

PERFORMANCE ANALYSIS OF A LOW-COST TRIANGULATION-BASED 3D CAMERA: MICROSOFT KINECT SYSTEM

J.C.K. Chow*, K.D. Ang, D.D. Lichti, and W.F. Teskey

Department of Geomatics Engineering, University of Calgary, 2500 University Dr NW, Calgary, Alberta, T2N 1N4, Canada
(jckchow, kdang, ddlichti, and wteskey)@ucalgary.ca

Commission V, WG V/3

KEY WORDS: 3D camera, RGB-D, accuracy, calibration, biometrics

ABSTRACT:



Recent technological advancements have made active imaging sensors popular for 3D modelling and motion tracking. The 3D coordinates of signalised targets are traditionally estimated by matching conjugate points in overlapping images. Current 3D cameras can acquire point clouds at video frame rates from a single exposure station. In the area of 3D cameras, Microsoft and PrimeSense have collaborated and developed an active 3D camera based on the triangulation principle, known as the Kinect system. This off-the-shelf system costs less than \$150 USD and has drawn a lot of attention from the robotics, computer vision, and photogrammetry disciplines. In this paper, the prospect of using the Kinect system for precise engineering applications was evaluated. The geometric quality of the Kinect system as a function of the scene (i.e. variation of depth, ambient light conditions, incidence angle, and object reflectivity) and the sensor (i.e. warm-up time and distance averaging) were analysed quantitatively. This system's potential in human body measurements was tested against a laser scanner and 3D range camera. A new calibration model for simultaneously determining the exterior orientation parameters, interior orientation parameters, boresight angles, leverarm, and object space features parameters was developed and the effectiveness of this calibration approach was explored.

1. INTRODUCTION

Passive photogrammetric systems have the disadvantage of high computation expense and require multiple cameras when measuring a dynamic scene. A single camera can be used when reconstructing a static scene, but more than one exposure station is needed. Terrestrial laser scanners are an attractive alternative for 3D measurements of static objects. These instruments can yield geospatial measurements at millimetre-level accuracy, at over a million points per second. Most importantly, they can determine 3D positions on textureless surfaces with no topography from a single scan. However, scanners cannot be used for measuring moving objects due to the scan time delay. With the aforementioned limitations, the development of 3D range cameras began (Lange & Seitz, 2001). Time-of-flight (TOF) 3D cameras measure the range between the sensor and the object space at every pixel location in real time with almost no dependency on background illumination and surface texture. Pulse-based range cameras can measure longer ranges but the distance accuracy is dependent on the clock accuracy (e.g. DragonEye). Although stronger laser pulses can be used for distance measurement, the frame rate of the sensor is limited due to the difficulty in generating short pulses with fast rise and fall time (Shan & Toth, 2008). This low data acquisition rate and requirement for expensive clocks can theoretically be mitigated by using AM-CW light (e.g. PMD, DepthSense, Fotonic, SwissRanger, and D-Imager). However, in reality a high integration time is required to reduce noise, which limits the frame rate for real-time applications (Kahlmann et al., 2006). One of the most important errors for phase-based 3D cameras stems from the internal scattering effect, where distance observations made to the background objects are biased by the strong signals returned by the foreground objects. This scene-dependent error can be difficult to model and is one of the main limiting factors for using this category of cameras in many applications (Mure-Dubois & Hugli., 2007). In addition, the 3D range cameras currently on

the market are still relatively expensive and have low pixel resolution.

On the other hand, the Kinect is a structured light (or coded light) system where the depth is measured based on the triangulation principle. The Kinect consists of three optical sensors: an infrared (IR) camera, IR projector, and RGB camera. The projector emits a pseudo-random pattern of speckles in the IR light spectrum. This known pattern is imaged by the IR camera and compared to the reference pattern at known distances stored in its memory. Through a 9x9 spatial correlation window, a disparity value is computed at every pixel location which is proportional to a change in depth. The Kinect emits speckles of 3 different sizes to accommodate for objects appearing at different depths and according to the manufacturer it produces a measurable range of 1.2m – 3.5m. More details about the inner mechanics of the Kinect can be found in (Freedman et al., 2010; and Konolige & Mihelich, 2010).

Before commencing user self-calibration of the Kinect, the off-the-shelf sensor error behaviour is tested and the results are summarized in Section 2. The performances of the Kinect for 3D object space reconstruction (more specifically for measuring the human body) is evaluated in Section 3. This system is not free of distortions; noticeable data artefacts and systematic errors in the Kinect have been reported in literature (Menna et al., 2011). The proposed mathematical model for performing self-calibration of the Kinect is presented in Section 4, followed by some empirical results illustrating the effectiveness of the new calibration method in Section 5. Please note, in the Preliminary Tests (Section 2) the Kinect data is acquired using either the Microsoft Kinect SDK Beta 1 (sub-headings marked with a ) or software named Brekel Kinect (Brekelmans, 2012) (sub-headings marked with a ). Thereafter, all Kinect data in subsequent sections is acquired using the Microsoft Kinect SDK Beta 1 (Microsoft, 2012) unless otherwise stated. Note, depending on the drivers and libraries used for data capture, different raw outputs are streamed by the Kinect.

* Corresponding author.

2. PRELIMINARY TEST RESULTS

An out-of-the-box Kinect system was tested for its distance measurement capabilities. A series of tests were carried out in an indoor environment and the results are presented below.

2.1 Warm-up Test

The distance measurement quality of 3D range cameras (Chiabrando et al., 2009) and laser scanners (Glennie & Lichti, 2011) have been shown to be affected by the warm-up time. In a 23.2°C, 883.9mb, and 36.8% humidity room, a white Spectralon target located approximately 1.1m from the Kinect was observed every 5 minutes over a period of 2 hours. The flat target was nominally orthogonal to the Kinect and a plane was fitted to the point cloud (Figure 1). During the first hour of warming up, the estimated normal distance to the best-fit plane changed by 1 cm. Based on Figure 2, it is advisable to warm-up the Kinect for at least 60 minutes prior to data capture. For all the results shown in this paper, the Kinect was turned on at least 90 minutes prior to data capture.



Figure 1: Data capture of a white planar Spectralon target using the Microsoft Kinect.

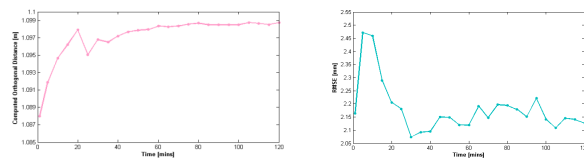


Figure 2: (a) Orthogonal distance to best-fit plane as a function of warm-up time. (b) Depth measurement noise as a function of warm-up time.

2.2 Ambient Light Test

A flat wall was imaged by the Kinect with the room's fluorescent lights turned on and off. This test was repeated with and without a strong light from a desk lamp illuminating the wall. The measured point cloud of the wall with and without white light illumination is shown in Figure 3. The depth measurements appear to be fairly robust against changes in the environment's lighting condition as demonstrated by the similarity between the computed least-squares best-fit plane parameters.

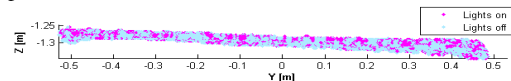


Figure 3: Top view of the point cloud of a wall with the background lights turned on and off.

2.3 Incidence Angle Test

In Soudarissanane et al. (2011) it has been reported that the noise of TOF distance measurements made with a laser escalates with increasing incidence angle. In a similar setup as shown in Figure 1, the Spectralon target was rotated about the vertical axis. The direction of the plane's normal was changed from nominally parallel to the optical axis of the camera to nearly orthogonal. The RMSE of the plane fitting as a function of the incidence angle is reported in Figure 4. Despite the fact that points are projected onto the scene and at large incidence angle they are elongated, through the correlation window

process the effect of the incidence angle on the depth measurement precision appears to be small.

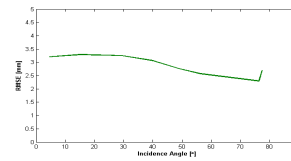


Figure 4: RMSE of plane fitting to planar Spectralon target as a function of the incidence angle

2.4 Radiometric Influence

TOF range measurements by laser scanners have shown dependency on the surface colour (Hanke et al, 2006), i.e. a black surface reflects less energy than a white surface, which results in a range bias. To determine if the Kinect's depth sensor can measure ranges independent of the surface texture, a white metallic plate with a black circle printed in the centre was imaged with the Kinect located at 1 m distance up to 3 m at increments of 0.5 m. The RGB image of the target and the reconstructed point cloud using the depth sensor of the Kinect is shown in Figure 5. The colours in Figure 5b depict the range to the target, and no significant discrepancies to the range measurements can be observed between the black and white regions. The RMSE of the plane fitting at various distances is comparable to the results from imaging a white wall (Figure 10).

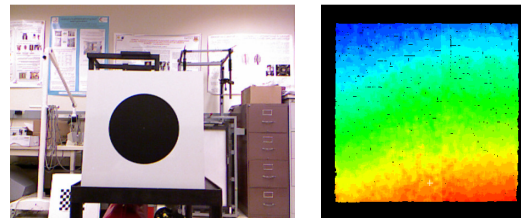


Figure 5: (a) RGB image of a black and white target. (b) Point cloud of the black and white image captured using the Kinect.

2.5 Distance Averaging

To reduce the random noise of depth measurements in TOF laser scanners and range cameras, it is common to take multiple distance observations for each point and use the average (Karel et al., 2010). To learn whether or not distance averaging is necessary, a planar wall located at approximately 1.5 m and 3 m from the Kinect was measured and the RMSE from the plane-fitting is plotted as a function of the number of depth images averaged. From Figure 6 it can be deduced that distance averaging at close-range is probably not necessary as the RMSE is only improved by approximately 0.1 mm even after averaging 100 images. However, at longer ranges such as 3 m, averaging 10 or more depth images improves the RMSE by a millimetre. For longer range applications that do not require 30 frames per second, distance averaging seems to be a viable solution for reducing the random errors in the depth measurements.

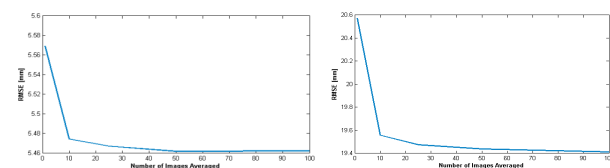


Figure 6: RMSE of plane-fitting as a function of pixel-by-pixel distance averaging at (a) 1.5 m and (b) 3 m

3. APPLICATIONS

To better understand the potential of using this low-cost gaming device for engineering-type applications, an experiment was conducted to explore the achievable accuracy in a common photogrammetric, biomedical, and Geomatics type problem. It is important to note that the Kinect results in this section are performed before any self-calibration.

3.1 3D Human Body Reconstruction

Precise 3D human body measurements are necessary for biometrics, animation, medical science, apparel design, and much more (Loker et al., 2005; Leyvand et al., 2011; and Weise et al., 2011). In this experiment the mannequin shown in Figure 7a was reconstructed using the Leica ScanStation 2 terrestrial laser scanner (Figure 7b), SwissRanger SR3000 TOF range camera (Figure 7c), and the Kinect (Figure 7d). Multiple point clouds were captured around the mannequin and registered using the iterative closest point (ICP) algorithm (Chen & Medioni, 1992) in Leica Cyclone to create a 3D model. By visual inspection, the laser scanner results are the most detailed, features such as the nose and eye sockets are easily identifiable. The results from the SR3000 are greatly degraded because of the internal scattering distortion (Jamtsho & Lichti, 2010). Distance measurements made to objects farther away from the camera are biased by signals returned from closer objects, which cause internal multipath reflection between the lens and CCD/CMOS sensor. Unlike systematic error such as lens distortions, internal scattering changes from scene to scene and can be a challenge to model mathematically. Even after calibrating the IOPs and range errors of the camera (Lichti & Kim, 2011), significant distortions of the mannequin can still be visually identified due to existence of foreground objects. In comparison, the Kinect delivered a more visually elegant model (Figure 7d). Smisek et al. (2011) quantitatively compared the Kinect with the newer SR4000 and also reported the Kinect as more accurate. One of the main reasons is because triangulation-based 3D cameras make direction measurements instead of TOF measurements so it does not experience any scattering. This has been confirmed empirically by the authors. The RMSE of the ICP registration for the Kinect point cloud alone was 3 mm. When compared to the model reconstructed by the ScanStation 2, a RMSE of 11 mm was computed using ICP after registration. In Figure 7e the cyan depicts the model from laser scanning and the pink is from the Kinect. Some systematic deviations between the two models can be observed and are expected because the Kinect has not yet been calibrated. However, the Kinect showed promising results even before calibration in this experiment and should be applicable to a wide range of 3D reconstruction tasks.

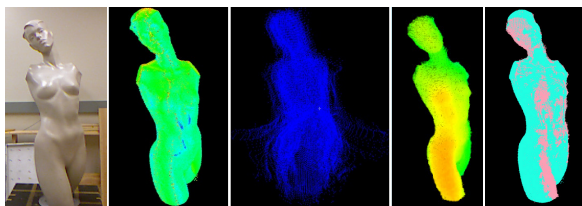


Figure 7: (a) RGB of the mannequin. Point cloud of the mannequin acquired using (b) ScanStation 2, (c) SR3000, and (d) Kinect. (e) ICP registered model from ScanStation 2 (cyan) and Kinect (pink).

4. MATHEMATICAL MODEL

The Kinect is composed of a RGB camera, IR camera, projector, multi-array microphone, and MEMs accelerometer. In this paper, only the optical sensors are considered. Computer

vision based calibrations for the Kinect are widely available (Burrus, 2012). Herrera et al. (2011) used checkerboard targets and a single plane to calibrate the Kinect. The method presented in this paper has a similar concept but instead of a fast computer vision approach, an accuracy-driven photogrammetric approach is taken. The IR camera and projector together form the depth sensor which gives distance information for every pixel in the IR camera. The calibration for the RGB camera can follow the conventional collinearity model shown in Equation 1. The EOPs, IOPs, and object space coordinates of the signalised targets can be determined simultaneously using least-squares adjustment. To model the systematic errors in both cameras and projector, Brown's distortion model (Brown, 1971) has been adopted (Equation 2).

$$\begin{aligned} x_{ij} &= x_p - c \frac{m_{11}(X_i - X_{oj}) + m_{12}(Y_i - Y_{oj}) + m_{13}(Z_i - Z_{oj})}{m_{31}(X_i - X_{oj}) + m_{32}(Y_i - Y_{oj}) + m_{33}(Z_i - Z_{oj})} + \Delta x \\ y_{ij} &= y_p - c \frac{m_{21}(X_i - X_{oj}) + m_{22}(Y_i - Y_{oj}) + m_{23}(Z_i - Z_{oj})}{m_{31}(X_i - X_{oj}) + m_{32}(Y_i - Y_{oj}) + m_{33}(Z_i - Z_{oj})} + \Delta y \end{aligned} \quad (1)$$

Where x_{ij} & y_{ij} are the image coordinate observations of point i in image j
 x_p & y_p are the principal point offsets in the x and y directions
 c is the principal point distance
 $X_i, Y_i, & Z_i$ are the object space coordinates of point i
 $X_{oj}, Y_{oj}, & Z_{oj}$ are the position of image j in object space
 $m_{11} \dots m_{33}$ are the elements of the rotation matrix (R) describing the orientation of image j
 Δx & Δy are the additional calibration parameters

$$\begin{aligned} \Delta x &= x'_{ij} (k_1 r_{ij}^2 + k_2 r_{ij}^4 + k_3 r_{ij}^6) + p_1 (r_{ij}^2 + 2x'_{ij})^2 + 2p_2 x'_{ij} y'_{ij} + a_1 x'_{ij} + a_2 y'_{ij} \\ \Delta y &= y'_{ij} (k_1 r_{ij}^2 + k_2 r_{ij}^4 + k_3 r_{ij}^6) + p_2 (r_{ij}^2 + 2y'_{ij})^2 + 2p_1 x'_{ij} y'_{ij} \end{aligned} \quad (2)$$

Where $k_1, k_2, & k_3$ describes the radial lens distortion
 p_1 & p_2 describes the decentring lens distortion
 a_1, a_2 describes the affinity and shear
 r_{ij} is the radial distance of point i in image j referenced to the principal point
 x'_{ij} & y'_{ij} are the image coordinates of point i in image j after correcting for the principal point offset

Co-registration of the RGB image and depth image is necessary to colourize the point cloud. This involves solving for the translational and rotational offsets (a.k.a. leverarm and boresight parameters) of the two cameras as well as their IOPs. Since the RGB camera and IR camera are rigidly mounted on the same platform, a boresight and leverarm constraint can be applied to strengthen the bundle adjustment (Equation 3).

$$\begin{bmatrix} \bar{x}_i - \Delta x \\ \bar{y}_i - \Delta y \\ -c \end{bmatrix}_{RGB} - \frac{1}{\mu_i} \mathbf{R}_{RGB}^R \mathbf{R}_{IR}^{R} \mathbf{R}_{Map} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}_{RGB} - \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix}_{IR}^{Map} - \mathbf{R}_{IR}^{Map} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}_{IR}^R = 0 \quad (3)$$

Where $b_x, b_y, & b_z$ are the leverarm parameters
 μ_i is the unique scale factor for point i
 R is the 3D rotation matrix

The depth sensor can also be modelled using the collinearity equations. However, the raw output from the depth sensor when using the Microsoft SDK is Z_i at every pixel location of the IR camera. This is fundamentally different from most Kinect error models published to date, which use open source drivers (e.g. libfreenect and OpenNI) that stream disparity values as 11 bit integers. Unlike in Menna et al. (2011) and Khoshelham & Oude Elberink (2012) whose observations are disparity values and the depth calibration is performed by solving the slope and bias of a linear mapping function that relates depth to disparity, an alternative calibration method suitable for the Microsoft SDK is proposed in this paper.

Since the Microsoft SDK does not give access to the IR image, traditional point-based calibration where signalized targets are observed is not applicable. Instead, a plane-based calibration

following the collinearity equation is adopted for calibrating the depth sensor. Since the Z coordinate of every point is provided, the corresponding X and Y coordinates can be computed in the IR camera's coordinate system based on the relationship shown in Equation 4. For the IR camera, image observations for each point are taken as the centre of every pixel. The IR camera uses the Aptina MT9M001 monochrome CMOS sensor, which has a nominal focal length of 6 mm and pixel size of 5.2 μm . The effective array size after 2x2 binning and cropping is 640 by 480 pixels with a pixel size of 10.4 μm .

$$X_i^{\text{IR}} = \frac{Z_i^{\text{IR}}}{c_{\text{IR}}} X_i^{\text{IR}}, \quad Y_i^{\text{IR}} = \frac{Z_i^{\text{IR}}}{c_{\text{IR}}} y_i^{\text{IR}} \quad (4)$$

Although the projector cannot "see" the image, it still obeys the collinearity condition and can be modelled as a reverse camera. It is assumed for every point $[X_i, Y_i, Z_i]_{\text{IR}}$ there is a corresponding $[x_{ij}, y_{ij}]_{\text{IR}}$ and $[x_{ij}, y_{ij}]_{\text{Pro}}$. The image coordinate observations can be determined by back-projecting the object space coordinates $[X_i, Y_i, Z_i]_{\text{Pro}}$ (which is equivalent to $[X_i, Y_i, Z_i]_{\text{IR}}$) into the projector's image plane using Equation 1. Very limited documentation about the projector is publically available so it is assumed that the projector has the same properties as the MT9M001 sensor. It is important to note that this assumption should not have major impacts on the effectiveness of the calibration. With the boresight and leverarm expressed relative to the IR camera, the functional model for calibrating the depth sensor is given in Equation 5.

$$\mu_{ij} \mathbf{R}_{\text{IR}}^{\text{Map}} \mathbf{R}_{\text{Pro}}^{\text{Map}} \begin{bmatrix} \bar{x}_i - \Delta x \\ \bar{y}_i - \Delta y \\ -c \end{bmatrix}_{\text{Pro}} + \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix}_{\text{IR}}^{\text{Map}} + \mathbf{R}_{\text{IR}}^{\text{Map}} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}_{\text{Pro}}^{\text{IR}} - \left(\mu_{ij} \mathbf{R}_{\text{IR}}^{\text{Map}} \begin{bmatrix} \bar{x}_i - \Delta x \\ \bar{y}_i - \Delta y \\ -c \end{bmatrix}_{\text{IR}} + \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix}_{\text{IR}}^{\text{Map}} \right) = 0 \quad (5)$$

Instead of solving for the scale factor μ_{ij} for every point, it is expressed as a function of the plane parameters ($a_k, b_k, c_k,$ and d_k), EOPs of the IR camera, IOPs of the applicable optical sensors, and the boresight and leverarm offsets for the RGB camera and projector. The functional model for the unique scale factor can be determined by solving for μ_{ij} in Equation 6, where X_i, Y_i, Z_i is defined by Equation 3.

$$[a_k \quad b_k \quad c_k] \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} - d_k = 0 \quad (6)$$

The proposed functional model minimizes the discrepancy of conjugate light rays at each tie point location while constraining every point reconstructed by the RGB camera and depth sensor to lie on the best-fit plane. Unlike most existing Kinect calibrations where the depth is calibrated independently of the bundle adjustment process, the proposed method solves for the EOPs, IOPs, boresights, leverarms, object space coordinates of targets measured by the RGB camera, and the plane parameters simultaneously in a combined least-squares adjustment model. The proposed method also takes into consideration the fact that the output depth/disparity values are a function of lens distortion. This mathematical model is kept as general as possible and should be applicable for other triangulation-based 3D cameras (e.g. Asus Xtion) and camera-projector systems.

5. EXPERIMENTAL RESULTS AND ANALYSES

5.1 Kinect Self-calibration

Two experiments were conducted to calibrate the Kinect. In the first calibration a single texturized plane was used for the calibration. In the second calibration, three roughly orthogonal planes were utilized. In both calibrations the datum was defined using inner constraints on the object space target coordinates and Baarda's data snooping was adopted. In the

first experiment a checkerboard pattern was projected onto a flat wall to provide some targets for calibrating the RGB camera (Figure 8a). If only the depth sensor needs to be calibrated, homogenous flat surfaces will suffice. Multiple convergent images were captured from different positions and orientations (Figure 8b) while ensuring the target field covers the majority of the image format (Fraser, 2012).



Figure 8: (a) Experimental design for calibrating the Kinect. (b) Network configuration of the calibration.

The corner points of the checkerboard pattern were extracted using the computer vision camera calibration toolbox from Caltech (Bouguet, 2010). Fifty depth images were averaged at each exposure station and points from the wall were semi-automatically extracted. Some statistics about this calibration are summarized in Table 1. It was initially assumed that relative to the IR camera, the projector has no rotational offsets and is located 7.5 cm away in the positive x-direction of the IR sensor, while the RGB camera is 2.5 cm away and has zero rotational offsets. The leverarm and boresight parameters determined in this adjustment are given in Table 2. To quantify the effectiveness of the calibration for the depth sensor, the misclosure of light rays (computed using Equation 5) at every object point used in the calibration was computed. From Table 3, it can be observed that after self-calibration the quality of fit between conjugate light rays are significantly improved.

| Parameter | Value |
|-------------------------|-------|
| # of exposure stations | 12 |
| # of unknowns | 232 |
| # of signalized targets | 48 |
| # of observations | 3456 |
| Average Redundancy | 0.94 |

| | IR-Projector | | IR-RGB | |
|----------------|----------------------------|----------------------|----------------------------|----------------------|
| | Value [$^{\circ}$ & m] | σ [" & mm] | Value [$^{\circ}$ & m] | σ [" & mm] |
| $\Delta\omega$ | -0.001 | 16 | -0.865 | 542 |
| $\Delta\phi$ | 0.018 | 34 | 0.071 | 560 |
| $\Delta\kappa$ | 0.003 | 19 | -0.959 | 1033 |
| Δb_x | 0.074 | 0.6 | 0.031 | 4.4 |
| Δb_y | 0.000 | 0.1 | 0.028 | 4.8 |
| Δb_z | 0.000 | 0.6 | 0.022 | 7.0 |

| | Before Calibration [mm] | After Calibration [mm] | % Improvement |
|-------------------|-------------------------------|------------------------------|---------------|
| RMSE _X | 0.56 | 0.18 | 68 |
| RMSE _Y | 0.88 | 0.32 | 64 |
| RMSE _Z | 0.02 | 0.01 | 64 |

From the single plane calibration, the estimation of the rotational and translational offsets between the IR camera and

RGB camera was quite poor. One of the biggest sources of uncertainty arises from the weak determination of the IR camera's EOPs since only one plane was deployed. In the second experiment, three differently oriented planes were used to calibrate the Kinect (Figure 9). A brief summary of the least-squares adjustment, the recovered boresights and leverarms, and the RMSE of the misclosure vectors are documented in Tables 4, 5, and 6, respectively. By adding two additional planes, even with fewer images, the precision of the 3D rigid body transformation parameters relating the IR and RGB camera is significantly improved.

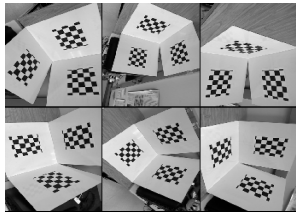


Figure 9: Images used for the Kinect multi-plane calibration

| Parameter | Value |
|-------------------------|-------|
| # of exposure stations | 6 |
| # of unknowns | 275 |
| # of signalized targets | 72 |
| # of observations | 2368 |
| Average Redundancy | 0.89 |

| | IR-Projector | | IR-RGB | |
|----------------|------------------|----------------------|------------------|----------------------|
| | Value [° & m] | σ [" & mm] | Value [° & m] | σ [" & mm] |
| $\Delta\omega$ | -0.002 | 29 | -0.775 | 362 |
| $\Delta\phi$ | -0.006 | 38 | -0.119 | 312 |
| $\Delta\kappa$ | -0.003 | 14 | -0.359 | 207 |
| Δb_x | 0.080 | 0.3 | 0.017 | 1.3 |
| Δb_y | 0.000 | 0.1 | -0.008 | 1.4 |
| Δb_z | 0.000 | 0.1 | 0.007 | 2.7 |

| | Before Calibration [mm] | After Calibration [mm] | % Improvement |
|-------------------|-------------------------------|------------------------------|---------------|
| RMSE _X | 0.49 | 0.37 | 23 |
| RMSE _Y | 0.42 | 0.24 | 41 |
| RMSE _Z | 0.62 | 0.29 | 53 |

Due to high correlations in both calibrations, only the IOPs of the RGB camera can be recovered at this point (Table 7). It is worth mentioning that the recovered y_p parameter from the first and second experiment corresponds to a shift of 31 and 32 pixels, respectively. This is similar to the empirical values reported by Khoshelham & Oude Elberink (2012) and the theoretical value of 32 pixels due to image cropping. But very different x_p and y_p values were reported in Menna et al. (2011).

Future work will attempt to estimate the IOPs of the IR camera either through stronger network configuration or by gaining access to the IR images using open source software such as RGBDemo (Burrus, 2012). As shown in Smisek et al. (2011), under certain illumination conditions the IR camera can observe signalised targets that are visible to the RGB camera. In that case, the conventional bundle adjustment with self-calibration can be applied to calibrate the IR camera. This should also

improve the boresight and leverarm estimation between the IR camera and RGB camera.

| | Single Plane | | Multi-Plane | |
|---------------------------|--------------|----------|-------------|----------|
| | Value | σ | Value | σ |
| x_p [mm] | 0.06 | 0.005 | 0.02 | 0.002 |
| y_p [mm] | -0.18 | 0.005 | -0.17 | 0.002 |
| c [mm] | 2.97 | 0.008 | 2.95 | 0.004 |
| k_1 [mm ⁻²] | 1.71e-2 | 4.47e-4 | 1.62e-2 | 4.04e-4 |
| k_2 [mm ⁻⁴] | -3.03e-3 | 1.14e-4 | -3.15e-3 | 1.03e-4 |

5.2 Precision versus depth

To verify the proposed sensor error model for calibrating the depth sensor, an independent experiment was made. In a 20' by 40' by 20' racquetball court under stable and controlled atmospheric conditions (20.0°C, 884.8mb and 48.8% humidity) a flat wall was observed with an uncalibrated Kinect situated at 1 m distance up to 10 m at increments of 0.5 m using the Brekel Kinect software. The RMSE from the least-squares plane fitting is shown in Figure 10 as the purple curve. Based on the proposed mathematical model in this paper, the same experiment was simulated with 100 well distributed gridded points observed on each plane and no systematic errors. It was assumed the IR camera and projector has a nominal baseline of 7.5 cm. Using Variance Component Estimations in the multi-plane self-calibration an image measurement precision of 0.8 μ m (1/13 pixel) was determined for the IR camera. Using these values, the RMSE of the fitted planes from the simulated data were computed and they follow closely the experimental depth error behaviour (orange curve in Figure 10). This depth error behaviour is similar to those reported in Khoshelham (2011).

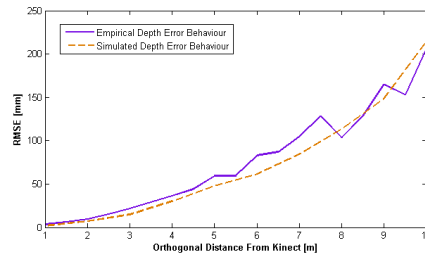


Figure 10: Empirical and theoretical error behaviour for the depth measurements of the Kinect.

6. CONCLUSION & FUTURE WORK

The Kinect system has demonstrated potential to be used for engineering applications. At long-range it may not be as accurate as a terrestrial laser scanner, photogrammetric system, or structured light system, but for the cost and portability it is delivering fairly high geometric accuracy at close-range. The Kinect was used to capture point clouds in a series of tests. Since this is a triangulation-based system, it does not exhibit data distortions that would otherwise be identifiable in TOF systems, of which the most pronounced range error is the internal scattering experienced by most TOF 3D range cameras. In one of the 3D reconstruction experiments, the uncalibrated Kinect system was capable of modelling a mannequin much more accurately than a calibrated SR3000. Before this gaming device should be used for any metric type applications the systematic errors should be properly modelled. A new calibration routine is presented in this paper which models the IR camera, RGB camera, and projector using the collinearity equation with boresight, leverarm, and point-on-plane

constraints. Although the IOPs of the IR camera cannot be recovered at this point, by updating the boresight and leverarm parameters the volume of uncertainty for each point determined by the depth sensor was improved significantly. Future calibrations will try to incorporate IR images of some target field to improve the IOPs, boresight, and leverarm observability.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, Alberta Innovates, Informatics Circle of Research Excellence, Terramatic Technologies Inc., and SarPoint Engineering Ltd. for funding this research. The authors are also grateful for Tammy Smith and Justin Waghray for assisting with data capture and processing the SR3000 and ScanStation 2 point clouds.

REFERENCES

- Bouguet, J. (2010, July 9). *Camera Calibration Toolbox for Matlab*. Retrieved March 30, 2012, from Computational Vision at CALTECH: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html
- Brekelmans, J. (2012). *Brekel Kinect*. Retrieved March 30, 2012, from http://www.brekel.com/?page_id=155
- Brown, D. (1971). Close-range camera calibration. *Photogrammetric Engineering*, 37 (8), 855-866.
- Burrus, N. (2012). *RGBDemo*. Retrieved March 30, 2012, from Manctl: <http://labs.manctl.com/rgbdemo/>
- Chen, Y., & Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, 145-155.
- Chiabrando, F., Chiabrando, R., Piatti, D., & Rinaudo, F. (2009). Sensors for 3D imaging: metric evaluation and calibration of a CCD/CMOS time-of-flight camera. *Sensors* (9), 10080-10096.
- Fraser, C. (2012). Automatic camera calibration in close-range photogrammetry. *ASPRS 2012 Annual Conference*. Sacramento, USA.
- Freedman, B., Shpunt, A., Machline, M., & Arieli, Y. (2010). *Patent No. US2010/0018123 A1*. United States of America.
- Glennie, C., & Lichti, D. (2011). Temporal stability of the Velodyne HDL-64E S2 scanner for high accuracy scanning applications. *Remote Sensing* 3(3), 539-553.
- Hanke, K., Grussenmeyer, P., Grimm-Pitzinger, A., & Weinold, T. (2006). First experiences with the Trimble GX scanner. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 5* (pp. 1682-1759). Dresden, Germany, Sept. 25-27: ISPRS Comm. V Symposium.
- Herrera, C., Kannala, J., & Heikkilä, J. (2011). Accurate and practical calibration of a depth and color camera pair. *14th International Conference on Computer Analysis of Images and Patterns*, (pp. 437-445). Seville, Spain.
- Jamtsho, S., & Lichti, D. (2010). Modeling scattering distortion of 3D range camera. *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences, XXXVIII (5)*, (pp. 299-304).
- Kahlmann, T., Remondino, H., & Ingensand, H. (2006). Calibration for increased accuracy of the range imaging camera SwissRangeeTM. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35 (5), 136-141.
- Karel, W., Ghuffar, S., & Pfeifer, N. (2010). Quantifying the distortion of distance observations caused by scattering in time-of-flight range cameras. *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences, XXXVIII (5)*, (pp. 316-321). Newcastle upon Tyne, UK.
- Khoshelham, K. (2011). Accuracy analysis of kinect depth data. *ISPRS Laser Scanning Workshop*. Calgary, Canada.
- Khoshelham, K., & Oude Elberink, S. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, vol. 12, 1437-1454.
- Konolige, K., & Mihelich, P. (2010, December 9). *Kinect Calibration: Technical*. Retrieved March 30, 2012, from Robot Operating System: http://www.ros.org/wiki/kinect_calibration/technical
- Lange, R., & Seitz, P. (2001). Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37 (3), 390-397.
- Leyvand, T., Meekhof, C., Yi-Chen, W., Jian, S., & Baining, G. (2011). Kinect identity: technology and experience. *Computer*, vol. 44, 94-96.
- Lichti, D., & Kim, C. (2011). A comparison of three geometric self-calibration methods for range cameras. *Remote Sensing*, vol 3 (5), 1014-1028.
- Loker, S., Ashdown, S., & Schoenfelder, K. (2005). Size specific analysis of body scan data to improve apparel fit. *Textile and Apparel Technology Management*, vol 4 (3), 103-120.
- Menna, F., Remondino, F., Battisti, R., & Nocerino, E. (2011). Geometric investigation of a gaming active device. *Videometrics, Range Imaging, and Applications XI*. Munich, Germany: SPIE Optical Metrology.
- Microsoft. (2012). *Kinect for Windows*. Retrieved March 30, 2012, from Microsoft: <http://www.microsoft.com/en-us/kinectforwindows/>
- Mure-Dubois, J., & Hugli, H. (2007). Optimized scattering compensation for time-of-flight camera. *Proc. SPIE: Two- and Three-Dimensional Methods for Inspection and Metrology V*, 6762, 67620H, 1-10.
- Shan, J., & Toth, C. (2008). *Topographic Laser Ranging and Scanning: Principles and Processing*. Boca Raton, Florida: CRC Press.
- Smisek, J., Jancosek, M., & Pajdla, T. (2011). 3D with kinect. *Consumer Depth Cameras for Computer Vision*, (pp. 1154-1160). Barcelona, Spain, November 12.
- Soudarissanane, S., Lindenbergh, R., Menenti, M., & Teunissen, P. (2011). Scanning geometry: Influencing factor on the quality of terrestrial laser scanning points. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(4), 389-399.
- Weise, T., Bouaziz, S., Li, H., & Pauly, M. (2011). Real time performance-based facial animation. *ACM Transactions on Graphics, Proceedings of the 38th ACM SIGGRAPH Conference*, vol 30 (4). Los Angeles, CA.