Personal Navigation and Indoor Mapping: Performance Characterization of Kinect Sensor-based Trajectory Recovery

Charles TOTH, Dorota BRZEZINSKA, USA
Allison KEALY, Australia and Guenther RETSCHER, Austria

Key words: Kinect™ sensor, indoor navigation, indoor mapping

SUMMARY

The Microsoft Kinect™ sensor has gained popularity in a large number of applications beyond its intended original design of being a 3D human interface device, including indoor mapping and navigation of pushcart and backpack sensor platforms. Indoor mapping and personal navigation systems are generally based on the multisensory integration model, as currently no sensor itself can provide a robust and accurate navigation solution. To assess the error budget as well as to support the design of such systems, the individual sensor error budgets should be known (estimated). In this paper, a performance analysis of the Kinect sensor is provided based on a series of indoor tests, where sufficient control, based on UWB trajectory reference, was available. The main goal of the study is to assess the trajectory reconstruction performance from Kinect imagery only, using widely available mainstream computer vision methods to process 2D and 3D image sequences. Test data was acquired by the Kinect sensor mounted on the top of a pedestrian backpack navigation prototype in forward looking orientation with a clear field of view, and a user walked a hallway loop in several patterns. The results were evaluated based on a UWB-based reference solution.

Personal Navigation and Indoor Mapping:
Performance Characterization of Kinect Sensor-based Trajectory Recovery

Charles TOTH, Dorota BRZEZINSKA, USA
Allison KEALY, Australia and Guenther RETSCHER, Austria

1. INTRODUCTION

Small-size and low-cost imaging sensors are widely used in a multitude of consumer devices, providing medium-quality, typically redundant data. Usual application of these devices has the potential to be extended and applied for mapping and navigation purposes. Besides typical imaging mobile devices such as cell phones or compact digital cameras, others such as mobile imaging sensors on vehicles and humans (personal navigation) can be adopted and used. For example, the Microsoft Kinect™ (Microsoft, 2014), which has been sold in the tens of millions, is relatively small and, although not typically mobile, can easily be carried by a human. One advantage of the Kinect™ is the availability and simultaneous use of both passive and active imaging sensors, significantly extending the possibilities of Kinect™ applications. Similar to other simple devices/sensors, the use of Kinect™ is also limited, particularly by the range of the active sensor and by data accuracy. In addition, low data accuracy is a logical consequence of the low cost of a sensor. The influence of low data quality, however, may be somewhat offset by data redundancies. In addition, its limited range is also restrictive, being typically a few meters for the Kinect™ active sensor. This, however, is generally acceptable for indoor mapping, especially in corridor environments where the distances between objects are generally short.

In mobile mapping, remotely sensed data are usually complemented by navigation sensor data to support platform georeferencing. In most outdoor applications, integrated GPS and IMU sensors are used for that purpose. The use of navigation sensors is not mandatory for such active sensors as laser or radar. In contrast, aerial images can be processed based just on ground control points (GCP), though the use of georeferencing is beneficial. Indoors, GPS cannot be used, posing challenges to any mapping in an unknown environment. This study aims at assessing the performance potential for indoor mapping of a low-cost sensor, Kinect™. This paper is focused on analyzing the 2D and 3D Kinect imagery matching performance that is realistically achievable under typical indoor conditions. In other words, no additional sensory data is used to reconstruct the platform trajectory.

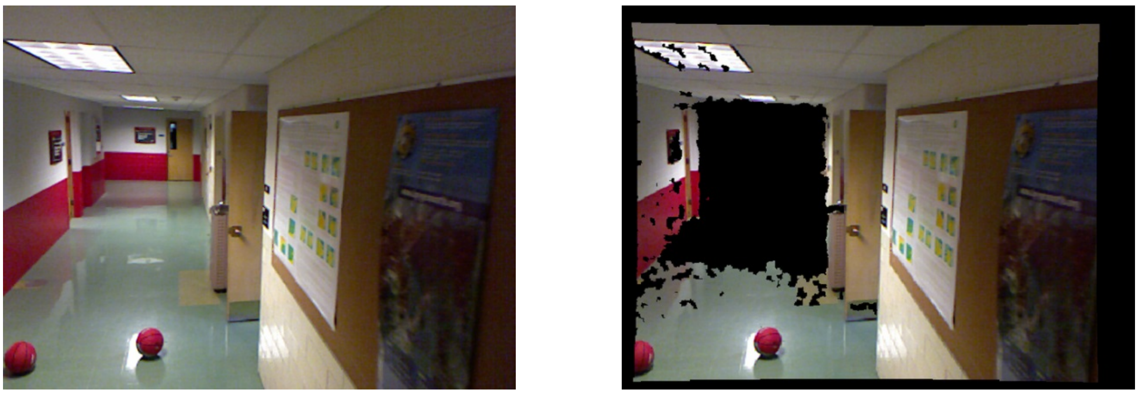
Simultaneous navigation and mapping based on the imagery is usually known as visual odometry (Scaramuzza and Fraundorfer, 2011) where the critical part of the computation is the matching of image frames. The Kinect™ sensor is an RGB-D (Red, Green, Blue, and Depth) camera, so there is no limitation of unknown scale of mono visual odometry. Different algorithms for frame matching in visual odometry based on the RGB-D camera models are proposed, see (Huang et al., 2011; Weinmann et al., 2011; Molnar and Toth, 2013; Whelan et al., 2013; Henry et al., 2014). In this work, a simple approach is proposed for Kinect™ data matching.

SYSTEM CONFIGURATION

The Kinect sensor is a motion sensing input device for the Xbox 360 video game console, originally developed by PrimeSense and later acquired by Microsoft. It allows the user to control and interact with the console by just giving voice and body gesture commands. Beside voice sensors (microphones), Kinect™ contains imagery sensors, including a basic RGB and an infrared (IR) sensor with an IR emitter, see Fig. 1. Detailed description of depth images generated by the Kinect™ sensor can be found in (Macknojia et al., (2012).

Figure 1: Kinect™ sensor.

The Kinect™ has three primary sensors: a 3D camera (IR-based active sensor), a conventional optical RGB sensor (2D camera), and a microphone array input. The device is USB-interfaced, similar to a webcam, and appears as a "black box" for the users. Note that very little is known of the sensors, internal components and processing methods are stored in the firmware. The emitter projects a structured light pattern of random points of light which is detected by the IR camera and then processed into a depth image. 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. Typical images are shown in Fig. 2; note the dark areas in Fig. 2b, showing that beyond a certain range 3D data recovery is not possible. The FOV of the 2D camera is 57° x 43°. The 3D camera can work at two resolutions with frame sizes of 640x480 and 320x240, respectively. The range data comes in 12-bit resolution, and the sensors' spatial relationship is shown in Fig. 1

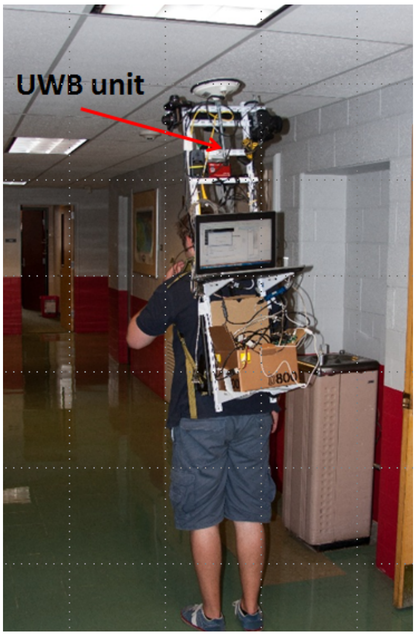


(a) (b)

Figure 2: Kinect™ images; (a) RGB and (b) colored 3D depth image.

Tests were performed with the Kinect™ installed on a backpack personal navigator prototype, developed earlier at The Ohio State University (OSU) Satellite Positioning and Inertial Navigation (SPIN) Laboratory (Grejner-Brzezinska et al., 2010; Toth et al., 2012; Zaydak et al., 2012). The Kinect™ sensor was mounted on the top of a backpack frame in a forward looking orientation with a clear field of view, and data was collected at 5-30 Hz as a user walked a hallway loop in several tests, see Fig. 3a. For reference, an UWB network was used, providing an overal trajectory accuracy about 10 cm (1σ), see Fig. 3b, see (Koppanyi et al., 2014). The OSU backpack personal navigator prototype contains about ten sensors that acquired data during the tests; most of this data was not used in this work.

(a)



(b)

Figure 3: OSU backpack personal navigator prototype; (a) Kinect™ and (b) UWB sensor.

2. 2D IMAGE BASED TRAJECTORY RECONSTRUCTION

To reconstruct the platform trajectory from the 2D image data, the spatial relationship between images should be established; basically finding the relative orientation between images, preferably for all possible image pairs and not only between consecutive ones. Once the image projection centers are known in the object space, the trajectory can be formed and transferred to any point or sensor location on the platform. The spatial relationship between 2D images can be described by the collinearity equations (Kraus, 1993), see Eq. 1, and standard photogrammetry solutions can provide the relative and/or exterior orientation of images by space resection. The proposed approach, the incremental trajectory reconstruction algorithm, is implemented as photo triangulation. The key aspect of any photogrammetric processing of the image is finding sufficient number of corresponding points and that the spatial distribution of the point is as even as possible.

$$x - x_0 = -c\frac{a_{11}(X - X_0) + a_{12}(Y - Y_0) + a_{13}(Z - Z_0)}{a_{31}(X - X_0) + a_{32}(Y - Y_0) + a_{33}(Z - Z_0)}$$

$$y - y_0 = -c\frac{a_{21}(X - X_0) + a_{22}(Y - Y_0) + a_{23}(Z - Z_0)}{a_{31}(X - X_0) + a_{32}(Y - Y_0) + a_{33}(Z - Z_0)} \quad (1)$$

In this study, the corresponding image points are obtained by using SIFT (Scale-Invariant Feature Transform) features (Lowe, 2004). Examples of SIFT keypoint locations on three consecutive images are shown in Fig. 4. The upper part depicts an easy case, where there are many features available for matching. In contrast, the lower part shows a sequence of images with less texture, and, consequently, the number of SIFT features are low. In addition, the distribution of the feature points is weak, as the points practically fall along a line. Note that the overlap is relatively small in that case. Note that the relatively poor performance is to a large extent due to the forward looking orientation of the camera, which makes intersection geometry weak.

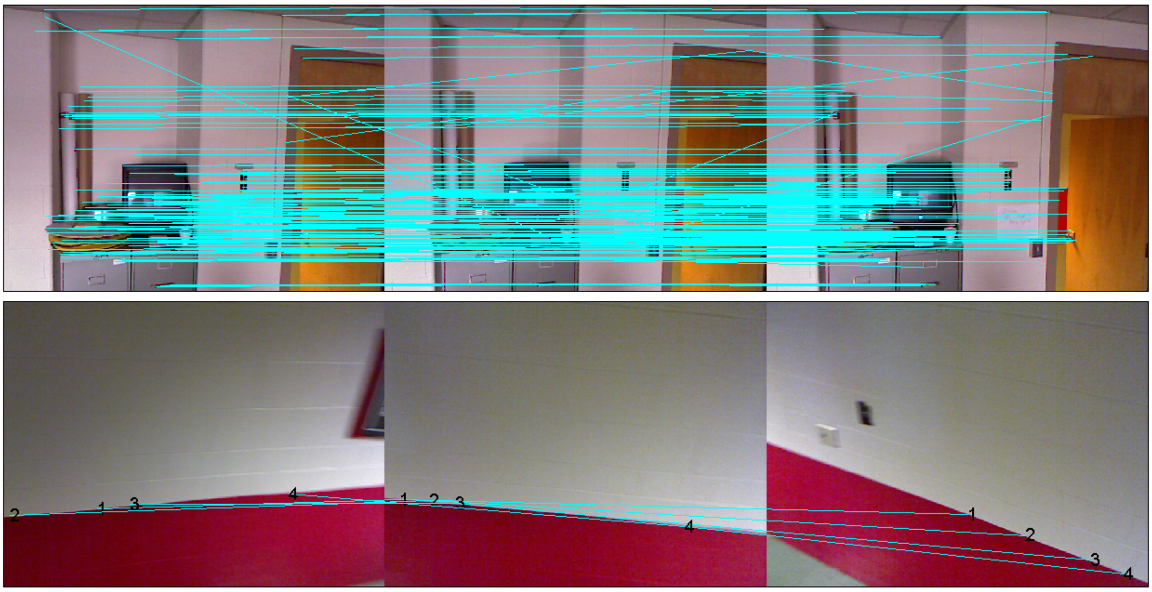
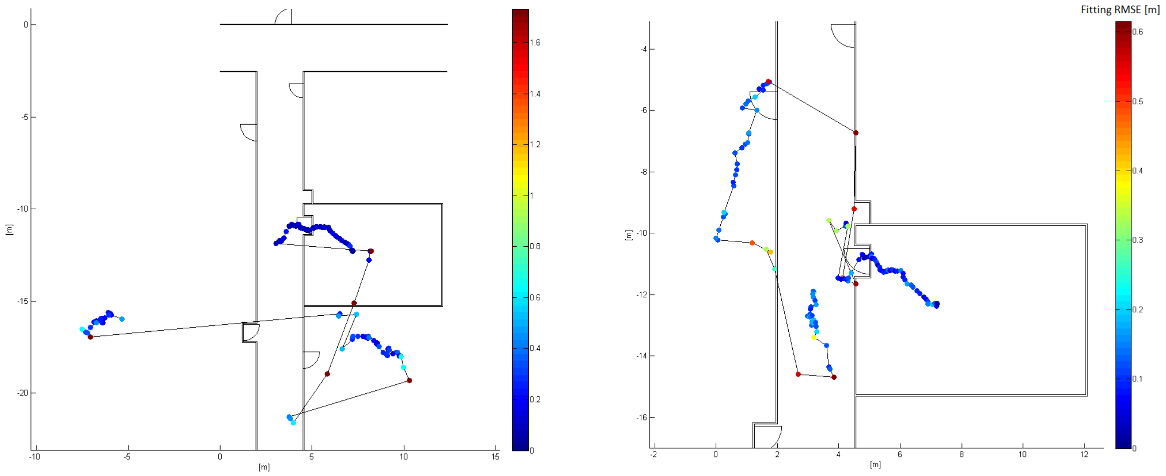


Figure 4: SIFT feature extraction and matching examples; top row shows normal situation, while the point distribution and image overlap are weak in the bottom row.

Incorrectly matched SIFT points are automatically eliminated in a simple, robust estimation procedure; points with residuals larger than three pixels are not used in the next iteration of the space resection process. RANSAC (RANdom SAmple Consensus) was also considered for that purpose, but experiences showed that the three pixel threshold works reliably for our data. SIFT performs well for image pairs of sufficient overlap and with good texture. Trajectory reconstruction performance is shown in Fig. 5a, clearly indicating that only shorter sections can be reconstructed based on 2D images at an acceptable level.

3. 3D IMAGE BASED TRAJECTORY RECONSTRUCTION

In this study, the ICP (Iterative Closest Point) algorithm, widely used in robotics and object reconstruction, is tested (Tokekar et al., 2009; Surmann et al., 2003; Newcombe et al., 2011). ICP can support various geometrical models, and the 3D similarity transform with unit scale was used in the experiments. Note that there are several other techniques to match 3D image sequences. Results from a method developed at OSU were reported in (Markiel, 2012). Since ICP is rather computationally intensive, the Kinect™ 3D images, the point clouds, were decimated to two 5x5x5 cm voxel cubes to speed up execution. Fig. 5 shows the reconstructed trajectories based on 2D and 3D Kinect™ image sequences. Clearly, the ICP-based solution represents a better performance compared to the 2D solution; note that the 2D image-based trajectory spreads more in space. But there are a few big jumps, where ICP was unable to provide a good match; the color bar on the right shows the RMS between point clouds. These problems are related to situations with significant changes in the image content, such as turning or sudden large movement between image captures. In summary, the trajectory reconstruction, based on ICP works fine when the RMS is small; this is mostly the case when the difference between the point clouds is small. Analyzing the cases in detail, where ICP fails, it can be stated that with a good initial estimate ICP works better, likely resulting in a correct solution.



(a) (b)

Figure 5: Trajectory reconstruction; (a) 2D image based and (b) 3D image based.

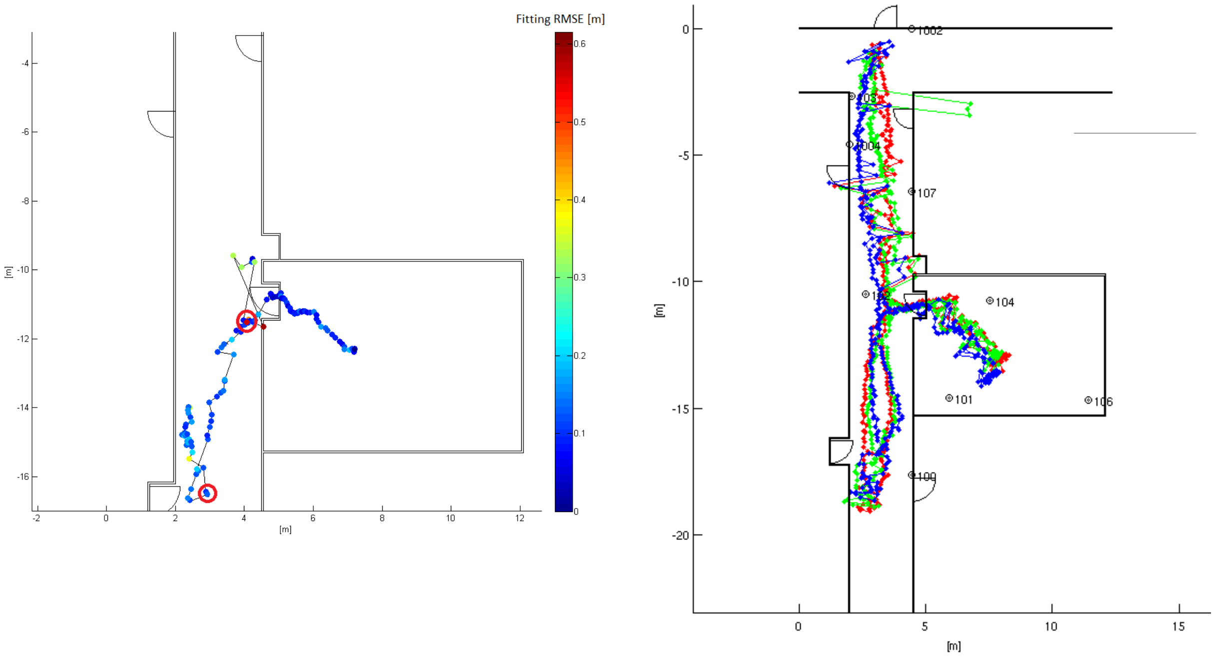
4. INTEGRATED SOLUTION BASED ON COMBINING 2D AND 3D IMAGES

Comparing where the 2D and 3D methods fail or succeed shows some complementarity. For example, at quick turns, the 2D generally provides good performance because of the better image geometry, while motion along straight lines can be better captured by ICP. The concept of combined matching presented in this work basically consists of first matching consecutive frames in 2D space, (which is realized by SIFT-based image matching), and then transferring matching points into the 3D space. Knowing the interior and relative orientation parameters of Kinect™ 2D and 3D sensors, it is possible to assign points from 2D space to matching points in 3D space and, consequently, perform point cloud registration as long as the translation and rotation parameters of each point cloud are known. The spatial relationship between two frames can be uniquely calculated based on three pairs of corresponding 3D points. Obviously, a higher number of matching points increases the reliability of the solution. However, it is not mandatory to use all possible matching points or to use techniques for dense matching that result in longer computation time, as it is more important to use points distributed as evenly as possible in the common area of the two frames.

During sensor calibration, the spatial relationships between the imaging sensors as well as the parameters of the sensors' interior orientation are estimated, and thus, the 2D and 3D image data can be easily related to each other. In the proposed approach, images are created based on the colored point cloud. One of the advantages of this solution is that after performing SIFT and finding 2D matching points, there is no need to search for conjugate points in 3D space as the 3D coordinates can be treated as additional features for each pixel of the artificial image. The combined trajectory reconstruction include the following main steps:

1. Image creation from the point cloud is generally simple, except for areas where there is no 3D data, see Fig. 2b
2. SIFT matching of reconstructed images; our implementation is based on applying SIFT individually to the three color bands to increase the number of feature points
3. Retrieving 3D information of matched points; to remove mismatched point pairs, the iterative least squares adjustment that determines the 6-parameter transformation has been extended by a weighting scheme, where the point weights are modified at each step
4. Robust estimation of transformation parameters; the basic ICP method for the 3D image sequence processing has been modified in several aspects to increase performance: (1) the last five 3D images are considered for matching to increase robustness by limiting the impact of a low quality image, and (2) the minimization criterion has been extended to include RGB information; this way not only the Euclidian distance but the color distance between images is included in the process
5. Trajectory reconstruction

Fig. 6 shows the result of the combined trajectory reconstruction with the reference solution obtained from the UWB network; note the UWB figure shows all the tests trajectories. The combined method clearly provides a good solution, except for two epochs, marked by red circles in Fig. 6a, where the trajectory could not be reconstructed and thus were manually connected. Once the trajectory is estimated, the point clouds can be merged. As an example, the final results of about 500 stitched point clouds with reconstructed trajectory is shown in Fig. 7. Clearly, the general shape of the corridor is preserved in the stitched point cloud. However, due to the incremental nature of the selected approach, some drift can be noticed. The trajectory was reconstructed quite well, though two jumps are present.



(a) (b)

Figure 6: Trajectory evaluation; (a) combined 2D and 3D image based trajectory reconstruction, and (b) UWB reference solutions.

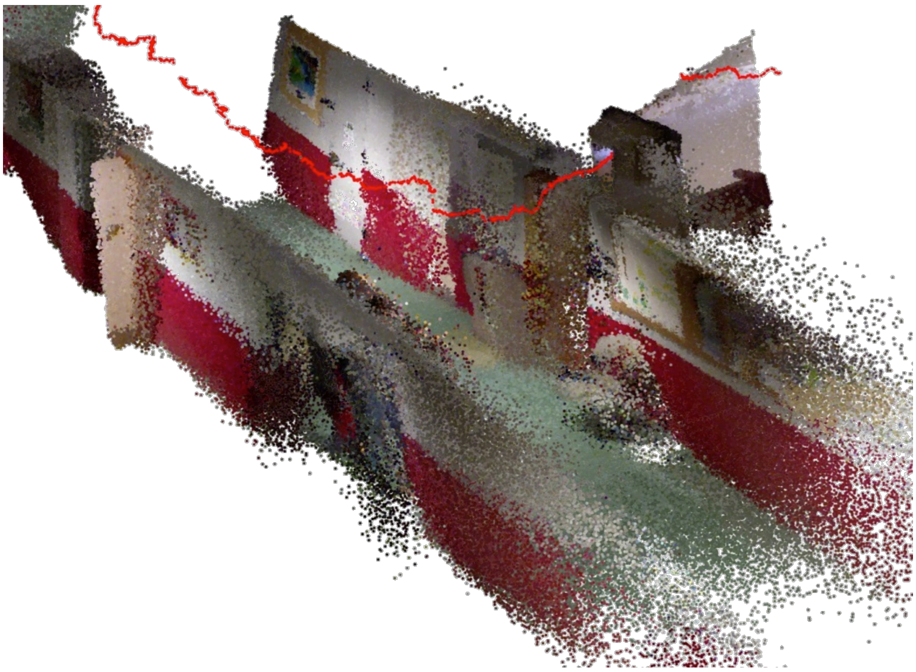


Figure 7: Stitched point cloud (ceiling removed) and reconstructed trajectory (red).

5. CONCLUSIONS

Initial results indicate that the method of combining Kinect™ 2D and 3D imagery for indoor navigation and mapping is feasible using low-cost RGB-D sensors. The test results revealed that trajectory reconstruction based on 2D imagery is generally unreliable, while matching 3D images (point clouds) provides somewhat better results, though no overall solution can be achieved in general. The proposed combined trajectory reconstruction, based on 2D and 3D images, properly estimated the trajectory and, consequently, produced a stitched point cloud that is a correct representation of the mapped area. The circumstances where the algorithm fails can be identified reliably. It must be emphasized that failures occurred only within specific conditions, such as fast turns, which seem to be unusual behavior during indoor navigation, and the overlap between images is small, so the matching may fail. Note that by introducing other sensor data, such as IMU data, these situations generally can be remedied. Finally, the proposed algorithm for point cloud stitching can be further improved by adding keyframes or implementing a Kalman filter to decrease the influence of the drift caused by the basic incremental approach.

REFERENCES

Github, https://github.com/avin2/SensorKinect, last access in January 2014.

Grejner-Brzezinska, D. A., J.N. Markiel, C.K. Toth, A. Zaydak, (2012): Getting Along: Collaborative Navigation in Transitional Environments, *GPS World*, November, pp. 43–92.

Henry, P., M. Krainin, E. Herbst, X. Ren, and D. Fox, (2014): RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments, in: *Experimental Robotics, Springer Tracts in Advanced Robotics*, Volume 79, pp. 477-491.

Huang, A.S., A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, (2011): Visual odometry and mapping for autonomous flight using an RGB-D camera, *International Symposium on Robotics Research (ISRR)*, pp. 1-16.

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.

Koppanyi, Z., Toth, C.K., Grejner-Brzezinska, D.A., and Jozkow, G. (2014): Performance Analysis of UWB Technology for Indoor Positioning, ION ITM, January 27-29, 2014, San Diego, CD ROM.

Kraus K. (1993), *Photogrammetry, Vol. I*: *Fundamentals and Standard Processes*, Dümmlers

Lowe, D. (2004): Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, 2004 pp. 91-110.

Macknojia, R., Chávez-Aragón, A., Payeur, P., Laganière, R., (2012): Experimental Characterization of Two Generations of Kinect's Depth Sensors, *IEEE International Symposium on Robotic and Sensors Environments*, pp. 150-155.

Markiel, J.N. (2012): Navigation in GPS Challenged Environments Based Upon Ranging Imagery, PhD dissertation, The Ohio State University.

Microsoft, (2014): Kinect™, http://www.xbox.com/en-us/kinect/, (Last accessed 10 March, 2014).

Molnar, B., Toth, C. (2013): Spherical Target Based Trajectory Recovery from Kinect Depth Imagery, Proceedings ASPRS 2013 Annual Conference, March 24-28, Baltimore, MD.

Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A., (2011): KinectFusion: Real-time dense surface mapping and tracking, *10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 127-136,

Scaramuzza, D., and F. Fraundorfer, (2011): Visual odometry [tutorial], *Robotics & Automation Magazine, IEEE,* 18(4): pp. 80-92.

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

Surmann, H., Nüchter, A., Hertzberg, J., (2003): An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments, *Robotics and Autonomous Systems*, Vol. 45, Issues 3–4, pp. 181-198.

Tokekar, P., Bhatawadekar, V., Fehr, D. Papanikolopoulos, N., (2009): Experiments in object reconstruction using a robot-mounted laser range-finder, Control and Automation MED '09, 17th Mediterranean Conference, pp. 946-951, 24-26 June 2009.

Toth, C.K., Jozkow, G., Koppanyi, Z., and Grejner-Brzezinska, D.A. (2012): Performance Analysis of the Kinect Sensor Trajectory Reconstruction, Proceedings of ION Pacific PNT 2013, April 22-25, 2013, Honolulu, HI, pp. 542-550.

Toth, K. C., Molnar, B., Zaydak, A., Grejner-Brzezinska, A. D, (2012): Calibrating the MS Kinect Sensor, ASPRS 2012 Annual Conference, Sacramento, USA

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.

Whelan, T., H. Johannsson, M. Kaess, J.J. Leonard, and J. McDonald, (2013): Robust real-time visual odometry for dense RGB-D mapping,. *2013 IEEE International Conference on Robotics and Automation*, pp. 5724-5731.

Zaydak, A., Toth, C.K., Grejner-Brzezinska, D.A., Molnar, B., Yi, Y., and Markiel, J.N. (2012): 3D Image-based Navigation in Collaborative Navigation Environment, ION GNSS, September 17-21, 2012, Nashville, pp. 2462-2468, CD ROM

BIOGRAPHICAL NOTES

Charles K. Toth is a Research Professor in the Department of Civil, Environmental and Geodetic Engineering, The Ohio State University. He received a M.Sc. in Electrical Engineering and a Ph.D. in Electrical Engineering and Geo-Information Sciences from the Technical University of Budapest, Hungary. His research expertise covers broad areas of spatial information systems, LiDAR, high-resolution imaging, surface extraction, modeling, integrating and calibrating of multi-sensor systems, multi-sensor geospatial data acquisition systems, 2D/3D signal processing, and mobile mapping technologies. He has published over 300 peer-reviewed journal and proceedings papers, and is the co-editor of the widely popular book on LiDAR: Topographic Laser Ranging and Scanning: Principles and Processing. He is the recipient of numerous awards, including the 2009 APSRS Photogrammetric Award, several Lumley Research Awards from OSU, and various best papers awards. He is ISPRS Technical Commission I President for the 2012-2016 Congress period, and was recently elected as ASPRS Vice President.

Dorota A. Grejner-Brzezinska is a Professor in Geodetic Science, and director of the Satellite Positioning and Inertial Navigation (SPIN) Laboratory at The Ohio State University. Her research interests cover GPS/GNSS algorithms, in particular, high precision positioning and navigation, such as DGPS and RTK, GPS/inertial and other sensor integration for navigation in challenged environments, sensors and algorithms for indoor and personal navigation, Kalman filter and non-linear filtering. She published over 180 peer reviewed journal and proceedings papers, numerous technical reports and five book chapters on GPS and navigation, and led over 20 sponsored research projects. She is ION Fellow, and the recipient of the 2005 ION Thomas Thurlow Award, the 2005 United States Geospatial Information Foundation (USGIF) Academic Research Award, and ESRI Award for Best Scientific Paper in Geographic Information Systems published in 2004.

Allison Kealy is a Senior Lecturer in The Department of Infrastructure Engineering at The University of Melbourne Australia. She holds an undergraduate degree in Land Surveying from The University of the West Indies, Trinidad, and a PhD in GPS and Geodesy from the University of Newcastle upon Tyne, UK. Allison's research interests include sensor fusion, Kalman filtering, high precision satellite positioning, GNSS quality control, wireless sensor networks and location based services. Allison is currently the co-chair of FIG Working Group 5.5 entitled Ubiquitous Positioning which is coordinated collaboratively with IAG Working Group 4.2.5.

Guenther Retscher is an Associate Professor at the Department of Geodesy and Geoinformation of the Vienna University of Technology, Austria. He received his Venia Docendi in the field of Applied Geodesy from the same university in 2009 and his Ph.D. in 1995. His main research and teaching interests are in the fields of engineering geodesy, satellite positioning and navigation, indoor and pedestrian positioning as well as application of multi-sensor systems in geodesy and navigation. Guenther is an IAG Fellow and chairs the IAG Sub-Commission 4.1 on 'Alternatives and Backups to GNSS' and the joint IAG 4.1 and FIG 5.5 Working Group on 'Ubiquitous Positioning Technologies and Techniques'.

CONTACTS

Charles Toth and Dorota Brzezinska
Department of Civil and Environmental Engineering and Geodetic Science
The Ohio State University
2070 Neil Avenue, Columbus, OH 43210-1275, USA
Tel: +1 614.292.7681 Fax: +1 614.292.2957
E-mail: toth@cfm.ohio-state.edu, dbrzezinska@osu.edu

Allison Kealy
Melbourne School of Engineering
The University of Melbourne
Victoria 3010 Australia, Australia
Tel: +61 3 8344 6804 Fax: +61 3 8344 4616
E-mail: a.kealy@unimelb.edu.au

Guenther Retscher
Department of Geodesy and Geoinformation
Vienna University of Technology
Gusshausstrasse 27-29 E120/5
1040 Vienna, Austria
Tel. +43 1 58801 12847 Fax +43 1 58801 12894
Email: guenther.retscher@tuwien.ac.at