CALIBRATING THE MS KINECT SENSOR

Charles K. Toth[a], Senior Research Scientist
Bence Molnar[b], PhD Candidate
Andrew Zaydak[a], PhD Student
Dorota A. Grejner-Brzezinska[a], Professor
[a]The Ohio State University
Columbus, OH 43210
[b]Budapest University of Technology and Economics
toth@cfm.ohio-state.edu

ABSTRACT

Flash LiDAR systems have been around for several years, but their use was limited, mainly due to their poor performance. The first generation products had range limitations and were quite sensitive to ambient light conditions, basically restricting their use to indoor applications. In addition, the high-noise level provided for very low accuracy and stability. As technology advanced, the Flash LiDAR sensors have shown gradual improvements in performance, resulting in an increasing number of applications. One of the most recently introduced inexpensive sensors is the Microsoft Kinect, originally designed to support gaming. Despite its low cost, the Kinect is very powerful, providing a relatively dense point cloud and high-definition video at high data rate. This paper reports about our investigation with this sensor, including calibration experiences and performance evaluation.

KEYWORDS: Flash LiDAR, point cloud, calibration

INTRODUCTION

Laserscanning systems have seen enormous growth in the geospatial data acquisition industry in the past decade, mainly because of the excellent performance offered by this technology and the explicitness of the 3D data; a review on LiDAR technology can be found in (Shen and Toth, 2008) and on advanced data processing in (Vosselman and Maas, 2010). First, airborne LiDAR, also called ALS (airborne laserscanning), was introduced in the late 90's. Then, terrestrial laserscanner (TLS) followed suite in the surveying community, and, finally, Mobile LiDAR, also called MLS (mobile laserscanning) arrived just a few years ago, and has seen phenomenal developments in recent years. All the sensor technologies used in these airborne and terrestrial applications are predominantly based on pulsed LiDAR and a smaller fraction is on continuous waveform (CW) LiDAR.

Flash LiDAR, also called Flash LADAR, is a substantially different sensing technology compared to pulsed and CW LiDAR techniques, as it is based on an sensor array, so a it can capture a whole 3D, also called depth or range, image with intensity data in a single step. Falsh LiDAR can use both basic solutions to emit laser, either a single pulse with large aperture will "flash" the area for a short time or in CW mode a continuous laser "light" provides steady illumination in the area. One of the first and early Flash LiDAR model, the SWR3000 (Kahlmann, 2006) is based on CW approach, offering an operating range up to 7.5 m and a frame rate of 15 Hz. All the Flash LiDAR systems are based on solid state semiconductor technologies, and there is a large variety of solutions and, consequently, operating parameters. Advanced modern systems can reach the 1,500 m range, which makes them deployable on airborne vehicles. Because of their smaller size, and less power requirements, they are attractive for mobile platforms. Flash LiDAR technology goes back to the late 90's, but started to reach maturity just recently. The issues are the limited power that should illuminate the area and the complex circuitry, typically avalanche photodiode detector (APD), needed to detect the few photons, backscattered from objects.

There are several Flash LiDAR systems in low-end category, while professional-grade systems are mainly manufactured by two companies, Advanced Scientific Concept, Inc., offering three products (ASC), and Raytheon Vision Systems, mainly focused on sensor developments (Bailey *et al.*, 2010).

KINECT SENSOR

The Kinect sensor is a motion sensing input device for the Xbox 360 video game console, originally developed by PrimeSense (PrimeSense), and acquired by Microsoft. The primary purpose is to enable users to control and interact with the Xbox 360 through a natural user interface using gestures and spoken commands without the need to touch a game controller at all. The Kinect has three primary sensors: a Flash LiDAR (3D camera), a conventional optical RGB sensor (2D camera), and microphone array input. The device is USB-interfaced, similar to a webcam, and appears as a “black box” for the users.

Very little is known of the sensors, internal components and processing methods stored in the firmware. The laser, IR, emitter projects a structured light pattern of random points to support 3D recovery. The 2D camera can acquire standard VGA, 640x480, and SXGA, 1280x1024, images at 30 Hz. The color formation is based on Bayer filter solution, transmitted in 32-bit and formatted in the sRGB color space. The FOV of the 2D camera is 57° x 43°. The 3D camera can work in two resolutions with frame sizes of 640x480 and 320x240. The range data comes in 12-bit resolution. The sensors’ spatial relationship is shown in Figure 1. The approximate difference between the laser emitter and detector that form a stereo par is about 7.96 cm, and the baseline between the 2D and 3D cameras is about 2.5 cm.

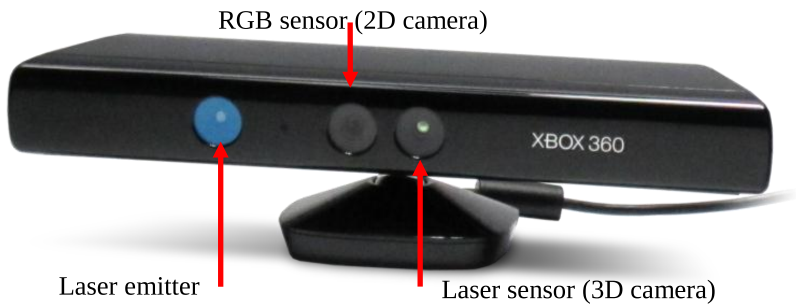
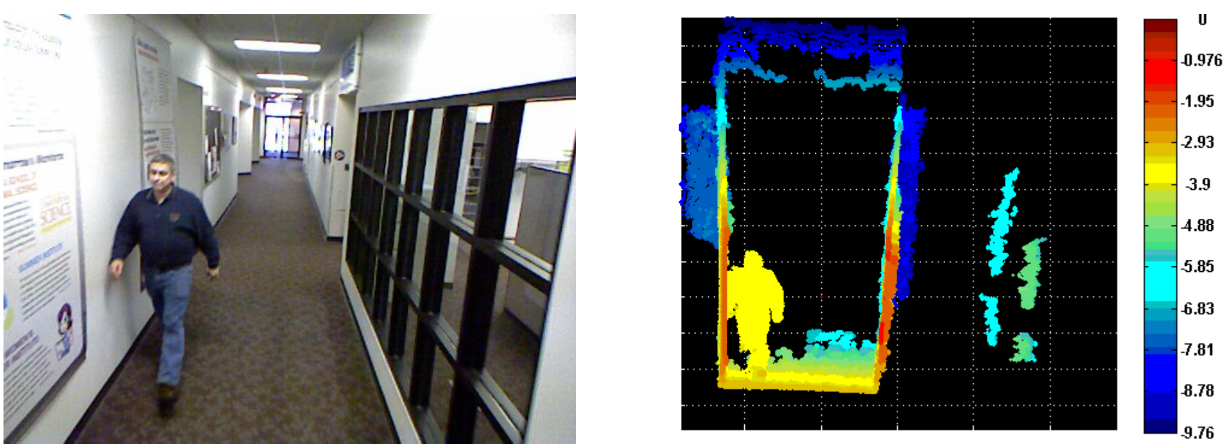


Figure 1. Kinect XBOX 3600 sensor, including 2D and 3D imaging sensors.



(a) (b)

Figure 2. Kinect 2D and 3D images.

Microsoft provides an SDK (Windows 7, Visual Studio 2010 Enterprise, and DirectX) to support application developments, including both polled and event-based access to the image data streams (Microsoft). In addition, skeletal tracking of up to two people is also supported. Kinect has a default measuring range of 0.8 m and to 3.9m (no ambiguity), which can be extended; our experiences indicate that up to 10 m range, reliable depth images can be acquired. The available open source drivers provide additional the opportunity to acquire raw data and a very powerful SDK is also available. In our investigation the SensorKinect driver (Github) was used with OpenNI (OpenNI) and all the subsequent processing was done in Visual Studio C++ and Matlab. A typical 2D and 3D image pair is shown in Figure 2.

CALIBRATION EXPERIENCES

In any mapping applications, the imaging sensors must be adequately characterized by a calibration process in order to achieve an optimal performance. During this process the sensor model parameters are estimated and the stochastic behavior of the sensed signal is analyzed. Based on the calibration parameter estimation, biases are removed, and the error budget of the sensor can be determined based on the statistical terms. The difference between 2D and 3D imaging sensor calibration methods is not really much, as in both cases, data is collected using reference scenes, which are based on a structured object space of known geometry and additional information, and then sensor model parameters are adjusted. Compared to conventional camera calibration, the characterization of the range measurement represents an additional task. Note that the projection of the optical system is identical in both cases, and, thus, requires the very well-established process of digital camera calibration. Though the Kinect sensor was exclusively developed for gaming, interest in many disciplines quickly started to rise and support tools as well as calibration/application were developed in a short time; (Khoshelham, 2011) is one of the earlier attempts to calibrate the Kinect sensor.

Our investigation is mainly focused only on the range calibration, as the depth accuracy is the critical aspect of our applications, including indoor mapping and navigation . In particular, the ranging accuracy dependence on the range is of particular interest. In the following, experiences obtained by calibrating a Kinect sensor are discussed.

Sensor Calibration

There are many tools available from the Kinect user's community to perform camera calibration. In general, the computer vision model, a calibration matrix-based linear model, is used, which is more than satisfactory for most of the applications. The processing itself is highly automated, and requires almost no user interaction. Figure 3a shows an image with the widely-used calibration target, and Figure 3b is an image where the intersection points of the regular grid are already extracted and marked. The data acquisition process follows the standard practice of acquiring images from different locations, at different angles, and different rotations around the camera optical axis.

(a) (b)

Figure 3. The number of detected points on target plane as a function of the object range.

The calibration results are specific to the selected resolution (high or low), and can be directly applied to the raw sensor data to obtain both 2D and 3D images. In addition to the geometrical modeling, the relative orientation of the sensors is also estimated in the calibration process. Note that range calibration allows only for a scale factor, as the sensor specific depth calibration is performed during manufacturing and is permanently stored in the sensor firmware.

Sensor Repeatability Test

The repeatability of the range measurement is an essential aspect of using depth imaging sensors, as it provides the assessment of the ranging precision in short term. To determine the sensor repeatability performance, a planar target was imaged from a distance ranging from 0.5 m to 5 m in 0.1 m steps. Figure 4 shows 3D (depth) images of the target from two different distances.

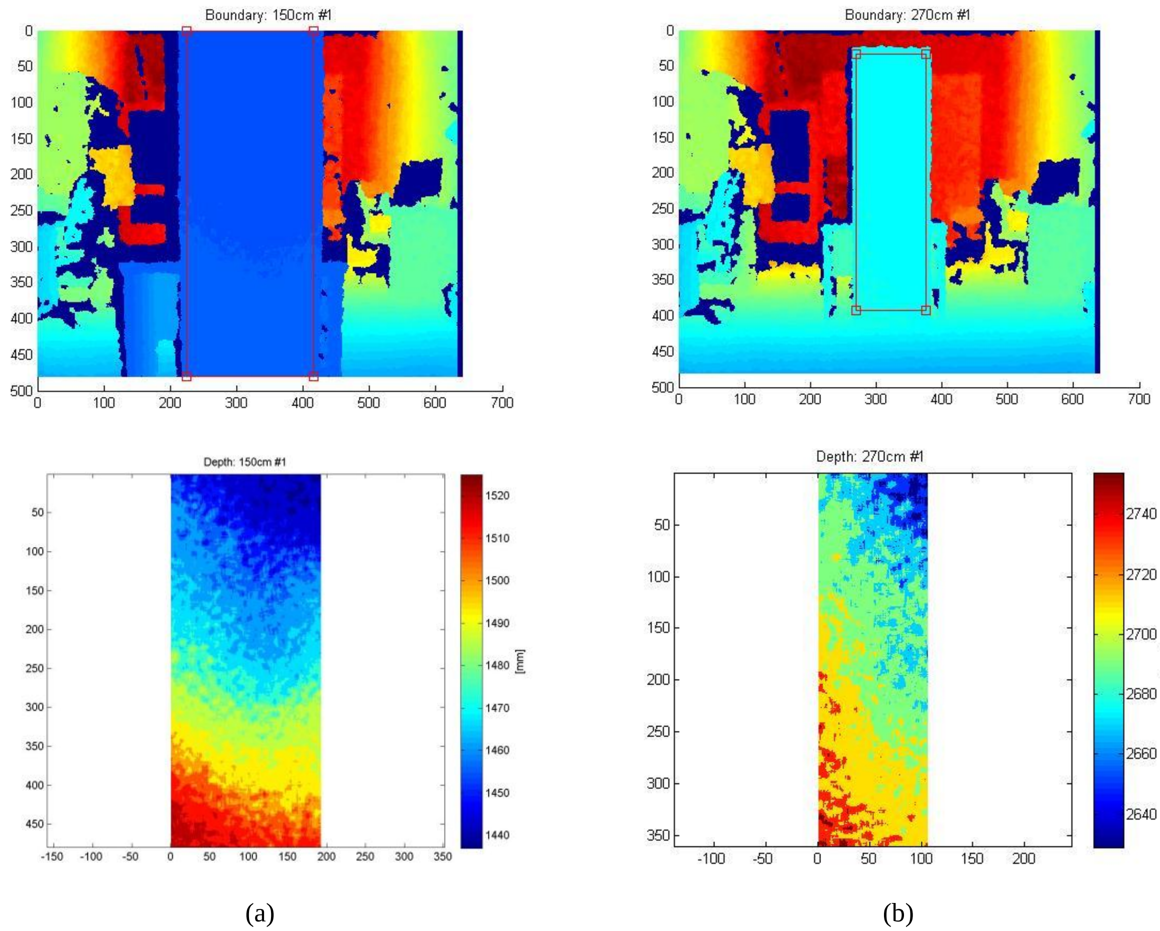


Figure 4. Pseudo color 3D images taken at 150 (a) and 270 (b) cm ranges; first row entire images and second raw images of the planar target extracted.

The measurement was repeated six times, so a total of 46 x 6 images were acquired and processed. The planar target has a size of 180 cm x 60 cm, so it's FOV in the image changes a lot. Consequently, the number of points obtained by the 3D sensor from the reference planar target varies over a large range, from 200K down to 10K, as shown in Figure 5.

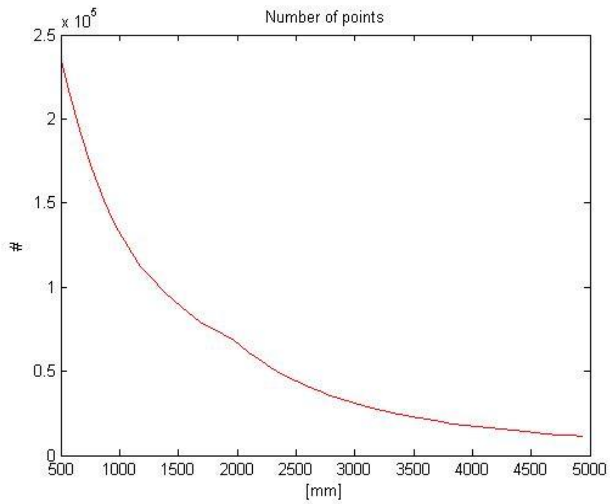
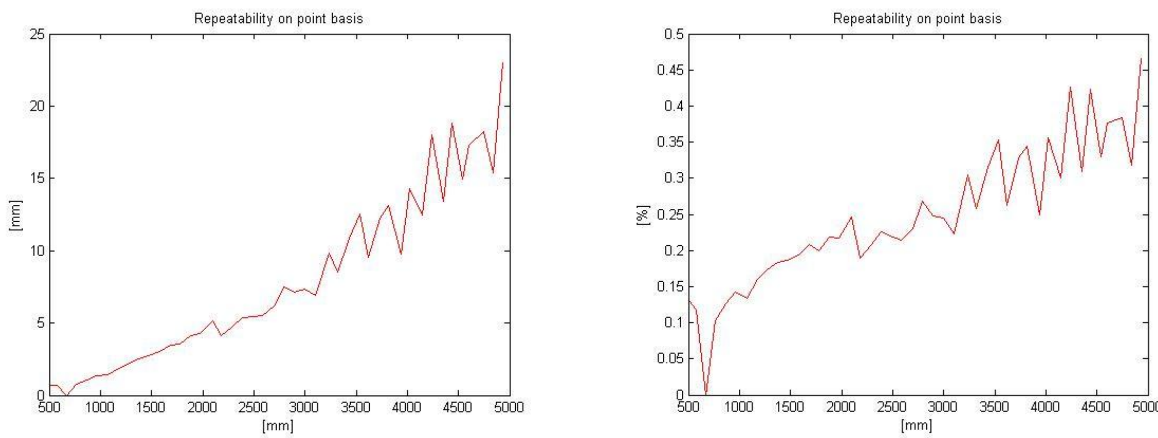


Figure 5. The number of detected points on target plane as a function of range.

In the first step, the standard deviation was computed on a point basis for each distance. The repeatability results, shown in Figure 6, clearly indicate a near linear dependency on the range. The overall performance for the whole range is lower than 0.5%, which is quite excellent compared to earlier Flash LiDAR results (Kahlmann *et al.*, 2006).



(a) (b)

Figure 6. Repeatability results; (a) absolute and (b) relative performance.

To achieve a better performance characterization, plane fitting was performed based on principal point component analysis and the fitting error was calculated in the plane normal direction. The fitting error, shown in Figure 7, has an interesting shape, as the curve has a local maximum in the central part of the range, near the ambiguity range. There is no obvious explanation for this character, except that the decreasing number of points can result in improving fits. In addition, this curve, as well as all error curves, looks sort of jagged or “discontinuous” which could be partially associated with rounding up the numbers during internal processing of the sensor firmware.

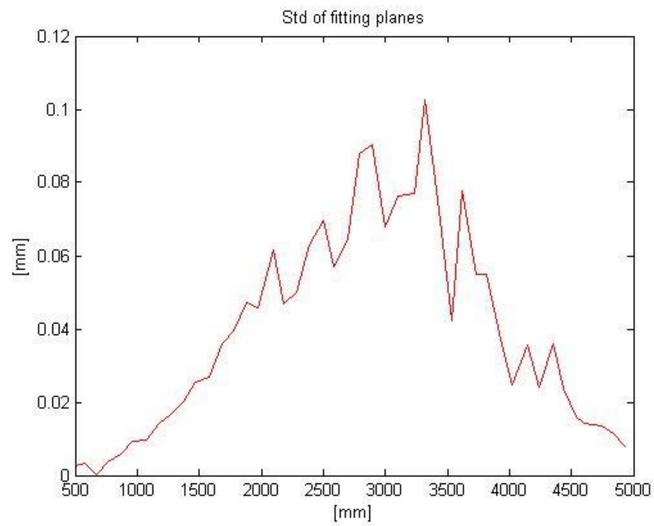
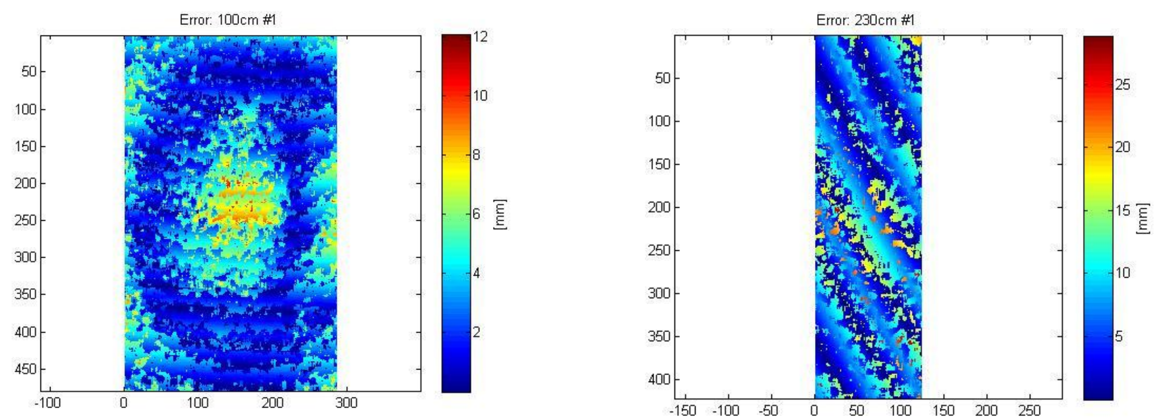


Figure 7. Standard deviation of plane fitting error in surface normal direction.

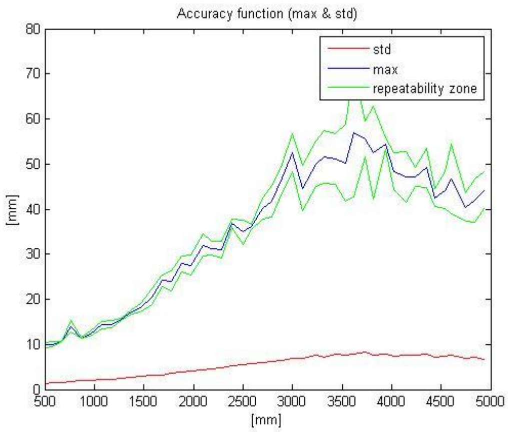
Figure 8 shows 2D error surfaces at two ranges, 1 m and 2.3 m. While the overall residual error numbers are small, their spatial distribution is somewhat unusual. Note that the circular pattern is caused by the distance calculation method, as described in (Khoshelham, 2011).



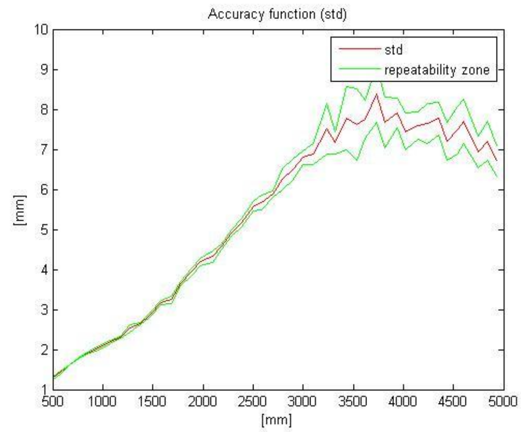
(a) (b)

Figure 8. Distribution of fitting errors at ranges (a) 1 m and (b) 2.3 m.

Based on the six measurement sets, the fitting plane residual errors were calculated and basic statistical parameters were determined, including maximums and STD, for each range. Figure 9 shows the results, including a maximum error envelope and the STD (9a) as well as well as only the STD with error envelope (9b). The results clearly indicate good accuracy performance, as at the shortest object distance, the STD is lower than 1 mm and the maximum error is 1 cm, while at 3.5 m (the ambiguity limit) the STD is 7 mm and the maximum is 5 cm. Theoretically, the STD function should be of quadratic form based on the used calculation method, yet the curve looks almost linear. Normalized for the range, the STD is about 0.2% of distance while the maximum error is about 1.6%, as shown in Figure 10.



(a)



(b)

Figure 9. STD of residual surface fitting errors.

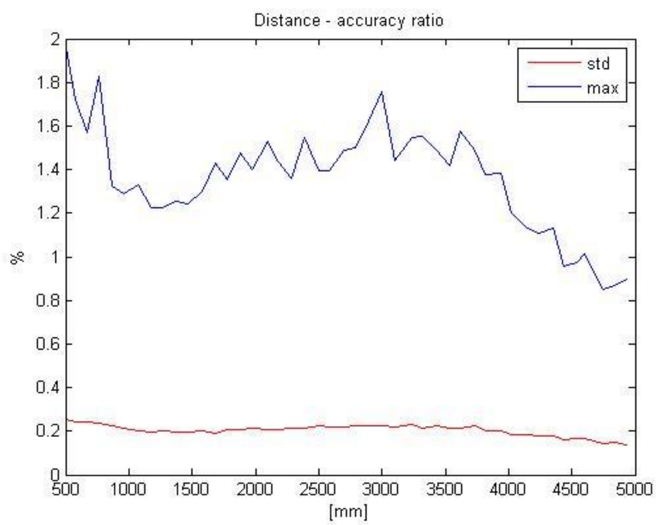


Figure 10. Normalized statistical parameters.

Application Performance Test

A Kinect sensor was integrated into a personal navigator (PN) sensor suite that was developed at The Ohio State University in late 2011 to support indoor mobile mapping and navigation research. The sensor configuration of the system is shown in Figure 11. The calibrated Kinect sensor has been used in several surveys, carried out in various buildings in December 2011, where it collected both 2D and 3D images. Typically, building corridors and to less extent rooms were mapped, see Figure 2. The objective of the data acquisition is to provide a rich multisensory data set for object space reconstruction and/or based on that navigating the sensor platform. The concept is mainly based on integrating 2D and 3D imagery to reconstruct the object space in both geometry and classifying objects. Though, the processing of data and the algorithmic developments will likely take longer time, some initial processing results are already available for the Kinect sensor. Based on 3D image sequences acquired in passing across a building, the trajectory of the Kinect sensor was reconstructed and compared to the floor blueprint. In this survey, the PN entered a single story building on the OSU campus from the North and left on the South. The 3D Kinect depth imagery based trajectory is overlaid on the Google Earth image of the building, see Figure 12. In overall, the trajectory follows closely the corridor, except in the center part, where there is an intersection with another corridor, not shown. The numerical results of the comparison indicate a difference between the reference and reconstructed trajectory as an about 1 m STD in horizontal position and a 0.4 m misclosure at the exit point. With integrating IMU data and, potentially, 2D imagery, as well as improving algorithmic performance, much smoother and more accurate trajectories are expected.

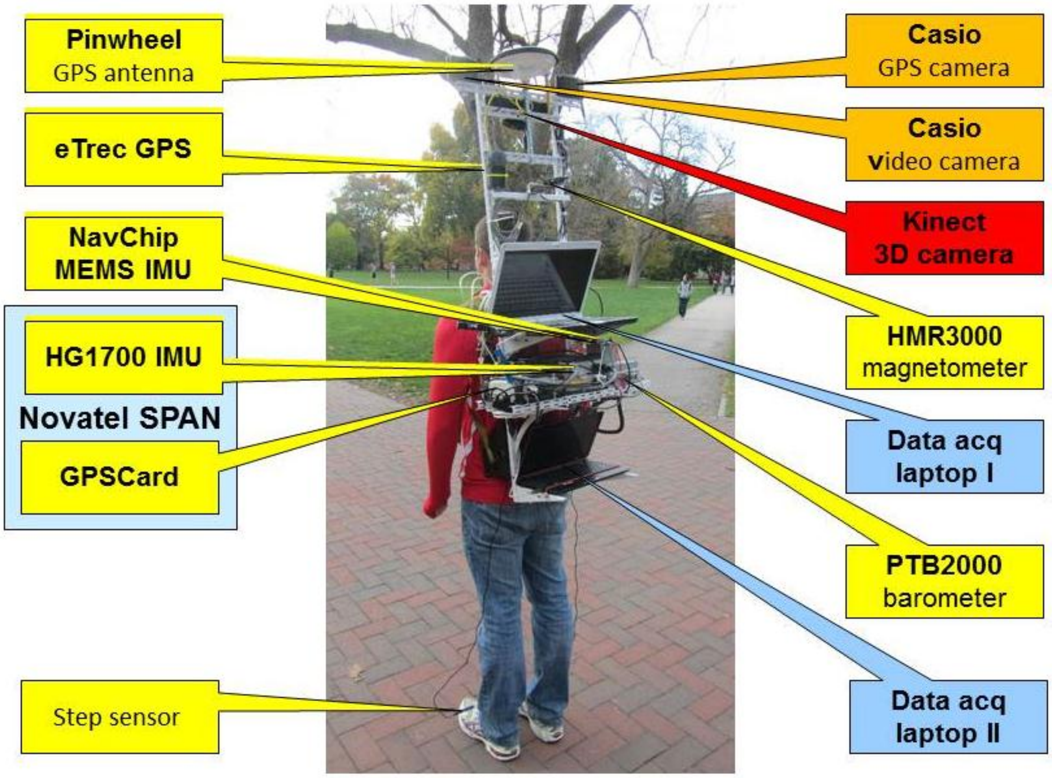


Figure 11. Personal navigator sensor configuration.

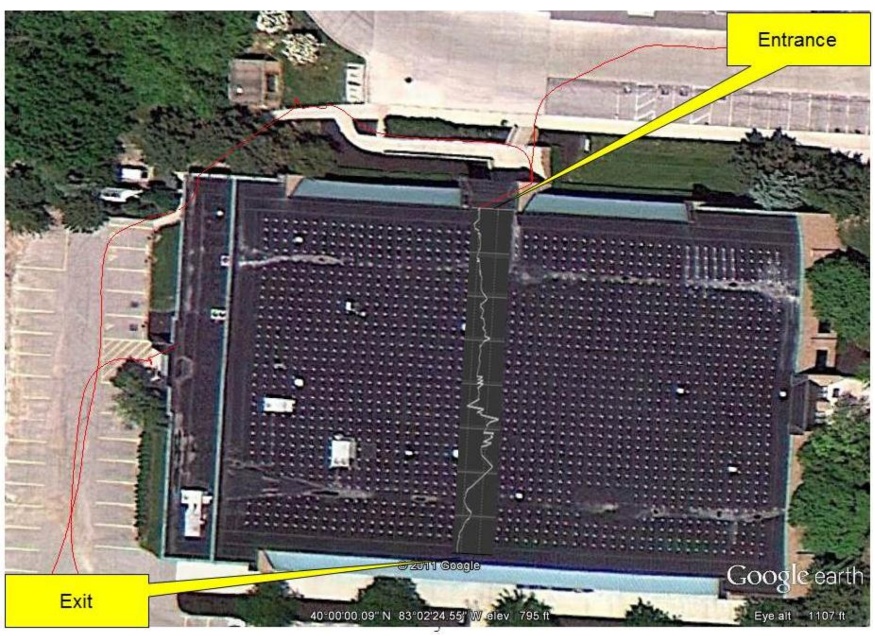


Figure 12. Kinect 3D-image sequence-based reconstruction of the platform motion.

CONCLUSIONS

In our limited experiences, the Kinect sensor has shown good and consistent performance. The calibration and repeatability tests confirmed that rather good quality 3D imagery can be acquired by this absolutely inexpensive sensor. The availability of several open source tools and the existence of an active user community make the integration of the Kinect sensor fairly simple, including basic data processing tasks too. While the Kinect is not a typical mapping sensor, its performance level makes it feasible to several applications, where the high accuracy of a TLS or MLS systems is not required.

REFERENCES

Advanced Scientific Concept, Inc., (ASC) http://www.advancedscientificconcepts.com/products/overview.html, last accessed in January 2012

Bailey, S., McKeag, W., Wang, J., and Jack, M., 2010. Advances in HgCdTe APDs and LADAR receivers, Infrared Technology and Applications XXXVI, Proc. SPIE 7660, 76603I.

Github, https://github.com/avin2/SensorKinect, last access in January 2012.

Kahlmann, T., Remondino, F., Ingensand, H., 2006. Calibration for increased accuracy of the range imaging camera SwissRanger, ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', Dresden, Germany, pp. 136-141.

Khoshelham, K., 2011. Accuracy Analysis of Kinect Depth Data: ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 38-5/W12, p. 6.

Konolige, K., Mihelich, P., 2010. Technical description of Kinect calibration, http://www.ros.org/wiki/kinect_calibration/technical, last accessed in January 2012.

Microsoft, 2010. Kinect. http://www.xbox.com/en-us/kinect/, last accessed in January 2012.

Microsoft, http://www.microsoft.com/en-us/kinectforwindows/resources/, last accessed in January 2012

OpenNI, http://openni.org/Documentation/ProgrammerGuide.html, last accessed in January 2012.

PrimeSense, (http://www.primesense.com), last accessed in January 2012.

Shan, J., Toth, C.-K., 2008. Topographic Laser Ranging and Scanning: Principles and Processing. Boca Raton, FL: Taylor & Francis.

Vosselman, G., Maas, H.-G., 2010. Airborne and Terrestrial Laser Scanning. Whittles Publishing, Caithness, Scotland, UK.

Weinmann, Ma., Weinmann, Mi., Hinz, S., Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. ISPRS Journal of Photogrammetry and Remote Sensing.

Weinmann, M., Wursthorn, S., Jutzi, B., 2011. Semi-automatic image-based co-registration of range imaging data with different characteristics: ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 38-3/W22, pp. 119-124.