

RESEARCH

Open Access



Multiple Kinect based system to monitor and analyze key performance indicators of physical training

Karolis Ryselis¹, Tautvydas Petkus², Tomas Blažauskas¹, Rytis Maskeliūnas³ and Robertas Damaševičius^{1*} 

*Correspondence:
robertas.damasevicius@ktu.lt
¹ Department of Software
Engineering, Kaunas
University of Technology,
51368 Kaunas, Lithuania
Full list of author information
is available at the end of the
article

Abstract

Using a single Kinect device for human skeleton tracking and motion tracking lacks of reliability required in sports medicine and rehabilitation domains. Human joints reconstructed from non-standard poses such as squatting, sitting and lying are asymmetric and have unnatural lengths while their recognition error exceeds the error of recognizing standard poses. In order to achieve higher accuracy and usability for practical smart health applications we propose a practical solution for human skeleton tracking and analysis that performs the fusion of skeletal data from three Kinect devices to provide a complete 3D spatial coverage of a subject. The paper describes a novel data fusion algorithm using algebraic operations in vector space, the deployment of the system using three Kinect units, provides analysis of dynamic characteristics (position of joints, speed of movement, functional working envelope, body asymmetry and the rate of fatigue) of human motion during physical exercising, and evaluates intra-session reliability of the system using test–retest reliability metrics (intra-class correlation, coefficient of variation and coefficient of determination). Comparison of multi-Kinect system vs single-Kinect system shows an improvement in accuracy of 15.7%, while intra-session reliability is rated as excellent.

Keywords: Kinect sensors, Motion tracking, Data fusion, Physical training, Smart health

Introduction

Physical training is an effective tool for health diagnostics, rehabilitation, and prevention of unwanted health problems such as obesity, hypertension or Parkinson's disease [2, 49, 58]. Kinetic therapy sessions can help both healthy people and people with motor dysfunctions to improve their control, strength, skills and the range of motion [22]. However, older and disabled people may have difficulties attending physical training sessions outside of their homes, while they rarely perform enough exercises at home due to low motivation. Motivation can be increased by exergames, which employ the advances in virtual reality, sensory and motion tracking technologies [17, 37, 41]. Motion sensor technology can be employed to monitor certain elements of physical training and provide motivation by a virtual personal trainer [62]. Analysis of human skeleton data in motion can be used to provide feedback on the incorrect body posture [59] or incorrect

sequence of exercises as well as to analyse the accuracy of movements [52], perform gait analysis for "Get Up and Go Test" in rehabilitation [9], capture long-term trends in human physical fitness [10], and perform ergonomic studies [50] as well as to facilitate in rehabilitation of stroke patients [64] and to recognize hand gestures [63]. Data obtained by using the Kinect sensor is recognized as of sufficient quality for using for diagnostics and rehabilitation [11, 15, 55] and can serve as an alternative to other high cost motion tracking systems such as Vicon and OptiTrack.

Therefore, there is a need for developing a reliable human motion observation system that could be used in home environment for providing feedback on physical training exercises performed by healthy or recovering subjects. The aim of the paper is to propose a novel method for fusion of data provided by three Kinect devices, and to evaluate the method experimentally using a multi-Kinect based virtual training system. The paper describes the proposed multi-Kinect sensor system, which uses three Kinect devices for monitoring in-home training and provide an accurate view on subject's performance and movement related state-of-health characteristics. Our contributions are detailed as follows:

1. A novel data fusion algorithm using algebraic operations in vector space,
2. The deployment of the system using three Kinect units,
3. Analysis of dynamic characteristics of human motion during physical exercising,
4. Evaluation of intra-session reliability of the system using test–retest reliability metrics (intra-class correlation coefficient, coefficient of variation and coefficient of determination).

The structure of the remaining parts of paper is as follows. "Related work" section discusses the related work. "Methods" section describes the three-Kinect based subject tracking method, presents the deployment of multiple Kinect units, and discusses the measured human skeleton performance measures. "Experiments and results" section provides the results of experiments and "Discussion" section discusses the results and outlines the limitations. Finally, "Conclusions" section presents conclusions.

Related work

Since its arrival in 2010, the Microsoft Kinect™ (Kinect) [53] technology have been used for various applications. Kinect combines optical video Red, Green, and Blue (RGB) camera and infrared (IR) radar based depth-sensing technologies for skeleton tracking and capturing of 3D motion. In 2014, a new and more precise Kinect sensor based on time-of-flight technology was introduced [47]. Kinect SDK 2.0 allows tracking of up to 25 body joints. With Kinect sensors able to detect human motions in real time, they offer possibilities for enhancing the physical and social well-being of people with restricted mobility, and assisted living environments for the elderly and people with disabilities [23].

One of the main drawbacks of the Kinect skeleton model that makes it difficult to directly apply for healthcare is the use of a non-anthropometric kinematic model, which allows for variable limb lengths [45]. The accuracy of Kinect may be improved by more precise estimation of anatomical features, using the best orientation of Kinect facing

the subject, or using multiple Kinect units. More complicated applications for analysis of complex human movement sequences require to use multiple cameras to capture orthogonal views of the same subject in order to extract the motion information and assure an objective evaluation of the training progress. For example, the studies have reported the use of two [16], three [31, 65, 66], four [44, 51], or even five [39] Kinect sensors for estimating joint positions.

Combination of data from multiple Kinect devices requires the solution of several technical problems such as fusion of inconsistent and noisy depth measurements, and estimates of 3D joints' positions. Using point clouds and depth information obtained from multiple cameras and performing object detection on colour images can improve the detection of a person using a combination of multiple Kinects [57]. Different variants of deployment of Kinect devices can be used for obtaining the 3D model of skeleton, for example, by using different Kinect devices to capture different parts of a human body [7], to capture depth data and RGB data from different viewpoints [16], to aggregate tracked data by weighting [4], to solve occlusion problems by data fusion [31]. A human pose recognition system utilizing a combination of body pose estimation and tracking using ridge body parts features from the joints points of the skeleton model, capable of achieving the mean recognition rate of 91.19%, is described in [23]. The same team presented a real-time tracking system for body parts pose recognition utilizing the ridge data of depth maps to estimate 3D body joint angles using the forward kinematic analysis [25]. A bag of features approach to re-identifying people among different view-independent multi camera tracks can achieve higher than 90% classification rate [21].

Accurate skeleton reconstruction from multiple sensors requires specific calibration procedures. Calibration procedures for multiple Kinect sensors with at least three acquisitions (point cloud fusion) are considered in [12]. By optimizing the re-projection error and setting weights to the external cameras in different locations, a joint calibration method of multiple devices is presented in [34]. Kim et al. [30] combine joint depth data retrieved from multiple sensors by transforming the coordinate systems in point clouds into a single coordinate system using the iterative closest point method. Chen et al. [8] combine the joint coordinates acquired by two Kinect devices to a common coordinate system and apply a heuristic skeleton fusion algorithm to reconstruct convinced human pose. Het et al. [20] adopted the information weighted consensus filter (IWCF) method based on proper weighting the prior and measurement information for human skeleton fusion from multiple view. Removing noisy effects from the background and tracking human silhouettes using temporal continuity constraints of human motion information can further improve the results [24].

The problem of accurately tracking the 3D motion of a monocular camera in a known 3D environment and dynamically estimating the 3D camera location is described in [32], suggesting a fully automated landmark-based camera calibration to initialize the motion estimation and employ extended Kalman filtering techniques to track landmarks and to estimate the camera location. Kalman filter also was used in other studies such as [13, 33, 39, 44, 48] for reducing the noise in the acquired signals. Several studies have demonstrated that Kalman filter has achieved the best denoising performance when compared to other filter-based approaches [14]. Other types of filters such as double exponential smoothing filter [65], median filtering [67],

fourth-order low-pass Butterworth filter [40]. also have been used. However, we use of filtering methods have not been demonstrated to increase the accuracy for multi-Kinect systems [10].

The reliability of Kinect V2 is not lower as that of other (high cost) motion tracking systems and Kinect can be used as a reliable and valid clinical measurement tool [33, 46, 68]. However, such studies typically used simple poses such as standing, walking, sit down and stand up, and for more complex poses such as performing different kinds of physical exercises, the accuracy reliability still could improved using multiple Kinect sensors rather than a single sensor, which was demonstrated to fail, for example, for tracking a lying person [43].

Several studies analysed the use of multiple Kinect sensors for human tracking [5, 39, 54, 56], however, these studies were oriented at tracking multiple skeletons at the same time, and all experiments were performed using standard poses when a subject is standing on both feet and performing movements in front of cameras. None of these studies were validate for uncommon poses such as a person lying on the ground. The summary of the related work is presented in Table 1.

Methods

In this section, we describe our proposed human subject tracking method based on the use of three Kinect sensors. The method is based on the fusion of data received from three Kinect devices and includes the alignment of the Kinect coordinate systems using algebraic operations in vector space. Hereinafter we describe the application of our method by detailing the required deployment of Kinect devices. We finalize this section with the description of skeleton performance measures implemented for the assessment of human limb performance during a physical training session.

Table 1 Summary of related work on multi-Kinect systems

References	Number of Kinect sensors used	Data pre-processing (filtering) methods used	Data (skeleton) fusion methods used
[3]	3	Mamdani Fuzzy Inference Scheme (FIS)	Maximization of energy sum function over candidate positions
[8]	2	Not used	Coordinate transform with joint state inference
[12]	4	Not used	Point cloud matching
[16]	2	Not used	Point cloud registration in a maximum a posteriori framework
[20]	4	Information weighted-consensus filter (IWCF)	Interactive multiple model (IMM)
[31]	3	Not used	Iterative Closest Point (ICP) algorithm & Rigid Transformation (RT) to reference frame
[39]	5	Kalman filter	Weighted Measurement Fusion
[56]	3	Kalman filter	Trajectory matching
[57]	4	Median filter	Iterative Closest Point (ICP) algorithm
[65]	3	Double exponential smoothing	Weighted position averaging

Proposed data fusion algorithm

Suppose we have Kinect sensors K_1, K_2, \dots, K_n that monitor an intersecting volume of space. The position of each Kinect sensor p is denoted by coordinates $C_{np} = (x_p, y_p, z_p)$. Each sensor has its own coordinate system CS_p and all the data sent by that sensor are provided in this coordinate system. For simplicity, we consider only two sensors K_1 and K_2 , and two reference joints J_1 and J_2 . We transform the local coordinate systems obtained from different cameras into a single global coordinate system using linear algebra operations in vector space [35]. Let us denote the transformation that transforms data from a Kinect sensor p to a common coordinate system as T_p . Then the final coordinates of point q from sensor p in a common coordinate space are $(x_{fq}, y_{fq}, z_{fq}) = T_p(x_{pq}, y_{pq}, z_{pq})$.

The transformation consists of two steps:

1. Rotate sensor coordinate space so that its xOz plane matches the floor plane. It is needed because each sensor is oriented at different angles to the floor.
2. Rotate and move coordinate space so that it matches common coordinate space.

Step 1 is needed because each sensor is oriented at different angles to the floor. Fortunately, the Kinect sensor detects and reports the floor plane. Given the fact that all sensors monitor the intersecting volume of space, in most cases all sensors will stand on the same floor plane. The goal of the first transformation is to modify each sensor's coordinate space so that its xOz plane is the same as the floor plane. Note that the y axis points upwards in the Kinect's coordinate system (Fig. 1).

Suppose that floor plane equation in sensor's p coordinate system is.

$$A_p x + B_p y + C_p z + D_p = 0. \tag{1}$$

Then the normal vector for the plane is $\vec{P}_p = \begin{bmatrix} A_p \\ B_p \\ C_p \end{bmatrix}$. The desired normal vector for this plane is $\vec{N} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, because it represents the desired xOz plane. Then there is a matrix T_{p1} that could be applied to vector \vec{P}_p to get vector \vec{N} : $\vec{P}_p T_{p1} = \vec{N}$. The

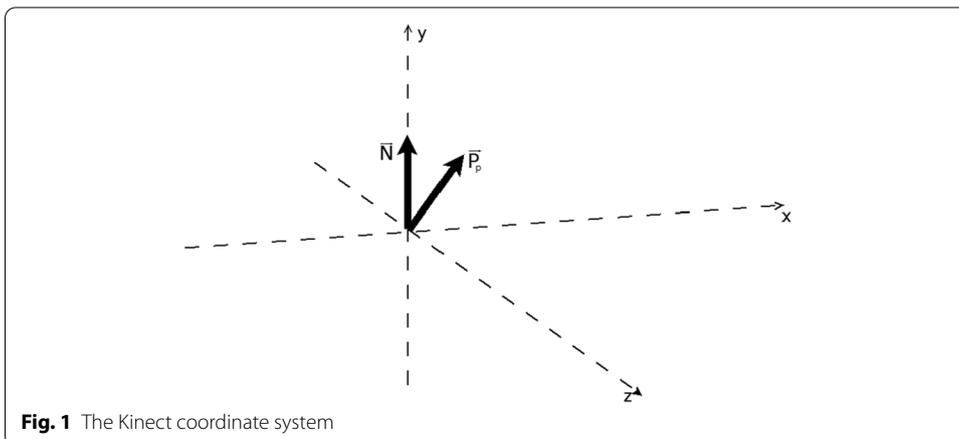


Fig. 1 The Kinect coordinate system

transformation could be applied to whole sensor’s point space. After this transformation sensor stays above the floor at the distance D , so it must be subtracted from the result we got after the transformation. Thus, the final transformation to transform the sensor’s coordinate system so that its position and orientation matches the floor is:

$$A_{tp} = A_p T_{p1} - (0, D_p, 0) \tag{2}$$

here A_p is original sensor’s p space, T_{p1} —transformation matrix, D_p —free coefficient from floor plane equation and A_{tp} is transformed sensor’s p space.

After this transformation, all sensors lie on the same plane and are oriented with no tilt. This simplifies further transformations as we only need to study a two-dimensional case only.

Suppose we have two sensors K_1 and K_2 and two reference joints J_1 and J_2 . Let us use the origin of the K_1 sensor coordinate system as the base. We can select any point in the space monitored by both sensors, say, J_3 , and two vectors $\overrightarrow{K_1 J_3}$ and $\overrightarrow{K_2 J_3}$ (see Fig. 2). The first vector’s coordinates in coordinate space CS_1 are the coordinates of point J_3 in this coordinate system. The same holds true for the second vector and CS_2 . The vector connecting both origins of coordinate spaces is $\overrightarrow{K_1 K_2}$. It is easy to see that $\overrightarrow{K_2 K_1} = \overrightarrow{K_2 J_3} - \overrightarrow{K_1 J_3}$. The vector CS_2 must be shifted by to match CS_1 .

First, we must find the angle between the coordinate systems of both sensors. Let us denote the vector $\overrightarrow{J_1 J_2}$ as \vec{J} . This vector has different coordinates in each sensor’s coordinate system. Let us choose the polar coordinate system. Then the system vector’s coordinates are (R_1, φ_1) for sensor K_1 and (R_2, φ_2) for sensor K_2 . The angle φ_1 shows the angle between sensor’s K_1 abscissa axis and vector \vec{J} and φ_2 shows the angle between sensor’s K_2 abscissa and the same vector \vec{J} . Let us rotate the vector \vec{J} by the value of $-\varphi_1$. This would change the polar rotation coordinate of the vector in both sensors’ coordinate systems by this value. Then the resulting vector’s direction matches the sensor’s x axis direction and the new angle between K_2 abscissa and \vec{J}

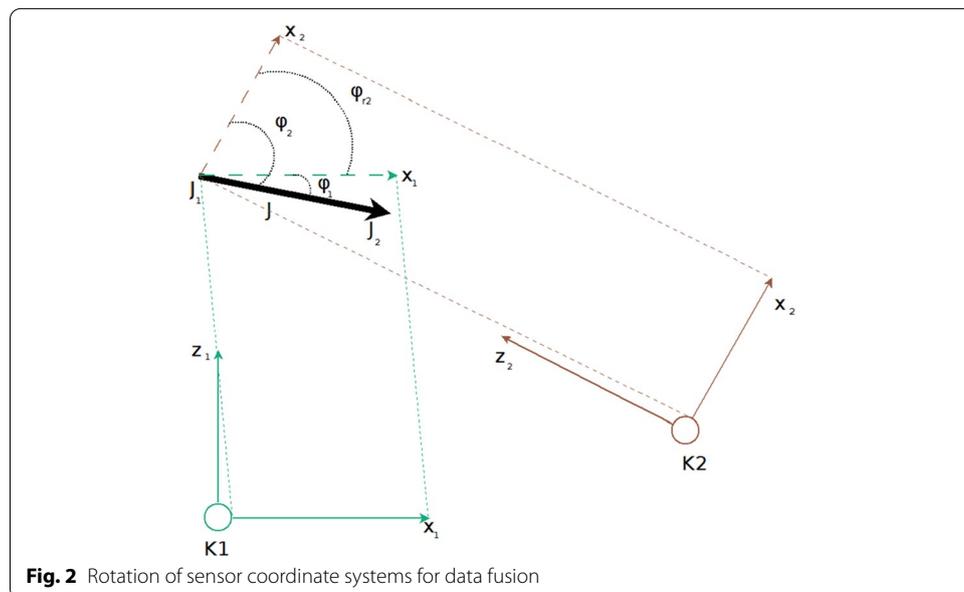


Fig. 2 Rotation of sensor coordinate systems for data fusion

is $\varphi_2 - \varphi_1$. As both rotated vector and K_1 x axis point the same direction, this is also the angle φ_{r2} between the coordinate systems of K_1 and K_2 . To find the value of φ_{r2} , we need to find the values of φ_1 and φ_2 as follows: $\varphi_{rp} = \varphi_p - \varphi_1$, where

$$\varphi_i = \begin{cases} \arccos\left(\frac{x_i}{\sqrt{x_i^2+z_i^2}}\right), & \text{if } x_i \geq 0 \\ 2\pi - \arccos\left(\frac{x_i}{\sqrt{x_i^2+z_i^2}}\right), & \text{if } x_i < 0 \end{cases}, \tag{3}$$

To apply the transformation, the rotation could be done in the polar coordinate system and then transformed into the Cartesian coordinates. If the original coordinates of a point J_q are $[x_q, y_q]$, in the polar coordinate system, they become $[R_q, \varphi_q] = [\sqrt{x_q^2 + y_q^2}, \varphi_q]$. Then we need to rotate this by angle φ_{r2} and the resulting vector is $[R_q, \varphi_q + \varphi_{r2}]$ which, in square coordinate system, is equal to $[R_q \cos(\varphi_q + \varphi_{r2}), R_q \sin(\varphi_q + \varphi_{r2})]$ (Fig. 2) as follows:

$$B_{t2} = [R_q \cos(\varphi_q + \varphi_{r2}), R_q \sin(\varphi_q + \varphi_{r2})] \text{ for } \forall q \in B_{t1}, \tag{4}$$

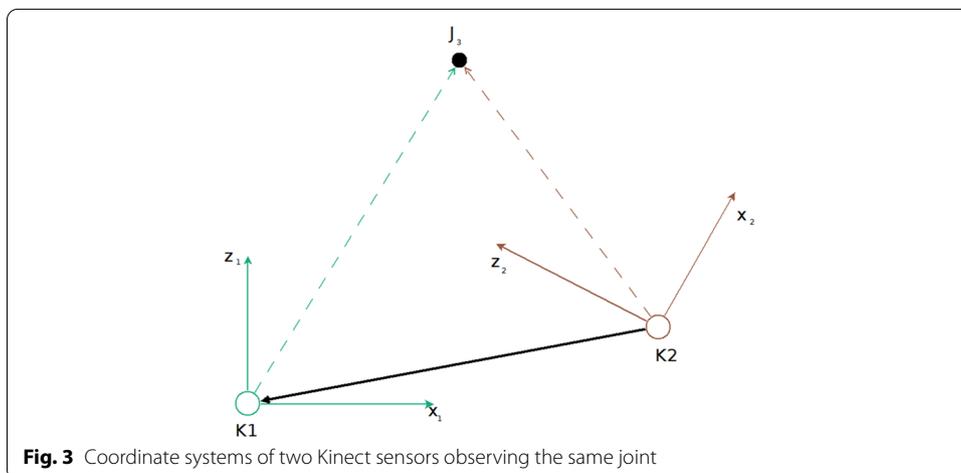
here $R_q = \sqrt{x_q^2 + y_q^2}$.

Once we have applied the transformations T_{p1} and rotation φ_{r2} , we need to move both sensors' coordinate systems' origins to the same point. After these transformations, the sensors will be oriented parallel to floor, on the same height, facing the same direction and on the same point in space. Thus, the coordinate systems of both sensors will be the same.

Suppose that the coordinates of J_3 (see Fig. 3) are $[x_{13}, y_{13}]$ in the coordinate space CS_1 and $[x_{23}, y_{23}]$. Then the required transformation vector is

$$T_{21} = [x_{23}, y_{23}] - [x_{13}, y_{13}], \tag{5}$$

In general case, we compare sensor K_p against K_1 . The required transformation is:



$$T_{p1} = [x_{p3}, y_{p3}] - [x_{13}, y_{13}], \quad (6)$$

Thus, the transform of a set of points B_p from sensor's K_p coordinate space CS_p to the sensor's K_1 coordinate space CS_1 as B_1 is as follows:

$$B_{t1} = B_p T_{p1} - [0, D_p, 0], \quad (7)$$

here B_p is the original coordinate space of sensor p , T_{p1} – the transformation matrix, D_p – the free coefficient from floor plane equation and B_{tp} is the transformed coordinate space of sensor p . Select any vector J_3 of two points known by both sensors with coordinates $[x_{13}, y_{13}]$ in CS_1 and $[x_{p3}, y_{p3}]$ in CS_p :

$$T_{p3} = [x_{p3}, 0, y_{p3}] - [x_{p3}, 0, y_{p3}], \quad (8)$$

$$B_1 = B_{t2} + T_{p2} \quad (9)$$

If sensors do not move during monitoring, the position of sensors does not need to be re-evaluated after each calculation. The parameters φ_{rp} and T_{p2} can be pre-calculated using the same methods as described above, and the transform is simplified as follows:

$$B_{t1} = B_p T_{p1} - [0, D_p, 0], \quad (10)$$

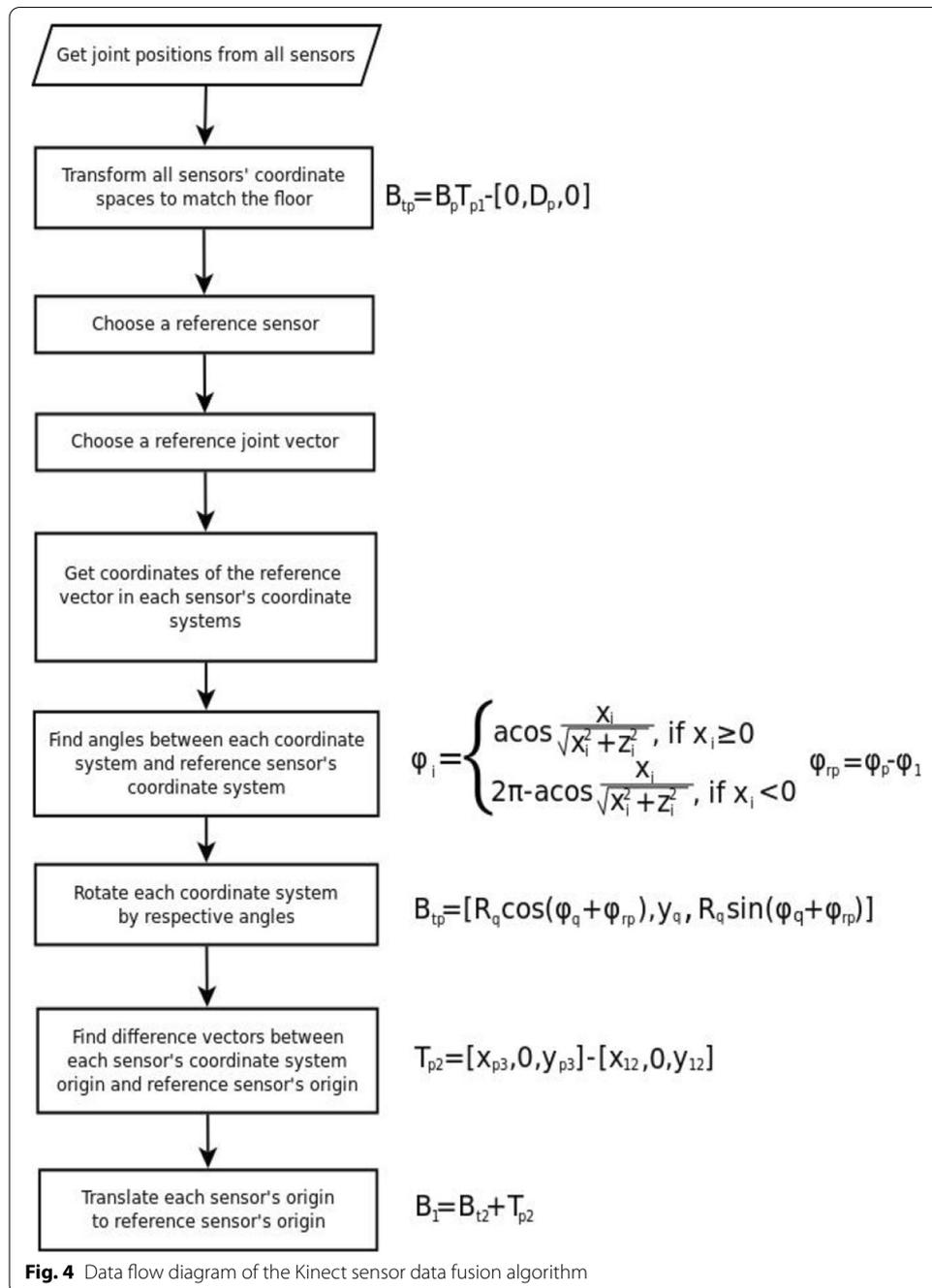
$$B_{t2} = [R_q \cos(\varphi_q + \varphi_{r2}), y_q, R_q \sin(\varphi_q + \varphi_{r2})] \quad \text{for } \forall q \in B_{t1}, \quad (11)$$

here $R_q = \sqrt{x_q^2 + y_q^2}$ and $B_1 = B_{t2} + T_{p2}$.

This transformation could be applied to any number of sensors. The base sensor K_1 must be chosen and data from each other sensor K_p could be transformed to coordinate space CS_1 using the suggested algorithm one by one. The algorithm does not require to know the positions of sensors in advance, so any configuration of the Kinect sensors could be used. However, due to noisy input and camera capture errors, the obtained skeletal joints in the global coordinate system may not coincide perfectly. Therefore, the averages of joint coordinates are used to best represent the skeleton. The aggregated human skeleton is computed from the average positions of joints. The calculations are summarized as an algorithm in a data flow diagram in Fig. 4.

Deployment of kinect units

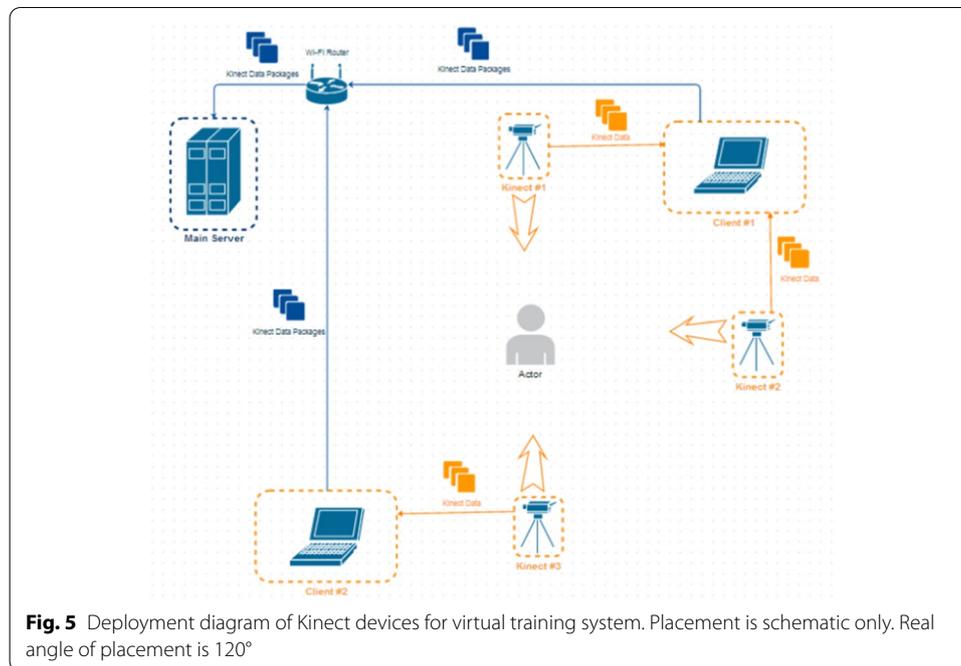
In Fig. 5, the deployment of three Kinect V2 devices for the capture of human skeleton positions is given. The subject is assumed to be standing in the middle of the room, while Kinect devices are located around him at 120° angles with respect to each other while keeping within the typical range of Kinect sensors (1.2–3.5 m). The system uses three Kinect sensors and three Client Personal Computers (PCs) for sensor data reading and processing. Each Kinect sensor device is connected to its own computer. The system also has the Wi-Fi Router for transmission of data between computers and Main Server. The data streamed is packetized and contains RGB and depth stream data. The data is sent to the Main Server, where the data is aggregated, stored and processed. Since each Kinect device is connected to its own Client computer, the lag of the system does not exceed



the lag of a single Kinect unit system (60–80 ms). A total latency of the system, which included calculation of skeleton key performance indicators (KPIs), is between 60 and 80 ms (mean = 70.8 ms), which was determined using the USB mouse based method as described in [36].

Skeleton performance measures

To evaluate the quantitative performance of human skeleton during motion activities the systems provides several types of metrics (or KPIs) as follows: evolution of joint



movement amplitudes and velocities [6], position of joints, angles at joints, functional working envelope (FWE), velocity of joints, rate of fatigue [18], mean velocity of the hand, normalized mean speed, normalized speed peaks, shoulder angle, and elbow angle [38]. The angle at joint is calculated as the scalar product between the segments (links) that connect at a given joint. For example, to compute the elbow angle, the scalar product is calculated between the normalized forearm and upper arm vectors. The rate of fatigue is calculated as the average difference in the joint movement velocity in the first versus the last half of a training session [29]. FWE defines the volume generated using all possible points touched by a considered body limb.

Experiments and results

Data collection and processing

The data for the experiments was collected from 28 healthy subjects (16 males and 12 females) with no reported motoric disorders, aged 22–36 years (mean 25.6 ± 1.8), height 1.68–1.92 m. All subjects were informed about the purpose study and participated in the tests freely. Data collection was approved by the local ethics committee and strictly followed the principles of the Helsinki declaration.

We have set up three Kinect devices (as described in Section 3.2) that send wirelessly the registered joint data to a computer that performs the required computations to compose full human skeleton and analyse motion sequences. The subjects were informed to move within 1–3 m of distance with respect to the Kinect sensors so that the data would not be overly affected by low resolution of depth measurements and noise [61].

We have collected the recordings of the Kinect skeleton data and performed data fusion using custom software written in C#. We have recorded the 25 joints of skeleton data for each person, while the position of three Kinect sensors throughout our experiment has not changed. Three thousand frames of video capture were used in each

experiment. The captured data was post-processed by aligning all the modalities in the temporal domain.

For time synchronization, we have adopted the solution proposed in [42], which uses the precision time protocol (PTP) allowing to synchronize computers in a network with millisecond accuracy. The timestamps of the captured data frames were used to align the data streams from both Kinect devices in time.

Once the full skeleton data is obtained, further analyses have been performed to evaluate recognition accuracy and reliability. Finally, the skeleton data (positions of joints) were further analysed using MATLAB (MathWorks, Inc., Natick, MA) to calculate the individual skeleton KPI values and evaluate system's reliability.

Physical exercise protocol

For a physical exercise sequence, we adopted a training protocol described in [61]. The training protocol consisted of three parts: (1) Warm-Up (10 min): slow movements of main body muscle groups followed by static and dynamic stretching exercises, (2) Active exercise (45 min): continuous muscle activities with an increasing level of difficulty and intensity, starting with a short walk, alternated with step exercises on the platform. Then, the subjects perform upper-limb lifts and lower limb flexions and extensions (knee lifts, side and forward-backward leg lifts, leg curls), repeated over time, and (3) Recovery (10 min): various postural control and spine mobility exercises.

Posture analysis

Table 2 shows the aggregated data for several standard and non-standard human postures. Each posture was measured for 20 s, during which the subject was required to stand still. The best and worst recognized human joints with their recognition error are given, and the entire visibility of human skeleton is evaluated.

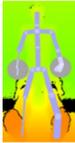
Assessment of accuracy

We assessed the accuracy of the developed multi-Kinect system using the marker tracking approach described in [60]. Reflective markers made of a polystyrene foam with a sticky back surface were attached to the joints of the human body (except hands) and tracked using a Vicon motion capture system (Vicon, Oxford, UK) with a sampling rate of 120 Hz. The Vicon tracking system was controlled by a different computer. Time synchronization between Vicon and our system was performed using cross-covariance of both data streams. The spatial coordinates of the reflective markers captured by Vicon were interpolated using cubic spline interpolation and downsampled to the original Kinect frame rate of 30 Hz. Then the coordinates were transformed to the Kinect coordinate system, assuming that X is assigned to the walking direction, Y is assigned to the vertical axis, and Z is the depth axis, and used as ground truth for comparing the accuracy of the proposed multi-Kinect system and a single Kinect system facing the subject. The results of comparison are presented Fig. 6. The overall results show an improvement of 15.7% in accuracy while using the multi-Kinect system. The result is statistically significant ($p < 0.001$ using the Student's paired t-test).

Table 2 Accuracy of the three-Kinect sensor based recognition of analysed human poses

Pose picture	Pose title/average error	Best recognized segment and its error	Worst recognized segment and its error	Human skeleton visibility
	Standing facing camera 10%	Right ankle–right hand (0.4%)	Left wrist–left hand (36%)	100% Left arm recognized incorrectly. Right leg is not fully tracked
	Standing sideways to the camera 12%	Spine mid–right shoulder (0.1%)	Mid hip–left hip (54%)	72% Hands are disproportionate. Right shoulder recognized incorrectly
	Squatting while holding legs 15%	Right ankle–right foot (1.5%)	Right ankle–right hand (44%)	68% Legs recognized incorrectly. Fingers are disproportionate
	Reaching toes standing 35%	Right hip–right knee (10.4%)	Spine mid–left shoulder (102%)	76% Hands are in wrong position. Head and spine recognized incorrectly
	Lying face down 17%	Right hip–right knee (1.4%)	Left wrist–left hand (70%)	68% Proportions of body parts are correct, but positioning is not correct
	Standing upside down 24%	Left hip–left knee (1.0%)	Neck–spine mid (71%)	Recognized as standing normally instead of upside down
	Standing on one leg. Bending forward and holding another leg 21%	Neck–spine mid (0.4%)	Left wrist–left hand (133%)	88% Hands are disproportionate
	Laying legs up 31%	Right knee–right ankle (0.7%)	Spine mid–right shoulder (106%)	64% Skeleton is completely warped: hands instead of feet. Other body parts are also in wrong places
	Stretching 32%	Spine mid–right shoulder (1.4%)	Right ankle–right hand (62%)	88% Body is not fully tracked–legs are tracked only above knees. Shoulder line is too high
	Low start pose 25%	Neck–spine mid (2.9%)	Right ankle–right hand (63%)	80% Legs are incorrect. Spine and neck are not accurately recognized

Table 2 (continued)

Pose picture	Pose title/average error	Best recognized segment and its error	Worst recognized segment and its error	Human skeleton visibility
	Sitting, legs straight 18%	Spine mid–right shoulder (1.4%)	Left wrist–left hand (53%)	96% Body part lower than shoulders are tracked incorrectly (points are in incorrect positions)
	Bending sideways 20%	Right hip–right knee (0.1%)	Spine mid–left shoulder (113%)	96% Most of the points are misaligned with actual pose
	Bend stretching while laying 21%	Spine mid–left shoulder (1.1%)	Left knee–left ankle (51%)	80% Both legs and arms are tracked on arms

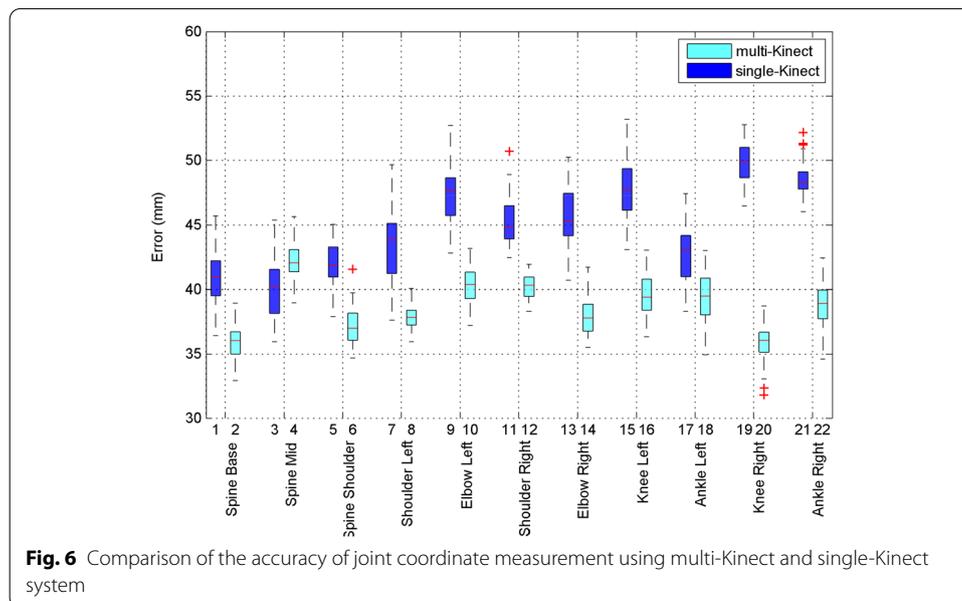
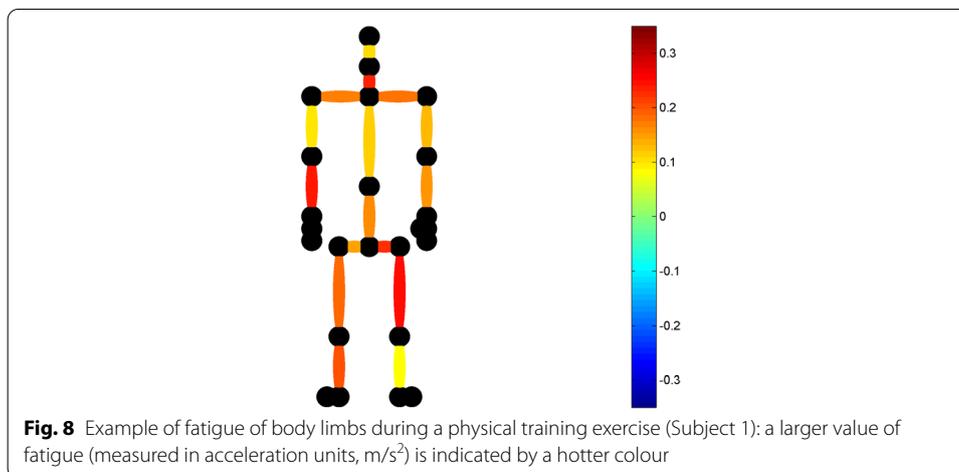
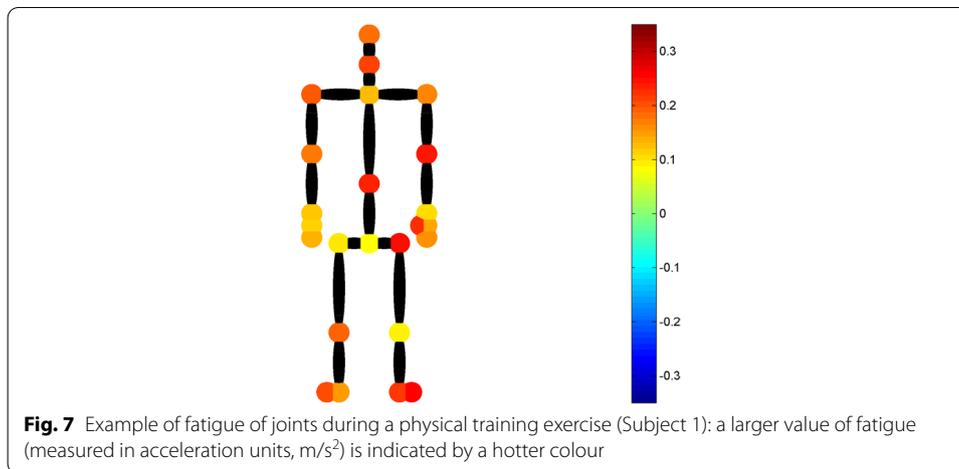


Fig. 6 Comparison of the accuracy of joint coordinate measurement using multi-Kinect and single-Kinect system

Analysis of dynamic characteristics of skeleton motion

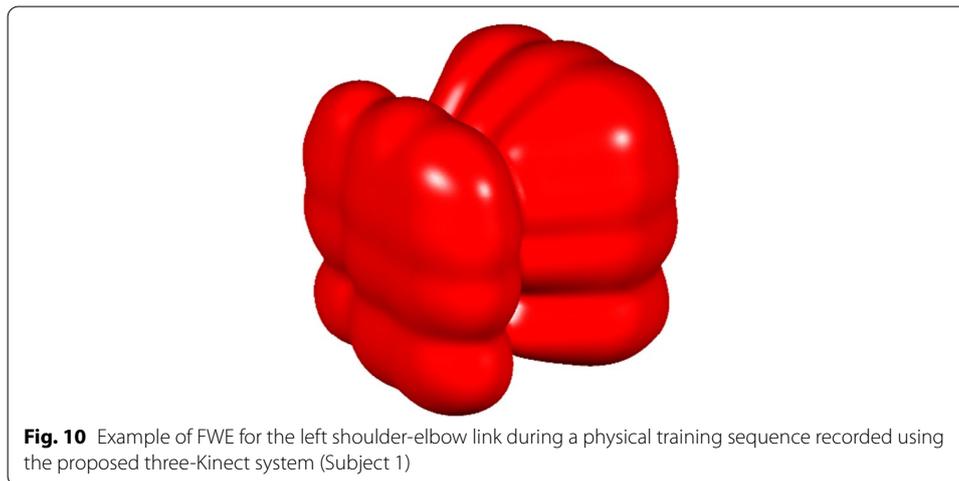
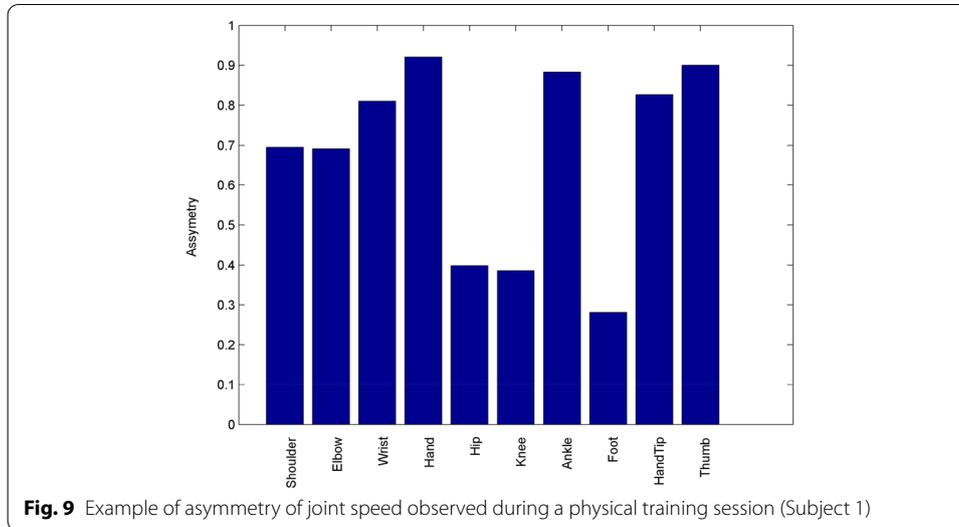
Following [27], we use the movement characteristics (amplitude, velocity) as a proxy variable to evaluate relative human fatigue during a physical training session. To analyse the dynamic characteristics of skeleton motion during the training exercise, the evolution of the speed of joints, which is computed as the distance travelled by the analysed joint in the time interval, is monitored. Figure 7 shows a graphical representation of joint fatigue calculated as the decrease of joint velocity in the second half of the training session with respect to the joint velocity in the first half of the session.

The travelled distance of each body limb that connects two joints of the body is also be used to evaluate relative fatigue during the training session (Fig. 8). This information can be used by a physiotherapist to adjust the training sequence or rehabilitation procedure.



The asymmetries in the joint movement amplitudes and speed between the left side and the right side of the body are important for monitoring the correctness of execution of training sequence as well as for rehabilitation of traumas. In some cases, such asymmetries can indicate some neurological disorders such as Huntington's disease due to rigidity of limbs. Here we calculate the asymmetry of the body movements as the ratio between the maximal speed of left side and right side joints achieved during the training session. The example of results is presented in Fig. 9. Note that we did not calculate mean values for all subjects due to individual differences in subjects, which make the averaging of values meaningless.

The FWE of a joint is calculated by collecting the positions of a joint in a 3D coordinate space. Then a probability density function (PDF) estimate of position points in the 3D space is calculated as the multiply of probability densities in each dimension. Finally, an isosurface is drawn at a specific threshold value of 3D PDF. The threshold value is calculated for the envelope to contain 95% of data points. The example of FWE for the shoulder-elbow link during a training exercise is given in Fig. 10. The volume and surface area of FWE can be used as a KPI for further analysis of human performance characteristics when performing physical motion tasks.



Evaluation of reliability

The reliability of human skeleton KPIs, i.e. normalized mean limb length (NML), normalized mean joint speed (NMS), normalized speed peaks (NSP) (as defined by [38]) were assessed using intra-class correlation coefficient (ICC), coefficient of variation (CoV) and coefficient of determination (R-squared) measures as suggested in [3]. Here normalized mean limb length (NML) is the mean value of the length of each body limb (link between adjacent joints) L_{mean} divided by its maximum value L_{max} . Normalized mean joint speed (NMS) is the mean value of the speed of each joint over time window V_{mean} divided by its maximum value V_{max} . Speed peaks are points where acceleration crosses the zero value and changes its sign. NSP is defined as the number of speed peaks divided by the number of data samples N .

$$NML = \frac{L_{mean}}{L_{max}} \quad (12)$$

$$NMS = \frac{V_{mean}}{V_{max}} \quad (13)$$

$$NSP = \frac{\text{Number of speed peaks}}{N} \quad (14)$$

The coefficient of variation (CoV) is a standardized measure of dispersion that is defined as the ratio of the standard deviation to the mean as follows:

$$CoV = \frac{\sigma(X)}{\mu(X)} \quad (15)$$

here σ is standard deviation, and μ is mean of sample X .

The coefficient of determination (R-squared) is the proportion of the variance in the dependent variable derived from the second session that can be predicted from the same variable derived from the first session. It is defined as squared correlation of data between first and second samples:

$$R^2 = \left(\frac{CoV(rX_1, rX_2)}{\sigma(rX_1)\sigma(rX_2)} \right)^2 \quad (16)$$

here rX_1 and rX_2 are ranked sequences of samples of X_1 and X_2 , and cov is the covariance.

The intra-session variabilities were analysed. Intra-session variability concerns the measurements taken during the same session, where a session was divided into two sub-sessions of equal length. The mean value and standard deviation as well as the intra-session test–retest reliability of the results expressed by ICC, R-squared and CoV are presented in Table 3. The results show that the performance indices, NMLL, NMS and NSP, all have more than 0.75 ICC values (excellent, according to (Lin, 1989)), and more than 0.8 R-squared (substantial, according to [19]) values together with acceptable CoV values.

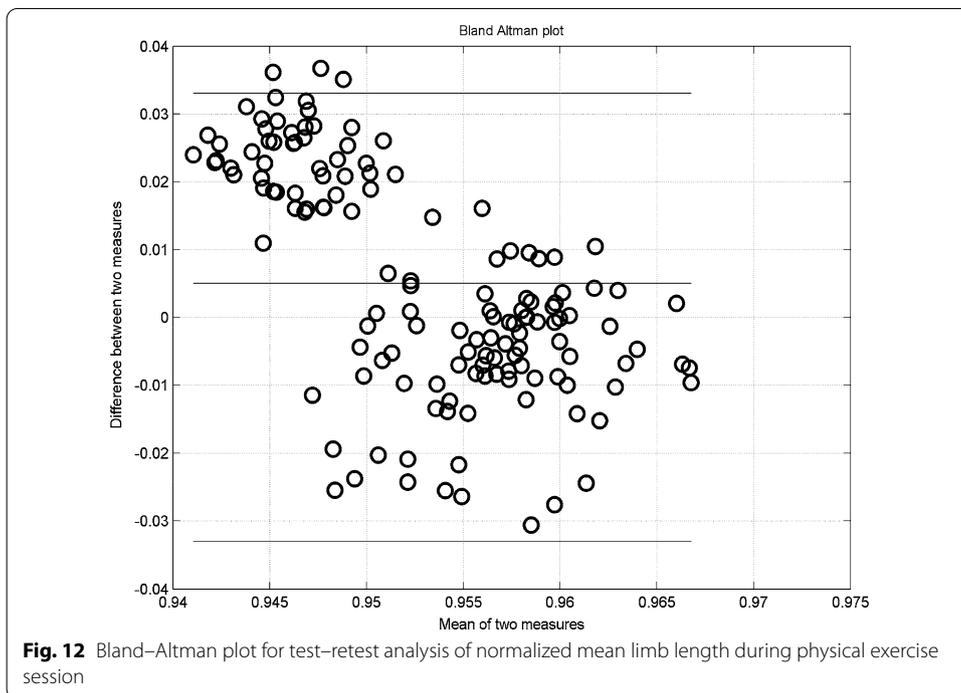
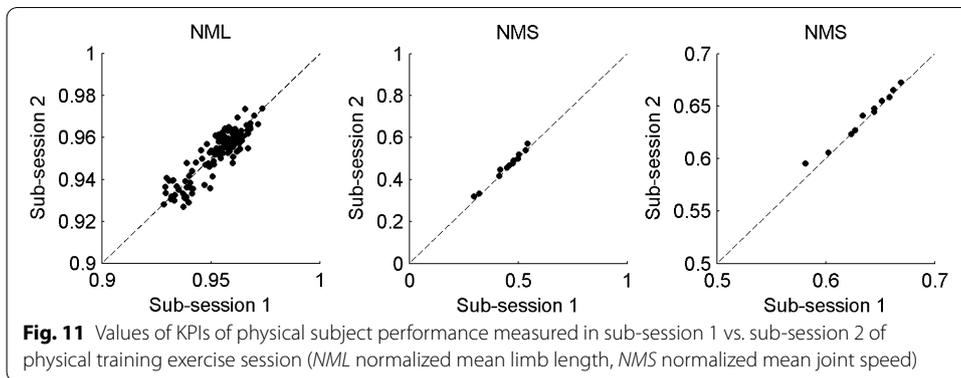
The subjects were informed to perform the same set of movements as uniformly as possible during the physical training session. The scatter plot of each KPI was plotted for first sub-session vs second sub-session as shown in Fig. 11. Good consistency of data requires that the values be located close to the identity line. To compare consistency, the coefficient of determination (R-squared) was calculated with respect to the identity line and is shown in Table 3.

The Bland–Altman Limit of Agreement (LoA) analysis was also performed and showed high correspondence between the measurements taken in the first and second halves of a

Table 3 Values of intra-session reliability of skeleton key KPIs obtained using three-Kinect system

KPI	Mean \pm SD	ICC	R-squared	CoV
NML	0.953 \pm 0.011	0.892	0.797	0.011
NMS	0.461 \pm 0.079	0.767	0.8611	0.171
NSP	0.639 \pm 0.025	0.903	0.802	0.038

NML normalized mean limb length, NMS normalized mean joint speed, NSP normalized speed peaks, SD standard deviation, ICC intra-class correlation coefficient, CoV coefficient of variation



session (see Fig. 12). Given two data samples X_1 , and X_2 , the Bland–Altman plot represents each data value as a point in the 2D coordinate space with coordinates [1]:

$$\left(\frac{x_1 + x_2}{2}, X_1 - X_2 \right) \tag{17}$$

here $x_1 \in X_1$, and $x_1 \in X_2$ are data values.

LoA are expressed both in absolute terms and as a proportion of the group mean. The majority of samples for NML values are within the 95% confidence limits.

Discussion

The Kinect sensor technology for human body tracking has limitations. Low accuracy of single face-oriented Kinect camera prevents from using it as a serious tool for physiotherapy, data collection and providing medical feedback about the patient’s performance

during the therapy sessions. Accuracy of Kinect drops when it is used in cluttered areas and the camera is not placed directly in front of the user. Inadequate calibration of sensors, overexposure or badly oriented calibration objects, specific properties of object surface, occlusions by other body parts or objects decrease the Kinect sensor's accuracy, too. The analysis of complex and non-standard human postures and motions such as squatting, sitting and lying using a single Kinect sensor has low recognition accuracy. The reconstructed human joints are asymmetric and have unnatural lengths while recognition error exceeds the error of recognizing standard body positions. Therefore, using a single Kinect device lacks of reliability required in sports medicine and rehabilitation procedures. In order to achieve higher accuracy or usability one needs to use multiple Kinects simultaneously. Using multiple Kinect devices arranged to track a subject from all sides allows to solve joint occlusion problem (which does not allow correct estimation of poses), to obtain higher joint recognition accuracy comparable with that of other similar known multi Kinect systems (see, e.g., Jalal et al. [26], and to derive valuable performance measures, which could evaluate the state of subject's skeletal systems and its evolution during physical training exercises.

We have achieved comparatively low error rates for poses, where one or several joints are concluded (e.g., standing on one leg—21%, lying face down—17%, squatting while holding legs—15%), which can not be recognized using a single subject-facing Kinect device due to low skeleton visibility. For example, in order to recognize a lying subject after the fall, Kepski and Kwolek [28] use an overhead mounted single Kinect device facing the floor, which obviously can not recognize other daily activity poses such as standing. Whereas in [16], the ratio of joint outliers (cases where the pose estimation fails), reaches up to 46%, depending upon orientation of the Kinect camera with respect to the subject. The comparison of results achieved using the proposed multi-Kinect system with a single Kinect system using reflective markers and the data captured by Vicon system as ground truth showed that a multi-Kinect sensor system provides more accuracy than a single Kinect sensor system.

Capability to evaluate individual motions of specific joints using skeleton KPIs allows to track their progress, provide feedback on additional physical training effort required or detect situations where a subject doesn't react well to the assigned training program. The physiotherapist can analyse evolution of physiological parameters such as angular amplitudes of limbs and movement speed of joints in a training session and across multiple sessions. For example, large decrease in joint speed and amplitudes of limb movements suggests that the patient becomes tired too quickly. This information can help the trainer to adjust the training program. FWE can help a therapist track performance and/or identify some specific mobility problems of a subject. A larger volume is likely to indicate an increased functional ability, while a less wide FWE can indicate a joint dysfunction or increased fatigue [18]. The asymmetry of maximum amplitudes and velocities achieved for a left and a right arm or leg may indicate a health problem, and can assist physiotherapists in analysing and monitoring the training progress by providing a quantitative estimate for the quality of motion and balance. KPIs could be used as valuable measures for patient rehabilitation as well.

To assess reliability of the results, we used the descriptive statistics and test-retest method followed by Bland-Altman statistical analysis as suggested in [55] based on

the review of studies in the domain. Our results are in-line with the results achieved by other authors (see Springer & Yogev Seligmann [55]: the ICC values were excellent, R-squared values were substantial, and the CoV values were acceptable CoV, while the Limits of Agreement according to the Bland–Altman analysis were within 5%.

The proposed method contributed towards the solution of multi-sensor data fusion problems, which are relevant when applying low-cost sensor solutions such as Kinect. The limitations of the study include a comparatively small number of subjects participating in the study. Another limitation is that Kinect v2 sensors are gradually retired and replaced by Azure Kinect, which is a next version of the Kinect technology. However, since there are still few studies performed with Azure Kinect, we hope that our study will make a valuable contribution towards the development and analysis of cloud-connected multiple sensors operating in assisted living environments. Validating the results of this study using Azure Kinect will be a subject of further research.

Conclusions

We have presented a novel solution for fusing skeletal representation data from multiple Kinect devices to provide a more complete coverage of a user, especially for uncommon poses such as lying or squatting. By suitably deploying Kinect sensors in the desired room, we can solve the limited visibility angle problem and recognize human joints regardless of the orientation angle: if one sensor is unable to recognize the human skeleton correctly, another sensor can recognize and provide more accurate information for the estimation of his/her physical performance during the physical training exercises.

By using a more accurate aggregated representation of human skeleton, the system can monitor the evolution of joints during motion tasks and calculate quantitative measures (KPIs), which provide a more accurate view on physical human performance while exercising. The reliability of the obtained KPIs has been validated using test–retest reliability metrics (ICC, R-squared, CoV). By monitoring the evolution of skeleton joints and calculating quantitative KPIs for the training sequence executed, such as the position of joints, speed of movement, functional working envelope, body asymmetry and the rate of fatigue (or reduced functional capability), the performance of a subject during in-home training can be evaluated by his/her therapist and/or trainer and the training programme can be adjusted accordingly.

Authors' contributions

Funding acquisition, TB; investigation, KR and TP; Methodology, TB and RM; Resources, TB; Software, KR and TP; Supervision, TB; Validation, TB and RM; Visualization, RD; Writing—original draft, KR and TP; Writing—review & editing, RM and RD. All authors read and approved the final manuscript.

Funding

This research received no external funding.

Availability of data and materials

Data is available on request.

Competing interests

The authors declare no conflicts of interest.

Human studies

Research on human subjects was approved by an Institutional Review Board of the Faculty of Informatics of Kaunas University of Technology.

Author details

¹ Department of Software Engineering, Kaunas University of Technology, 51368 Kaunas, Lithuania. ² UAB Telesoftas, 44150 Kaunas, Lithuania. ³ Department of Multimedia Engineering, Kaunas University of Technology, 51368 Kaunas, Lithuania.

Received: 10 April 2019 Accepted: 5 December 2020

Published online: 14 December 2020

References

1. Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. *Statistician* 32:307–317. <https://doi.org/10.2307/2987937>
2. Alves MLM, Mesquita BS, Morais WS, Leal JC, Satler CE, dos Santos Mendes FA (2018) Nintendo wii™ versus xbox kinect™ for assisting people with parkinson's disease. *Percept Mot Skills* 125(3):546–565. <https://doi.org/10.1177/0031512518769204>
3. Asteriadis S, Chatzitofis A, Zarpalas D, Alexiadis DS, Daras P (2013) Estimating human motion from multiple Kinect sensors. In: 6th international conference on computer vision/computer graphics collaboration techniques and applications (MIRAGE'13), Article 3, 6 p.
4. Baek S, Kim M (2015) Dance experience system using multiple kinects. *Int J Future Comput Commun* 4(1):45–49
5. Camalan S, Sengul G, Misra S, Maskeliūnas R, Damaševičius R (2018) Gender detection using 3d anthropometric measurements by kinect. *Metrol Meas Syst* 25(2):253–267. <https://doi.org/10.24425/119568>
6. Cary FC, Postolache O, Girão PM (2014) Kinect Based System and Serious Game Motivating Approach for Physiotherapy Assessment and Remote Session Monitoring. In: Proceedings of international conference on sensing technology—ICST 2014;1:1–5.
7. Chen G, Li J, Wang B, Zeng J, Lu G, Zhang D (2015) Reconstructing 3D human models with a Kinect. *Comput Anim Virt Worlds* 27(1):72–85. <https://doi.org/10.1002/cav.1632>
8. Chen N, Chang Y, Liu H, Huang L, Zhang H (2018) Human pose recognition based on skeleton fusion from multiple kinects. *CCC*. 2018:5228–5232. <https://doi.org/10.23919/ChiCC.2018.8483016>
9. Cippitelli E, Gasparrini S, Spinsante S, Gambi E (2015) Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view. *Sensors* 15:1417–1434
10. Clark RA, Mentiplay BF, Hough E, Pua YH (2019) Three-dimensional cameras and skeleton pose tracking for physical function assessment: a review of uses, validity, current developments and kinect alternatives. *Gait Posture* 68:193–200. <https://doi.org/10.1016/j.gaitpost.2018.11.029>
11. Clark RA, Pua YH, Oliveira CC, Bower KJ, Thilarajah S, McGaw R, Mentiplay BF (2015) Reliability and concurrent validity of the Microsoft Xbox one kinect for assessment of standing balance and postural control. *Gait Posture* 42(2):210–213. <https://doi.org/10.1016/j.gaitpost.2015.03.005>
12. Córdova-Esparza DM, Tervenb JR, Jiménez-Hernández H, Herrera-Navarro AM (2016) A multiple camera calibration and point cloud fusion tool for Kinect V2. *Sci Comp Program* 143:1–8
13. Das P, Chakravarty K, Chowdhury A, Chatterjee D, Sinha A, Pal A (2018) Improving joint position estimation of kinect using anthropometric constraint based adaptive kalman filter for rehabilitation. *Biomed Phys Eng Express* 4:3. <https://doi.org/10.1088/2057-1976/aaa371>
14. Edwards M, Green R (2014) Low-latency filtering of kinect skeleton data for video game control. In: Proceedings of the 29th international conference on image and vision computing New Zealand, IVCNZ 2014, pp. 190–195.
15. Galna B, Barry G, Jackson D, Mhiripiri D, Olivier P, Rochester L (2014) Accuracy of the microsoft kinect sensor for measuring movement in people with Parkinson's disease. *Gait Posture* 39(4):1062–1068. <https://doi.org/10.1016/j.gaitpost.2014.01.008>
16. Gao Z, Yu Y, Zhou Y, Du S (2015) Leveraging two kinect sensors for accurate full-body motion capture. *Sensors* 15:24297–24317
17. Garcia-Agundez A, Folkerts A, Konrad R, Caserman P, Tregel T, Goosses M, Kalbe E (2019) Recent advances in rehabilitation for parkinson's disease with exergames: A systematic review. *J Neuroeng Rehabil* 16:1. <https://doi.org/10.1186/s12984-019-0492-1>
18. Gauthier S, Cretu AM (2014) Human movement quantification using Kinect for in-home physical exercise monitoring. In: IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA), 6–11.
19. Hair JF, Sarstedt M, Hopkins L, Kuppelwieser VG (2014) Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *Eur Business Rev* 26(2):106–121
20. He H, Liu G, Zhu X, He L, Tian G (2019) Interacting multiple model-based human pose estimation using a distributed 3D camera network. *IEEE Sens J* 19(22):10584–10590. <https://doi.org/10.1109/JSEN.2019.2931603>
21. Huang Q, Yang J, Qiao Y (2012) Person re-identification across multi-camera system based on local descriptors. In: 2012 Sixth international conference on distributed smart cameras (ICDSC), 1–6.
22. Hulsteen RM, Johnson TM, Ridders ND, Mellecker RR, Barnett LM (2015) Children's movement skills when playing active video games. *Percept Mot Skills* 121(3):767–790. <https://doi.org/10.2466/25.10.PMS.121c24x5>
23. Jalal A, Kamal S, Kim D (2014) A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* 4(7):11735–11759
24. Jalal A, Kamal S, Kim D (2015) Shape and Motion Features Approach for Activity Tracking and Recognition from Kinect Video Camera. In: 2015 IEEE 29th international conference on advanced information networking and applications workshops, 445–450. <https://doi.org/10.1109/WAINA.2015.38>

25. Jalal A, Kim Y (2014) Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data. In: 11th IEEE international conference on advanced video and signal based surveillance (AVSS), 119–124. <https://doi.org/10.1109/AVSS.2014.6918654>
26. Jalal A, Uddin MZ, Kim TS (2012) Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE T Consum Electr* 58(3):863–871
27. Karg M, Venture G, Hoey J, Kulic D (2014) Human movement analysis as a measure for fatigue: a hidden markov-based approach. *IEEE Trans Neural Syst Rehabilitation Eng* 22(3):470–481. <https://doi.org/10.1109/tnsre.2013.2291327>
28. Kepski M, Kwolek B (2014) Fall detection using ceiling-mounted 3D depth camera. *Int Conf Comput Vision Theory Appl* 2014:640–647
29. Khoshelham K, Elberink SO (2012) Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* 12:1437–1454
30. Kim Y, Baek S, Bae BC (2017) Motion capture of the human body using multiple depth sensors. *ETRI J* 39:181–190. <https://doi.org/10.4218/etrij.17.2816.0045>
31. Kitsikidis A, Dimitropoulos K, Douka S, Grammalidis N (2014) Dance analysis using multiple Kinect sensors. In: 2014 international conference on computer vision theory and applications (VISAPP), 789–795.
32. Koller D, Klinker G, Rose E, Breen D, Whitaker R, Tuceryan M (1997) Real-time vision-based camera tracking for augmented reality applications. In: ACM Symposium on Virtual reality software and technology, VRST '97, 87–94.
33. Li C, Fahmy A, Sienz J (2019) An augmented reality based human-robot interaction interface using kalman filter sensor fusion. *Sensors* 19:20. <https://doi.org/10.3390/s19204586>
34. Liao Y, Sun Y, Li G, Kong J, Jiang G, Jiang D, Liu H (2017) Simultaneous calibration: a joint optimization approach for multiple kinect and external cameras. *Sensors* 17(7):1491
35. Lipschutz S (2012) *Linear Algebra*, 5th edn. McGraw-Hill Education, New York
36. Livingston MA, Sebastian J, Ai Z, Decker JW (2012) Performance measurements for the Microsoft Kinect skeleton. In: 2012 IEEE virtual reality workshops (VRW), 2012, 119–120.
37. Mateo F, Soria-Olivas E, Carrasco JJ, Bonanad S, Querol F, Pérez-Alenda S (2018) HemoKinect: a Microsoft Kinect V2 based exergaming software to supervise physical exercise of patients with hemophilia. *Sensors* 18:2439
38. Mobini A, Behzadipour S, Saadat M (2015) Test–retest reliability of Kinect’s measurements for the evaluation of upper body recovery of stroke patients. *Biomed Eng Online* 14:75
39. Moon S, Park Y, Ko DW, Suh IH (2016) Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering. *Int J Adv Robot Syst* 13(2):65. <https://doi.org/10.5772/62415>
40. Moreira GM, Giovanini LHF, de Castro MPR, Nogueira GN, Boumer TC, Manffra EF (2019) Filtering motion signals from microsoft kinect® in the context of stroke rehabilitation. *Res Biomed Eng* 35(3–4):265–270. <https://doi.org/10.1007/s42600-019-00029-8>
41. Mugueta-Aguinaga I, Garcia-Zapirain B (2017) FRED: exergame to prevent dependence and functional deterioration associated with ageing. A pilot three-week randomized controlled clinical trial. *Int J Environ Res Public Health* 14:1439
42. Müller B, Ilg W, Giese MA, Ludolph N (2017) Validation of enhanced kinect sensor based motion capturing for gait assessment. *PLoS ONE* 12(4):e0175813. <https://doi.org/10.1371/journal.pone.0175813>
43. Naeemabadi M, Dinesen B, Andersen O, Najafi S, Hansen J (2018) Evaluating Accuracy and Usability of Microsoft Kinect Sensors and Wearable Sensor for Tele Knee Rehabilitation after Knee Operation. In: 11th international joint conference on biomedical engineering systems and technologies. *Biodevices* 1:128–135. <https://doi.org/10.5220/0006578201280135>
44. Núñez JC, Cabido R, Montemayor AS, Pantrigo JJ (2017) Real-time human body tracking based on data fusion from multiple RGB-D sensors. *Multimed Tools Appl* 76(3):4249–4271. <https://doi.org/10.1007/s11042-016-3759-6>
45. Obdržálek Š, Kurillo G, Ofli F, Bajcsy R, Seto E, Jimison H, Pavel M (2012) Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In: Annual international conference of the IEEE engineering in medicine and biology society, 1188–1193.
46. Otte K, Kayser B, Mansow-Model S, Verrel J, Paul F, Brandt AU, Schmitz-Hübsch T (2016) Accuracy and reliability of the kinect version 2 for clinical measurement of motor function. *PLoS ONE* 11(11):e0166532. <https://doi.org/10.1371/journal.pone.0166532>
47. Pagliari D, Pinto L (2015) Calibration of kinect for Xbox one and comparison between the two generations of microsoft sensors. *Sensors* 15:27569–27589
48. Palmieri P, Melchiorre M, Scimmi LS, Pastorelli S, Mauro S (2021) Human arm motion tracking by kinect sensor using kalman filter for collaborative robotics. In: *Mechanisms and machine science*. Springer International Publishing. pp. 326–334. https://doi.org/10.1007/978-3-030-55807-9_37
49. Pin-Barre C, Laurin J (2015) Physical exercise as a diagnostic, rehabilitation, and preventive tool: influence on neuroplasticity and motor recovery after stroke. *Neural Plast* 608:581
50. Plantard P, Auvinet E, Pierres ASL, Multon F (2015) Pose estimation with a kinect for ergonomic studies: evaluation of the accuracy using a virtual mannequin. *Sensors* 15:1785–1803
51. Ruchay AN, Dorofeev KA, Kolpakov VI (2018) Fusion of information from multiple Kinect sensors for 3D object reconstruction. *Computer Optics* 42(5):898–903. <https://doi.org/10.18287/2412-6179-2018-42-5-898-903>
52. Saenz-de-Urturi Z, Garcia-Zapirain SB (2016) Kinect-based virtual game for the elderly that detects incorrect body postures in real time. *Sensors* 16:704
53. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A (2011) Real-time human pose recognition in parts from a single depth image. *IEEE Comput Vis Pattern Recognit* 56:1297–1304
54. Shuai L, Li C, Guo X, Prabhakaran B, Chai J (2017) Motion capture with ellipsoidal skeleton using multiple depth cameras. *IEEE Trans Vis Comput Graph* 23(2):1085–1098
55. Springer S, Yogeve Seligmann G (2016) Validity of the kinect for gait assessment: a focused review. *Sensors* 16(2):194. <https://doi.org/10.3390/s16020194>

56. Sun SW, Kuo CH, Chang PC (2018) People tracking in an environment with multiple depth cameras: a skeleton-based pairwise trajectory matching scheme. *J Vis Commun Image Represent* 35:36–54
57. Susanto W, Rohrbach M, Schiele B (2012) 3D Object Detection with Multiple Kinects. In: *Computer Vision – ECCV 2012*. In: Workshops and demonstrations. LNCS, vol. 7584, 93–102.
58. Tan D, Pua Y, Balakrishnan S, Scully A, Bower KJ, Prakash KM, Clark RA (2019) 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* 57(2):369–377. <https://doi.org/10.1007/s11517-018-1868-2>
59. Tariq M, Majeed H, Beg MO, Khan FA, Derhab A (2019) Accurate detection of sitting posture activities in a secure IoT based assisted living environment. *Future Gener Comp Sy* 92:745–757. <https://doi.org/10.1016/j.future.2018.02.013>
60. Timmi A, Coates G, Fortin K, Ackland D, Bryant AL, Gordon I, Pivonka P (2018) Accuracy of a novel marker tracking approach based on the low-cost Microsoft Kinect v2 sensor. *Med Eng Phys* 59:63–69. <https://doi.org/10.1016/j.medengphy.2018.04.020>
61. Todde F, Melis F, Mura R, Pau M, Fois F, Magnani S, Tocco F (2016) A 12-Week vigorous exercise protocol in a healthy group of persons over 65: study of physical function by means of the senior fitness test. *Biomed Res Int* 2016:1–6
62. Ubert T, Forberger S, Gansefort D, Zeeb H, Brand T (2017) Community capacity building for physical activity promotion among older adults—a literature review. *Int J Environ Res Public Health* 14:1058
63. Vaitkevičius A, Tarozza M, Blažauskas T, Damaševičius R, Maskeliūnas R, Woźniak M (2019) Recognition of American sign language gestures in a virtual reality using leap motion. *Appl Sci* 9:3. <https://doi.org/10.3390/app9030445>
64. Webster D, Celik O (2014) Systematic review of Kinect applications in elderly care and stroke rehabilitation. *J. Neuro-Eng, Rehabil*, p 11
65. Wu Y, Gao L, Hoermann S, Lindeman RW (2018) Towards robust 3D skeleton tracking using data fusion from multiple depth sensors. In: 10th international conference on virtual worlds and games for serious applications, VS-Games 2018, <https://doi.org/10.1109/VS-Games.2018.8493443>
66. Yang L, Zhang L, Dong H, Alelaiwi A, El Saddik A (2015) Evaluating and improving the depth accuracy of Kinect for Windows v2. *IEEE Sens J* 15(8):4275–4285
67. Yang K, Peng L, Tong L, Liu R, Liu B (2019) An assessment method for upper limb rehabilitation training using kinect. In: 8th annual IEEE international conference on cyber technology in automation, control and intelligent systems, CYBER 2018, 949–953. <https://doi.org/10.1109/CYBER.2018.8688256>
68. Yang Y, Pu F, Li Y, Li S, Fan Y, Li D (2014) Reliability and validity of kinect rgb-d sensor for assessing standing balance. *IEEE Sens J* 14(5):1633–1638. <https://doi.org/10.1109/JSEN.2013.2296509>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
