Gait & Posture 68 (2019) 193–200

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives

Ross A. Clark[a,*], Benjamin F. Mentiplay[b], Emma Hough[a], Yong Hao Pua[c]

[a] School of Health and Sports Sciences, University of the Sunshine Coast, 90 Sippy Downs Drive, Sippy Downs, QLD, 4556, Australia
[b] La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Bundoora, VIC 3086, Australia
[c] Department of Physiotherapy, Singapore General Hospital, Singapore

ARTICLE INFO

Keywords:
Depth camera
Joint angles
Validity
Reliability
Walking
XBox

ABSTRACT

Background: Three-dimensional camera systems that integrate depth assessment with traditional two-dimensional images, such as the Microsoft Kinect, Intel Realsense, StereoLabs Zed and Orbecc, hold great promise as physical function assessment tools. When combined with point cloud and skeleton pose tracking software they can be used to assess many different aspects of physical function and anatomy. These assessments have received great interest over the past decade, and will likely receive further study as the integration of depth sensing and augmented reality smartphone cameras occurs more in everyday life.

Research Question: The aim of this review is to discuss how these devices work, what options are available, the best methods for performing assessments and how they can be used in the future.

Methods: Firstly, a review of the Microsoft Kinect devices and associated artificial intelligence, automated skeleton tracking algorithms is provided. This includes a narrative critique of the validity and clinical utility of these devices for assessing different aspects of physical function including spatiotemporal, kinematic and inverse dynamics data derived from gait and balance trials, and anatomical assessments performed using the depth sensor information. Methods for improving the accuracy of data are examined, including multiple-camera systems and sensor fusion with inertial monitoring units, model fitting, and marker tracking. Secondly, alternative hardware, including other structured light and time of flight methods, stereoscopic cameras and augmented reality leveraging smartphone and tablet cameras to perform measurements in three-dimensional space are summarised. Software options related to depth sensing cameras are then discussed, focussing on recent advances such as OpenPose and web-based methods such as PoseNet.

Results and Significance: The clinical and non-laboratory utility of these devices holds great promise for physical function assessment, and recent developments could strengthen their ability to provide important and impactful health-related data.

1. Introduction

The Microsoft Kinect was the first mass produced, three-dimensional (3D) camera that possessed a price point making it available to almost any consumer. When combined with automated skeleton tracking software incorporating artificial intelligence, this device provided unparalleled access to a low-cost, marker-less platform for assessing the kinematic and spatiotemporal aspects of physical function in healthy and clinical populations. This review will discuss:

1) How the Microsoft Kinect provided low-cost, widely available access to advanced 3D camera technology. This section will summarise the two models that were designed and sold for gaming purposes, how they functioned and their current status.
2) While the Microsoft Kinect SDK was a powerful tool for extracting segment position and 3D joint angles, much of the data was highly questionable with respect to validity and reliability. The types of assessments that have incorporated skeleton tracking and raw camera sensor data will be examined, including comments on the validity, reliability and clinical utility. A summary of the benefits and limitations of marker-less tracking is also provided.
3) How methods such as multiple depth cameras, model-based

* Corresponding author.
E-mail address: rclark@usc.edu.au (R.A. Clark).

https://doi.org/10.1016/j.gaitpost.2018.11.029

assessments, marker-tracking and sensor fusion have been used to improve data quality.

4) With the cessation of the Kinect as a mass-produced system specifically designed for gaming, what other devices are available? This will include common systems such as structured light, time of flight and stereoscopic cameras, in addition to recent advances using standard cameras and augmented reality.
5) The different software methods for estimating human poses and obtaining skeleton tracking information. This will focus on recent advances including Openpose and PoseNet. For this review, pose recognition will be defined as skeleton tracking.
6) Finally, with the production of the Kinect V2 ceasing, what does the future hold in this field? This review ends with a brief outline of potential future uses of depth cameras in the field of clinical assessments.

2. What Kinect hardware and software is available?

The Microsoft Kinect was designed as a gaming device to counter the “exergaming” peripherals released by Nintendo (for example, the Wii Balance Board and Wii Hand Controllers) and Sony (for example, the PlayStation EyeToy). While commercial gaming controllers that required more than just button manipulation had existed decades prior to these devices (for example, the Nintendo Power Glove), the Nintendo Wii devices were the first time that these systems had sold in large quantities. From a biomechanical and research perspective the Nintendo devices were the first to be used for physical function assessment of a wide range of tasks. This included the Wii Balance Board being used as a balance and weight bearing asymmetry assessment and feedback device [1–3], and the hand controllers used for gesture recognition [4] and to assess gait speed [5]. Arguably, the key difference between the Nintendo devices and the Microsoft Kinect was that the latter represented a true evolution in technology. The Nintendo devices were low-cost, mass-produced variations of existing devices that people in biomechanics laboratories were accustomed to. For example, the Wii Balance Board was a replica of a force platform, and the Wii Hand Controller used an inertial monitoring unit (IMU) and light tracker to determine its position in 3D space. The Kinect V1 was much different:

2.1. Kinect V1

The Kinect V1 (also known as the Xbox 360 Kinect) used an infrared structured light 2D camera to create a 3D depth representation of the space in front of it. The structured light pattern matching method was quite simplistic, with clusters of infrared dots observed and the size of the pattern used to determine distance from the camera. For this reason, the structured light method is often referred to as speckle pattern matching.

To highlight how this works, the closer to an object that the pattern is reflected on, the smaller the pattern appears to the camera (Fig. 1). Because the segment is closer to the camera it does not allow the speckle pattern to disperse as widely, and thus the pattern appears more clustered. Each dot within a pattern can be reused in many different patterns, and as such a relatively precise depth position for each pixel of the camera can be calculated based on thousands of clusters. By performing simple trigonometry, it was possible to determine the position and distance in 3D space for each pixel. Because we know the distance to an object from the camera, and we know that at that distance from the camera each pixel has a width of X and height of Y, we can determine displacement in all three axes. To allow for specific sections of the depth image to be assessed, the image data collected from the two-dimensional (2D) video camera in the Kinect could be overlayed on the depth data. By doing this the 3D position in space of individual pixels could be determined that corresponded to points of interest such as anatomical landmarks.

2.2. Kinect V2

With the exception of an additional 2D infrared image, the data derived from the Kinect V2 (also known as the Xbox One Kinect) was consistent with the Kinect V1 in that a 2D video image, 3D depth image and skeleton and facial tracking information could be extracted from the SDK. However, the Kinect V2 used a different, time-of-flight method for creating the depth map which promised increased resolution and accuracy. In its simplest form, this method measures the time it takes a light source to bounce back from a reflected target, and given the speed of light is constant the distance to the objects surface can be calculated. Advances in single beam time-of-flight systems related to autonomous driving have made them an excellent, low-cost method for assessing gait speed [6], however they do not provide depth map capability. To achieve this, the Kinect V2 used an advanced matricial time of flight camera, for further details on how this method functions refer to other technical papers [7,8].

2.3. Skeleton tracking with the Kinect systems

One of the most exciting aspects of the Kinect was the ability to automatically identify anatomical landmarks in close to real-time using artificial intelligence. Briefly, this allowed for the extraction of joint centres and segment orientation, providing the ability to calculate joint kinematics and spatiotemporal aspects of movement. The Kinect SDK has a very simple interface with extensive documentation available online, and was built for the most common depth camera. Therefore, it is unsurprising that it is the most commonly used method for assessing skeleton tracking. Based on a publication from the Microsoft Research team [13], the algorithm for identifying the anatomical landmarks was created by training a randomised decision forest algorithm using a subset of 100,000 depth scans of a variety of movements including running, dancing, driving, kicking and navigating menus. For a complete description of the methodology refer to the paper by Shotton et al

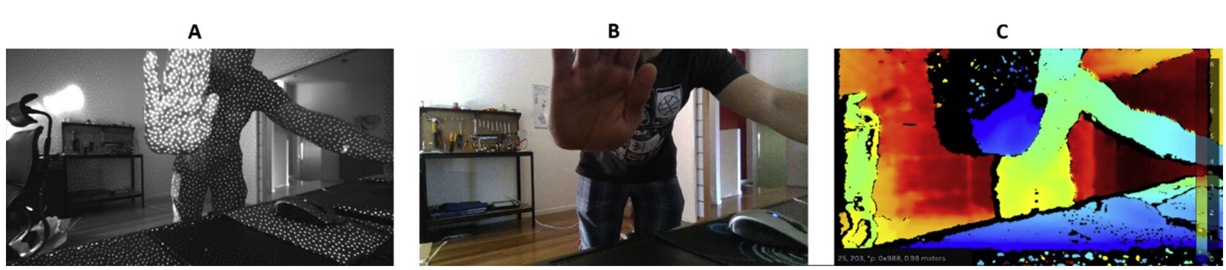

Fig. 1. An example of how infrared structured light is used to create a 3D reconstruction of the scene in front of the depth camera. In A) the patterns of infrared dots can be seen, with dot patterns on body segments closer to the camera (eg. hand) appearing closer together and those on more distant body parts (eg. thighs) appearing more separated. B) shows the image recorded from the standard 2D colour video camera, and C) show the depth map reconstruction. In this case, C) is derived from the infrared dot pattern, with pixels appearing blue being closer to the camera and a linear progression to red pixels being the furthest away (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

[13]. While the Microsoft SDK received widespread use, its restriction to the Kinect system has also now made it currently obsolete. This may change in the future with the Project Azure Kinect, however it is unlikely the SDK will remain in its current format given that the Project Azure Kinect is designed as a web platform. It may also be problematic, for example in situations such as hospitals and remote areas where internet access may be limited.

The OpenNI/NI2 and NiTE platform was one of the first used for performing skeleton tracking with the original Kinect [14]. OpenNI allowed interfacing with the Kinect hardware, while the Primesense NiTE software provided the pose estimation information. It is unclear if the skeleton tracking algorithm provided data similar in accuracy to the Kinect SDK. One study compared data using these two methods and compared them to a custom, side-on algorithm created using depth data [15]. Their results (subjectively assessed based on Figure 10 in the article) suggest that the Kinect SDK may perform better or on par with the OpenNI method when tracking most joints, with the exception of the head during a Tinetii Test. However, this study had numerous limitations. This included the use of surface markers placed on the lateral surface of the subjects and captured using the infrared camera in the Kinect as the criterion reference, which do not represent joint centres in 3D space and would bias findings against the front-on algorithms. With support for the OpenNI and Primesense NiTE software ceased soon after it was purchased by Apple in 2013, the future utility of it is unclear.

2.4. Comparing the Kinect V1 and V2

2.4.1. Depth map accuracy

Interestingly, the Kinect V2 did not always outperform the Kinect V1 in terms of depth map accuracy despite its more recent release, higher resolution and time of flight technique. Whilst the Kinect V2 did not have the exponential inaccuracy associated with the Kinect V1 the further the object was away from the camera [9], it was susceptible to inaccuracies based on colour and operating temperature that were not as detrimental for the Kinect V1 [10]. Importantly, under some specific conditions the Kinect V1 was found to be more precise than the V2 [10], and authors went so far as to provide guidelines on which model to use based on the desired application [8]. However, this was confounded by evidence that differences in performance existed even within some model variants of the same Kinect version [11].

2.4.2. Skeleton tracking

There is no strong evidence that the Kinect V2 outperformed the Kinect V1 with respect to accuracy of the skeleton tracking, however it is feasible that the more consistent sampling rate and accuracy of the depth camera of the V2, and alterations over time to the SDK, may have improved performance. One study by Reither et al. [12] compared the two Kinect versions to a criterion reference (an eight camera marker-based system) and observed that the V1 device underestimated reaching displacements but was similar to the V2 for kinematics, however the relative consistency that each device had with the criterion reference was overall very similar. This may have been related to the displacement error associated with the V1′s depth camera, which is used in the trigonometry calculation to calculate distance on each axis.

3. What assessments can be performed – and are they valid?

3.1. Balance and Functional Tasks

Of all the physical function assessments that the Kinect and other depth cameras hold promise for, potentially the most clinically feasible and useful one is as an automated balance and postural control assessment tool. While the Nintendo Wii Balance Board, which is a force platform, has shown great potential as a low-cost clinical balance assessment device [17], it and more advanced laboratory-based systems such as three-dimensional motional analysis (3DMA) or force platforms have a number of limitations that are overcome by the Kinect. This includes:

1 In contrast to force platforms, they do not require the person to step onto a raised platform which itself is a fall hazard, or have the expense of requiring the sensor to be embedded into the floor. Additionally, individual segments can be tracked to examine specific postural movements, for example, the anterior-posterior sway of the trunk relative to the ankle joint centre during a forward limits of stability test. In contrast, force platforms only provide composite measures, and therefore it is difficult to identify specific postural strategies during balance tasks.
2 Where depth image accuracy is not affected by distance to the object (for example with the Kinect V2) the test does not need to be performed in an identical location each time. This provides testing flexibility, and because most systems can track multiple people at once, allows simultaneous patient assessment.
3 In contrast to 3DMA systems or IMUs, nothing needs to be placed on the patient. This significantly reduces assessment time.

Given these potential advantages, many studies have examined the Kinect for assessment of balance and postural control [18–23]. One of the first studies to validate the Kinect as a physical function assessment tool was performed using balance tests [18], and since then many studies have examined the validity of both the Kinect V1 and V2 [24]. Overall, the findings indicate that for relatively static or slow-moving tests performed with a fixed base of support, as the task movement increases the validity improves [18,24]. This is to be expected, as the Kinect algorithm includes filtering properties to smooth the data and therefore small (eg. < 2 mm) oscillations are unlikely to be correctly recorded. This is an important consideration if using the Kinect or other depth cameras to assess balance and postural control, as simple eyes open two-legged balance in a healthy cohort may not provide accurate information, whereas the validity of the same test in a population with advanced Parkinson’s disease may be satisfactory.

The Kinect has also been used to assess more dynamic balance tests, such as the functional reach and sit-to-stand tasks [25], and composite tasks such as the timed up and go [26]. Studies indicate that the Kinect is particularly useful for assessing trunk motion during trials such as the functional reach or limits of stability. For a test such as the timed up and go, it can be used to break down the individual components of the assessment into variables such as the time taken to stand up, step lengths, gait speed, time taken to turn and the total task time [26]. One prior study performed in a cohort of participants post-stroke found that the length of the first step during the TUG provides additional information over and above the total TUG time [26]. While this could potentially provide useful information about a patient’s expected long-term outcomes, it is unclear if this is worthwhile given the inaccuracy associated with some of the variables extracted from the Kinect. This is particularly important when considering the turning movement, as the Kinect cannot accurately record postural movement when the person performs the turn. Given the boom in low-cost IMUs currently available, and the ease of programming IMU chips with on-board calibration and kinematic processing algorithms (for example, the Bosch BNO080 chipset), the Kinect and other marker-less tracking systems may not be the optimal method to instrument these assessments.

3.2. Posture

Using the depth imaging camera, it is possible to create 3D maps of anatomical landmarks. This can be achieved in multiple ways, including:

1 Using depth image frames aligned with the object of interest. For example, a study by Mentiplay et al. [27] assessed static foot posture

using a Kinect by aligning the Kinect camera with the different axes of the foot. The anatomical landmarks were then identified on the image, and the 3D position of the landmarks extracted from the depth data. The data extracted from the Kinect showed typically excellent validity in comparison with a criterion reference 3DMA system, and could be used to predict dynamic foot function [28]. This technique has also been used for other anatomical assessments, including for assessing spinal curvature [29], patient size [30], and lung function associated with respiratory volume [31]. This method allows for rapid setup and assessment but cannot provide information about landmarks that are not directly facing the camera.

2 Using multiple depth cameras to create a 3D model of the object of interest. This can be achieved by setting up the depth cameras so that they can provide unobstructed data that views all faces of the object. For example, a depth camera could be setup in each corner of the room and focused toward the midpoint. Given the known, static positions of the cameras it is possible to map the object and create a 3D model. This method is likely to be the most accurate for full anatomical assessments [32], however requires a large space and is best suited to permanently mounted hardware. To the authors' knowledge this technique is not regularly used in clinical settings, however it has shown promise for mapping anatomical landmarks in mannequins [33] and humans [34,35].

3 Using a single depth camera and moving it or the object to create a full 3D reconstruction. This is perhaps the most exciting method, as it allows for the creation of a 3D model without the need for multiple cameras [36]. This can be achieved using methods such as simultaneous location and mapping (SLAM), which stitches the consecutively recorded frames together to build a 3D model. Microsoft released their own software which could achieve this (3D Scan), and which provided subjectively impressive results. It is beyond the scope of this review to describe this method in detail, however it does have great potential in future anatomical and postural assessment studies.

While research in clinical populations using these methods is limited, it does have great potential in some populations for monitoring outcomes and disease progression. For example, in conditions such as scoliosis or camptocormia these data could be used to examine change over time in spinal curvature. Future research is needed to examine the long-term reliability and sensitivity to change for monitoring disease progression.

3.3. Overground gait

The greatest potential for gait assessment using 3D cameras would be for overground walking, as it could have utility as a rapid screening tool for clinical use if the outcome measures were shown to be valid. However, early research and a review published in 2016 [37] reported that the Kinect skeleton tracking algorithm was poor with respect to most kinematic variables, and only some spatiotemporal variables such as step length and step width were valid for use.

3.3.1. Kinematic analysis

While exhaustive comparisons of the different software (eg. OpenNI vs. Microsoft SDK) and/or hardware (Kinect V1 vs. V2) methods have not been performed, the available evidence from our own team and others indicates that at best the joint angle data extracted from the Kinect must be carefully chosen and validated for specific use cases [38,39]. Despite this disclaimer, in our experience not all kinematic data is invalid. We have observed that trunk angles can be highly accurate, particularly if calibration offsets are used [40], and segments such as the elbow and shoulder appear to have face validity for specific tasks or in populations with pronounced impairments (for example, associated reactions of the biceps). It is often reported that the pattern of the kinematic data from the Kinect resembles criterion reference data, but the magnitude and subtle aspects are confounded by error [41,42].

3.3.2. Spatiotemporal analysis

In contrast, the spatiotemporal information relative to foot positioning and aspects of pelvis motion such as medial-lateral and vertical displacement and forward progression velocity (i.e. gait speed) appears reasonably robust and could potentially be useful in mass- and rapid-screening settings [39]. Multiple studies have reported that foot position related variables have excellent validity with criterion reference systems such as marker-based 3DMA and pressure mats like the GAITRite using the Microsoft Kinect SDK [39,43–46] and custom algorithms based on depth data [47]. Given the potential importance of variables such as step length, asymmetry and width this may provide useful data for identifying people at risk of poor long-term outcomes. Timing-related variables however have not been shown to have such strong validity [39,44], potentially due to reasons including: 1) identifying ground contact or toe-off is very difficult given that the automated skeleton tracking algorithm struggles to identify the foot during the swing phase; and 2) the camera samples at a relatively low (30 Hz) and inconsistent rate, which can result in large timing inaccuracy if errors of even a single frame are observed. These errors could potentially be somewhat overcome by methods such as cyclical pattern matching (for example autocorrelation) that do not rely on identifying a discrete point in time but uses the overall pattern to identify movement, and interpolating the data to increase the sample rate. However, before using temporal information it must be determined whether this additional inaccuracy outweighs the potential for these temporal variables to provide unique information over and above that provided by the other spatial measures and overall gait speed.

3.3.3. Inverse dynamics

Based on the inconsistent kinematic agreement between the Kinect and criterion reference systems, it is not surprising that studies which have used the Kinect for inverse dynamics analysis have reported modest findings [48,49]. For example, Plantard et al. [49] asked 12 healthy males to perform simple tasks consisting of lifting either nothing or an empty cardboard box. During these trials the participants were observed with a Kinect and a criterion reference 3DMA system. Joint angles for the shoulder and elbow were derived, with the addition of inverse dynamics calculated to determine the torque at each joint. Similar to prior kinematic studies, the authors' observed that the general patterns for both joint angles and torques were consistent between the systems using cross-correlation analysis. However, the root mean square error expressed as a percentage for the joint torques was very large, with none of the variables recording under 10% error and the majority exceeding 20% between systems. Given the options for low-cost (< $USD500) force platform technology available at present (for example, the Pasco PASPORT), the benefits of attempting to predict ground reaction forces from the Kinect are unclear.

3.4. Treadmill gait

Walking on a treadmill is well suited to skeletal tracking, as the person remains within a confined location and repeated gait cycles can be obtained. However, evidence is conflicting regarding the validity of spatiotemporal and kinematic data from the Kinect during treadmill walking. One study indicated that step length, step width, step time and stride time have excellent agreement when compared with a criterion reference marker based system, and that sagittal plane angles of the hip and knee also had very good agreement [50]. However, that study had a very small sample size (N = 10) and as such wide confidence intervals for statistical analysis. Another small sample size study (N = 9) reported that the Kinect could provide some accurate kinematic data for the trunk [51]. In contrast, two prior studies [52,53], both with a larger sample size (N = 20 in both studies), reported that while the skeleton tracking data somewhat follows the trend for the joint trajectories of the knee and hip, there is substantial error in the magnitude and as such it cannot be recommended for assessing these kinematic variables.

4. Methods of optimising results

4.1. Model-based vs raw data

One method that has been proposed as a means of improving the accuracy of data derived from the Kinect is to apply model fitting [54,55]. This essentially fits the data acquired to the expected pattern of the data to reduce noise. While studies have found that this can improve the accuracy of the data [54,55], it also poses the problem that by fitting the observed data to a normative pattern it may reduce the ability to detect real gait impairments. Further research is required to determine if this method of model fitting improves accuracy or overly normalises the signal.

4.2. Marker-tracking with the kinect

The combination of a colour and depth camera that could provide overlayed images had great potential for the creation of marker-tracking systems. Put simply, the depth camera could be used to remove all data from the colour image that was outside the capture volume of the movement being performed. This improved the viability of performing colour-based marker tracking in uncontrolled environments, for example tracking red markers on a person when the wall behind them may have red colours present on a poster. If a standard camera was used this would require colour masking and algorithms to exclude (if possible) the extraneous colour, which are often difficult to implement automatically. Once this information is removed using the depth camera, colour matching to identify the marker in 2D space on the standard video image is performed. From there, the pixel locations corresponding to the marker position could be extracted and averaged from the depth camera to locate the position in 3D space of each marker that is tracked. As an example, in our own early research with the Kinect (unpublished) we used spherical fitting with sub-pixel resolution to track illuminated balls placed on the subject, and found that single leg squatting kinematics were highly consistent with data derived from a Vicon system. Numerous other authors have created marker-based tracking systems using the Kinect and other depth camera systems, both with automatic and manual marker identification, and found generally good results (for example, [56,57]). However, the actual use case for this sort of system appears limited due to numerous factors including: 1) the high likelihood of marker occlusion because of the single camera position resulting in extensive time dedicated to post-processing, and 2) this method is unlikely to provide benefits over other body-worn methods such as IMUs or low-cost multi-camera systems such as the Optitrack, which would likely be far more accurate. Therefore, prior studies in this area will not be a major focus of this review.

4.3. Multi-camera and fusion systems

While each of the depth cameras discussed in this review has the capacity for 3D imaging, there are potential advantages to integrating multiple depth cameras for assessment purposes. These include:

1. Reducing occlusion: A single depth camera can only assess what is directly in front of it, and therefore important data is often missed. This is obvious in cases such as turning around, where the depth camera cannot see the front of the person when they face away from it and therefore no information can be gathered about facial features. Importantly though, a single depth camera also cannot directly assess anatomical segments if another body part is in between it and the depth camera. For example, if facing the depth camera and performing a deep squat movement the thigh may not be seen. As such, if a pose estimation algorithm is used it must infer the joint positions, and this is likely to have unacceptable error. Having multiple depth cameras situated around a room may overcome this.
2. Improving accuracy by averaging results: Pose estimation algorithms often have high levels of jitter due to the uncertainty of estimating joint positions. Therefore, data often must be heavily filtered using either the estimation algorithm itself (eg. the jitter radius settings in the Microsoft Kinect SDK) or in post-processing. This error could be reduced by using multiple cameras with different fields of view all assessing the same movement. From these data the pose assessed from each device can be estimated, and if there are sufficient cameras, rules can be set, such as to ignore inferred markers. The data for each joint can then be smoothed using techniques such as means/medians, or more advanced methods such as Kalman filtering [58].
3. Extending the field of view: A major limitation of many depth cameras is the limited field of view of the distance from the camera. This is particularly evident with respect to pose estimation with the Kinect, as neither version can detect skeleton information at distances of 4.5 m or more. While this was satisfactory for gaming, balance and treadmill assessments, it hindered the use of these systems for examining overground gait. Some studies overcame this limitation by integrating multiple depth cameras along the same walkway to extend the tracking range [59].

The limitations of a multi-device system are however apparent, and include:

1. Setup difficulty: Implementing and calibrating multiple depth cameras in a dedicated laboratory is feasible, however doing this in a clinical or mass-screening setting may not be. Unlike multi-camera 3DMA systems, which can have the cameras mounted relatively high above the desired capture volume, because of the limited field of view and relatively low precision of depth cameras they require mounting within a few metres of where the movement will occur. If this is a shared space with other rehabilitation tasks being performed it may not be feasible to have permanently mounted cameras. Additionally, factors such as power supply requirements and cabling length make the implementation of multiple depth cameras difficult.
2. Hardware demands: The Kinect cameras, in particular the Kinect V2, were very hardware intensive which made multiple device systems difficult to implement. For example, the Kinect V2 required an individual USB3.0 host controller per device. Few computers or laptops came with more than one host controller per system (note: the multiple USB ports on a computer typically route to the same host controller), making this method almost impossible to implement on a laptop and requiring modification to desktop computers. Additionally, the computational demands of streaming, interpreting and logging multiple depth and image frames required powerful computer hardware.
3. Camera crosstalk: Many of the depth camera systems are negatively impacted if their field of view is overlapped by another depth camera. This is particularly evident when using structured light cameras such as the original Kinect V1, because it relies on identifying patterns of dots that can be interfered with if overlapped.
4. Limited evidence of improving accuracy: While it appears logical that increasing the number of cameras would improve data accuracy, which some studies have reported [60], the evidence to support this is not strong. For example, one study estimated pose using a skeleton tracking technique consisting of a) a single, front-on Kinect; b) the average joint position estimated using 5 calibrated Kinects; and c) an optimised Kalman filter framework which used the data from the 5 calibrated Kinect sensors and prescribed weighting factors to each camera relative to the movement to estimate joint positions [61]. Despite the hardware, software and setup complexity

of the multi-Kinect setups the results were not always superior. Based on the figure provided in the article, while the multi-Kinect method performed better when occlusions were present (such as when spinning around), the reported error for assessments including running, crossing arms and legs, and bowing from the waist (similar to the functional reach test) appeared superior for a single-Kinect compared to multi-Kinect systems when a mean filter was used, and similar to the results of a Kalman filtering method.

Many of these limitations may be overcome with new, less demanding devices such as the Orbecc Persee and Intel Euclid which are discussed later in this review. These devices perform computation onboard and allow wireless data transfer. Therefore, it is feasible that in the near future a 10+ depth camera system could be implemented outside of research laboratories. These limitations may also be overcome by fusing the skeleton tracking data with kinematic data recorded using an IMU, which has been shown in prior research to improve accuracy [62]. However, the use case for such multi-camera and fusion systems appears very narrow, as the costs and complexity would likely exceed multi-camera marker-tracking and multi inertial monitoring unit systems. This is particularly evident for the IMU-fusion systems, because if body worn sensors are required this removes many of the benefits of these systems. Future research is required to determine whether the drawbacks of a multi-camera or IMU-fusion system outweigh the improved accuracy, if any exists.

5. Evidence of utility

While studies have assessed the validity of the Kinect for overground gait assessment, including in clinical populations such as Parkinson's disease [63] and stroke [64], few have examined whether the data it provides could be beneficial in a clinical setting. Some studies have shown the ability of Kinect derived data to discriminate between healthy and clinical populations. For example, Dolatabadi et al. [65] collected data in a healthy control cohort and a group who had previously had a stroke, and reported that machine learning techniques could correctly classify a gait pattern into normal or pathological. While it is unclear if this is more accurate than something as simple as a subjective assessment performed by a clinician, it does indicate that without any human input gait patterns can be accurately classified based on Kinect data. The direct applications for this are unclear, as further research is needed in more populations. However, future evolutions of this technology could result in the automated detection of as yet undiagnosed conditions, or identify abnormalities in gait patterns associated with potential threatening behaviour such as concealed weaponry.

While instrumenting gait analysis using the Kinect is far simpler than performing a marker-based gait assessment, it does not provide as much data richness as many of the outcome measures must be disregarded due to poor validity. Therefore, studies have examined whether the limited data available from the Kinect provides additional, important information. Studies by Tan et al. [63] in Parkinson's disease and Clark et al. [64] in stroke have used regression techniques to examine whether adding the Kinect to clinical gait assessments can provide additional unique information over and above simple gait speed. Both studies reported that Kinect-derived variables such as the length of the first step during a timed up and go and vertical pelvis displacement during a gait assessment may provide unique information about a patient's physical function. However, these findings should be considered preliminary as neither study went as far as to show that these additional benefits could improve patient care by predicting poor outcomes. Further research in this area is needed to determine if this method of instrumentation provides any benefits.

6. Alternative hardware to the Kinect V1 and V2

Whilst there has been a cessation of production of the Kinect V1 and V2, other devices are filling this void. The alternatives to the Kinect V1 and V2 typically fall under the following methods:

6.1. Other structured light and time of flight camera systems

These devices use similar techniques for structured light (Kinect V1) or time of flight (Kinect V2) assessment of distance from the camera. There are many different systems which incorporate these methods, including the LIPSedge DL and ASUS Xtion.

6.2. Stereoscopic camera systems

Arguably the simplest method of creating a 3D spatial reconstruction, having a system with two or more cameras set at fixed distances from each other allows for techniques such as trigonometry, direct linear transformation and non-linear transformation to be implemented to recreate a 3D space. This method has been used for decades, and is utilised as the criterion reference by gait analysis systems such as VICON and Optitrack to track markers and quantify human movement in gait labs throughout the world. It can however also be used to create depth maps, with this stereovision method having great potential benefits over infrared speckle or time-of-flight systems. These include advances in camera technology (eg. 8 K and 16 K cameras) potentially allowing more precise measurements, modifiable lenses, and removing the issues with using infrared light sources such as difficulties performing outdoor assessments. This potentially allows for assessment of depth over much longer distances in outdoor and light-uncontrolled environments. For example, the Zed (Stereolabs, U.S.A.) incorporates dual 4-megapixel cameras and purports to provide accurate spatial reconstruction data for distances of up to 20 m in outdoor settings. The Realsense D400 (Intel, U.S.A.) series of cameras incorporates both stereovision (via dual 2.1-megapixel cameras) and an infrared structured light projector, which can allow for both methods to be used together to potentially improve 3D spatial reconstruction accuracy. However, its accuracy at the time of writing in comparison with the Kinect cameras is unknown.

6.3. Augmented reality systems

At the time of writing, these systems are very much new and unproven technology, however they show immense promise for the more widespread use of cameras for real-world 3D assessments. Each of the two major smartphone operating systems (Apple iOS and Android) have recently released augmented reality applications for performing distance measurements using a single camera on a smartphone or tablet. The method used to perform this is similar to the structured light method of the Kinect V1, and consists of:

1) Identifying a textured surface on the phone camera, which allows a pattern to be identified.
2) Moving the camera on the axis away from and towards the surface, resulting in the pattern getting smaller (moving away) and larger (moving towards).
3) Using the cameras IMU to estimate its movement in 3D space and the distance it moved during step 2, and because the camera has a fixed field of view and consequently the size of the detected pattern should change linearly based on distance from the camera, an approximation of 3D positions of each pixel on the surface can be estimated.

The accuracy at present is questionable, and appears to be highly device and calibration dependent as it relies on the camera knowing its position in 3D space. However, in a real-world environment limited to hand-controlled calibration (i.e. manually moving the phone) we have observed the accuracy of the Android ARCore used with the app ARuler on a Samsung Galaxy S7 to often be < 1 cm for measuring distances of < 1 m. This accuracy may be sufficient for many physical function assessment tasks, particularly if the aim is to perform them in a variety of non-standardised settings such as in a home or community environment. For example, a standard forward reach test could be performed in a patient's home by simply pointing the video camera at the wall. The patient then performs the test with their hand tracking along the wall, and the forward-reach distance is simply the distance between the start and end of the test measured on the camera. Thus, this removes the need to attach a tape measure or mark the start and end position of the movement on the wall and measure the distance. However, future research is needed to determine the validity of AR derived measurements in a variety of settings before they can be recommended for use.

7. Other skeleton tracking software

If the 3D camera system is to be used for fully automated identification of body segments, as per the skeleton tracking methods used with the Kinect, the algorithms and software used are at least as important as the quality of the spatial reconstruction acquired from the camera. The following section summarises some of the different software methods for automatically identifying anatomical landmarks from 2D and 3D images. It is not meant to be exhaustive, however will cover promising recent advances in the field.

7.1. Recent advances in skeleton tracking – OpenPose and posenet

OpenPose is capable of identifying anatomical landmarks using just 2D camera images [16]. It is a very promising method for performing skeletal tracking on a large scale, with the creators describing the algorithm as achieving near real-time performance regardless of the number of people in the image. Subjective assessment of available demonstration videos indicates that the technique is quite robust. Additionally, it is capable of 3D tracking using multiple calibrated 2D images, which may allow for marker-less assessment of 3D kinematic and spatiotemporal data. However, while it appears promising and at the time of writing is being continually developed, at present the accuracy of this method for assessing either spatiotemporal or kinematic data is unknown.

One of the more impressive recent advances in skeletal tracking, PoseNet running on TensorFlow.js allows for 2D skeletal tracking to be performed via a web browser and camera on computers, tablets and phones irrespective of their operating system (for an example, refer to https://storage.googleapis.com/tfjs-models/demos/posenet/camera.html). While this does not incorporate 3D imaging at present, the potential for a platform independent, browser-based skeleton tracking system for telerehabilitation and home monitoring is considerable.

8. What does the future hold?

One of the most frustrating and exciting aspects of working with technology is the rapid pace at which advances occur. The use of 3D cameras shifted from a tiny, niche market prior to the Kinect to one that sold many millions of devices per annum with the Xbox One and mandatory Kinect V2. With the increase in applications of 3D cameras, for example to create models for 3D printing and to provide spatial mapping for augmented reality headsets (at the time of writing both the Magic Leap One (Magic Leap, U.S.A.) and the next iteration of the Microsoft HoloLens were promoted as containing depth scanning cameras), the uptake and usability of these devices may increase dramatically. However, the specific technology used for future 3D camera systems is unclear. While time of flight, structured light and synchronised cameras are the most commonly used methods currently, the significant investment made by two of the largest companies – Apple and Alphabet (Google's parent company) – into creating augmented reality algorithms to allow standard phone cameras to measure distance in 3D space cannot be ignored. This is particularly evident with the cancellation by Google of Project Tango, initially a phone with an inbuilt depth sensing camera, and focus by the company on ARCore which fuses the data from a single camera with information from IMUs.

9. Conclusion

In conclusion, 3D cameras and pose recognition software have great promise as physical function assessment tools. However, the benefits of these systems must be weighed against the inherent inaccuracy. In some instances it may be a case of "good enough is good enough", however for many uses 3D cameras will provide invalid data. As new technology and software methods are released to leverage the power of 3D cameras, robust research must be performed to validate findings before they are implemented as physical function assessment tools.

Conflict of interest

None of the authors have a conflict of interest

Credit author statement

Ross Clark: Conceptualization, Methodology, Supervision, Writing – Original Draft. **Benjamin Mentiplay:** Writing – Review & Editing. **Emma Hough:** Investigation, Validation, Writing _ Review & Editing. **Yong-Hao Pua:** Writing – Review & Editing.

Acknowledgements

Author RAC was funded by a National Health and Medical Research Council Career Development Fellowship (#1090415). The funding body was not involved in the study.

References

[1] J. Foo, K. Paterson, G. Williams, R. Clark, Low-cost evaluation and real-time feedback of static and dynamic weight bearing asymmetry in patients undergoing in-patient physiotherapy rehabilitation for neurological conditions, J. Neuroeng. Rehabil. 10 (2013).
[2] R.A. Clark, R. McGough, K. Paterson, Reliability of an inexpensive and portable dynamic weight bearing asymmetry assessment system incorporating dual Nintendo Wii Balance Boards, Gait Posture 34 (2011) 288–291.
[3] R.A. Clark, A.L. Bryant, Y. Pua, P. McCrory, K. Bennell, M. Hunt, Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance, Gait Posture 31 (2010) 307–310.
[4] T. Schlömer, B. Poppinga, N. Henze, S. Boll, Gesture recognition with a wii controller, TEI'08 - Second International Conference on Tangible and Embedded Interaction - Conference Proceedings, (2008), pp. 11–14.
[5] R.A. Clark, K. Paterson, C. Ritchie, S. Blundell, A.L. Bryant, Design and validation of a portable, inexpensive and multi-beam timing light system using the Nintendo Wii hand controllers, J. Sci. Med. Sport 14 (2011) 177–182.
[6] R.A. Clark, Y.H. Pua, K.J. Bower, L. Bechard, E. Hough, P. Charlton, et al., Validity of a low-cost laser with freely available software for improving measurement of walking and running speed, J. Sci. Med. Sport (2018) in press.
[7] R. Horaud, M. Hansard, G. Evangelidis, C. Menier, An overview of depth cameras and range scanners based on time-of-flight technologies, Mach. Vis. Appl. 27 (2016) 1005–1020.
[8] H. Sarbolandi, D. Lefloch, A. Kolb, Kinect range sensing: structured-light versus time-of-flight kinect, Comput. Vis. Image Underst. 139 (2015) 1–20.
[9] T. Mallick, P.P. Das, A.K. Majumdar, Characterizations of noise in Kinect depth images: a review, IEEE Sens. J. 14 (2014) 1731–1740.
[10] O. Wasenmüller, D. Stricker, Comparison of Kinect v1 and v2 Depth Images in Terms of Accuracy and Precision. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), (2017), pp. 34–45.
[11] N.M. DiFilippo, M.K. Jouaneh, Characterization of different microsoft kinect sensor models, IEEE Sens. J. 15 (2015) 4554–4564.
[12] L.R. Reither, M.H. Foreman, N. Migotsky, C. Haddix, J.R. Engsberg, Upper extremity movement reliability and validity of the Kinect version 2, Disabil. Rehabil. Assist. Technol. 13 (2018) 54–59.
[13] J. Shotton, T. Sharp, A. Fitzgibbon, A. Blake, M. Cook, A. Kipman, et al., Real-Time

human pose recognition in parts from single depth images, Commun. ACM 56 (2013) 116–124.
[14] E.A. Suma, B. Lange, A. Rizzo, D.M. Krum, M. Bolas, FAAST: the flexible action and articulated skeleton toolkit, Proceedings - IEEE Virtual Reality, (2011), pp. 247–248.
[15] E. Cippitelli, S. Gasparrini, S. Spinsante, E. Gambi, Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view, Sensors (Switzerland) 15 (2015) 1417–1434.
[16] Z. Cao, T. Simon, S.E. Wei, Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition (2017) 1302–1310.
[17] R.A. Clark, B.F. Mentiplay, Y.H. Pua, K.J. Bower, Reliability and validity of the Wii Balance Board for assessment of standing balance: a systematic review, Gait Posture 61 (2018) 40–54.
[18] R.A. Clark, Y.H. Pua, K. Fortin, C. Ritchie, K.E. Webster, L. Denehy, et al., Validity of the Microsoft Kinect for assessment of postural control, Gait Posture 36 (2012) 372–377.
[19] B. Dehbandi, A. Barachant, A.H. Smeragliuolo, J.D. Long, S.J. Bumanlag, V. He, et al., Using data from the Microsoft Kinect 2 to determine postural stability in healthy subjects: a feasibility trial, PLoS One 12 (2017).
[20] M.A. Eltoukhy, C. Kuenze, J. Oh, J.F. Signorile, Validation of static and dynamic balance assessment using microsoft kinect for young and elderly populations, IEEE J. Biomed. Health Inform. 22 (2018) 147–153.
[21] S.S. Galen, V. Pardo, D. Wyatt, A. Diamond, V. Brodith, A. Pavlov, Validity of an interactive functional reach test, Games Health J. 4 (2015) 278–284.
[22] T. Tamei, Y. Orito, H. Funaya, K. Ikeda, Y. Okada, T. Shibata, Kinect-based posturography for in-home rehabilitation of balance disorders, APSIPA Trans. Signal Inf. Process. 4 (2015).
[23] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, D. Li, Reliability and validity of kinect RGB-D sensor for assessing standing balance, IEEE Sens. J. 14 (2014) 1633–1638.
[24] R.A. Clark, Y.H. Pua, C.C. Oliveira, K.J. Bower, S. Thilarajah, R. McGaw, et al., Reliability and concurrent validity of the Microsoft Xbox one Kinect for assessment of standing balance and postural control, Gait Posture 42 (2015) 210–213.
[25] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, L. Rochester, Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease, Gait Posture 39 (2014) 1062–1068.
[26] S. Vernon, K. Paterson, K. Bower, J. McGinley, K. Miller, Y.H. Pua, et al., Quantifying individual components of the timed up and go using the kinect in people living with stroke, Neurorehab. Neural Repair 29 (2015) 48–53.
[27] B.F. Mentiplay, R.A. Clark, A. Mullins, A.L. Bryant, S. Bartold, K. Paterson, Reliability and validity of the Microsoft Kinect for evaluating static foot posture, J. Foot Ankle Res. 6 (2013).
[28] K.L. Paterson, R.A. Clark, A. Mullins, A.L. Bryant, B.F. Mentiplay, Predicting dynamic foot function from static foot posture: Comparison between visual assessment, motion analysis, and a commercially available depth camera, J. Orthop. Sports Phys. Ther. 45 (2015) 789–798.
[29] J. Quek, S.G. Brauer, J. Treleaven, R.A. Clark, The concurrent validity and intrarater reliability of the Microsoft Kinect to measure thoracic kyphosis, Int. J. Rehabil. Res. 40 (2017) 279–284.
[30] T.S. Cook, G. Couch, T.J. Couch, W. Kim, W.W. Boonn, Using the microsoft kinect for patient size estimation and radiation dose normalization: Proof of concept and initial validation, J. Digit. Imaging 26 (2013) 657–662.
[31] C. Sharp, V. Soleimani, S. Hannuna, M. Camplani, D. Damen, J. Viner, et al., Toward respiratory assessment using depth measurements from a time-of-Flight sensor, Front. Physiol. 8 (2017) 65.
[32] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, A.E. Saddik, Evaluating and improving the depth accuracy of Kinect for Windows v2, IEEE Sens. J. 15 (2015) 4275–4285.
[33] S.-W. Hsiao, Y.-C. Tsao, Applying multiple kinect on the development of a rapid 3D mannequin scan platform, international journal of mechanical, aerospace, industrial, Mechatron. Manuf. Eng. 10 (2016) 1198–1203.
[34] J. Tong, J. Zhou, L. Liu, Z. Pan, H. Yan, Scanning 3D full human bodies using kinects, IEEE Trans. Vis. Comput. Graph. 18 (2012) 643–650.
[35] E. Auvinet, J. Meunier, F. Multon, Multiple depth cameras calibration and body volume reconstruction for gait analysis, 2012 11th International Conference on Information Science, Signal Processing and their Applications (2012) 478–483.
[36] P. Henry, M. Krainin, E. Herbst, X. Ren, D. Fox, RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments, Int. J. Rob. Res. 31 (2012) 647–663.
[37] S. Springer, G.Y. Seligmann, Validity of the kinect for gait assessment: A focused review, Sensors (Switzerland) 16 (2016).
[38] R.P. Kuster, B. Heinlein, C.M. Bauer, E.S. Graf, Accuracy of KinectOne to quantify kinematics of the upper body, Gait Posture 47 (2016) 80–85.
[39] B.F. Mentiplay, L.G. Perraton, K.J. Bower, Y.H. Pua, R. McGaw, S. Heywood, et al., Gait assessment using the Microsoft Xbox one Kinect: concurrent validity and inter-day reliability of spatiotemporal and kinematic variables, J. Biomech. 48 (2015) 2166–2170.
[40] R.A. Clark, Y.H. Pua, A.L. Bryant, M.A. Hunt, Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining, Gait Posture 38 (2013) 1064–1066.
[41] J.J. Liddy, H.N. Zelaznik, J.E. Huber, S. Rietdyk, L.J. Claxton, A. Samuel, et al., The efficacy of the Microsoft KinectTM to assess human bimanual coordination, Behav. Res. Methods 49 (2017) 1030–1047.
[42] R. Tanaka, H. Takimoto, T. Yamasaki, A. Higashi, Validity of time series kinematical data as measured by a markerless motion capture system on a flatland for gait assessment, J. Biomech. 71 (2018) 281–285.
[43] G. Baldewijns, G. Verheyden, B. Vanrumste, T. Croonenborghs, Validation of the kinect for gait analysis using the GAITRite walkway, 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2014) 5920–5923.
[44] E. Dolatabadi, B. Taati, A. Mihailidis, Concurrent validity of the Microsoft Kinect for Windows v2 for measuring spatiotemporal gait parameters, Med. Eng. Phys. 38 (2016) 952–958.
[45] M. Gabel, R. Gilad-Bachrach, E. Renshaw, A. Schuster, Full body gait analysis with kinect, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2012) 1964–1967.
[46] J. Latorre, R. Llorens, C. Colomer, M. Alcañiz, Reliability and comparison of Kinect-based methods for estimating spatiotemporal gait parameters of healthy and post-stroke individuals, J. Biomech. 72 (2018) 268–273.
[47] E.E. Stone, M. Skubic, Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing, Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference, (2011), pp. 6491–6494.
[48] M. Eltoukhy, C. Kuenze, M.S. Andersen, J. Oh, J. Signorile, Prediction of ground reaction forces for Parkinson's disease patients using a kinect-driven musculoskeletal gait analysis model, Med. Eng. Phys. 50 (2017) 75–82.
[49] P. Plantard, A. Muller, C. Pontonnier, G. Dumont, H.P.H. Shum, F. Multon, Inverse dynamics based on occlusion-resistant Kinect data: Is it usable for ergonomics? Int. J. Ind. Ergon. 61 (2017) 71–80.
[50] M. Eltoukhy, J. Oh, C. Kuenze, J. Signorile, Improved kinect-based spatiotemporal and kinematic treadmill gait assessment, Gait Posture 51 (2017) 77–83.
[51] T.W. Macpherson, J. Taylor, T. McBain, M. Weston, I.R. Spears, Real-time measurement of pelvis and trunk kinematics during treadmill locomotion using a low-cost depth-sensing camera: a concurrent validity study, J. Biomech. 49 (2016) 474–478.
[52] X. Xu, R.W. McGorry, L.S. Chou, J.H. Lin, C.C. Chang, Accuracy of the Microsoft Kinect™ for measuring gait parameters during treadmill walking, Gait Posture 42 (2015) 145–151.
[53] A. Pfister, A.M. West, S. Bronner, J.A. Noah, Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis, J. Med. Eng. Technol. 38 (2014) 274–280.
[54] S. Motiian, P. Pergami, K. Guffey, C.A. Mancinelli, G. Doretto, Automated extraction and validation of children's gait parameters with the Kinect, Biomed. Eng. Online 14 (2015).
[55] D. Pratyusha, C. Kingshuk, C. Arijit, C. Debatri, S. Aniruddha, P. Arpan, Improving joint position estimation of Kinect using anthropometric constraint based adaptive Kalman filter for rehabilitation, Biomed. Phys. Eng. Express 4 (2018) 035002.
[56] V. Bonnet, C. Azevedo Coste, L. Lapierre, J. Cadic, P. Fraisse, R. Zapata, et al., Towards an affordable mobile analysis platform for pathological walking assessment, Rob. Auton. Syst. 66 (2015) 116–128.
[57] B.F. Mentiplay, K. Hasanki, L.G. Perraton, Y.H. Pua, P.C. Charlton, R.A. Clark, Three-dimensional assessment of squats and jumps using the Microsoft Xbox one Kinect: reliability and validity, J. Sports Sci. (2018) in press.
[58] P.N. Pathirana, S. Li, H.M. Trinh, A. Seneviratne, Robust real-time bio-kinematic movement tracking using multiple kinects for tele-rehabilitation, Ieee Trans. Ind. Electron. 63 (2016) 1822–1833.
[59] D.J. Geerse, B.H. Coolen, M. Roerdink, Kinematic validation of a multi-Kinect v2 instrumented 10-meter walkway for quantitative gait assessments, PLoS One 10 (2015).
[60] Y. Kim, S. Baek, B.C. Bae, Motion capture of the human body using multiple depth sensors, Etri J. 39 (2017) 181–190.
[61] S. Moon, Y. Park, D.W. Ko, I.H. Suh, Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering, Int. J. Adv. Robot. Syst. 13 (2016) 65.
[62] F. Destelle, A. Ahmadi, N.E. O'Connor, K. Moran, A. Chatzitofis, D. Zarpalas, et al., Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors, Eur. Signal Proc. Conf. (2014) 371–375.
[63] D. Tan, Y.H. Pua, S. Balakrishnan, A. Scully, K. Bower, K. Prakash, et al., Automated analysis of gait and modified timed up and go using the Microsoft Kinect in people with Parkinson's disease: associations with physical outcome measures, Med. Biol. Eng. Comput. (2018) in press.
[64] R.A. Clark, S. Vernon, B.F. Mentiplay, K.J. Miller, J.L. McGinley, Y.H. Pua, et al., Instrumenting gait assessment using the Kinect in people living with stroke: reliability and association with balance tests, J. Neuroeng. Rehabil. 12 (2015).
[65] E. Dolatabadi, B. Taati, A. Mihailidis, Automated classification of pathological gait after stroke using ubiquitous sensing technology, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2016) 6150–6153.