6th Congress of Alps-Adria
Acoustics Assosiation
16.-17. October 2014
Graz, Austria

Austrian Acoustics Association
Österreichische Gesellschaft für Akustik

Alps Adria Acoustics Association

ISBN: 978-3-200-03921-6

EVALUATION OF THE SPATIAL RESOLUTION ACCURACY OF THE FACE TRACKING SYSTEM FOR KINECT FOR WINDOWS V1 AND V2

Clemens Amon, Ferdinand Fuhrmann

Joanneum Research (clemens.amon@joanneum.at, ferdinand.fuhrmann@joanneum.at)

Abstract: *In this paper we present the evaluation results for estimating the area of detection as well as the rotation accuracy of two versions of the face tracking system of the Microsoft Kinect sensor. The tested sensors are the Kinect for Windows V1 and V2 sensors in the use with the respective Software Development Kit. The tests were performed using an artificial head which was moved along a grid. The aim of this work is to prove and extend the manufacturer's data regarding the technical specifications for the particular device. Results show that the Kinect for Windows V2 provides a greater area of detection as well as a higher head rotation accuracy than its previous version, the Kinect V1.*

Keywords: face tracking, Microsoft Kinect, spacial accuracy, area of detection, rotation accuracy

1. INTRODUCTION

In multi media computing as well as for rehabilitation purposes, the demand for cost-effective head position tracking devices is increasing. The growing interest within the scientific scene in gaming tracking systems, like the Nintendo Wiimote, the Sony EyeToy or the Microsoft Kinect sensor emphasises this trend [1, 2]. Especially in the field of electro acoustics, methods to accurately track a moving object at low latency are required for optimal sound reproduction at the desired point in the room [3]. In this work we evaluated the face tracking system of the 2010 released Microsoft Kinect sensor V1 and the 2014 pre-released Microsoft Kinect sensor V2 in terms of their detection area in the x-y plane. Furthermore, we determined the head rotation accuracy on different measurement points (MPs) within the area of detection of both sensors.

The Kinect V1 is a depth sensor based on the structured-light technique, which is based on the emission of light beams in a known pattern, which are projected to the environment in the field of view of a camera. The recorded deformation of the light beams in reference to the known pattern provides information about the structure of the recorded scene [4]. In contrast to this technique, the Kinect V2 sensor is based on the Time-of-flight (TOF) principle. Short infrared (IR) light bursts are sent out, the reflections of the signal are recorded and the delay between the emitted and the received signal is calculated. Hence, these delays also involve 3D room information [5]. Apart from the applied technique to get depth information of the scene, the two devices differ in their specifications regarding the depth sensor range, image resolution, the audio sampling frequency (both devices are equipped with a 4 channel microphone array), the field of view and the systems latency. Table 1 lists the specifications of the two devices in detail.

Component	V1	V2
Technique	Structured-light	Time-of-flight
Depth Sensor	1.8 to 3.5 m	1.3 to 3.5 m
IR Depth Image	320 x 240	512 x 424
Colour Image	640 x 480	1920 x 1080
Infra-red Image	no IR	512 x 424
Audio Stream	16 kHz, 16-bit	48 kHz, 16-bit
Field of View hor:	57 degrees	70 degrees
Field of View ver:	43 degrees	60 degrees
Minimum Latency	102 ms	20-60 ms

Table 1: Kinect for Windows V1 and V2 Specifications [6, 7, 8].

2. EXPERIMENTAL SETUP

To compare the two devices, we conducted experiments for estimating the detection area and the rotation accuracy, both in the horizontal plane. Measurements are performed under standardised and constant light conditions in a ITU-R BS 1116-1 conform hearing room [9] using an artificial head, a measurement grid, and a turntable for rotation accuracy tests. The light properties for the halogen spots and diffuse light setup under which all tests are performed are listed in table 2.

Light System	Halogen Spots and Diffuse Light
ISO	100
Aperture	2.8
Shutter Speed	1/25 s
Exposure Value [10]	7.8

Table 2: Light setting - listening room Joanneum Research (JR)

In sections 3 and 4 we present the measurement setup and routines as well as the results of the evaluation of the area of detection and the rotation accuracy for both sensors. Section 5 finally concludes this article.

3. DETECTION AREA

The measurements regarding the area of detection are performed by moving the artificial head along the grid lines in the x direction (see Figure 1). The coordinates at the extreme, where the Kinect sensor looses the position of the head are recorded. It should be noted, that the two sensors use different algorithms to detect the head and the face. The Kinect V1 sensor is able to detect a face without torso, while the Kinect V2 sensor needs a torso to identify the head and the face. Figure 2 shows the grid, the position of the Kinect sensors and the MPs 01 - 11 where rotation accuracy measurements were performed. The dark shaded purple area marks the evaluated area of detection in the x/y-plane for the Kinect V1 sensor and the lighter blue area marks the area of detection in the x/y-plane for the Kinect V2 sensor. It can be seen, that the measured detection area for the Kinect V1 sensor is slightly smaller than specified by the manufacturer (horizontal field of view). We measured a field of view of $\sim 46°$. For the Kinect V2 Sensor the results are in accordance with the manufacturer's specification.

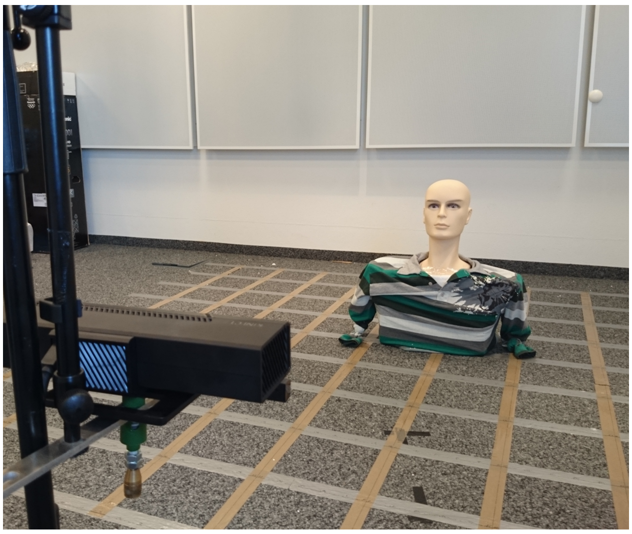

Figure 1: Artificial head placed on the measurement grid.

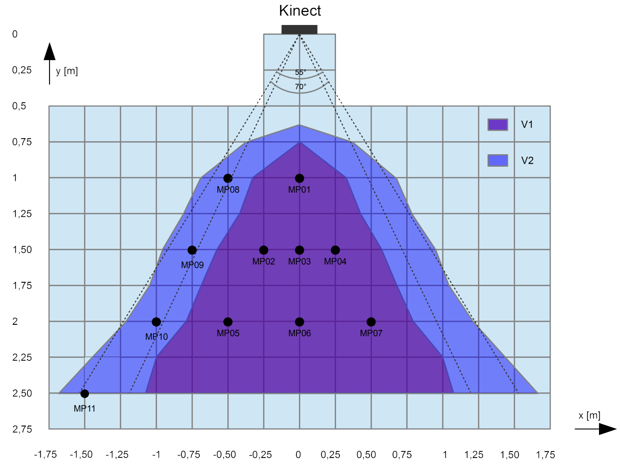

Figure 2: Measurement grid with MPs 1 - 11 and the manufacturer specified horizontal field of view.

4. ROTATION ACCURACY

The rotation accuracy of the two different sensors was measured by rotating the artificial head using an Outline ET250-3D automatic turntable. Starting at the $0°$-position the artificial head was clock-wise and counter-clock-wise rotated in steps of $1°$ until the detection of the face was lost. The detected angle at each turntable rotation step was recorded. These measurements were done at MP01, MP02, MP05 and MP06 for both sensors, at MP03 - MP05 for the Kinect V1 sensor and at MP08 - MP011 for the Kinect V2 sensor. MP04 and MP07 are not further mentioned, as the measurement results are identical with the results at MP02 and MP05, mirrored along the x-axis. Figure 3 shows the measured angle values φ_{out} over the turntable positions φ_{in}. It can be seen, that the Kinect V2 detects the face in a higher range of rotation (V2: $-60°$ - $55°$, V1: $-42°$ - $57°$). The deviation of the points from the optimal values is shown in Figure 4. The head rotation measured of the Kinect V2 in most turntable rotation positions show a considerably lower error than the values at the same positions measured with the Kinect V1 sensor. This improvement of the Kinect V2 compared to the Kinect V1 can be seen in the results of all MPs inside the area of detection of the Kinect V1. Table 3 lists the mean errors of two different rotation ranges for the two sensors at four MPs. It is evident that the Kinect V2 sensor detects head rotations in all tested positions more accurately than the Kinect V1 sensor.

Device	Range	Error [°] at			
		MP01	MP02	MP05	MP06
Kinect V1	$\pm30°$	2.0	2.4	4.8	2.4
	$\pm40°$	2.8	2.8	5.8	3.0
Kinect V2	$\pm30°$	0.7	2.4	2.3	2.3
	$\pm40°$	1.0	2.5	3.2	2.3

Table 3: Average rotation accuracy error in degrees at MPs 01, 02, 05 and 06 in rotation ranges $\pm30°$ and $\pm40°$. Errors are calculated in two different rotation ranges of $\pm30°$ and $\pm40°$ from the frontal orientation of the head.

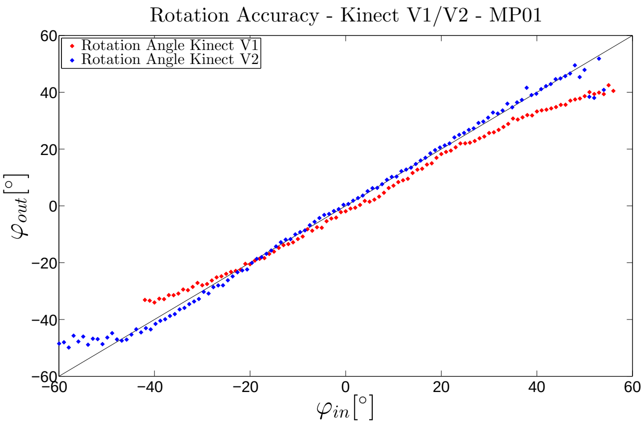

Figure 3: Rotation accuracy at MP01, Kinect Sensor V1 and V2.

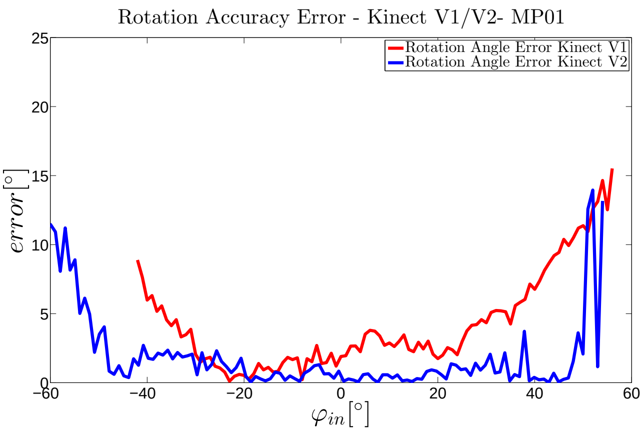

Figure 4: Rotation accuracy error at MP01, Kinect Sensor V1 and V2.

In Addition, we performed four measurements using the Kinect V2 outside the area of detection of the Kinect V1 sensor. The results of the accuracy measurement at MP08 and MP11 are presented in Figure 5. At MP08 the rotation is detected similar to the results at MP01, except a noticeable bend in the rotation positions around the $0°$-position. Furthermore, it can be seen that the range as well as the accuracy drifts to one direction if the head is positioned aside the middle line, off the frontal line to the sensor ($x \neq 0$). The results of the measurements at MP11 show a considerably offset of up to $+10°$ for rotation positions where the head faces the sensor. The accuracy decreases rapidly for counter-clock-wise steps ($< 0°$) in sensor averted positions. Here, the before mentioned offset as well as the drift in accuracy and range of detection are evident.

5. CONCLUSION

We evaluated the Kinect V1 sensor and its follow-up model the Kinect V2 sensor in terms of their area of detection in the horizontal plane. We compared our results with the manufacturer's specifications regarding the field of view. While we could confirm the manufacturer's specifications for the Kinect V2, we measured a smaller field of view of the Kinect

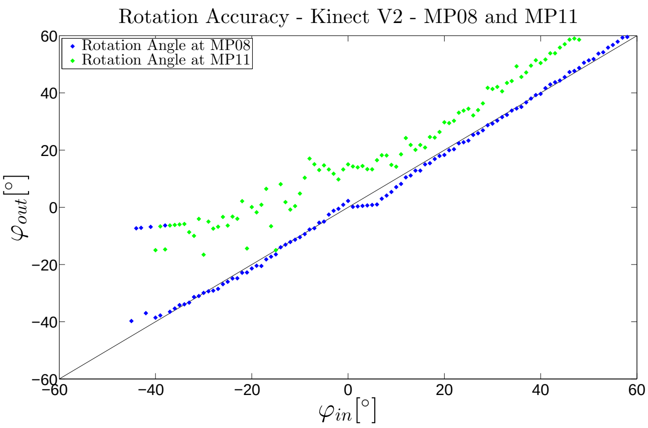

Figure 5: Rotation accuracy at MP08 and MP11, Kinect sensor V2.

V1 sensor. Furthermore, we performed tests regarding the rotation accuracy of the two sensor devices. Compared to the Kinect V1, the Kinect V2 performs better in nearly all tested measurement positions. In general, it can be stated that the Kinect V2 sensor features significant improvements compared to its previous model. We see limitations caused by the relatively low refresh rate of 30 frames per second (fps). Professional tracking systems offer refresh rates of up to 600 fps at comparably image resolution [11, 12]. Especially in acoustic tracking applications even small movements of the tracked face could yield a lack or misinterpretation of sound source localisation as the new position is updated to slow. Microsoft offers with the Kinect V2 a cost-efficient and powerful multi-modal sensor device.

REFERENCES

[1] D. Wiebusch, M. Fischbach, A. Strehler, M. E. Latoschik, G. Bruder, and F. Steinicke, "Evaluation von Headtracking in interaktiven virtuellen Umgebungen auf Basis der Kinect," in *Proceedings of the GI-Workshop VR/AR*, 2012.

[2] Microsoft kinect vs playstation move vs nintendo wii. [Online]. Available: http://www.pocket-lint.com/news/104272-kinect-vs-move-vs-wii

[3] N. Z. J. Hiipakka and T. Takala, "The importance of head movements for binaural room synthesis," in *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, 2001.

[4] Z. Zhang, "Microsoft Kinect sensor and its effect," *MultiMedia, IEEE*, vol. 19, no. 2, pp. 4–10, Feb 2012.

[5] Collaboration, expertise produce enhanced sensing in xBox One. [Online]. Available: http://blogs.microsoft.com/blog/2013/10/02/coll-expertise-produce-enhanced-sensing-in-xbox-one/

[6] Kinect for Windows sensor components and specifications. [Online]. Available: http://msdn.microsoft.com/

en-us/library/jj131033.aspx

[7] XBox Kinect 2 specs. [Online]. Available: http://www.develop-online.net/news/next-xbox-leak-reveals-kinect-2-specs/0114096

[8] Everything Kinect 2. [Online]. Available: http://123kinect.com/everything-kinect-2-one-place/43136/

[9] Methods for the subjective assessment of small impairments in audio systems. [Online]. Available: http://www.itu.int/rec/R-REC-BS.1116/en

[10] Lichtmessung mit der Digitalkamera. [Online]. Available: http://www.bagru.info/reports/barrierefreies_institut/lichtmessug_mit_der_digitalkamera.pdf

[11] T-series - precise motion capture in any volume anywhere. [Online]. Available: http://www.vicon.com/System/TSeries

[12] Precision position tracking for virtual reality. [Online]. Available: http://www.worldviz.com/products/ppt