Comparison of depth cameras for three-dimensional Reconstruction in Medicine

CHIU, Chuang-Yuan, THELWELL, Michael, SENIOR, Terry <http://orcid.org/0000-0002-3049-5724>, CHOPPIN, Simon <http://orcid.org/0000-0003-2111-7710>, HART, John <http://orcid.org/0000-0002-6142-4515> and WHEAT, Jonathan <http://orcid.org/0000-0002-1107-6452>

Published version

CHIU, Chuang-Yuan, THELWELL, Michael, SENIOR, Terry, CHOPPIN, Simon, HART, John and WHEAT, Jonathan (2019). Comparison of depth cameras for three-dimensional Reconstruction in Medicine. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 233 (9), 938-947.

Article type:

Original Article

Corresponding Author:
Chuang-Yuan Chiu, Centre for Sports Engineering Research, Sheffield Hallam University, Broomgrove Road, Sheffield S10 2NA, UK
Email: c.chiu@shu.ac.uk

Comparison of Depth Cameras for 3D Reconstruction in Medicine

Chuang-Yuan Chiu [1], Michael Thelwell [1], Terry Senior [1], Simon Choppin [1], John Hart [1], Jon Wheat [2]

[1] Centre for Sports Engineering Research, Sheffield Hallam University, Sheffield, UK

[2] Academy of Sport and Physical Activity, Sheffield Hallam University, Sheffield, UK

Abstract

KinectFusion is a typical 3D reconstruction technique which enables generation of individual 3D human models from consumer depth cameras for understanding body shapes. The aim of this study was to: compare 3D reconstruction results obtained using KinectFusion from data collected with two different types of depth camera (time-of-flight and stereoscopic cameras) and compare these results with those of a commercial 3D scanning system to determine which type of depth camera gives improved reconstruction. Torso mannequins and machined aluminium cylinders were used as the test objects for this study. Two depth cameras, Microsoft Kinect V2 and Intel Realsense D435, were selected as the representatives of time-of-flight and stereoscopic cameras, respectively, to capture scan data for the reconstruction of 3D point clouds by KinectFusion techniques. The results showed that both time-of-flight and stereoscopic cameras, using the developed rotating camera rig, provided repeatable body scanning data with minimal operator-induced error. However, the time-of-flight camera generated more accurate 3D point clouds than the stereoscopic sensor. Thus, this suggests that applications requiring the generation of accurate 3D human models by KinectFusion techniques should consider using a time-of-flight camera, such as the Microsoft Kinect V2, as the image capturing sensor.

Keywords: 3D Scanning; Imaging; Sensor; Reconstruction; KinectFusion; Depth Camera; Accuracy; Reliability

1 Introduction

Consumer depth cameras, such as the Microsoft Kinect and Intel Realsense cameras, were introduced to the market during the past decade. They are cost-effective, portable and can capture both colour and depth in real-time [1, 2]. Consequently, consumer depth cameras have been widely used in 3D reconstruction [1].

Understanding body shapes (through 3D scanning) is an important application of 3D reconstruction as it enables rapid measurement of the human body with minimal physical contact [3]. Furthermore, 3D scanning enables more complex body measures (e.g. body surface area and segmental volume) to be collected directly [4]. This technique generates individual 3D human models that can be used in a range of applications, including anthropometric surveys [5, 6], virtual fitting of clothing [7], sports performance prediction [8], biomechanical analyses [9-12] and medical diagnosis [13-15].

KinectFusion [16] is a typical 3D reconstruction technique which enables the generation of individual 3D human models from consumer depth cameras for different purposes. This technique requires capturing images of a participant from different perspectives, with a distance of 1 - 1.5 metres between the camera and the individual. A series of algorithms merge the images collected by the depth cameras to generate an individual 3D human model [17, 18]. Ng, Hinton [19] compared body measurements (girths, surface areas and body volumes) obtained from KinectFusion using a Microsoft Kinect V2 sensor and found the results were highly correlated to those acquired using reference methods ($R^2 > 0.9$). Recently, researchers have applied this technique using different types of depth cameras to generate 3D human models for various applications, including 3D printing [20], clothes fitting [18], computer animation [21] and body measurements [22].

Structured light, time-of-flight, and stereoscopic cameras are the three representative types of depth cameras. Structured light cameras capture a projected light pattern and determine the distance between the camera and the object by observing the deformation of the pattern [23]. Time-of-flight based cameras calculate this distance by measuring the travel time-of-flight of the signal emitted from the projector and received by the sensors [23]. A stereoscopic device uses multiple cameras to capture images of an object and find corresponding points between image pairs to estimate the distance between the camera and object of interest [24]. The different principles used to generate depth images means that the accuracy and repeatability of depth detection from different depth sensors varies between devices [23, 25-27].

In the past, researchers have typically used Microsoft Kinect version 1 (Microsoft Kinect V1) and version 2 (Microsoft Kinect V2) as the representative structured light and time-of-flight cameras, respectively, to conduct comparison studies [23, 25-27]. Gonzalez-Jorge, Rodríguez-Gonzálvez [25, 26] showed that the time-of-flight based Microsoft Kinect version 2 captured more accurate and repeatable depth images at close range (1-2 m) than the structured light based Microsoft Kinect version 1. However, the 3D reconstruction results obtained from the time-of-flight and stereoscopic cameras have not been compared. The stereoscopic cameras, such as Intel Realsense sensors, have improved image resolution and capture frame rate compared to the Microsoft Kinect V2 sensors [24, 28], which might be a benefit for generating accurate and repeatable 3D models by KinectFusion techniques.

Given the multiple potential applications of 3D body scanning, the accuracy and repeatability of low-cost solutions is of interest to a range of research communities. Therefore, the aim of this study was to compare the 3D reconstruction results obtained using KinectFusion for scan data collected via two different types of depth cameras (time-of-flight and stereoscopic cameras). In addition, the results were compared with scan data acquired from a reference commercial 3D scanning system to determine which type of depth camera gives improved reconstruction results.

2 Materials and Methods

Four cylinders and three torso mannequins were used as the test objects in this study as shown in Figure 1. The dimensions of these objects were measured using a large caliper list in Table 1. A commercial 3D surface imaging system (3dMD; 3dMD LLC, Atlanta, USA) was calibrated and used to obtain 3D scan data of the test objects. Because the sizes of mannequins were slightly larger than the 3dMD scanning regions, multiple scans were applied to the torso mannequins and aligned by the functions of Meshlab (version 2016.12) [29] to generate reference torso meshes. According to the manufacturer and previous literature [30], the error of this system is less than 0.5 mm.

[insert Figure 1.]

Table 1 Dimensions of the test objects in this study

Scanning Objects	Width/ Diameter (cm)	Length/ Diameter (cm)	Height (cm)
Torso 1	46.0	20.0	59.3
Torso 3	36.6	24.0	59.6
Torso 2	29.6	17.2	55.9
Cylinder 1	8.7	8.7	32.1
Cylinder 2	11.2	11.2	42.1
Cylinder 3	16.0	16.0	42.2
Cylinder 4	22.5	22.5	40.0

A bespoke rotating camera rig with a stationary central platform was developed which enabled a depth camera to be mounted at an adjustable capture distance from the test object, shown in Figure 2. During the scanning process, an operator manually pushes the wheel-mounted rotating arm of the frame 360^o around the test object, allowing the depth camera to capture depth images of the test object from various directions. A Microsoft Kinect V2 was selected as the representative time-of-flight camera in this study as it can project powerful illumination and generate higher quality depth maps that other newer options (e.g. Lips DL or Asus Xtion Pro 2) as found during pilot testing. Intel Realsense D435 was used as the representative stereoscopic camera as it was the latest model at the time of conducting the tests of this study. In addition, Intel Realsense D435 provides a wider field of view than Intel Realsense D415. The wider field of view can minimize the capturing distance which might improve the accuracy and reliability. The technical specifications and depth camera settings used in this study are listed in Table 2. During each scanning trial, only one camera captured images to avoid any camera interference, as shown in Figure 2. The scanning time for each trial was approximately 10 seconds. The Microsoft Kinect V2 or the Intel Realsense sensors captured approximately 200 and 600 frames for each trial, respectively. Because of the stability of power supply, hardware compatibility, and file saving speed, the captured frame numbers were less than the theoretical values (300 for Microsoft Kinect V2 and 900 for Intel Realsense D435).

Table 2 The technical details and depth camera settings used in this study.

Sensor	Microsoft Kinect V2	Intel Realsense D435
Principles of depth measurement	Time-of-flight	Stereoscopic camera
Theoretical field of view	70.6°×60.0°[31]	91.2°×65.5°[32]
Image resolution	512×424	848×480
Set capture frame rate	30 frames per second	90 frames per second
Approximated real capture frame rate	20 frames per second	60 frames per second

[insert Figure 2.]

The depth cameras were mounted at a height of approximately 40 cm, so that the sensor axis was aimed at the centre of the scanning object. Four capture distances were used for each type of RGB-D sensor, with the sensors mounted at 75cm, 100 cm, 125cm and 150cm from the test object.

Two trained operators conducted four repeated scanning trials for each type of depth camera and capture distance to determine the inter-operator repeatability of the manually driven rotation scanning system. The test protocol consisted of both operators collecting scan data with one sensor and then collecting data using the another sensor. In other words, both operators conducted the scanning process at each distance without changing the depth camera being used. In total, 448 scanning trials (seven objects × two depth cameras × four capture distances × two operators × four repeated trials) were performed.

Images captured from the two depth cameras were then used as input for the KinectFusion techniques [16] to generate 3D point clouds. The resolution of KinectFusion was set as 256 voxels per metre for all scanning trials, which is similar to previous work [18]. To determine the effect of resolution, 128, 384 and 512 voxels per metre for all scanning trials were also applied for the scanning trials with 100cm capture distances. To understand the effect of a shorter scanning time on reconstruction quality, the image sets collected during the scanning trials at 100 cm capture distance were resampled using a ‘two-frame interval’ as input to the KinectFusion technique. The two-frame interval can be used to simulate halving the sensor frame rate or capturing time. The 3D scan data from the 3dMD system and the 3D point clouds generated with KinectFusion techniques were edited with bespoke software that applied random sample consensus algorithms [33] and the density filters to select a region of interest (i.e. deleting scanning stage, floor, and the rotation platform, etc.) as shown in Figure 3.

[insert Figure 3.]

Three-dimensional (3D) reconstruction results for data collected with the 3dMD system, Microsoft Kinect V2 and Intel Realsense D435 systems were compared using point-to-point distance. The iterative closest point algorithm was applied to align the KinectFusion point clouds to the reference 3D point cloud (obtained using 3dMD)before the point-to-point distances were calculated. The accuracy of KinectFusion for each device ($s \in \{\text{Kinect}, \text{D435}\}$), each distance ($d \in$ {(75 cm, 100cm ,125 cm, 150cm)}), each resolution ($v \in \{128\ \text{voxels/m},\ 256\ \text{voxels/m}, 384\ \text{voxels/m},\ 256\ \text{voxels/m}\}$), each frame interval ($f \in \{1\ \text{frame},\ 2\ \text{frames}\}$) was represented by the mean and standard deviation of point-to-point difference between the KinectFusion reconstruction, and the reference model ($M^{s,d,v,f}$, $\sigma^{s,d,v,f}$) defined by equation (1) and (2).

$$M^{s,d,v,f} = \frac{\sum_{c=1}^{7}\sum_{o=1}^{2}\sum_{t=1}^{4} m_{c,o,t}^{s,d,v,f}}{56} \tag{1}$$

$$\sigma^{s,d,v,f} = \sqrt{\frac{\sum_{c=1}^{7}\sum_{o=1}^{2}\sum_{t=1}^{4}\left(m_{c,o,t}^{s,d,v,f} - M^{s,d,v,f}\right)^2}{56}} \tag{2}$$

where c states the number of the scanning object, o is the number of the operator, t represents the trial number and $m_{c,o,t}^{s,d,v,f}$ is the mean point cloud distance between the reference and KinectFusion output in single trials defined by equation (3) [34].

$$m_{c,o,t}^{s,d,v,f} = \max\left(\frac{\sum_{i=1}^{n^{ref}} \left\| y_{nearest}^{obj}(x_i^{ref}) - x_i^{ref} \right\|}{n^{ref}}, \frac{\sum_{i=1}^{n^{obj}} \left\| y_{nearest}^{ref}(x_i^{obj}) - x_i^{obj} \right\|}{n^{oboj}}\right) \tag{3}$$

where n^{ref} is the point number of the 3dMD point cloud, $x_1^{ref}, \ldots, x_n^{ref}$ are the point on the 3dMD point cloud, $y_{nearest}^{obj}(x_i^{ref})$ is the nearest point to x_i^{ref} on the KinectFusion point cloud, n^{obj} is the point number of the KinectFusion point cloud, $x_1^{obj}, \ldots, x_n^{obj}$ are the point on the KinectFusion point cloud, $y_{nearest}^{ref}(x_i^{obj})$ is the nearest point to x_i^{obj} on the 3dMD point cloud.

To determine intra-operator and inter-operator repeatability, similar methods were used to align and calculate the distance between pairs of KinectFusion 3D point clouds captured in repeated trials with the same capture distances and depth camera devices. The intra-operator repeatability for each device (s), each capture distance (d), each resolution (v), and each frame interval (f) was represented by the mean and standard deviation of the point-to-point distance between the KinectFusion outputs obtained in the repeated trials $(((o_1, t_1), (o_2, t_2)) \in \{[(1,1), (1,2)], [(1,3), (1,4)], [(2,1), (2,2)], [(2,3), (2,4)]\})$. The inter-operator repeatability for each device (s), each distance (d), each resolution (v), and each frame interval (f) was represented by the mean standard deviation of the point-to-point distance between the KinectFusion Output from the inter operator repeated trials $(o_1, t_1), (o_2, t_2)) \in \{[(1,1), (2,1)], [(1,2), (2,2)], [(1,3), (2,3)], [(1,4), (2,4)]\})$.

3 Results

The mean point-to-point differences of KinectFusion for Microsoft Kinect V2 at 100 cm, 125 cm and 150cm were less than 6.5 mm, while the mean point-to-point differences of the Realsense D435 were larger than 6.5 mm at all capture distances (Table 3). Excluding the data captured at 75 cm using Microsoft Kinect, the accuracy of both Microsoft Kinect V2 and Realsense D435 improved with an decrease in capture distance.

The intra-operator repeatability for both the Microsoft Kinect V2 and Realsense D435 at all capture distances was less than 5mm. Similarly, the inter-operator repeatability for both depth cameras was less than 5mm in most cases. When the capturing distance increased, the intra-operator and inter-operator error for both depth cameras increased even though the capture durations for all scanning trials were approximately 10 seconds.

Table 3 Accuracy and repeatability for different capturing distance with a fixed resolution (256 voxels/m) in mean ± standard deviation.

Sensor	Distance (cm)	Accuracy (mm)	Intra-repeatability	Inter-repeatability

			(mm)	(mm)
Kinect	75	7.51±2.95	2.85±1.45	2.73±1.33
Kinect	100	4.60±1.25	1.46±0.41	1.60±0.42
Kinect	125	4.90±1.74	1.78±1.18	1.90±1.14
Kinect	150	6.21±3.12	2.00±0.94	2.02±0.74
D435	75	6.96±0.40	1.77±0.30	1.93±0.33
D435	100	7.59±1.66	2.09±0.54	2.18±0.54
D435	125	8.50±3.05	3.74±1.76	3.76±1.48
D435	150	11.75±3.27	4.93±3.18	5.32±3.20

The accuracy for Microsoft Kinect V2 decreased (point-to-point error increased) while the resolution increased. The accuracy for Realsense D435 remained at similar level (128 voxel/m) with higher resolution applied (Table 4). For both depth cameras, the intra-repeatability and inter-repeatability improved with increased resolution.

Table 4 Accuracy and repeatability for different resolutions with a fixed capturing distance (100 cm) in mean ± standard deviation.

Sensor	Resolution (voxels/m)	Accuracy (mm)	Intra-repeatability (mm)	Inter-repeatability (mm)
Kinect	128	4.22±0.59	1.77±0.54	2.11±0.59
Kinect	256	4.60±1.25	1.46±0.41	1.60±0.42
Kinect	384	6.39±1.79	1.55±0.67	1.70±0.65
Kinect	512	6.95±1.99	1.34±0.51	1.40±0.41
D435	128	7.12±1.07	2.43±0.73	2.74±0.54
D435	256	7.59±1.66	2.09±0.54	2.18±0.54
D435	384	6.95±1.86	1.96±0.68	2.05±0.64

D435	512	6.99±1.89	1.80±0.60	1.85±0.55

While increasing the capturing speed (frame interval = 2), the accuracy of the Microsoft Kinect decreased, while the accuracy of D435 increased (Table 5). The repeatability of the Microsoft Kinect V2 decreased with increasing capturing speed, while the repeatability of the Realsense D435 remained consistent with increasing capturing speed.

Table 5 Accuracy and repeatability for different frame interval with a fixed capturing distance (100 cm) and a resolution (256 voxels per meter) in mean ± standard deviation.

Sensor	Frame interval (frame)	Equivalent Capturing Speed (second/round)	Accuracy (mm)	Intra-repeatability (mm)	Inter-repeatability (mm)
Kinect	1	10	4.60±1.25	1.46±0.41	1.60±0.42
Kinect	2	5	5.02±2.64	2.49±3.22	2.50±2.60
D435	1	10	7.59±1.66	2.09±0.54	2.18±0.54
D435	2	5	6.80±0.94	2.00±0.40	2.01±0.39

4 Discussion

The aim of this study was to compare the 3D reconstruction results obtained from KinectFusion when using two different types of consumer depth cameras (time-of-flight and stereoscopic cameras). In addition, the results were compared with scan data acquired from a commercial 3D scanning system to determine which type of depth camera gives improved KinectFusion reconstruction results.

Using Microsoft Kinect V2 sensor gave improved accuracy and repeatability compared to the Realsense D435 sensor at most capture distances. One possible reason for this might be that the time-of-flight camera (Microsoft Kinect V2) generates higher quality depth maps than the stereoscopic camera (Realsense D435) as shown in Figure 4. The noise present in the depth images might introduce error while applying KinectFusion techniques to reconstruct the 3D data. As shown in Figure 3, 4, and 5 the noise present around the neck/shoulders in the depth images caused errors in the 3D reconstruction of the neck/shoulder region.

[insert Figure 4.]

[insert Figure 5.]

According to previous studies [30, 35], the error in point cloud distances might cause in excess of 10 times this error in anthropometric measurements. For example, the nominal accuracy (point-to-point difference) of 3dMD is around 0.2 mm [30], but the error in anthropometric measurement could be as high as 2.0 mm [35]. Thus, the observed difference in accuracy between the Microsoft Kinect V2 and Realsense D435 (around 2.0 mm for the cases with the best accuracy) could cause 2.0 cm of variation in girth measurements. Information fusion techniques [36-38], are powerful solutions for decreasing the effect of noise, uncertainty, and external disturbance and could be applied to improve the camera pose estimation and 3D reconstruction results if improved accuracy was required in the future applications.

It seems that the higher-quality depth maps captured by Microsoft Kinect V2 generated improved 3D reconstruction results. Though the Realsense D435 captured high-resolution depth maps at a higher frame rate, its 3D reconstruction results were not as good as those generated by the Microsoft Kinect V2. However, the software development kit of Realsense D435 provides a wider range of options to alter the camera settings (e.g. the intensity of projected light) to adapt to various environments (e.g. indoor or outdoor environment). In this study, only the default pre-sets were used. Hence, future studies could improve the depth map quality generated using the Realsense D435 sensor by optimising these camera settings for a specific set of conditions to enhance the accuracy of 3D reconstruction results generated by KinectFusion.

Pagliari and Pinto [26] suggested that the errors in depth detection with the Microsoft Kinect V2 increased linearly with an increase in capture distance, which might explain the $M^{s,d,256,1}$ of 3D reconstruction increasing between 100 cm and 150 cm. The Microsoft Kinect V2 cannot capture objects at distances less than 50 cm, which cause the poor accuracy and repeatability of Kinect V2 at 75 cm, since the sensor cannot detect the sphere balls accurately and causes issues for camera pose estimation while applying KinectFusion. The accuracy and reliability of depth detection by the RealSense D435 sensor decreased from 75 cm to 150 cm, so the errors in depth detection with Realsense D435 might also grow with an further increase in capture distance. Therefore, it is suggested that a time-of-flight camera with an appropriate capture distance should be used for applications that require accurate and reliable reconstruction of 3D human scan data using KinectFusion techniques.

Using high resolution (384, 512 voxels/m) when applying KinectFusion with the Microsoft Kinect V2 caused the accuracy to decrease. By contrast, the accuracy of the Realsense D435 remained at similar levels in various voxel resolutions. The possible reason might be that the high-resolution reconstruction is sensitive to the flying pixels which are generated by the Microsoft Kinect V2. Using high resolution for KinectFusion Reconstruction can generate dense 3D point clouds which might decrease the point-to-point distance in repeated trials; this might explain the enhanced repeatability shown in trials that applied a higher resolution setting. Therefore, it is highly recommended that using a good flying pixels filter as the pre-process of the high-resolution KinectFusion reconstruction in order to obtain accurate and repeatable results.

Previous studies have typically used a turning table to rotate participants 360^o and capture images from all directions using a fixed camera [22, 39]. However, scanning procedures with this approach often take more than 30 seconds to complete. Typically, people cannot hold their breath consistently for this period without moving. Also, older and younger users tend to move while standing on a turning table during the scanning procedure. The bespoke rotating camera rig with a stationary central platform developed for this study enables a more rapid scanning procedure (around 10 seconds), which is similar to the scanning time for some commercial scanning systems used for whole-body measurements (e.g. Hamamatsu BLS 9036 [4], Vitus^{smart} XXL 3D body scanner [40]). Furthermore, the results in this study show that both intra-operator and inter-operator repeatability were similar for both depth cameras (< 3 mm in most cases), meaning the developed rotation platform enabled consistent image capturing of participants from all directions. It is probable that the use of the rotating camera rig reduced the influence of operator error during scanning.

While halve the number of frames for reconstruction (frame interval = 2), the accuracy and repeatability of Microsoft Kinect V2 decreased but the accuracy of Realsense D435 improved. A possible reason might be that KinectFusion restricts the permissible camera rig rotation velocity. Increasing the rotation velocity with Microsoft Kinect V2 caused the KinectFustion reconstruction from sparse frames which led to worse accuracy and repeatability. However, increasing the rotation velocity with Realsense D435 could avoid some error accumulation of camera pose estimation and obtain accurate 3D reconstruction results. Hence, the operator should restrict the permissible rotation velocity to maintain the accuracy and repeatability of 3D reconstruction while using Microsoft Kinect V2 and Realsense D435. Although using the lower rotation velocity with Microsoft Kinect V2 might improve the reconstruction for a rigid body, people tend to move when the capturing time increases. Therefore, it is worth checking whether the time-of-flight sensors which can capture high-quality images in high frequency to improve the reconstruction results. The capture rate of Microsoft Kinect V2 (around 20 fps) is less than the theoretical values (30 fps). Further studies that use Microsoft Kinect V2 should control the stability of power supply, hardware compatibility, and file saving speed to optimize the capturing rates for improved 3D reconstruction results. As the small differences within this acceptable range of rotation velocity still existed and caused some errors between the trials. Further development of the capture system, such as automation of the rotating camera rig, might be required to optimize the intra-repeatability and inter-repeatability of this scanning platform for future research in health applications.

In this study, four cylinder objects and three torso mannequins were used to represent different body segments and shapes. Comparing to non-rigid objects (e.g. human participants), using these rigid objects can determine the point-to-point distance and show that the Microsoft Kinect V2 based system can provide accurate and repeatable 3D reconstruction results. However, the accuracy and reliability of anthropometric measurement (e.g. waist girths) of this kind of system from human participants is still unknown. Further studies should be conducted to examine whether this system can be used for obtaining anthropometric data from human participants.

5 Conclusions

This study compared the 3D reconstruction results obtained from KinectFusion, using two different types of depth cameras (time-of-flight and stereoscopic cameras) with the results acquired from a reference method, a commercial 3D scanning system. When used as part of the rotating rig with a stationary central platform developed for this study, both time-of-flight and stereo cameras enable repeatable 3D body scan data to be collected with minimal influence of operators on the results. However, the time-of-flight depth camera generated more accurate 3D point clouds than the sensor using stereoscopic techniques. Thus, this suggests that applications requiring high-quality depth maps and the generation of accurate 3D human models by KinectFusion techniques should consider using a time-of-flight camera, such as the Microsoft Kinect V2, as the sensor for capturing depth image.

Funding

The authors would like to thank the Centre for Sports Engineer Research to support purchasing the components of the rotation platforms.

Conflict of Interest Statement

The authors declare that there is no conflict of interest with regard to this paper for any author.

References

1. Zollhöfer M, Stotko P, Görlitz A, et al. State of the Art on 3D Reconstruction with RGB-D Cameras. *Computer Graphics Forum* 2018; 37: 625-652. DOI: doi:10.1111/cgf.13386.
2. Han J, Shao L, Xu D, et al. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. *IEEE Transactions on Cybernetics* 2013; 43: 1318-1334. DOI: 10.1109/TCYB.2013.2265378.
3. Treleaven P and Wells J. 3D body scanning and healthcare applications. *Computer* 2007; 40: 28-34.
4. Stewart AD. Kinanthropometry and body composition: A natural home for three-dimensional photonic scanning. *J Sports Sci* 2010; 28: 455-457. DOI: 10.1080/02640411003661304.

5. Robinette KM, Daanen H and Paquet E. The CAESAR project: a 3-D surface anthropometry survey. In: *3-D Digital Imaging and Modeling, 1999 Proceedings Second International Conference on* 1999 1999, pp.380-386.

6. Wells JCK, Treleaven P and Charoensiriwath S. Body shape by 3-D photonic scanning in Thai and UK adults: comparison of national sizing surveys. *Int J Obes* 2012; 36: 148-154.

7. Lin Y-L and Wang M-JJ. The development of a clothing fit evaluation system under virtual environment. *Multimedia Tools and Applications* 2016; 75: 7575-7587. journal article. DOI: 10.1007/s11042-015-2681-7.

8. Schranz N, Tomkinson G, Olds T, et al. Is three-dimensional anthropometric analysis as good as traditional anthropometric analysis in predicting junior rowing performance? *J Sports Sci* 2012; 30: 1241-1248. DOI: 10.1080/02640414.2012.696204.

9. Warburton T, Lyttle A, James R, et al. DIFFERENTIATION OF THE ENERGETIC PARAMETERS BETWEEN MALE AND FEMALE ELITE POLE VAULTERS. In: *ISBS-Conference Proceedings Archive* 2016.

10. Clarkson S, Choppin S, Hart J, et al. *Calculating Body Segment Inertia Parameters from a Single Rapid Scan Using the Microsoft Kinect*. 2012, p.153-163.

11. Bullas AM, Choppin S, Heller B, et al. Validity and repeatability of a depth camera-based surface imaging system for thigh volume measurement. *J Sports Sci* 2016: 1-7. DOI: 10.1080/02640414.2016.1149604.

12. Pandis P and Bull AM. A low-cost three-dimensional laser surface scanning approach for defining body segment parameters. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 2017; 231: 1064-1068. DOI: 10.1177/0954411917727031.

13. Sun J, Xu B, Lee J, et al. Novel Body Shape Descriptors for Abdominal Adiposity Prediction Using Magnetic Resonance Images and Stereovision Body Images. *Obesity* 2017; 25: 1795-1801. DOI: 10.1002/oby.21957.

14. Ridner SH, Montgomery L, Hepworth J, et al. Comparison of upper limb volume measurement techniques and arm symptoms between healthy volunteers and individuals with known lymphedema. *Lymphology* 2007; 40: 35-46.

15. Wheat JS, Choppin S and Goyal A. Development and assessment of a Microsoft Kinect based system for imaging the breast in three dimensions. *Med Eng Phys* 2014; 36: 732-738. DOI: http://dx.doi.org/10.1016/j.medengphy.2013.12.018.

16. Izadi S, Kim D, Hilliges O, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. *Proceedings of the 24th annual ACM symposium on User interface software and technology*. Santa Barbara, California, USA: ACM, 2011, p. 559-568.

17. PengPeng H, Taku K, Duan L, et al. 3D textile reconstruction based on KinectFusion and synthesized texture. *International Journal of Clothing Science and Technology* 2017; 29: 793-806. DOI: doi:10.1108/IJCST-01-2017-0007.

18. Wang Z, Zhong YQ, Chen KJ, et al. 3D Human Body Data Acquisition and Fit Evaluation of Clothing. *Advanced Materials Research* 2014; 989-994: 4161-4164. DOI: 10.4028/www.scientific.net/AMR.989-994.4161.

19. Ng BK, Hinton BJ, Fan B, et al. Clinical anthropometrics and body composition from 3D whole-body surface scans. *Eur J Clin Nutr* 2016; 70: 1265-1270. Original Article. DOI: 10.1038/ejcn.2016.109.

20. Li H, Vouga E, Gudym A, et al. 3D self-portraits. *ACM Trans Graph* 2013; 32: 1-9. DOI: 10.1145/2508363.2508407.

21. Hu P, Komura T, Holden D, et al. Scanning and animating characters dressed in multiple-layer garments. *Vis Comput* 2017; 33: 961-969. journal article. DOI: 10.1007/s00371-017-1388-3.

22. Miyeon L, Dong IY and Sungmin K. Development of low cost three-dimensional body scanner using depth perception camera. *International Journal of Clothing Science and Technology* 2017; 29: 857-867. DOI: doi:10.1108/IJCST-02-2017-0010.

23. Sarbolandi H, Lefloch D and Kolb A. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer Vision and Image Understanding* 2015; 139: 1-20. DOI: https://doi.org/10.1016/j.cviu.2015.05.006.

24. Keselman L, Woodfill JI, Grunnet-Jepsen A, et al. Intel(R) RealSense(TM) Stereoscopic Depth Cameras. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 21-26 July 2017 2017, pp.1267-1276.

25. Gonzalez-Jorge H, Rodríguez-Gonzálvez P, Martínez-Sánchez J, et al. Metrological comparison between Kinect I and Kinect II sensors. *Measurement* 2015; 70: 21-26. DOI: https://doi.org/10.1016/j.measurement.2015.03.042.

26. Pagliari D and Pinto L. Calibration of Kinect for Xbox One and Comparison between the Two Generations of Microsoft Sensors. *Sensors* 2015; 15: 27569.

27. Wasenmüller O and Stricker D. Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. In: Cham, 2017, pp.34-45. Springer International Publishing.

28. Runyang Z, Xueshi G and Geng W. Applications of RGB-D data for 3D reconstruction in the indoor environment. In: *2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)* 12-14 Aug. 2016 2016, pp.375-378.

29. Cignoni P, Callieri M, Corsini M, et al. MeshLab: 3D triangular meshes processing and editing. *Astrophysics Source Code Library* 2017.

30. Li BYL, Mian AS, Liu W, et al. Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In: *2013 IEEE Workshop on Applications of Computer Vision (WACV)* 15-17 Jan. 2013 2013, pp.186-192.

31. Fankhauser P, Bloesch M, Rodriguez D, et al. Kinect v2 for mobile robot navigation: Evaluation and modeling. In: *2015 International Conference on Advanced Robotics (ICAR)* 27-31 July 2015 2015, pp.388-394.

32. Siena FL, Byrom B, Watts P, et al. Utilising the Intel RealSense Camera for Measuring Health Outcomes in Clinical Research. *J Med Syst* 2018; 42: 53. journal article. DOI: 10.1007/s10916-018-0905-x.

33. Derpanis KG. Overview of the RANSAC Algorithm. *Image Rochester NY* 2010; 4: 2-3.

34. Girardeau-Montaut D. CloudCompare version 2.6. 1 user manual. 2017.

35. Bullas AM. *The importance of complex anthropometrics in the assessment of cyclists*. Sheffield Hallam University, 2017.

36. Du G, Zhang P and Li D. Human–Manipulator Interface Based on Multisensory Process via Kalman Filters. *IEEE Transactions on Industrial Electronics* 2014; 61: 5411-5418. DOI: 10.1109/TIE.2014.2301728.

37. Du G, Zhang P and Liu X. Markerless Human–Manipulator Interface Using Leap Motion With Interval Kalman Filter and Improved Particle Filter. *IEEE Transactions on Industrial Informatics* 2016; 12: 694-704. DOI: 10.1109/TII.2016.2526674.

38. Du G and Zhang P. A Markerless Human–Robot Interface Using Particle Filter and Kalman Filter for Dual Robots. *IEEE Transactions on Industrial Electronics* 2015; 62: 2257-2264. DOI: 10.1109/TIE.2014.2362095.

39. Bourgeois B, Ng BK, Latimer D, et al. Clinically applicable optical imaging technology for body size and shape analysis: comparison of systems differing in design. *Eur J Clin Nutr* 2017. Original Article. DOI: 10.1038/ejcn.2017.142.

40. Chiu C-Y, Pease DL and Sanders RH. The effect of pose variability and repeated reliability of segmental centres of mass acquisition when using 3D photonic scanning. *Ergonomics* 2016: 1-6. DOI: 10.1080/00140139.2016.1161245.

Figure Captions

Figure 1 The test objects used in this study. From left to right: Torso 1, Torso 3, Torso 2, Cylinder 1, Cylinder 2, Cylinder 3, Cylinder 4.

Figure 2 A bespoke central-stationary platform with a rotating camera rig was developed and used in this study. A sensor (Microsoft Kinect V2) was mounted on the rotating camera rig (highlighted) and connected to a laptop. The diagram (bottom) shows the bespoke the rotating camera rig (without central-stationary platform).

Figure 3 (a) The aligned 3D point cloud obtained from 3dMD (Torso 1). (b) The 3D point cloud which applied random sample consensus algorithms for plane segmentation still contained a few isolated points. (c) The 3D point cloud applied the density filters to delete isolated points. (d) The reference 3D point clouds of the test object obtained from the 3dMD scanning system (Torso 3). (e) The processed 3D point clouds captured by Microsoft Kinect V2 (Torso 3). (f) The processed 3D point clouds captured by Intel Realsense D435 (Torso 3).

Figure 4 Left: A sample image with minimal noise around the test object captured by Microsoft Kinect V2. Right: A sample image with much noise around the test object captured by Realsense D435.

Figure 5 The point-to-point difference shown on the reference 3D point cloud (top row: Data obtained with Microsoft Kinect V2; bottom row: Data obtained with Realsense D435; red: vertex distance near 0; blue: vertex distance near 30 mm; unit: mm).