A Comparison of Depth Sensors for 3D Object Surface Reconstruction

Vinod Gutta[1], Edward D. Lemaire[2,3], Natalie Baddour[3], and Pascal Fallavollita[4]

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada – K1N 6N5
[2] Center for Rehabilitation Research and Development, Ottawa Hospital and Research Institute, Ottawa, Ontario, Canada – K1H 8M2
[3] Department of Mechanical Engineering, University of Ottawa, Ottawa, Ontario, Canada – K1N 6N5
[4] Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada – K1N 6N5

Abstract— **The ability of depth cameras like Kinect to capture a scene's depth information in three-dimensions, along with 2D color RGB images, in real-time makes marker-less human motion capture a potential option for applications such as rehabilitation, robotics, education, etc. Various depth sensor technologies are commercially available, and selecting the appropriate depth sensor is highly dependent on the desired application. This research compared Kinect V2, Astra Pro, and RealSense D415 depth sensing technologies for object surface reconstruction within an interior daily living environment. Intel RealSense D415 was resistant to interference with multiple sensors and point cloud data at 1m range was more accurate than Kinect V2 and Astra Pro.**

Keywords— **Depth camera, marker-less, time-of-flight, structured light, stereoscopic.**

I. Introduction

Human motion capture with optical systems can be classified into two categories, marker-based and marker-less. For marker-based systems, a person wears either active markers that emit light or passive markers that reflect light, and three-dimensional (3D) marker positions are tracked in realtime. Examples of marker-based optical tracking systems are Vicon, Qualisys, and Codamotion.

For marker-less motion capture no markers are placed on the human body. Images are captured while a person moves in the system's field of view. Motion capture is based on processing these captured images.

Marker-less approaches typically use multiple RGB cameras or depth cameras. Finding 3D human body information from multiple RGB cameras requires extensive image processing and computation. Depth sensors based on infra-red light are independent of ambient lighting conditions in the scene and can also provide 3D data in real-time. These depth sensors may be convenient and easier to use than RGB cameras. Kinect for Xbox One (Kinect V2) has been used in the literature for marker-less gait analysis, but has not progressed to use in practice.

Human stride duration depends on walking speed [1]. Based on analysis from [2], 0.5 m/s is considered slow walking and speeds greater than 1.6 m/s are fast walking. Gait phases can have a minimum duration of 0.12s (fast walking at 1.75m/s, loading response and "pre-swing").

Depth sensors such as the Kinect V2 operate at 30fps, which implies that a frame is captured every 33.33ms. Depth sensors working at 30fps can identify all gait sub-phases, even during fast walking. However, proper identification of foot-off, which happens in 30ms for a 0.671m/s walking speed, is theoretically not possible [4]. Even at low walking speeds, the Kinect V2 cannot consistently track foot and toe-off moments [4, 5]. Furthermore, to avoid occlusion while walking and increase the accuracy of tracking, multiple sensors are required [6].

The purpose of this research was to find a suitable depth sensor to overcome the limitations of the Kinect V2 sensor for human foot 3D construction during fast walking. As a part of this study, three depth sensors were investigated: Kinect V2 (time-of-flight), Orbbec Astra Pro (structured light), Intel RealSense D415 (stereoscopic). The research outcomes provide a basis for developing new marker-less human movement analysis approaches.

II. Kinect V2, Astra Pro, RealSense D415

Table 1 shows typical specifications among Kinect V2, Astra Pro, and RealSense D415 sensors. Most RealSense D415 specifications are superior to those of the Kinect V2 and Astra Pro. RealSense D415 can be configured to run at various resolutions and speeds, and has a high color resolution of 1920 × 1080 and depth resolution of 1280 × 720.

A. Kinect V2

Kinect V2 has a wider field of view and uses time-of-flight technology. In brief, Infrared (IR) rays are projected onto an object and these rays reflect to a camera array after hitting an object. The object's depth is calculated based on IR time-of-flight [7].

B. Astra Pro

Astra Pro uses structured light technology to determine depth information. Known features are projected onto an object and observed using an IR camera. 3D construction is based on image correlation and triangulation [8].

C.

C. RealSense D415

Active IR stereoscopic technology is used to find depth data. An IR projector projects a texture pattern onto the scene, to find more matching features between stereo IR images. Depth is estimated based on the disparity between keypoints in stereo images [9].

III. METHODOLOGY

Experiments were performed with Kinect V2, Astra Pro, and RealSense D415 to compare depth image quality and 3D point cloud quality. Point clouds were generated directly from the depth cameras.

A. Single sensor depth images

A standard-size basketball of radius 119.3mm and two boxes with dimensions of 342mm x 162mm x 115mm were placed on the ground plane. Depth sensors were positioned approximately 1m from these objects (Fig. 1). Depth images of these objects were captured with a single sensor and 100 frames were averaged. This paper focused on sphere analysis.

Table 1 Specification comparison

	Kinect V2	**Orbbec Astra Pro**	**Intel RealSense D415**
Technology	Time of flight	Structured Light	Stereoscopic
Dimensions	Length: 250 mm Width: 66mm Height: 67mm	Length: 165mm Width: 30mm Height: 40mm	Length: 99mm Width: 20mm Height: 23mm
Color Resolution	1920 x 1080 at 30fps	640 x 480 at 30fps	848 x 480 at 60fps
Depth Resolution	512 x 424 at 30fps	640 x 480 at 30fps	848 x 480 at 60fps
Max. depth speed	30fps	30fps	Up to 90fps
Depth range	0.5m – 5 m	0.6m – 8m	0.3m – 10m
Field of view	70.6° x 60°	60° x 49.5°	69.4° x 42.5°
Multiple sensors per computer	No	Yes	Yes
External Power supply	Yes	No	No

B. Multiple sensor depth images

The experimental setup remained the same as for the single sensor assessment, but with one additional depth sensor opposite to the first sensor placed at 1m from the objects (Fig. 1). Depth images were captured with both sensors active at the same time and averaged over 100 frames.

C. Sphere detection from point cloud

Background removal was necessary to segment the ball from the scene. The minimum background method was applied using 500 depth frames captured without the ball and box objects in the scene [10]. Then, the basketball was placed in front of the sensor on the ground plane. 100 depth frames were captured and background subtraction was applied to the depth data. A point cloud was created from each frame, containing only points belonging to the basketball. These points were fitted to a RANSAC sphere model [11].

Variables included the actual basketball radius (R) estimated radius (R_{est}), number of points in point cloud (N), number of points inliers to RANSAC sphere (N_m), estimated center (C_{est}), number of points farther than R from C_{est} (N_+), number of points nearer than R from C_{est} (N_-), distance of i^{th} model inlier point from C_{est} (d^i_m), distance of i^{th} farther point from C_{est} (d^i_+), and distance of i^{th} nearer point from C_{est} (d^i_-).

σ_+ is standard deviation towards the exterior of a sphere, σ_- is standard deviation towards the interior of sphere, and R_{rms} is the root mean square error. Results provided in Table 2 were the average of 100 iterations.

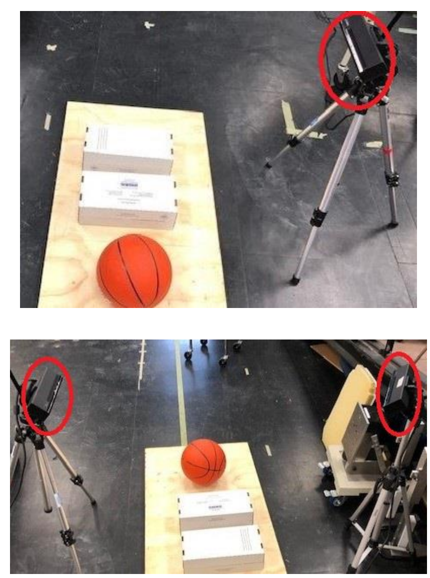

Fig. 1 Top: single sensor setup, Bottom: two sensors setup.

The standard deviations and root mean square error are given by

$$\sigma_+ = \sqrt[2]{\frac{\sum_{n=1}^{N_+} (R - d_+^n)^2}{N_+}} \quad (1)$$

$$\sigma_- = \sqrt[2]{\frac{\sum_{n=1}^{N_-} (R - d_-^n)^2}{N_-}} \quad (2)$$

$$R_{rms} = \sqrt[2]{\frac{\sum_{n=1}^{N_m} (R - d_m^n)^2}{N_m}} \quad (3)$$

IV. Results

A. Depth images

The depth images from the sensors were threshold to 2000 mm. Fig. 2 - 6 show depth camera images, where each pixel is a depth, scaled from 0-2000 mm to 0-255 for display. Black pixels indicate either no depth data available from the sensor or depth data range is greater than 2000 mm.

The depth images when using a single Kinect V2 (Fig. 2) had clean, sharp edges for both the rectangular shaped object and spherical object. Astra pro's depth images had many depth data gaps at object edges. RealSense D415's depth images had less missing depth data than Astra pro.

For two sensors capturing data simultaneously, Astra pro had extensive interference, resulting in missing depth data (i.e., dark pixels in Fig. 3). Kinect V2 had less interference with 2 sensors but the interference region was not consistent (Fig. 4). When averaged over 100 frames, the black pixel areas were reduced for Kinect (Fig. 5). Averaging would only be applicable for static objects.

The stereoscopic-based Intel RealSense D415 sensors did not have interference, even when both sensors capture depth data simultaneously (Fig. 6).

B. Point cloud

Intel RealSense D415's delivered a denser point cloud because of its high depth resolution. Error in estimated sphere radius was less than 1mm for both Astra Pro and RealSense D415 depth sensors. Intel RealSense D415 point cloud quality outperformed the other depth sensors in every aspect, other than the standard deviation of points towards the interior of the sphere (Table 2).

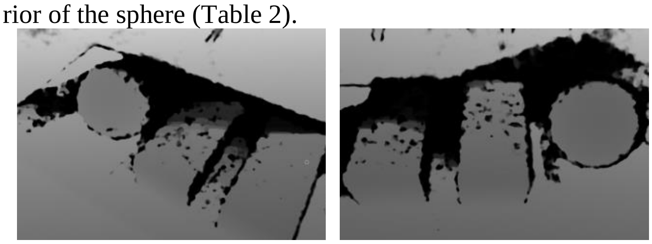

Fig. 3 Averaged depth images from two Astra Pro sensors

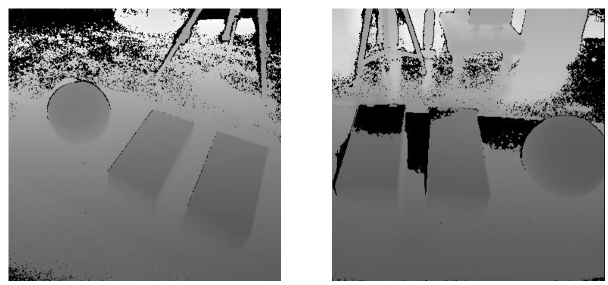

Fig. 4 Depth images from two Kinect sensors

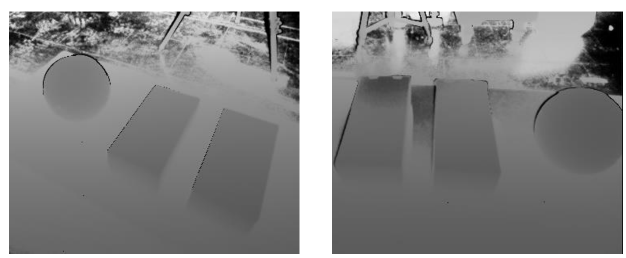

Fig. 5 Averaged depth images from two Kinect sensors

Fig. 2 Single Kinect V2 sensor averaged depth image

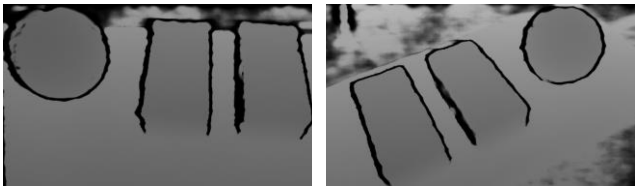

Fig. 6 Averaged depth images from two RealSense D415 sensors

Table 2 Sphere detection

	Kinect V2	**Orbbec Astra Pro**	**Intel RealSense D415**
N	8132.33	18589.8	26708
N_m	7654.43	18482.6	26708
R_{est}	113.83	120.25	118.74
R_{rms}	7.15	1.88	1.68
σ_+	5.79	1.86	1.37
σ_-	7.27	0.93	1.78

V. Conclusions

To construct 3D objects, data from multiple sensors are simultaneously required in order to avoid occlusion. The Kinect V2 and the Astra Pro sensors had interference when multiple sensors were used at the same time. The Intel RealSense D415 sensor did not display evidence of interference when multiple sensors were used simultaneously, and delivered a dense point cloud at 60fps. Therefore, the RealSense 415 is a feasible technology to construct 3D objects and also delivered less point cloud errors than the Astra Pro and Kinect Xbox One sensors at a range of 1m. The RealSense D415 dimensions, no external power supply requirement, and ability to use multiple sensors per computer are other advantages for a marker-less human movement analysis application.

Acknowledgment

This project is a part of the “CREATE-BEST” program, which is funded by the Natural Sciences and Engineering Research Council of Canada.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Zijlstra W, Hof AL (2003) Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait Posture 18(2):1-10 DOI 10.1016/S0966-6362(02)00190-X
2. Bohannon RW, Williams Andrews A (2011) Normal walking speed: A descriptive meta-analysis. Physiotherapy 97(3):182-189 DOI 10.1016/j.physio.2010.12.004
3. Hebenstreit F, Leibold A, Krinner S, Welsch G, Lochmann M, Eskofier BM (2015) Effect of walking speed on gait sub phase durations. Hum Mov Sci 43:118-124 DOI 10.1016/j.humov.2015.07.009
4. Goncalves RS, Hamilton T, Krebs HI (2017) MIT-Skywalker: On the use of a markerless system. 2017 Int Conf Rehabil Robot 205-210 DOI 10.1109/ICORR.2017.8009247
5. Kharazi MR, Memari AH, Shahrokhi A (2015) Validity of microsoft kinect TM for measuring gait parameters. 2015;(November):25-27.
6. Yang L, Yang B, Dong H, Saddik A El (2016) 3-D Markerless tracking of human gait by geometric trilateration of multiple kinects. IEEE Syst J 2016:1-11 DOI 10.1109/JSYST.2016.2553518
7. Sarbolandi H, Lefloch D, Kolb A (2015) Kinect range sensing: Structured-light versus Time-of-flight kinect. Comput Vis Image Underst 139:1-20 DOI 10.1016/j.cviu.2015.05.006
8. Orbbec Astra Pro. (2019) https://orbbec3d.com/products/. Accessed January 16, 2019.
9. Intel RealSense D400 Series. (2019) https://realsense.intel.com/stereo-depth-vision-basics/. Accessed January 16, 2019.
10. Stone E, Skubic M (2011) Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. Proc 5th Int ICST Conf Pervasive Comput Technol Healthc 2011:71-77 DOI 10.4108/icst.pervasivehealth.2011.246034
11. RANSAC Sphere - PCL (2019) http://docs.pointclouds.org/1.8.1/group__sample__consensus.html. Accessed January 16, 2019.

Corresponding Author: Vinod Gutta
Institute: University of Ottawa
Address: Ottawa, Canada
Email: vgutt079@uOttawa.ca