NATURAL USER INTERFACE SENSORS FOR HUMAN BODY MEASUREMENT

J. Boehm

University College London, Department of Civil, Environmental and Geomatic Engineering, Gower Street, London
WC1E 6BT UK

Commission V, WG V/4

KEY WORDS: Human Body Measurement, Range Camera, Kinect, NUI

ABSTRACT:

The recent push for natural user interfaces (NUI) in the entertainment and gaming industry has ushered in a new era of low cost three-dimensional sensors. While the basic idea of using a three-dimensional sensor for human gesture recognition dates some years back it is not until recently that such sensors became available on the mass market. The current market leader is PrimeSense who provide their technology for the Microsoft Xbox Kinect. Since these sensors are developed to detect and observe human users they should be ideally suited to measure the human body. We describe the technology of a line of NUI sensors and assess their performance in terms of repeatability and accuracy. We demonstrate the implementation of a prototype scanner integrating several NUI sensors to achieve full body coverage. We present the results of the obtained surface model of a human body.

1. INTRODUCTION

Human body measurement has an established history of measurement systems and applications over the past 30 years. Applications are varied and range from medical applications to applications in fashion and entertainment. Measurement systems are typically purpose-built optical scanners. The optical measurement principles employed by existing commercial solutions are laser line triangulation, active triangulation using white light pattern projection and monocular vision. An overview of systems and principles is given by (D'Apuzzo, 2005). Literature reports prices of commercial scanners ranging from $35,000 up to $500,000, which has prevented the widespread use of these systems so far.

Natural User Interfaces (NUI) have been promoted for some years as the natural successor and addition to Touch User Interfaces (TUI) and Graphical User Interfaces (GUI). The idea is to free the user from having to hold an input device such as a mouse or a stylus or interact on a predefined surface such as a touch screen. Instead the user's natural gestures such as waving and pointing are to be recognised and interpreted as input. Different sensor systems from monocular cameras to time-of-flight cameras have been suggested to capture a user's gestures.

However it was not until Microsoft's introduction of the Kinect as a NUI controller to their video game console Xbox 360 that a NUI sensor became widely available at a consumer price. The impact on the market was immediate. One million Kinect sensors were sold in just 10 days after the launch (Microsoft Corp., 2010). Adding to these numbers more than 10 million units were sold within the first 5 month.

This easily makes it the 3D sensor with the highest number of units sold at probably the lowest price, which has dropped below $99 by the time of writing. While originally intended only for the use with Microsoft's video game console, the sensor soon attracted applications beyond gaming. However, since the sensor is tuned to recognize the human body and its pose, applications to measure the human body are the most evident.

Within this paper we will demonstrate the application of a NUI sensor to human body measurement. We will describe the sensor characteristic beyond its specifications given by the manufacturer. In order to determine its fit for purpose we also report on our tests of the sensor's repeatability and accuracy. While tests of Kinect-like sensors have been performed before, we add to these in that we test not only single units but a whole set of units to show variations due to production tolerances. We report on our prototype implementation of an 8 sensor set-up and show first data sets captured with the system.

2. RELATED WORK

Weiss et al. (2011) have proposed a single sensor body scanner for home use based on the Microsoft Kinect. In order to capture the full body the user has to move into different poses in front of the fixed sensor. When users are moving into different poses their body shapes are obviously changing. Approaches based on the single fixed sensor principle thus have to accommodate for the changes in shape. The authors use a body model named SCAPE which considers 3D body shape and pose variations. The full 3D model thus is not a direct result of sensor readings but a combination of sensor readings and an underlying body model.

Newcombe and Davison (2010) have developed a structure from motion (SFM) approach to integrate depth maps from a moving Kinect sensor. The system has been further developed into the KinectFusion system (Newcombe et al., 2011). A single sensor is slowly moved around an object or a scene to fully capture it. The main contribution is the real-time capability of the system, which allows a user to interactively build (capture) a full scene. The downside to capturing whole body models is that due to the nature of the SFM approach, displacements between frames should be small to allow for optimal alignment. Thus motion is slow and it takes some time to capture a full body model during which the captured human may not move. Other notable contributions of this work include the innovative representation of the scene as volumetric elements and the introduction of bilateral filtering to depth maps from a NUI sensor.

3. SENSOR CHARACTERSITICS

While Microsoft were the first to introduce a NUI sensor at a consumer price to the mass market, it is important to understand that they did not develop the sensor completely on their own. The Kinect is a complex combination of software for gesture recognition, sound processing for user voice locating and a 3D sensor for user capture. The actual 3D sensor contained in the Kinect is based on a system developed by PrimeSense and implemented in a system on a chip (SOC) marketed by PrimeSense under the name PS1080.

PrimeSense has licensed this technology to other manufacturers as well, among them ASUS and Lenovo. A PrimeSense based sensor is currently available in different products or in different packages: the PrimeSense Developer Kit, the Microsoft Kinect and the ASUS Xtion (see Figure 1).

PrimeSense describe their 3D sensor technology as "LightCoding" where the scene volume is coded by near infrared light. Without internal details being available the system can be characterized as an active triangulation system using fixed pattern projection. The fixed pattern is a speckle dot pattern generated using a near infrared laser diode. The triangulation baseline in-between projector and camera is approximately 75 mm.

PrimeSense only give few specification of their reference sensor design listed in Table 1. Notably any accuracy in z direction is missing. Such performance criteria have to be established in dedicated test, which we will come to in the next section. One of the specifications given however is quite striking. While the point sampling of a single depth frame is quite low at only VGA resolution (640 x 480), this number has to be seen in relation to the frame rate. If we multiply the number of points of a single frame with the frame rate of 30 frames per second we

Field of View (Horizontal, Vertical, Diagonal)	58° H, 45° V, 70° D
Depth Image Size	VGA (640x480)
Operation range	0.8m-3.5m
Spatial x/y resolution (@2m distance from sensor)	3mm
Maximal image throughput (frame rate)	60fps

Table 1. Specifications of the PrimeSense reference sensor design as given by the manufacturer.

receive a sampling rate of 9216000 points per second. This outperforms current terrestrial laser scanners by an order of magnitude.

When we consider the maximum opening angle of the sensor which in one direction is 58 degree (this determines the field of view) it becomes clear by simple geometry that in order to cover the full body of an average sized person on one side we need a stand-off distance of approximately 1.8 m. Firstly when we want coverage from all sides, for example from four sides, this would lead to a very big footprint of a multi-sensor system. But secondly we also must take some basic photogrammetric rule of thumb into consideration.

As mentioned above the triangulation base is only 75 mm, which naturally limits the distance which can be measured reliably. While an exact limit to the base to height ratio cannot be given in the general case, but needs to be established on a case to case basis, we can assume as a rule of thumb that the base to height ratio should not fall below 1:16 (refer for example to Waldhäusl and Ogleby, 1994 or Luhmann, 2000). With the given base of 75 mm we should therefore not exceed a distance of approximately 1.2 m.

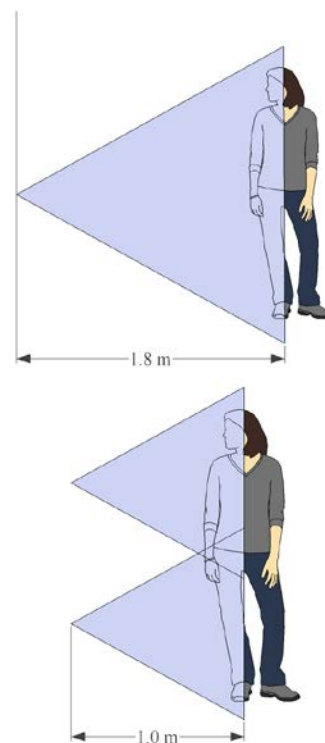Figure 1. Overview of different products available using PrimeSense's 3D NUI sensor technology

Figure 2. Graphical comparison of one sensor at a larger stand-off distance (top) or two sensors on top of each other at a shorter stand-off distance (bottom).

Figure 2 depicts the two different situations of either a longer stand-off distance or a shorter one. The shorter stand-off distance requires at least two sensors to be stacked on top of each other to achieve full coverage. However it gives the advantage of better triangulation accuracy and a more compact setup.

4. SENSOR TESTS

Since the manufacturer does not specify sensor repeatability and accuracy in depth, these quantities have to be established in suitable tests. Such tests have been carried out by different research labs, for example by (Menna et al., 2011). We have designed our own test strategy which separates repeatability (or precision) and accuracy. In addition we do not perform the tests on a single unit of one sensor model only, but we test several units in order to establish variations due to manufacturing tolerances. We also consider interference generated from additional sensors which overlap the field of view of the sensor under test.

As the rough photogrammetric estimates described above have shown, the sensor cannot be expected to provide reliable depth measurements at long distances. Thus we keep the distances reasonable for all tests following. We aim at measuring objects at approximately 1 m distance.

4.1 Repeatability

We test repeatability by observing two spheres in the field of view of a sensor over time. The sequence of depth measurements is recorded and later single frames of the recording are extracted and evaluated. The quantity we measure is the distance of the two spheres which of course is kept constant over the duration of the measurements. Since we are only interested in repeatability there is no need for a reference value of the distance.

Figure 3 shows the setup for this test. The sensor tested is the leftmost sensor. The distance of the sensor to the spheres is approximately 1 m. The distance of the spheres is approximately 0.5 m.

Two further sensors are added to test interference. Since the

Figure 3. Test set up for repeatedly measuring the distance of two spheres, both with and without interference from other sensors.

sensor uses static pattern projection, two sensors potentially generate some interference, when their field of view overlaps. This interference can occur in two forms. For one when two projectors illuminate a common area the brightness, roughly speaking, doubles. This can create sensor saturation and as a consequence creates a blind spot on the sensor or a gap in the depth measurement. This occurs most often on highly reflective surfaces. The API to the PrimeSense NUI sensor allows adapting sensor gain to compensate for this. However this is not a trivial procedure and is highly dependent on the scene. The second form of interference which we are interested in occurs when the projected dot patterns overlap and the sensor actually miss-matches the sensed pattern with the stored pattern. This situation occurs less frequently and it is almost unpredictable if it occurs at all or how strong the effect is.

In order to quantify this effect we place a second sensor at a distance of 0.5 m to the right of the sensor under test and
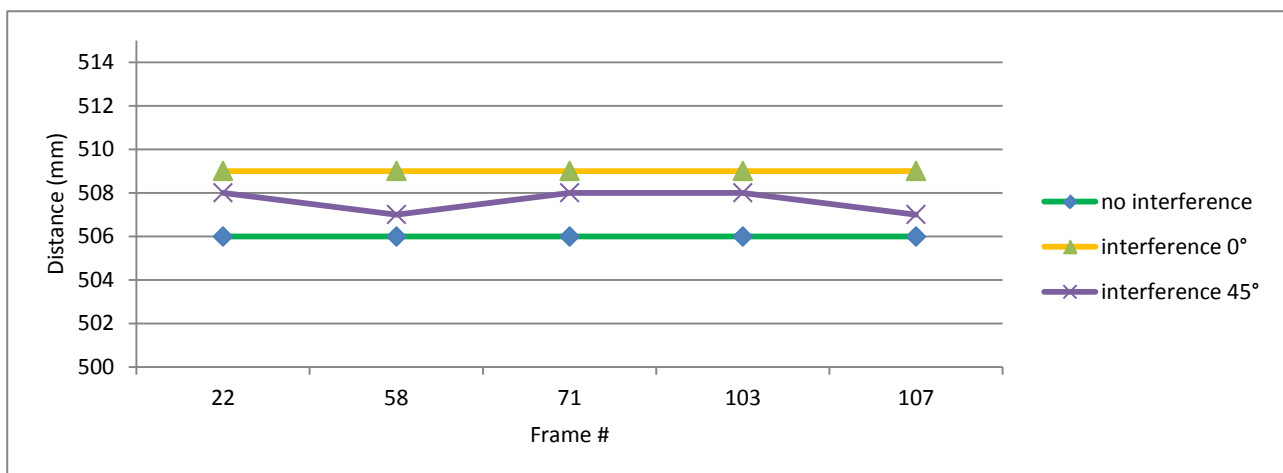

Figure 4. Repeatability of the measurement of the distance of two spheres. Three different scenarios are tested: no interference, i.e. only one sensor is switched on, interference from a sensor at 0 degree tilt angle and interference from a sensor at 45 degree tilt angle.

oriented in the same direction (0 degree). A third sensor is placed at 1 m distance to the right of the sensor under test with a viewing direction tilted 45 degrees towards the first sensor.

Figure 4 shows the graph resulting from the tests performed. When only a single sensor is used, i.e. there is no interference, the results are constant at 506 mm to the mm. This indicates that the measurement of the distance of two spheres is perfectly repeatable. These results are very encouraging for using this sensor in a measurement task.

When a second sensor is activated the distance changes, albeit only slightly. If the viewing direction of the second sensor is at 0 degrees with respect to the viewing direction of the sensor under test, the graph shows again a constant distance measurement, but at an offset of 3 mm. This deviation is in the order of magnitude of the sensor accuracy that we expect at this distance. If the viewing direction of the second sensor is at 45 degrees the measurements deviate again only slightly alternating from 1 to 2 mm offset.

4.2 Accuracy

Our tests for absolute accuracy are based on the VDI/VDE guideline for acceptance test and verification of optical measuring systems 2634 part 2 (VDI, 2002). However not all aspects of the guideline could be met due to practical reasons. The guideline defines a test of a series of measurements of sphere distances, referred to as sphere spacing. Figure 5 shows the suggested arrangement of test length in a measurement volume of the guideline.

We built a pyramidal structure of five spheres, which provides 10 test lengths varying from 0.7 m to 1.2 m. The pyramid has a base of 1 m x 1 m and a height of 0.5 m. We can test length aligned with the directions from the corners of the base to the tip of the pyramid, along the sides of the base and along the diagonals of the base. While this is in part similar to the guideline's suggestion, it does not fulfil all aspect. We have established reference values for the sphere's distances using a phase-based terrestrial laser scanner. Out of experience we assume these values to be accurate to 1 mm. Clearly this does not replace the calibration certificate required by a full test according to the guideline.

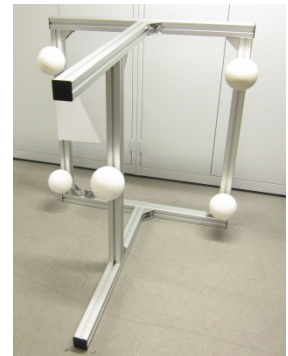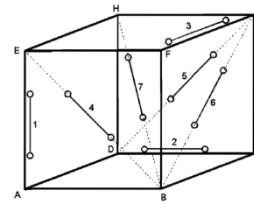We have performed these tests for 10 different units of the

Figure 6. VDI 2634 suggestion for an ideal arrangement of test length in the measurement volume (left). Realized pyramidal structure with 10 test length (right).

ASUS Xtion Pro sensor model. From the 10 test lengths we have recorded the maximum (positive) deviation to the reference length established by the laser scanner and the minimum (negative) deviation. The graph in Figure 7 summarizes the test results.

Clearly these results are much more disappointing than the results of the repeatability. Some sensor units have a sphere spacing error of almost 15 mm. The span from maximum to minimum deviation for some sensor units is 20 mm. However it is interesting to observe how the units differ. Some of them have a smaller span from minimum to maximum, in one instance less than 5 mm and are distributed well around 0 (see for example sensor #10). Others have larger spans and are clearly biased.

These results suggest that it is worth to test each individual unit and exclude units performing under par. Considering the low price of a unit, one might want to select the best units from a larger batch of sensors. The results are also a caution for not having too high expectations on the accuracy of these consumer grade sensors.

5. PROTOTYPE IMPLEMENTATION

Based on the estimates using photogrammetric considerations and the sensor performance tests, we can design a prototype for a scanner which provides full body coverage by integrating the
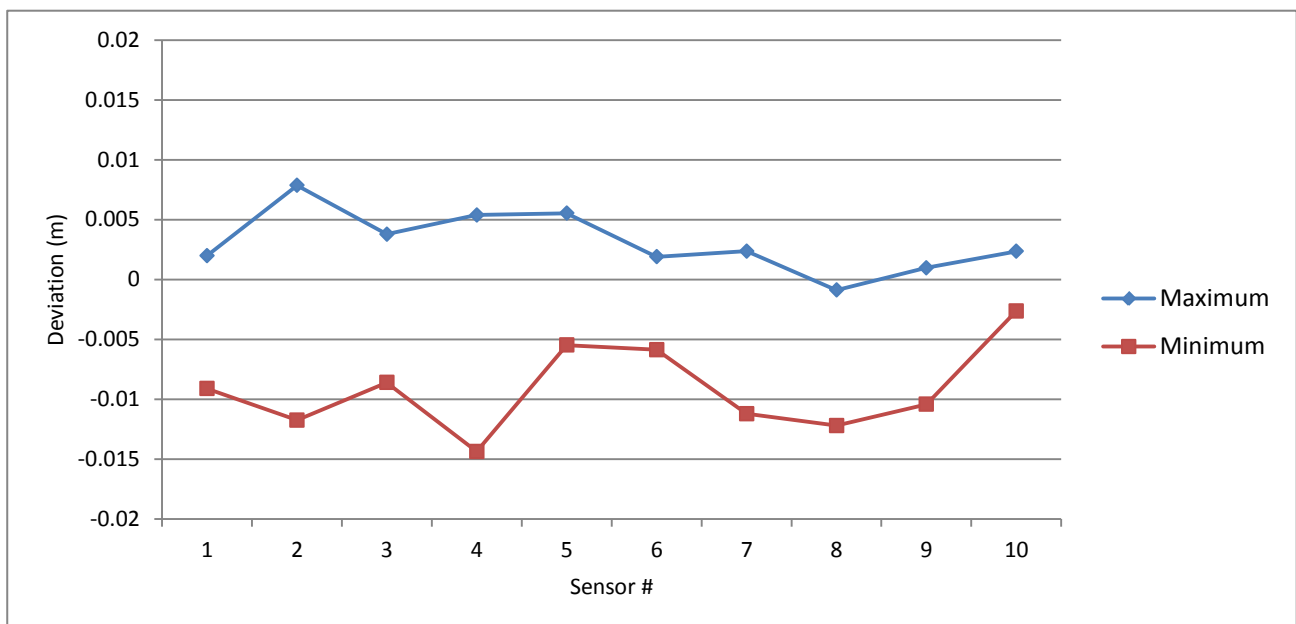

Figure 5. Maximum and minimum sphere spacing error of 10 different sensor units.

Figure 7. A prototype of a full body scanner integrating eight ASUS Xtion Pro NUI sensors in the corners of a cubical frame.

measurements of multiple NUI sensors. Figure 8 shows the frame of a cube, where 8 sensors have been mounted near the corners of the cube. The sensors are oriented so that they are targeting a central volume of measurement, where the person is to stand. Figure 8 also shows the calibration object representing a reference coordinate system, which is used to align the sensors to a common reference frame. This calibration is performed once when the sensor system is installed.

Using the point clouds from the 8 sensors and relying on the common reference frame it is obviously easy to integrate the separate point clouds into a single point cloud. Since each separate point cloud is delivered in the form of a raster representing the sensor matrix, each point cloud can be easily triangulated separately and for visualization purposes the meshes can be overlaid. Figure 10 shows such an overlay of meshes of a full body model from different points of view.

As discussed above the NUI sensor used in this prototype, as generally all the sensors based on the PS1080, contains a substantial amount of noise. This becomes particularly visible when the point cloud is meshed and rendered with an oblique light source. (Newcombe et al., 2011) have suggested using bilateral filtering, originally developed by (Tomasi and Manduchi, 1998), on the PS1080 depth map to reduce this noise. We use the speed optimized implementation of the bilateral filter from (Paris and Durand, 2006). Figure 9 shows the effect the filtering has on the raw mesh at different filter settings. The mesh from the unfiltered points is shown on the
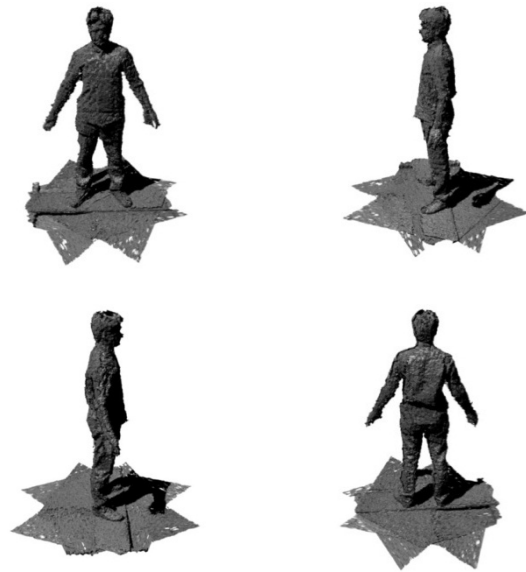

Figure 9. A full body model integrated from 8 sensor readings of unfiltered point clouds.

left, the middle shows a modest filtering and the right shows strong filter coefficients.

6. CONCLUSIONS

The NUI sensor based on the PrimeSense PS1080 has been shown to be well suited for human body measurement. This comes as no surprise since the sensor has been specifically designed to recognise human gestures. Using this consumer product has the overwhelming advantage of reducing sensor costs by several orders of magnitude compared to purpose-built sensors.

However we have shown that some care has to be taken when designing a system based on low-cost NUI sensors. Specifically the distance of sensor to object has to be adapted in order not to compromise sensor performance. Even at optimal distances we have to accept some level of inaccuracies and noise as the sensor test have shown. This has to be considered when a specific application is targeted. For visualization specialized filters exist to reduce the noise and produce visually pleasing surfaces.

Figure 8. A mesh of the original unfiltered point cloud shows the noise of the sensor (left). Applying a bilateral filter can produce visually fair surfaces at varying smoothness levels, depending whether modest (middle) or strong (right) filter parameters are chosen.

7. REFERENCES

D'Apuzzo, N., 2005. Digitization of the human body in the present-day economy, Proc. SPIE 5665, 252, DOI:10.1117/12.581305

Luhman, T., 2000. Nahbereichsphotogrammetrie, Wichmann.

Menna, F., Remondino, F., Battisti, R. and Nocerino, E., 2011. Geometric investigation of a gaming active device, Proc. SPIE 8085, 80850G.

Microsoft Corp., 2010. Kinect for Xbox 360 Hits Million Mark in Just 10 Days, Press Release, Redmond, Washington Nov. 15

Newcombe, R.A. and Davison, A.J., 2010. Live dense reconstruction with a single moving camera, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1498 – 1505.

Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S. and Fitzgibbon, A., 2011. KinectFusion: Real-Time Dense Surface Mapping and Tracking, ISMAR '11 Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127 – 136.

Paris, S. and Durand, F., 2006. A fast approximation of the bilateral filter using a signal processing approach, Proceedings of the European conference on computer vision.

Tomasi, C. and Manduchi, R., 1998. Bilateral filtering for gray and color images. In Proceedings of the International Conference on Computer Vision (ICCV).

VDI, 2002. VDI/VDE 2634 Part 2 - Optical 3D measuring systems – Optical systems based on area scanning

Waldhäusl, P. and Ogleby, C., 1994. 3x3-Rules for Simple Photogrammetric Documentation of Architecture, International Archives of Photogrammetry and Remote Sensing, Volume XXX, Part 5, Melbourne, 1994, pp. 426-429.

Weiss, A., Hirshberg, D., Black, M. J., 2011. Home 3D body scans from noisy image and range data, IEEE International Conference on Computer Vision, Barcelona, pp. 1951-1958.