Microsoft Kinect for THz Sensor Management

Philip ENGSTRÖM, Maria AXELSSON, Mikael KARLSSON
Swedish Defence Research Agency (FOI), Linköping, Sweden

Abstract

Rapid stand-off detection of hidden threats is highly desired in many security and surveillance applications. 3D THz scanning technologies provide high resolution 3D imagery, which can be used to detect hidden objects and classify these. However, due to sensor limitations, accurate subject pose tracking is needed. In this paper we present a proof-of-concept for a support system for THz sensor management based on structured light and laser scanning. A laser scanner from SICK is used for accurate 2D-position tracking and a Microsoft Kinect is used for 3D pose estimation. We show how the technologies complete each other and how data can be fused for increased robustness and accuracy.

Keywords: 3d body scanning, laser scanning, Kinect, THz, screening, sensor management

1. Introduction

Detection of person-borne concealed threats, e.g. weapons or explosive devices, from a distance is desired in many security and surveillance applications. Today, it is common with portal-based THz or X-ray screening systems at airports or government buildings, but a stand-off technique could provide an early warning and add flexibility to the screening, like screening of a passing crowd entering a sports arena.

Stand-off 3D THz imaging to detect concealed treats is currently under development [1,2]. A THz sensor can provide high resolution 3D range data of each subject showing layers of clothes and any concealed items. The objects can be detected in the 3D data as anomalies like protruding surfaces on the body. However, efficient scene scanning and management of the THz sensor is required because of it being a scanning sensor technology and it has limitations such as a narrow field of view and a relatively long image acquisition time. The portal systems available on the market today only work on cooperative subjects, i.e., non-moving. With more complex scenarios, a sensor management system is needed for planning and control of the THz sensor [3]. All subjects in the scene need to be tracked and positioned in real time. Pose tracking is also needed as the recorded THz data has to be corrected for the subject movements during the acquisition.

High detection rate of anomalies is the foremost goal. The detection rate depends on the data quality, which in turn is heavily dependent on how well the management system is able to capture the subject's movements. The pose tracking can also be used to estimate scan completeness, and for stitching data from separate scans.

In this paper we present a proof-of-concept for THz sensor management using the consumer product Microsoft Kinect, combined with a laser scanner from SICK. The laser scanner is used to detect subjects at long range, before they come into the THz sensor field-of-view and the structured light sensor is used for 3D positioning and 3D pose estimation to direct the THz sensor to the next point of interest.

The paper is structured as follows: In Section 2 we present our system approach. Section 3 covers conducted experiments. Section 4 the results, and Section 5 the conclusions.

2. Our System Approach

Most surveillance today is based on visual cameras, and much research is being done on tracking people in such images. However, for tracking the precise pose of person we believe that other sensor technologies are much better suited for the task. We propose a combination of structured light and laser scanning for pose tracking in real time with the OpenNI middleware [4]. In our setup the structured light sensor is a Microsoft Kinect [5] and the laser scanner a SICK LMS-511 [6] as shown in the middle image of Figure 1. These sensors merely serve as representatives for their technologies. OpenNI is open source and serves as a good base for future development. The idea is to use structured light at close range to produce depth maps good enough for accurate pose tracking. The laser scanning is used at long range for subject tracking and for increased robustness.

Fig. 1. Left, the IR pattern of Kinect (NIR image). Middle, the Kinect placed in front of the LMS-511. Right, the laser pulse of LMS-511 (NIR image).

2.1. Structured Light - Microsoft Kinect

Structured light is a 3D imaging technique that relies on projecting a known pattern onto the scene (left image in Figure 1). The pattern deforms when hitting objects in the scene and since the original pattern is known it is possible to calculate depth and surface information of objects in the scene.
The main advantage is its simplicity, which enabled Microsoft to mass produce the Kinect sensor for the consumer market. The Kinect produce a 0.3 megapixel dense depth map with a few centimeters accuracy within 5m distance. Its disadvantages are the sensitivity to lighting conditions and the relatively low accuracy.

2.2. Laser Scanning - SICK LMS-511

Laser scanners commonly use the detection method pulsed time of flight where the range is calculated from Equation 1, where *D* is range (distance), *c* the speed of light and *T* the measured time.

$$D = \frac{(c * T)}{2} \qquad (1)$$

The SICK LMS series operate using this principle and the right image in Figure 1 shows how the laser is moved across the scene. The major advantage with time of flight is accuracy, and the LMS-511 outputs 1140 range measurements along a 190 degree “semicircle” at 25Hz with centimeter accuracy up to 80m. The amount of measurements is tiny compared to the 300 000 depth values from the Kinect, but on the other hand the field of view is wider and the accuracy much higher. The LMS-511 is capable at scanning up to 100Hz but with fewer measurements. Figure 2 illustrates how the spatial resolution varies with speed and range. A 0.4m wide target will for example only result in one measurement at 30m when scanning at 100Hz. It is important to select a laser scanner with properties suitable for the given scenario.

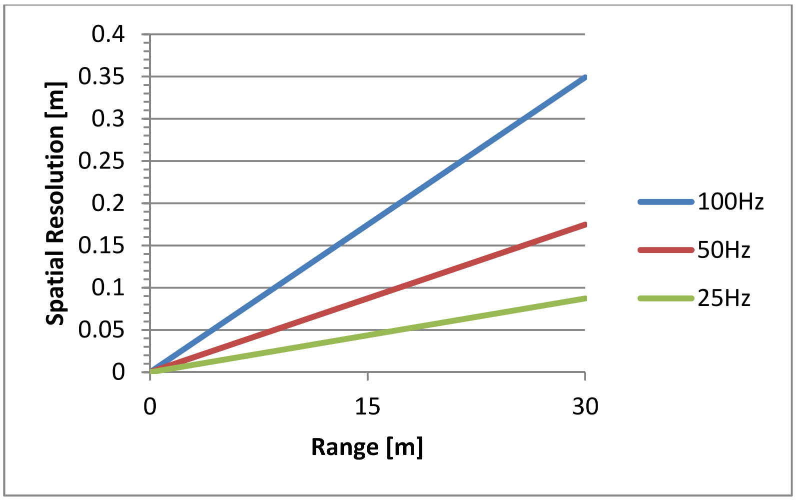

Fig. 2. The spatial resolution of SICK LMS-511 at 100, 50, and 25Hz.

2.3. OpenNI

OpenNI is a non-profit organization that was founded shortly after the launch of the Kinect. OpenNI provides an open source alternative to Microsoft's software and drivers for the Kinect. The open source software includes several features such as full body tracking, i.e., pose tracking. This pose tracking is slightly less capable than the one provided by Microsoft, but on the upside it is fully extendible since it is licensed under LGPL [7] (GNU Lesser General Public License). Figure 3 show a depth map provided by OpenNI and a tracked pose.

The main issue with the pose tracking is the time required for calibration, i.e., the time needed for the system to determine the current pose. This is especially difficult when the subject is in motion, and this is the main motivation for adding tracking, i.e., the laser scanner. Our goal is to enable faster and robust calibration by providing tracking information based on laser scanning.

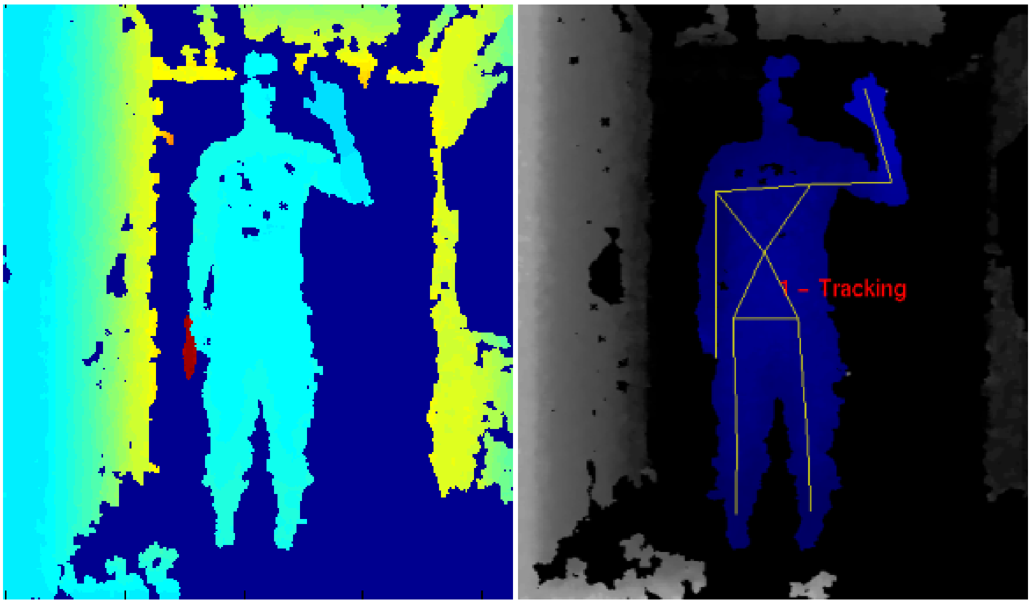

Fig. 3. Left, color coded depth map from Microsoft Kinect. Right, pose tracked with OpenNI.

3. Experiments

First a series of static experiments were conducted for the purpose of validating the sensors on static scenarios. After analyzing the outcomes, a series of dynamic experiments with a walking subject were conducted.

3.1. Static Evaluation

The static experiments were conducted in long corridor without windows, which enabled controlled lighting conditions. Objects were placed in the scene at distances from 2m up to 10m. The particular objects were selected to pinpoint strengths and weaknesses of the sensors, and are depicted in Figure 4. The first object was an open square cone with well-defined angles, and this was used to verify that the sensors can handle surfaces at an angle. The second object was a manikin, which was thoroughly measured both with and without clothing. The third object was a large square board with four areas with different reflectivity.

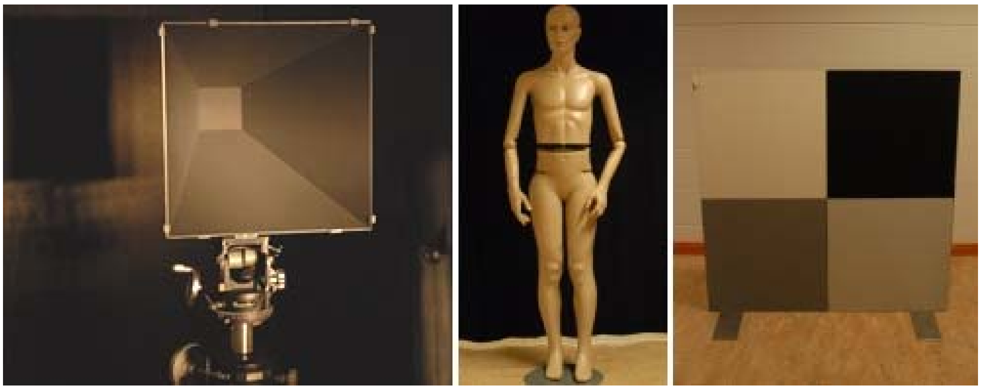

Fig. 4. The objects selected for the static evaluation of the sensors.

3.2. Dynamic Evaluation

The dynamic experiments were conducted in a long corridor with windows as seen in Figure 5. All tests involved a test subject walking away from the camera, turning, and walking back, always walking straight in front of the system. The turning point varied between 10m and 25m. The speed of the laser scanner was set at 25Hz, 50Hz, and 100Hz. For most experiments only the depth image of the Kinect was recorded but some recordings also included the visual camera. The synchronization between the sensors was done manually with an object blocking both sensors field of view before and after each test. Some individual tests with each of the sensors were also conducted for verification purposes.

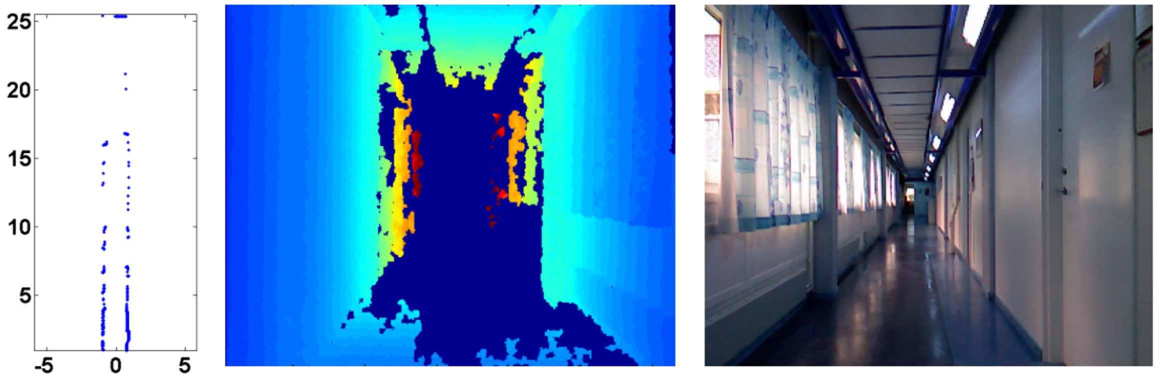

Fig. 5. Left, laser range measurements of the empty corridor. Middle color coded depth map from Microsoft Kinect. Right, the visual image of Microsoft Kinect depth.

4. Results

4.1. Static Evaluation

The laser scanner data is very accurate and fairly easy to interpret, but it is important to configure and choose the equipment so that it fits the specific need of the scenario. For example in our scenario with a long narrow corridor, a field of view of 190 degrees is more or less pointless since only a fraction of it is used. A narrower field of view and increased angular resolution would have been better, but even at the lowest setting the data in our case was acceptable.

With the Kinect several unexpected findings were made, both positive and negative. On the plus side the range was better than expected. According to specifications measurements are only accurate up to 5m but you can get reliable data up to 10m given the right conditions, i.e., lighting conditions and target reflectance. Low reflecting targets may however be a problem even below 5m also under optimal lighting conditions. The left image of Figure 6 shows a target consisting of 4 plates with 81%, 5%, 18% and, 49% reflectivity at 830nm (the wavelength of the laser illuminator of the Kinect). The right image of Figure 6 show the depth map and it is clear that 5% reflectivity is too low. The LMS-511 has its threshold at 3% reflectivity.

Another issue for the Kinect is transparent or glossy materials. Figure 7 shows an office scene with two coffee mugs on a desk. The Kinect is unable to capture the depth of the desk surface due to glossiness and incident angle. The issue with transparent materials is illustrated by placing the same mugs in a plastic bag which results in further loss of depth values.

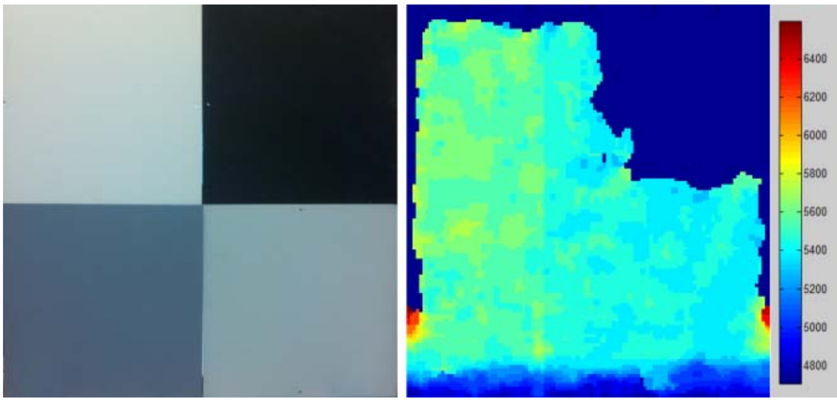

Fig. 6. Left, a flat plate with four fields with different reflectivity and on the right a Kinect depth map of the plate at 5.5 m. The Kinect has obvious problems with low reflecting targets.

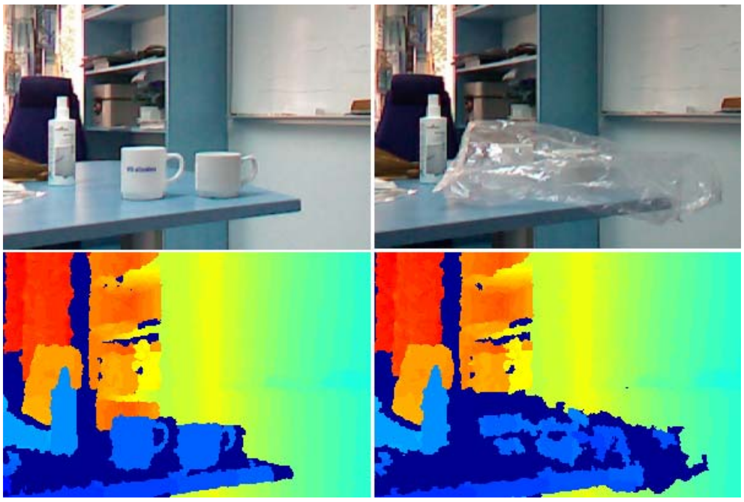

Fig. 7. Top, visual image of mugs without and with surrounding plastic. Bottom, Kinect depth image of the corresponding scenes.

An issue with the Kinect data, in particular at long range, is noise. A common method to counter noise is to average consecutive frames and this is quite effective with Kinect data especially if it can be done on a stable region. Because of the method used to calculate depths, regions with high contrast in depth tend to bleed into each other. For example in Figure 6, the lower left part of the field with low reflectance gets some depth values from the neighboring fields. This bleeding is not deterministic and will vary from frame to frame.

Table 1 below shows how the standard deviation is lowered by averaging 5 frames at different distances. The first 3 rows involve averaging the whole plate and the 4th row averaging on a smaller region away from edges. As long as the edge problem is handled, the standard deviation at long distances can be reduced by up to 50%.

Table 1. Reducing Kinect depth noise by averaging consecutive frames (mm).

Distance	Std Sick	Std K	Std K5	Imprv.
7600	9,09	141,73	117,57	17%
5500	8,46	68,05	58,88	13%
3600	6,37	25,05	21,01	16%
7600 small region	*9,09*	*64,22*	*35,42*	*45%*

4.2. Dynamic Evaluation

The laser scanning data, under normal circumstances, allows the subject to be tracked with centimeter accuracy. This is of course dependent on the properties of the laser scanner and also on its placement. In our scenarios, we choose to place the scanner straight in front of the subject and horizontal. This is the easiest scenario since the laser will be measuring the distance to the same region of the subject as long as the walking direction is unchanged. Depending on the scenario other positions might be better. There is also an option to tilt the sensor, but this would make the tracking considerably harder.
The scanner horizontal straight in front will give robust data as long as the subject is fairly smooth where the laser measures. A protruding object such as a camera hanging on the stomach might cause noise if the laser unpredictably hits and misses the object. More experiments are needed to decide how much of a problem this is and how to deal with it. One solution could be a secondary laser scanner with a different placement.

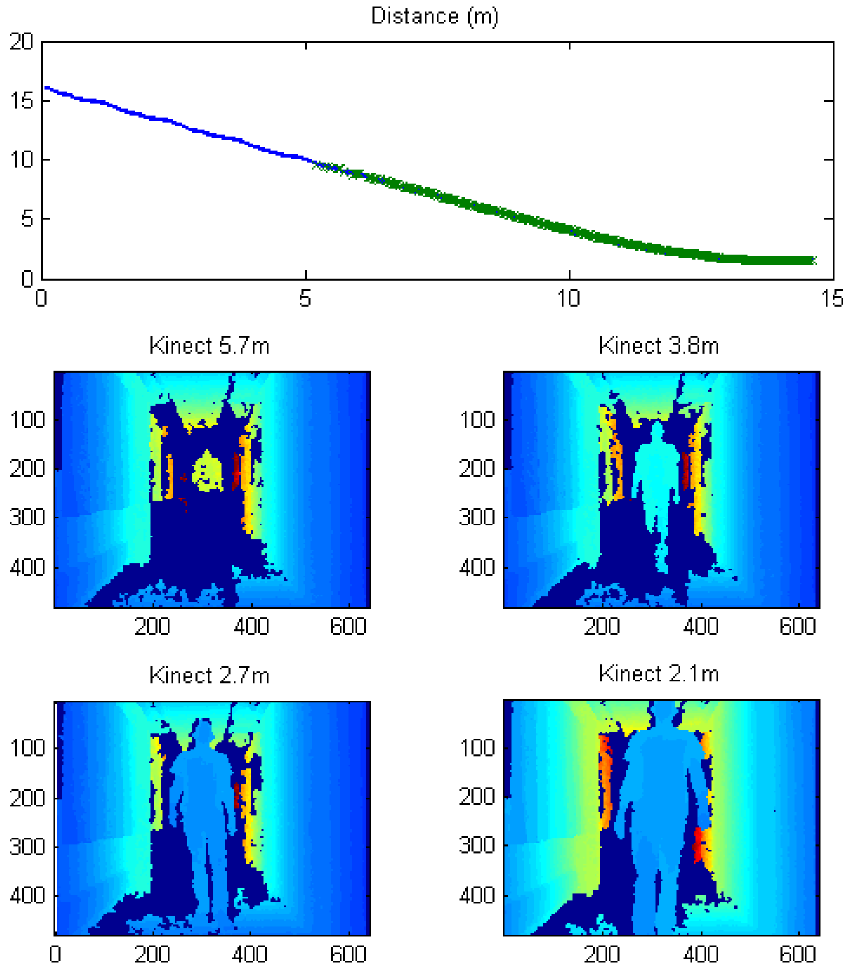

Fig. 7 Top, distance to target as function of time, blue is LMS-511 data and green is Kinect data. Below, 4 depth maps from the Kinect at 5.7m, 3.8m, 2.7m, and, 2.1m.

Figure 7 depicts a typical scenario with at subject entering the scene beyond 15m away and walking straight towards the system. The blue line in the plot is the distance based on LMS-511 data, and the green is the distance based on Kinect data. The target is always visible in the laser scan data and becomes visible in the Kinect data at approximately 10m. Figure 7 also include 4 depth maps at 5.7m, 3.8m, 2.7m, and 2.1m. At 5.7m only a small part of the upper body is visible, at 3.8m some parts of the legs are still missing, and, at 2.1m the field of view becomes an issue (head and feet out of field of view). The average walking speed of a young human is 1.5m/s [8], which means that the Kinect will have at most a couple of seconds worth of data. This is the motivation for speeding up the pose calibration process.

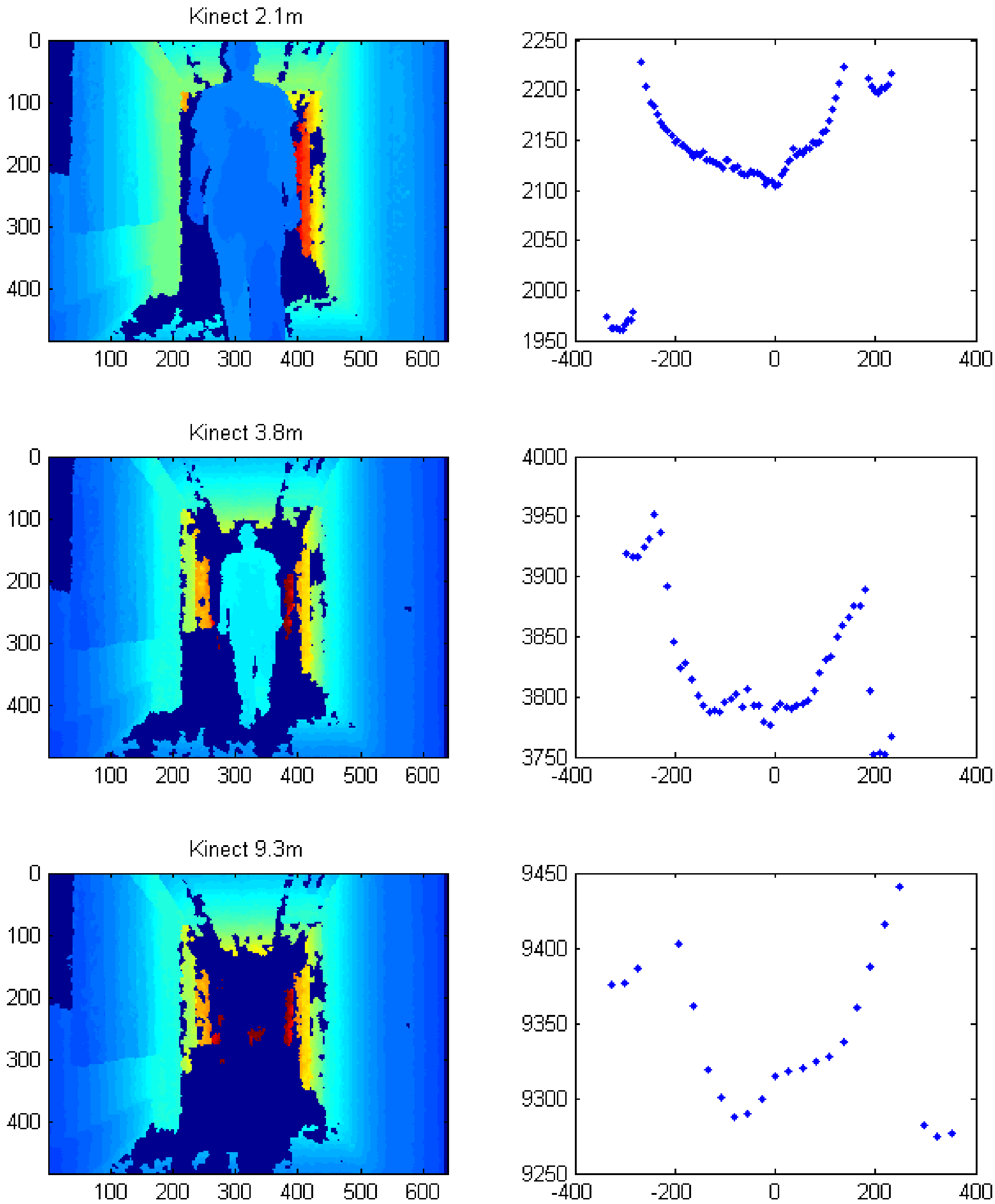

Fig. 8 Left, Kinect depth maps at 2.1m, 3.8m, and, 9.3m.
Right, the corresponding LMS-511 measurements at the same distance and time.

Figure 8 shows a comparison between depth maps and lased scans at different ranges. The LMS-511 was set to 25Hz, i.e., its maximal angular resolution for these scans. At 9.3m we got 23 measurements, of which 17 were hits on the body. It is easy to see that this data provides more information of the subjects pose than the Kinect depth map at the same distance, especially when analyzing the data over time.

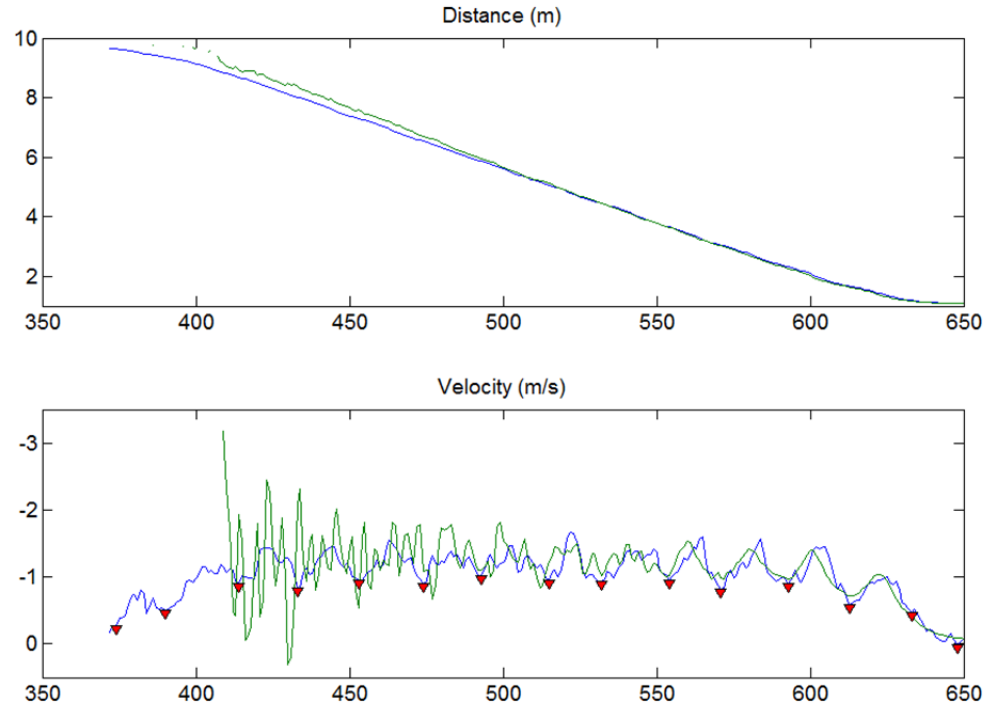

Fig. 9. Upper, distance to target. Lower, target velocity. Green line is Kinect and blue is LMS-511.

The upper diagram in Figure 9 shows the distance to target in mm at sample 350 to 600 for one of the experiments. The blue line is the distance measured by the LMS-511 and the green is the distance measured by Kinect (mean distance of the central region). All experiments resulted in more or less the same diagrams, with the Kinect overestimating the distance at ranges beyond 5m.
The lower diagram of Figure 9 shows the derivative of the distance divided by the frame rate, at sample 350 to 600. It becomes very clear how noisy the Kinect data is at long range and it is first at distances below 5m that the minima start to overlap. The red triangles have been marked automatically by finding the local minima and they coincide with the subjects steps, i.e., when the velocity is low.

5. Discussion and Conclusions

We have presented a proof of concept for a THz sensor management system based on structured light and laser scanning. The fact that structured light is a suitable solution for pose tracking has already been established by the success of Microsoft's Kinect. We have shown the strengths and limitations of the technique and proved that it would be difficult to build a robust system solely on structured light. For most objects and conditions the technique works well but the problems with low reflecting targets and sensitivity to bright light would result in missed detections and inaccurate tracking in several scenarios. However, when combined with robust tracking from a laser scanner it looks very promising. The 2D tracking provides a graceful degradation when structured light fail and the data is accurate enough for making pose assumption. The data collected during the dynamic experiments show this, for example in Figure 9, where the subject's steps have been automatically detected. Another possible extension is arm tracking using the distribution of the laser measurements.
All tests so far have only involved on single target, but extending the scenario to two persons is straightforward as long as they do not occlude each other. It might even be possible to track even more persons, and also to deal with some occlusions, but the robustness of the system would then be impaired and this have been a key aspect in this work.
The data from the sensors have been saved and processed offline but OpenNI's pose tracking algorithm itself work in real time and we see no issues with making a complete management system that performs in real time.
Future work includes modifying the pose tracking initialization of OpenNI to make use of the available tracking. An alternative is to develop standalone pose tracking based on the fused data. It is also relevant to compare the pose tracking of OpenNI with the pose tracking included in Microsoft's software.

References

1. Ken B. Cooper, Robert J. Dengler, Nuria Llombart, Bertrand Thomas, Goutam Chattopadhyay and Peter H. Siegel, "THz Imaging Radar for Standoff Personnel Screening," Terahertz Science and Technology, IEEE Transactions on , vol.1, no.1, pp.169-182, Sept. 2011
2. Michael C Kemp, "A review of millimetre-wave and terahertz technology for detection of concealed threats," Infrared, Millimeter and Terahertz Waves, 2008. IRMMW-THz 2008. 33rd International Conference on , vol., no., pp.1-2, 15-19 Sept. 2008
3. Maria Axelsson, Mikael Karlsson and Staffan Rudner, "A multi-camera positioning system for steering of a THz stand-off scanner", Proc. SPIE 8022, 80220L (2011);
4. OpenNI (accessed 2012): http://openni.org/About.aspx
5. Microsoft Kinect (accessed 2012): www.microsoft.com/enus/kinectforwindows
6. SICK LMS-511 (accessed 2012): mysick.com/saqqara/im0037514.pdf
7. LGPL (accessed 2012): http://www.gnu.org/licenses/lgpl.html
8. Nick Carey, “Established Pedestrian Walking Speeds”, Portland State University, http://www.westernite.org/datacollectionfund/2005/psu_ped_summary.pdf (accessed 2012)