (u, v) will there be corresponding 3D points in $\mathbf{G}$ and $\mathbf{Z}$, and thus the angular correction fields $\Omega(u_i, v_i)$ and $\Phi(u_i, v_i)$ are not fully defined. In order to obtain a directional compensation estimate for the rays through al pixels, we therefore interpolate the per-ray estimates $\Omega(u_i, v_i)$ and $\Phi(u_i, v_i)$ across the entire image plane to yield dense angular correction fields $\Omega(u, v)$ and $\Phi(u, v)$.

Correction of Distance Along Ray After rigid and directional alignment, the measurement ray directions of space I and space II closely match. However, due to the systematic measurement bias along the ray (Sect. 4.2), the lengths of the rays may not correspond. To compensate this we apply a transform $\mathcal{B}$ which adds to each ray distance a depth offset, $\tau(d_m(u, v))$. As shown in Fig. 5(a), we can approx-

Figure 7. Sequence recorded with 2 depth and 3 video cameras: (left) color-coded depth maps aligned in common world space; camera arrangement shown in foreground (data were median filtered,depth shadow of person seen on distant geometry) - (middle) same time step now showing entire geometry projectively textured from all video camera views (triangles at depth boundaries removed) - (right) one video input frame for comparison.

imate $\tau(d_m(u,v))$ by a 6-degree polynomial which yields a faithful yet easy-to-compute model for correcting systematic depth inaccuracies.

Given the above generative compensation model we can correct the measured depth data from all Swissrangers and align them in world space using the extrinsic parameters obtained via traditional optical calibration (space I), Fig. 1 and Fig. 7.

6. Results and Discussion

We have applied our multi-view calibration procedure to different setups of 2-3 depth cameras and 3-5 video cameras. In both the 2-depth camera, Fig. 7, and 3-depth camera case, Fig. 1, the dynamic depth maps nicely align and the frames from the calibrated video cameras can simultaneously be back-projected onto the 3D geometry. Fig. 8 shows an overhead rendering of the reconstructed scene with 2 depth sensors. From this perspective one can see that both the geometry on the back wall and the geometry of the person in the foreground align well. For visualization, we slightly filter the depth maps to minimize random noise impact. For rendering, we employ OpenGL and display each depth map as individual triangle mesh after filtering elongated triangles at occlusion boundaries. The inter-twisted color pattern in overlapping depth maps which is visible in Fig. 1 and Fig. 7(a) is due to random noise in the data. Therefore, this pattern should not be mistaken for a sign of incorrect alignment.

As shown in Fig. 6 our proposed TOF bias compensation is an essential step without which 3D depth maps would never properly align in 3D. In this figure, we set the back clipping planes such that the background of the room is discarded during rendering.

We also performed a quantitative evaluation of the reduction of 3D Euclidean error relative to the ground truth measurements. Tab. 1 illustrates the high impact of our method for one depth camera by showing the mean error reduction, minimal and maximal errors, as well as error standard deviations after each step in the bias correction procedure. For this data set, roughly 2500 ground truth points were measured. Overall, a reduction from over 6 cm average error to around 1 cm average error is achieved which is the typical range observed for all depth

	min	max	mean	stddv
Raw	0.0031	0.0977	0.0494	0.0202
Rigid	0.00033	0.1249	0.0353	0.0176
Rigid + dir	0.00048	0.1247	0.0314	0.0191
full	4.8422e-04	0.0647	0.0136	0.0088

Table 1. Depth measurement accuracy (3D Euclidean distance in m) before compensation, after rigid alignment, after rigid+directional alignment, and after the full compensation pipeline. In total, around 2500 corresponding points in space I and space II were analyzed. The first two columns show the minimal and maximal error observed in the whole data set. Columns 3 and 4 show the error mean and standard deviation after each step which illustrates the significant improvement in accuracy achieved with our approach.

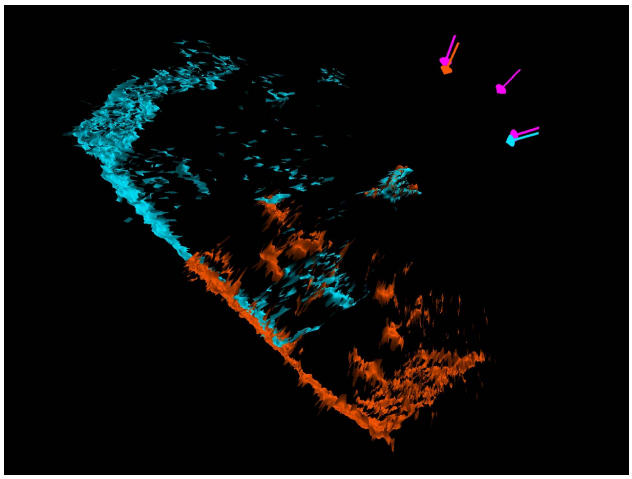
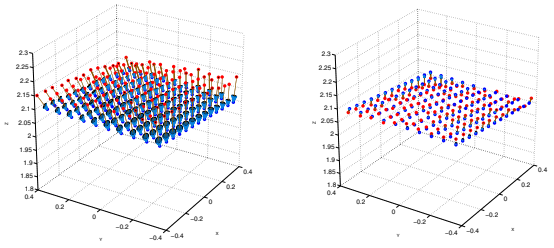


Figure 8. Aligned depth maps captured with two depth cameras shown from an overhead view. The maps of the walls in the background and the person in the foreground are well aligned (the isolated geometry in-between stems from boxes standing in the scene's background).



(a) (b)

Figure 9. Visual illustration of the 3D error between space II measurements (red spheres) and space I ground truth (blue spheres). The arrows illustrate the size of the Euclidean error and the direction of the offset. (a) shows the significant misalignment between ground truth checkerboard data and sensor output without systematic bias correction. (b) shows the strongly reduced error after our compensation method has been applied.

cameras. Please note that ground truth data are always captured with all sensors switched on. Fig. 9 also shows visually the achieved error reduction relative to ground truth data captured at one checkerboard position. While the arrows in Fig. 9(a) show that the unprocessed measurements of the Swissranger strongly deviate from ground truth, after full correction the 3D error has been significantly reduced, Fig. 9(b).

We would like to note that currently, our calibration is only valid for one fix set of recording conditions. Since we are working under fixed indoor conditions and focus on scenes of similar speed, we have not extensively examined dependencies on different sensor integration times, sensor temperature or changing external lighting conditions.

Furthermore, the core of this paper is not the algorithmic solution of the multi-view sensor fusion problem, i.e. the development of algorithms to improve the quality of depth maps at individual time steps after alignment and bias correction. The investigation of this problem is planned as part of future work. Therefore, currently all renderings shown are from largely unprocessed data coming out of the sensors. We only performed simple median and small kernel Gaussian filtering to remove the most severe random noise peaks and better visualize the bias compensation effect. In the future, we plan to further refine our model of random noise and its dependencies which will play a more important role during multi-view fusion and quality improvement of individual frames.

7. Conclusion

We have presented the design of a multi-view time-of-flight sensor fusion recording system. Detailed analysis of depth measurement inaccuracies enabled us to compensate systematic TOF measurement errors and calibrate all depth and video data into a common frame. Our calibration procedure for depth cameras is highly practical and easily ties in with standard optical camera calibration procedures. Starting from the now aligned data, we will in future investigate new ways for improved dynamic shape and texture reconstruction of arbitrary dynamic scenes.

References

[1] D. Anderson, H. Herman, and A. Kelly. Experimental characterization of commercial flash ladar devices. In *International Conference of Sensing and Technology*, November 2005.

[2] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. In *Proc. of SIGGRAPH'03*, pages 569–577, 2003.

[3] M. Hebert and E. Krotkov. 3d measurements from imaging laser radars: How good are they? *IVC*, 10:170–178, 1992.

[4] B. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journ. of the OSA*, 4(4):629–642, 1987.

[5] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration for increased accuracy of the range imaging camera swissrangertm. In *Proc. of IEVM*, 2006.

[6] M. Lindner and A. Kolb. Lateral and depth calibration of pmd-distance sensors. pages II: 524–533, 2006.

[7] M. Lindner and A. Kolb. Calibration of the intensity-related distance error of the pmd tof-camera. In *SPIE: IRCV XXV*, volume 6764, 2007.

[8] M. Lindner, A. Kolb, and K. Hartmann. Data-fusion of PMD-based distance-information and high-resolution RGB-images. In *Int. Sym. on Signals Circuits & Systems (ISSCS), session on Algorithms for 3D TOF-cameras*, pages 121–124. IEEE, 2007.

[9] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-based visual hulls. In *Proceedings of ACM SIGGRAPH 00*, pages 369–374, 2000.

[10] MESA. http://www.mesa-imaging.ch/, 2008.

[11] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc. An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution. In *Proc. SPIE: Optical Design and Engineering*, pages 534–545, 2004.

[12] H. Rapp. Experimental and theoretical investigation of correlating TOF-camera systems. Master's thesis, University of Heidelberg, Germany, 2007.

[13] K. Strobl, W. Sepp, S. Fuchs, C. Paredes, and K. Arbter. http://www.vision.caltech.edu/bouguetj/calib_doc/, 2007.

[14] M. Waschbuesch, S. Wuermlin, and M. Gross. 3D video billboard clouds. In *Proc. Eurographics*, 2007.

[15] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph.*, 24(3):765–776, 2005.

[16] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM TOG (Proc. SIGGRAPH'04)*, 23(3):600–608, 2004.