On using the Microsoft Kinect™ sensor in the analysis of human motion

M.J. Malinowski⁹, E. Matsinos*, S. Roth⁸

⁹Institute of Mechatronic Systems, School of Engineering, Zurich University of Applied Sciences (ZHAW), Technikumstrasse 5, CH-8401 Winterthur, Switzerland

⁸Institute of Applied Information Technology, School of Engineering, Zurich University of Applied Sciences (ZHAW), Steinbergasse 13, CH-8401 Winterthur, Switzerland

*E-mail: evangelos[DOT]matsinos[AT]zhaw[DOT]ch, evangelos[DOT]matsinos[AT]sunrise[DOT]ch

Abstract

The present paper explores the possibility of using the Microsoft Kinect™ sensor in the analysis of human motion; we attempt the validation of the output of the original version of the sensor on the basis of a marker-based system which is assumed to provide the reference solution (baseline, ‘ground truth’). The similarity between the two outputs is assessed after comparing a number of waveforms, representing the variation within the gait cycle of quantities which are commonly used in order to characterise (and model) motion. The data acquisition involved a commercially-available treadmill and five velocity settings: walking data were acquired at 5 km/h, running data at 8, 10, 11, and 12 km/h. The analysis revealed three problems with such an application of the Kinect sensor: the systematic underestimation of the knee angle by about 25%, the appearance of artefacts in the motion of the lower leg of the subject, and the inability of the sensor to capture reliably the details regarding the asymmetry of the motion for the left and right parts of the subject’s body.

PACS: 87.85.gj; 07.07.Df

Key words: Kinect, motion analysis, biomechanics, treadmill

1 Introduction

Microsoft Kinect™ (hereafter, simply ‘Kinect’) [1], a low-cost, portable motion-sensing hardware device, was developed by the Microsoft Corporation (Microsoft, USA) as an accessory to the Xbox 360 video-game console (2010).
The sensor is a webcam-type, add-on peripheral device, enabling the operation of Xbox via gestures and spoken commands. In 2011, Microsoft released the software-development kit (SDK) for Kinect, thus enabling the development of applications in several standard programming languages. The first upgrade of the sensor (‘Kinect for Windows v2’), both hardware- and software-wise, tailored to the needs of Xbox One, has recently become available for general development and use. The present paper aims at investigating the possibility of involving the original version of the sensor in the analysis of motion data of subjects walking or running ‘in place’ (on a commercially-available treadmill). If successful, Kinect could become an interesting alternative to marker-based systems (MBSs) in capturing data for motion analysis.

Regarding medical applications of the Kinect sensor, e.g., in physiotherapy in the home environment, a number of products are available, e.g., by ‘Home Team’ (Massachusetts, USA, https://www.hometeamtherapy.com) and ‘Reflexion’ (California, USA, http://www.reflexionhealth.com). In particular, it is known that ‘Reflexion’ aims at increasing the success rates in rehabilitation; the participation of the US Navy in the tests of their product attests to the importance of the availability of such solutions.

The validation of the output of the Kinect sensor in static measurements or in case of slow movements was a popular subject in the recent past. Data acquired from 20 healthy subjects, performing three simple tasks (forward reach, lateral reach, and single-leg, eyes-closed standing balance) were analysed in Ref. [2]. In their work, the authors drew attention to some systematic effects in the Kinect output, in particular concerning the sternum.

Even more optimistic were the results obtained (and the conclusions drawn) in Ref. [3], which investigated the accuracy in the determination of the joint angles from data acquired with a Kinect sensor and with an MBS; an inclinometer was assumed to provide the reference or baseline solution (also called ‘ground truth’). The authors concluded that the differences in accuracy and reliability between the two measurement systems were small, thus enabling the use of Kinect as “a viable tool for calculating lower extremity joint angles”. In fact, the Kinect results and those obtained with the inclinometer were found to agree to better than 2°. Interestingly, the authors also made a point regarding the depth measurements with Kinect, which are subject to increasing uncertainty with increasing distance from the sensor, reaching a maximal value of about 4 cm at the most distal position of the sensor, which (according to the specifications) should not exceed about 4.5 m. Naturally, such a dependence introduces bias in the analysis of walking and running data acquired with a treadmill. These effects are present regardless of the largeness of the subject’s stray motion in depth; for example, the depth uncertainties are different at the two extreme positions of the lower leg, i.e., ahead of and behind the walker/runner.

2
Another medical application of Kinect was investigated in Ref. [4], namely its use for home-based assessment of movement symptoms in patients with Parkinson’s disease. In that study, a number of tasks were performed by 19 subjects, ten of which comprised the control group; parallel data were acquired with Kinect and with an MBS. The authors reported that the Kinect results were generally (but not in all cases examined) found to correlate well with those obtained with the MBS for a variety of movements.

Regarding the use of Kinect in medical/health-related applications, complacency and optimism were impaired after the paper of Bonnechère et al. [5] appeared. The authors recorded data from 48 subjects, performing four simple tasks (shoulder abduction, elbow flexion, hip abduction, and knee flexion), and compared the results of different sessions pursuing both their reproducibility within each measurement system, as well as an assessment of the differences between the two systems. The authors concluded that the lower body is not tracked well by Kinect. The conclusions of Ref. [5] constitute rather disturbing news in terms of applications of the sensor in monitoring walking and running behaviour in a medical/health-related environment.

The literature on the biomechanics of motion is extensive. Some selected works, relevant to the present study, include Refs. [6,7,8,9,10,11]. Earlier scientific works are cited therein, in particular in the review articles [9,10,11].

- Cavagna and Kaneko [6] studied the efficiency of motion in terms of the mechanical work done by the subject’s muscles.
- Cavanagh and Lafortune [7] studied the ground reaction forces in running at about 16.2 km/h, as well as the motion of the ‘centre of the pressure distribution’ during the stance phase of the right foot of 17 subjects. Relatively large variability is seen in their results, partly due to the different characteristics in the motion of rear-foot and mid-foot strikers, partly reflecting the extent of their database in terms of running experience, weekly training, and (perhaps, more importantly) habitual individual behaviour. The vertical component of the ground reaction force, which may be as large as three times the subject’s body weight, showed sizeable variability.
- Cairns, Burdett, Pisciotta, and Simon [8] analysed the motion of ten competitive race-walkers in terms of the ankle flexion, of the knee and hip angles, as well as of the pelvic tilt, obliquity, and rotation. The work discussed the main differences between walking and race-walking, and provided explanations for the peculiarity of the motion in the latter case, invoking the goal of achieving higher velocities (than in normal walking) while maintaining double support with fully-extended knee and suppressing the vertical undulations of the subject’s centre of mass (CM).
- Öunpuu [9] discussed important aspects of the biomechanics of gait, including the variation of relevant physical quantities within the gait cycle. That work may be used as a starting point for those in seek of an overview in
the topic. It must be borne in mind that the subjects used in Ref. [9] were children.

- Novacheck [10] also provided an introduction to the biomechanics of motion. Figs. 5 and 6 of that work contain the variation of the important angles (projections on the coronal, sagittal, and transverse planes) within the gait cycle, at three velocity settings: 4.32 km/h (walking), 11.52 km/h (running), and 14.04 km/h (sprinting). Fig. 9 therein provides the variation of the joint moments and powers (kinesics) in the sagittal plane within the gait cycle.
- In a subsequent article [11], Schache, Bennell, Blanch, and Wrigley investigated the inter-relations in the movement of the lumbar spine, pelvis, and hips in running, aiming at optimising the rehabilitation process in case of relevant injuries.

To the best of our knowledge, the present work constitutes the first study addressing the possibility of using Kinect in motion analysis. The material has been organised in six sections. In Section 2, the two measurement systems used in our study (i.e., the original Kinect sensor and the MBS) are described in detail. The definitions of some important quantities used in the description of the motion, as well as the details regarding the data analysis, are provided in Section 3. In Section 4, we describe the experimental part of the study. The results, obtained from the analysis of the data of Section 4, are presented in Section 5. We discuss the implications of our findings in the last section.

2 Description of our motion-capturing systems

To enable the analysis of the data with the same software application in the present study, the output of the MBS is transformed into Kinect format, using reasonable associations between the Kinect nodes and the markers of the MBS; due to the removal of the constant offsets in the data analysis (see Subsection 3.2.3), the exact matching between the Kinect nodes and the locations of the MBS markers is not essential.

2.1 The original Kinect sensor

The skeletal data (‘stick figure’) of the Kinect output, for each time frame involved in the sampling, comprises 20 three-dimensional (3D) vectors of spatial coordinates, i.e., measurements of the \((x,y,z)\) coordinates of the 20 nodes which the original Kinect sensor associates with the axial and appendicular parts of the subject’s skeleton. In coronal (frontal) view of the subject (sensor view), the Kinect coordinate system is defined with the \(x\) axis (medial-lateral) pointing to the left (i.e., to the right part of the body of the subject being
viewed), the $y$ axis (vertical) upwards, and the $z$ axis (anterior-posterior) away from the sensor, see Fig. 1.

The nodes 1 to 4 are main-body nodes, identified as HIP.CENTER, SPINE, SHOULDER.CENTER, and HEAD. The nodes 5 to 8 relate to the left arm: SHOULDER_LEFT, ELBOW_LEFT, WRIST_LEFT, and HAND_LEFT; similarly, the nodes 9 to 12 on the right arm are: SHOULDER_RIGHT, ELBOW_RIGHT, WRIST_RIGHT, and HAND_RIGHT. The eight remaining nodes pertain to the legs, the first four to the left (HIP_LEFT, KNEE_LEFT, ANKLE_LEFT, and FOOT_LEFT), the remaining four to the right (HIP_RIGHT, KNEE_RIGHT, ANKLE_RIGHT, and FOOT_RIGHT) leg of the subject. The nodes of the Kinect sensor may be seen in Fig. 2.

Parallel to the video image, Kinect captures an infrared image, produced from the reflected radiation of a laser grid generated by the infrared emitter (seen on the left in Fig. 1); captured with a CCD camera, this infrared image yields the information on the depth $z$ of the objects being viewed by the sensor. The sampling frequency in the Kinect output (both for the video and the skeletal data) is 30 Hz.

The description of the algorithm, used in the determination of the 3D positions of the skeletal joints of the subject being viewed with Kinect, may be found in Ref. [12]. Candidate values for the 3D positions of each skeletal joint are obtained via the elaborate analysis of each depth image separately. These positions may be used as starting points in an analysis featuring the temporal and kinematic coherence in the subject’s motion; it is not clear whether such a procedure has been hardcoded in the preprocessing (hardware processing) of the captured data. Shotton et al. define 31 body segments covering the subject’s body, some of which are used in order to localise skeletal joints, some to fill the gaps or yield predictions for other joints. In the development of their algorithm, Shotton et al. generated static depth images of humans (of children and adults) in a variety of poses (synthetic data). The application of their method results in the extraction of probability-distribution maps for the 3D positions of the skeletal joints; their joint proposals represent the modes (maxima) in these maps. According to the authors, the probability-distribution maps are both accurate and stable, even without the imposition of temporal or kinematic constraints. It must be borne in mind that the ‘3D positions of the joints’ of Ref. [12] are essentially produced from the ‘3D positions of the projections of the joints onto the front part of the subject’s body’ after applying a ‘shift’ in depth (i.e., from the surface to the interior of the subject’s body), namely a constant offset ($\zeta_c$) of 39 mm (see end of Section 3 of Ref. [12]). Although the ‘computational efficiency and robustness’ of the

---

1 The left and right parts of the subject refer to what the subject perceives as the left and right parts of his/her body.
procedure are praised in Ref. [12], it remains to be seen whether results of similar quality can be obtained in dynamic applications (e.g., when the subject is in motion).

2.2 The Vicon output

The Vicon Tracker [13] (Vicon Motion Systems Ltd, UK) is a powerful object-tracking solution, comprising a number of cameras (seven were used in the context of the present project) which provide high-quality, low-latency data, sampled herein at 200 Hz. Being an MBS, the Vicon measurement system outputs the coordinates of reflective markers placed on the subject’s body. Information on the ‘Plug-in Gait’ placement scheme of the markers, followed herein, may be found in Refs. [14]. A total of 39 19-mm (diameter) markers were used (see Table 1), in a placement which is commonly known as ‘31 + 8’. The Vicon output was finally transformed into Kinect format using the following association scheme.

- The Kinect-equivalent HEAD was assigned to the midpoint of the marker positions LFHD and RFHD. The marker positions LBHD and RBHD, pertaining to the back of the head, were not used.
- The Kinect-equivalent SHOULDER_CENTER was taken to be the marker position CLAV. The marker positions C7 and RBAK, which are placed on the back part of the body, were not used.
- The Kinect-equivalent SPINE was estimated as an average of the marker positions T10, LPSI, and RPSI.
- The Kinect-equivalent SHOULDER_LEFT and SHOULDER_RIGHT were taken to be the marker positions LSHO and RSHO, respectively. Regarding the upper part of the body, the marker positions LUPA, LFRA, RUPA, and RFRA were not used.
- The Kinect-equivalent ELBOW_LEFT and ELBOW_RIGHT were taken to be the marker positions LELB and RELB, respectively.
- The Kinect-equivalent WRIST_LEFT and WRIST_RIGHT were assigned to the midpoints of the marker positions LWRA and LWRB, and of RWRA and RWRB, respectively.
- The Kinect-equivalent HAND_LEFT and HAND_RIGHT were taken to be the marker positions LFIN and RFIN, respectively.
- The Kinect-equivalent KNEE_LEFT and KNEE_RIGHT were taken to be the corrected (according to Ref. [15]) marker positions LKNE and RKNE, respectively.
- The Kinect-equivalent ANKLE_LEFT and ANKLE_RIGHT were taken to be the corrected (according to Ref. [15]) marker positions LANK and RANK, respectively.
- The Kinect-equivalent FOOT_LEFT and FOOT_RIGHT were taken to be
the marker positions LTOE and RTOE, respectively.

- The Kinect-equivalent HIP_LEFT and HIP_RIGHT positions were evaluated from those of the marker positions LASI, RASI, LPSI, and RPSI, according to Ref. [15]. Regarding the procedure set forth in that paper, a few comments are due. The positions of the hips were obtained therein using a model for the geometry of the pelvis, featuring three parameters \((\theta, \beta, \text{and } C)\), the values of which were obtained from a statistical analysis of radiographic data of 25 subjects; however, the values of these parameters are poorly known (see page 583 of Ref. [15]). A simple analysis of the uncertainties given in Ref. [15] shows that, when following that method, the resulting uncertainties in the estimation of the positions of the hips are expected to exceed about 10 mm in each spatial direction. As a result, the positions of the hips, calculated from the Vicon output according to that procedure, should not be considered as accurate as the rest of the information obtained from the Vicon system. More importantly, it is not evident how the movement of the pelvis reflects itself in the motion of the four markers which are used in the extraction of its 3D position and orientation; it is arguable whether any markers, placed on the surface of the subject’s body, capture the pelvic motion accurately.

- The Kinect-equivalent HIP_CENTER was estimated as an average of the Kinect-equivalent HIP_LEFT and HIP_RIGHT, and of the marker position STRN.

- Regarding the lower part of the body, the marker positions LTHI, LTIB, LHEE, RTHI, RTIB, and RHEE were not used.

With regard to the markers which are placed on the subject’s extremities, it must be borne in mind that their positions are also affected by rotations (around these extremities), not only by the translational motion of these extremities; the markers are placed at some distance from the actual rotation axes, coinciding with the longest dimension of the upper- and lower-extremity bones. For instance, rotating the left humerus by 90° around its long axis (assumed, for the sake of the argument, to align with the vertical axis \(y\)) will result in a movement of the marker LELB along a circular arc, thus affecting its \(x\) and \(z\) coordinates. On the other hand, the Kinect nodes are supposed to be placed on (or, in any case, closer to) the rotation axes; as a result, it is expected that they are less affected by such rotations. Given that the comparison of the outputs of the two measurement systems cannot take account of such effects, it can only be approximate.
3 Definitions and details on the data analysis

3.1 Definitions of some important angles characterising the motion

We will next describe how one may obtain from the output estimates of three important angles in the sagittal plane, representing the level of flexion of the trunk, of the hip, and of the knee. Estimates for the left and right parts of the body will be obtained for the hip and knee angles.

- **Trunk angle.** This angle is obtained from the \((y,z)\) coordinates of four points, comprising the nodes 1 (HIP_CENTER), 3 (SHOULDER_CENTER), and two midpoints, namely of the nodes 13 (HIP_LEFT) and 17 (HIP_RIGHT), and of the nodes 5 (SHOULDER_LEFT) and 9 (SHOULDER_RIGHT). An unweighted least-squares fit on the \((y,z)\) coordinates of these four points yields the slope \(\alpha\) (with respect to the \(y\) axis) of the optimal straight line. The trunk angle is defined as \(\theta_T = -\arctan(\alpha)\); \(\theta_T = 0^\circ\) in the upright position, positive for forward leaning.

- **Hip angle.** Two definitions of the hip angle have appeared in the literature: the angle may be defined with respect to the trunk or to the \(y\) axis; in the present paper, we adopt the latter definition. If the relevant hip coordinates are \((y_H,z_H)\) and those of the knee are \((y_K,z_K)\), the hip angle is obtained via the expression:

\[
\theta_H = \arctan \left( \frac{z_H - z_K}{y_H - y_K} \right). \tag{1}
\]

Two hip angles will be obtained: the left-hip angle \(\theta_{HL}\) uses the nodes 13 (HIP_LEFT) and 14 (KNEE_LEFT); the right-hip angle \(\theta_{HR}\) uses the nodes 17 (HIP_RIGHT) and 18 (KNEE_RIGHT).

- **Knee angle.** This is the angle between the femur (thigh) and the tibia (shank). Two definitions of the knee angle have appeared in the literature: the knee angle may be \(180^\circ\) or \(0^\circ\) in the extended position of the knee; we adopt the latter definition. It will shortly become clear why we make use of both the sine and the cosine of the knee angle:

\[
\beta_1 = \sin(\theta_K) = \frac{(y_A - y_K)(z_K - z_H) - (y_K - y_H)(z_A - z_K)}{L_f L_t} \tag{2}
\]

and

\[
\beta_2 = \cos(\theta_K) = \frac{(y_K - y_H)(y_A - y_K) + (z_K - z_H)(z_A - z_K)}{L_f L_t} \tag{3}
\]

As it is not clear at which depth (and on which basis) the Kinect sensor places node 2 (SPINE), this node should not be included in estimations involving the \(z\) coordinate.
where the coordinates of the ankle are denoted as \((y_A, z_A)\), and \(L_f\) and \(L_t\) are the projected lengths of the femur and the tibia onto the sagittal plane, respectively:
\[
L_f = \sqrt{(y_K - y_H)^2 + (z_K - z_H)^2}
\]
and
\[
L_t = \sqrt{(y_A - y_K)^2 + (z_A - z_K)^2}.
\]

We define the knee angle as:
\[
\theta_K = \begin{cases} 
\arccos(\beta_2), & \text{for } \beta_1 > 0 \\
\arcsin(\beta_1), & \text{otherwise}
\end{cases}.
\] (4)

Two knee angles will be obtained: the left-knee angle \(\theta_{KL}\) uses the nodes 13 (HIP\_LEFT), 14 (KNEE\_LEFT), and 15 (ANKLE\_LEFT); the right-knee angle \(\theta_{KR}\) uses the nodes 17 (HIP\_RIGHT), 18 (KNEE\_RIGHT), and 19 (ANKLE\_RIGHT).

We define four angles in the coronal plane: the lateral trunk, the lateral hip, the lateral knee, and the lateral pelvic angles; the lateral pelvic angle is also called pelvic obliquity. Estimates for the left and right parts of the body will be obtained for the lateral hip and lateral knee angles.

- **Lateral trunk angle.** The same four points, which had been used in the evaluation of the trunk angle in the sagittal plane, are also used in extracting an estimate of the lateral trunk angle; of course, the \((x, y)\) coordinates of these points must be used now. In addition to these nodes, node 2 (SPINE) has also been used. The lateral trunk angle is defined with respect to the \(y\) axis; \(\theta_{LT} = 0^\circ\) in the upright position, positive for tilting in the positive \(x\) direction (tilt of the subject to his/her right).

- **Lateral hip angle.** This angle describes hip abduction/adduction in the coronal plane. Similarly to the hip angle in the sagittal plane, two definitions of the lateral hip angle are possible: the angle may be defined with respect to the trunk or to the \(y\) axis; herein, we adopt the latter definition. If the relevant hip coordinates are \((x_H, y_H)\) and those of the knee are \((x_K, y_K)\), the lateral hip angle is obtained via the expression:
\[
\theta_{lH} = - \arctan \left( \frac{x_H - x_K}{y_H - y_K} \right).
\] (5)

Two lateral hip angles will be obtained: the lateral left-hip angle \(\theta_{lHL}\) uses the nodes 13 (HIP\_LEFT) and 14 (KNEE\_LEFT); the lateral right-hip angle \(\theta_{lHR}\) uses the nodes 17 (HIP\_RIGHT) and 18 (KNEE\_RIGHT).

- **Lateral knee angle.** This is the projection of the angle between the femur and the tibia onto the coronal plane.
\[
\theta_{lK} = \arcsin \left( \frac{(x_K - x_H)(y_A - y_K) - (x_A - x_K)(y_K - y_H)}{L_f L_t} \right),
\] (6)
where $L_f$ and $L_t$ are now redefined as the projected lengths of the femur and the tibia onto the coronal plane, respectively:

$$L_f = \sqrt{(x_K - x_H)^2 + (y_K - y_H)^2}$$

and

$$L_t = \sqrt{(x_A - x_K)^2 + (y_A - y_K)^2}.$$  

The angle is defined positive when, with respect to the femur direction, the ankle appears (in coronal view) ‘further away’ from the subject’s body. Of course, two lateral knee angles may be defined, corresponding to the left and right parts of the subject’s body, $\theta_{lKL}$ and $\theta_{lKR}$, respectively.

- **Pelvic obliquity.** This angle is defined as:

$$\theta_{lP} = \arctan \left( \frac{y_{HR} - y_{HL}}{x_{HR} - x_{HL}} \right),$$

where $(x_{HL}, y_{HL})$ and $(x_{HR}, y_{HR})$ are the $(x,y)$ coordinates of the left and right hips, respectively.

In relation to motion analysis, a few additional angles may be found in the literature: the pelvic tilt and the angle describing the plantarflexion/dorsiflexion of the foot are defined in the sagittal plane; the hip, pelvic, and foot rotations in the transverse plane. We do not believe that the Kinect output can yield reliable information on these quantities. The knee angle, obtained from the 3D vectors $(x_K - x_H, y_K - y_H, z_K - z_H)$ and $(x_A - x_K, y_A - y_K, z_A - z_K)$, will be called ‘knee angle in 3D’; it is easily evaluated using expressions analogous to Eqs. (2)-(4). In view of the fact that the angle between 3D vectors is invariant under rotations (SO(3) rotation group) and translations in 3D, the knee angle in 3D is independent of details regarding the alignment between the Kinect and the Vicon coordinate systems.

A few last comments are prompt.

1. The trunk angle $\theta_T$ is positive in walking and running; it is difficult to maintain balance if one leans backwards while moving forwards. However, the trunk angle, obtained from the Kinect output, is frequently negative. This is due to the fact that the nodes of the Kinect output, which enter the evaluation of $\theta_T$, do not represent locations on the spine.

2. Due to the properties of the knee joint, the knee angle is expected to satisfy the condition $\theta_K \geq 0$. In practice, even in the fully-extended position, $\theta_K$ remains (for many subjects) positive by a few degrees; knee hyperextension is a deformity. However, owing to the placement of the nodes by Kinect, the knee angle (estimated from the Kinect output) may occasionally come out negative. To examine further such cases, we have proposed the use of Eq. (4) in the evaluation of the knee angle.
(3) In Section 5, we provide evidence that the Kinect output for the coordinates of the hips is not reliable. Although the inaccuracy in the positions of the hips is expected to affect the determination of the hip and knee angles, other systematic effects (see Section 5) have a larger impact on these evaluations.

One possibility to avoid the effects, relating to the first two cases above, is to extract robust measures for the selected physical quantities from the data. For instance, one could use the variation of these quantities within the gait cycle or even their range of motion (RoM), i.e., the difference between the maximal and minimal values within the gait cycle. As long as an extremity moves as one rigid object, such measures (being essentially differences of two values) are not affected by a constant bias which may be present in the data.

3.2 Details on the data analysis

The motion of the subject on the treadmill is split into two components: the motion of the subject’s CM and the motion of the subject’s body parts relative to the CM. Of course, the accurate determination of the coordinates of the subject’s physical CM from the Kinect output is not possible. As a result, the Kinect-obtained CM should rather be considered to be one reference point, moving synchronously with the subject’s physical CM. Ideally, these two points are related via a simple spatial translation (involving an unknown, yet constant 3D vector) at all times; if this condition is fulfilled, the Kinect-obtained CM may be safely identified as the subject’s physical CM, because a constant spatial separation between these two points does not affect the evaluation of the important quantities characterising the motion. At all time frames, the coordinates of the CM are obtained from seven nodes, namely from the first three main-body nodes 1 to 3, from the shoulder nodes 5 and 9, as well as from the hip nodes 13 and 17. Being subject to considerable movement in walking and running motion, the node 4 (HEAD) is not included in the determination of the coordinates of the subject’s CM. Prior to further processing, the CM offsets \((x_{CM}, y_{CM}, z_{CM})\) are removed from the data; thus, the motion is defined relative to the subject’s CM at all times. (The angles, defined in Subsection 3.1, involve differences of corresponding coordinates; as a result, they are not affected by the removal of the CM offsets from the data.) The largeness of the subject’s ‘stray’ motion on the treadmill is assessed on the basis of the root-mean-square (rms) of the distributions of \(x_{CM}, y_{CM}, \) and \(z_{CM}\).

To investigate the stability of the motion over time, the data may be split into segments. In our data analysis, the duration of these segments may be chosen at will; herein, we have made use of 12 s segments in the analysis of
the Kinect-captured data. Within each of these segments, information which may be considered ‘instantaneous’ is obtained, thus enabling an examination of the ‘stability’ of the subject’s motion at the specific velocity setting (see Subsection 3.2.2). The symmetry of the motion for the left and right parts of the subject’s body may be investigated by comparing the corresponding waveforms (representing the variation of the relevant quantity within the gait cycle, see Subsection 3.2.3). Finally, the largeness of the motion of the extremities may be examined on the basis of the RoMs obtained from these waveforms. We subsequently address some of these issues in somewhat more detail.

### 3.2.1 Determination of the period of the gait cycle

Ideally, the period of the gait cycle $T$ is defined as the time lapse between successive time instants corresponding to identical postures of the human body (position and direction of motion of the subject’s body parts with respect to the CM). (Of course, the application of ‘identicalness’ in living organisms is illusional; no two postures can ever be expected to be identical in the formal sense.) We define the period of the gait cycle as the time lapse between successive most distal positions $z$ of the same lower leg, which is identified herein as the ankle; one could also use the midpoints of the ankle and foot nodes, yet our preliminary data analysis revealed frequent artefacts in the signals of the foot nodes. The arrays of time instants, at which the left or right lower leg is at its most distal position with respect to the instantaneous CM of the subject, are used in timing the waveforms corresponding to the left or right part of the subject’s body.

The period of the gait cycle is related to two other quantities which are used in the analysis of motion data.

- The stride length $L$ is the product of the velocity $v$ and the period of the gait cycle: $L = vT$.
- The cadence $C$ is defined as the number of steps per unit time; one commonly-used unit is the number of steps per min. It has been argued (e.g., by Daniels [16]) that optimally the minimal cadence in running motion should be 180 steps per min, implying a maximal period of the gait cycle of $2/3$ s.

### 3.2.2 Assessment of the stability of the motion

To examine the constancy of the period of the gait cycle throughout each session (according to our definition, each session involves one velocity of the treadmill belt), the values of the instantaneous period of the gait cycle are submitted to further analysis. The overall constancy is judged on the basis of a simple $\chi^2$ test, assessing the goodness of the representation of the input data by one overall average value; the resulting p-value is obtained from the
minimal value $\chi^2_{\text{min}}$ for the given number of degrees of freedom (DoF), i.e., for the number of data segments reduced by one unit.

### 3.2.3 Determination of the waveforms

Using the time-instant arrays from the analysis of the left and right lower-leg signals (as described in Subsection 3.2.1), each time series (pertaining to a specific node and spatial direction) was split into one-period segments, which were subsequently superimposed and averaged, to yield a representative movement for the node and spatial direction over the gait cycle. Finally, one average waveform for each node and spatial direction is obtained, representative of the motion at the particular velocity setting. The investigation of the asymmetry in the motion rests on the comparison of the waveforms obtained for corresponding left and right nodes, and spatial directions.

Average waveforms for all nodes and spatial directions, representing the variation of the motion of that node (in 3D) within the gait cycle, were extracted separately for the left and right nodes of the extremities; waveforms were also extracted for the important angles introduced in Subsection 3.1. As mentioned in Subsection 3.2.1, the time instant at which the left (right) lower leg of the subject was at its most distal position (with respect to the subject’s CM) marked the start of each gait cycle (as well as the end of the previous one), suitable for the study of the left (right) part of the subject’s body. In case that left/right (L/R) information is not available (as, for example, for the trunk angle), the right lower leg was used in the timing. All waveforms were subsequently 0-centred. The removal of the average offsets is necessary, given that the two measurement systems yield output which cannot be thought of as corresponding to the same locations of the subject’s anatomy. For instance, according to the ‘31 + 8’ placement scheme, the markers for the shoulder are placed on top of the acromioclavicular joints; the Kinect nodes SHOULDER_LEFT and SHOULDER_RIGHT match better the physical locations of the shoulder joints.

The left and right waveforms yielded two new waveforms, identified as the ‘L/R average’ (LRA) and the ‘right-minus-left difference’ (RLD); we also attempt the validation of the RLDs because we intend to use the Kinect output in order to detect asymmetrical features in the motion. The validation of the Kinect output rests on the comparison of the Kinect LRA and RLD waveforms with those obtained with the Vicon system.
3.3 Scoring options when comparing waveforms

The similarity of corresponding waveforms is judged on the basis of five scoring options: Pearson’s correlation coefficient, the Zilliacus error metric, the RMS error metric, Whang’s score, and Theil’s score. Assuming that a Kinect (0-centred) waveform is denoted by \( k_i \) and the corresponding Vicon (0-centred) waveform by \( v_i \), the five scoring options are defined in Eqs. (8)-(12) (for details on the original works, see Ref. [17]); all sums are taken from \( i = 1 \) to \( N \), where \( N \) stands for the number of bins used in the histograms yielding these waveforms. (In the present work, \( N = 50 \).)

Pearson’s correlation coefficient

\[
    r = \frac{\sum k_i v_i}{\sqrt{\sum k_i^2 \sum v_i^2}}
\]

Zilliacus error metric

\[
    d_z = \frac{\sum |k_i - v_i|}{\sum |v_i|}
\]

RMS error metric

\[
    d_{rms} = \frac{\sum (k_i - v_i)^2}{\sum v_i^2}
\]

Whang’s score

\[
    d_w = \frac{\sum |k_i - v_i|}{\sum |v_i| + \sum |k_i|}
\]

Theil’s score

\[
    d_t = \frac{\sum (k_i - v_i)^2}{\sum v_i^2 + \sum k_i^2}
\]

In case of identical (Kinect and Vicon) waveforms, \( r = 1 \); all other scores vanish (\( d_z = d_{rms} = d_w = d_t = 0 \)).

Evidently, Whang’s score is the symmeterised version of the Zilliacus error metric, whereas Theil’s score is the symmeterised version of the RMS error metric. Although the differences between the Zilliacus and the RMS error metric are generally small (as are those between Whang’s and Theil’s scores), we will retain all aforementioned scoring options. At fixed velocity of the treadmill belt, we investigate the variation of these scores for the nodes and spatial directions of the extremities. Tests will be made using the LRA waveforms, as well as those corresponding to the RLD waveforms. After studying the goodness of the association of the waveforms at fixed velocity, we will investigate velocity-dependent effects.

We are aware of the fact that other ways for testing the similarity of the outputs of different measurement systems have been put forth. For instance, some authors favour the use of the ‘coefficient of multiple correlation’ (CMC) [18,19,20,21]. Ferrari, Cutti, and Cappello [21] define the CMC as:

\[
    \text{CMC} = \left[ 1 - \frac{\sum_{i=1}^{P} \sum_{j=1}^{W} \sum_{k=1}^{N} (w_{ijk} - \bar{w}_{jk})^2/(WN(P-1))}{\sum_{i=1}^{P} \sum_{j=1}^{W} \sum_{k=1}^{N} (w_{ijk} - \bar{w}_{.k})^2/(WP(N-1))} \right]^{1/2}
\]

14
where the triple array \( w_{ijk} \) contains the entire data, i.e., \( PW \) waveforms of dimension \( N \) (\( N \) depends on the gait cycle in Ref. [21]); \( P \) is the number of measurement systems (in our case, \( P = 2 \), Vicon and Kinect) being used in the study (‘protocols’, in the language of Ref. [21]) and \( W \) denotes the number of waveforms obtained within each measurement system. The averages \( \bar{w}_{.jk} \) and \( \bar{w}_{.j.} \) in Eq. (13) are defined as:

\[
\bar{w}_{.jk} = \frac{1}{P} \sum_{i=1}^{P} w_{ijk},
\]

(14)

\[
\bar{w}_{.j.} = \frac{1}{N} \sum_{k=1}^{W} \bar{w}_{.jk}.
\]

(15)

Unlike Pearson’s correlation coefficient, ‘directional information’ for the association between the tested quantities is lost when using the CMC in the analysis. In its original definition [22], the CMC is bound between 0 and 1. However, the quantity CMC, obtained with Eq. (13), is frequently imaginary (the ratio of the triple sums may be larger than 1); this is due to the use of \( \bar{w}_{.j.} \), instead of the grand mean (along with the normalisation factor \( W(PN-1) \), instead of \( (WPN-1) \)), in the denominator of the expression. Importantly, it is not clear how the obtained CMC values relate to the goodness of the association between the tested waveforms. The association scheme of Ref. [23] is arbitrary; there is no theoretical justification for such an interpretation of the CMC results.

The basic problem in testing the similarity of the waveforms lies with the fact that the established tests in correlation theory enable the acceptance or the rejection of the hypothesis that the observed effects can be accounted for by an underlying correlation of ‘strength’ \( \rho_0 \), where \(-1 < \rho_0 < 1\). The test when \( \rho_0 = 0 \) involves the transformation:

\[
t = r \sqrt{\frac{N-2}{1-r^2}}.
\]

The variable \( t \) is expected to follow the \( t \)-distribution (Student’s distribution) with \( N-2 \) DoF. The tests when \( \rho_0 \neq 0 \) involve Fisher’s transformation; the details may be found in standard textbooks on Statistics. No tests are possible when \( \rho_0 = 1 \), i.e., when attempting to judge the goodness of the association of waveforms, if ideally the waveforms should be identical. The only tests which can be carried out in such a case are those involving \( \rho_0 = 0 \), i.e., investigating the presence of a statistically-significant correlation between the tested waveforms when the null hypothesis for no such effects is assumed to hold. In practice, the one-sided tests (we are not interested in an anticorrelation of the waveforms) for \( N-2 = 48 \) DoF result in the rejection of the null hypothesis at the significance level of 5% when \( r \gtrsim 0.2353 \) and at the significance level of 1% when \( r \gtrsim 0.3281 \).
Formal, well-defined (in the mathematical sense) ways to compare waveforms do exist. As a general remark, the application of such rigorous tests would result in significant discrepancies in many cases, even when a judgment based on a visual inspection of the tested quantities is favourable. a) One possibility would be to obtain the uncertainties in the histogram bins and make use of a $\chi^2$ function to assess the goodness of the association. The variability of the output across different sensors could also be assessed and this additional uncertainty could be taken into account in the tests. b) Another possibility would be to invoke analysis of variance (ANOVA), defining the reduced ‘within-treatments’ variation as

$$\bar{V}_w = \sum_{i=1}^{P} \sum_{j=1}^{W} \sum_{k=1}^{N} (w_{ijk} - \bar{w}_{i,k})^2 / (PN(W - 1))$$

and the reduced ‘between-treatments’ variation as

$$\bar{V}_b = \sum_{i=1}^{P} \sum_{j=1}^{W} \sum_{k=1}^{N} (\bar{w}_{i,k} - \bar{w}_{..k})^2 / (N(P - 1)) .$$

 Appearing in these expressions are two average waveforms: the average waveform obtained with measurement system $i$:

$$\bar{w}_{i,k} = \frac{1}{W} \sum_{j=1}^{W} w_{ijk}$$

and the grand-mean waveform:

$$\bar{w}_{..k} = \frac{1}{P} \sum_{i=1}^{P} \bar{w}_{i,k} .$$

The ratio $F = \bar{V}_b / \bar{V}_w$ is expected to follow Fisher’s distribution with $N(P - 1)$ and $PN(W - 1)$ DoF. The resulting p-value enables a decision regarding the acceptance or rejection of the null hypothesis, i.e., of the observed effects being due to statistical fluctuation. c) A third possibility would be to histogram the difference of corresponding waveforms obtained with the two measurement systems within the same gait cycle $j$; the decision on whether the final waveform is significantly different from 0 can be made on the basis of a number of tests, including $\chi^2$ tests for the constancy and shape of the result of the histogram. Nevertheless, to retain simplicity in the present paper, we have decided to make use in the data analysis of the simple scoring options introduced by Eqs.(8)-(12).

4 Data acquisition

The data acquisition involved one male adult (ZHAW employee), with no known motion problems, walking and running on a commercially-available
treadmill (Daum Electronic GmbH, Germany). The placement of the treadmill in the Movement Laboratory (Bewegungslabor) of the Institute of Physiotherapy (School of Health Professions, ZHAW), where the experimentation took place, was such that the motion of the subject be neither hindered nor influenced in any way by near-by objects. Prior to the data-acquisition sessions, the Kinect sensor was properly centred and aligned with the principal axes of the laboratory. The sensor was then left in the same position, untouched throughout the data acquisition. The dedicated calibration of the Vicon system (which is performed each time that the system is used) is expected to yield a coordinate system which is aligned with that of the laboratory.

The (original) Kinect sensor also provides information on the elevation (pitch) angle at which it is set. During our tests, we discovered that this information is not reliable, at least as far as the particular device used in our experimentation is concerned. To enable the accurate determination of the elevation angle of the Kinect sensor, we set forth a simple procedure. The subject stands (in the upright position, not moving) at a number of positions on the treadmill belt, and static measurements (e.g., 5 s of Kinect data) at these positions are obtained and averaged. The elevation angle of the Kinect sensor may be easily obtained from the slope of the average (over a number of Kinect nodes, e.g., of those pertaining to the hips, knees, and ankles) \((y,z)\) coordinates corresponding to these positions. The output data, obtained with the Kinect sensor, have been corrected (off-line) accordingly, to yield the appropriate spatial coordinates in the laboratory system. To prevent Kinect from re-adjusting the elevation angle during the data acquisition (which is a problematic feature), we attached the Kinect body unto a plastic structure mounted on a tripod.

Five velocity settings were used in the data acquisition: walking data were acquired at 5 km/h; running data were acquired at 8, 10, 11, and 12 km/h. At each velocity setting, the subject was given 1 min to adjust his movements comfortably to the velocity of the treadmill belt. The Kinect output spanned 2 min at each velocity setting. The variation of the distance between the subject and the Kinect sensor remained remarkably small throughout the data acquisition at all velocities, ranging between about 2.14 and 2.40 m, well within the limits for the use of the sensor set by the manufacturer. The Vicon system produces abundant output, hence (also given its expected superior accuracy) it was deemed sufficient to restrict the Vicon-related data acquisition to about 24 s for walking and 18 s at the remaining velocity settings \(^3\). The recording on the two measurement systems started almost simultaneously; in view of the fact that waveforms will be obtained from the output data, the ‘exact’ synchronisation of the two measurement systems is inessential.

\(^3\) We made use of 6 s segments in the analysis of the (transformed-into-Kinect format) Vicon data.
It must be mentioned that measurements had also been acquired at 9 km/h. However, the off-line analysis of these data revealed that two of the ‘useful’ markers, which were attached to the subject’s arms, had been detached during that data-acquisition session; as a result, it is not straightforward to carry out all the tests relating to these data. To avoid complexity (i.e., the use of different sample sizes in the tests), it was decided to exclude the 9 km/h data from the tests.

One additional remark is due. The Kinect sensor may lose track of the lower parts of the subject’s extremities (wrists, hands, ankles, and feet) for two reasons: either due to the particularity of the motion of the extremity in relation to the position of the sensor (e.g., the identification of the elbows, wrists, and hands becomes problematic in some postures, where the viewing angle of the ulnar bone by Kinect is small) or due to the fact that these parts of the human body are obstructed (behind the subject) for a fraction of the gait cycle. Assuming that these instances remain rare (e.g., below the 5% level of the available data), the missing values may be reliably obtained (interpolated) from the well-determined (tracked) data of the corresponding time series of measurements. In the case of the Kinect data acquired in the present study, the maximal number of the interpolated signals occurred in the 12 km/h data set. From a total of 3632 available time instants, the Kinect sensor lost track of the left foot of the subject on 84 occasions, representing about 2.3% of the corresponding time-series data. The left hand and the left ankle followed, with 66 and 45 occasions, respectively. Finally, the right ankle and the right hand were not tracked on 31 and 23 occasions, respectively. Although, when normalised to the total number of the available values, the untracked signals usually appear ‘harmless’ as they represent a small fraction of the total amount of measurements (e.g., the amount of these signals did not exceed the 0.37% level in the 12 km/h data set), particular attention must be paid in order to ensure that no node be significantly affected, as the interpolation will yield unreliable results in such a case.

The largeness of the ‘stray’ motion of the subject was assessed on the basis of the rms of the $x_{CM}$, $y_{CM}$, and $z_{CM}$ distributions. The three rms average values (over all velocity settings) from Kinect were: 21.1, 28.1, and 28.4 mm; the corresponding values from Vicon were: 19.0, 34.0, and 30.0 mm. Averages of the period of the gait cycle $T$, of the cadence $C$, and of the stride length $L$ are given in Table 2, separately for the two measurement systems at each velocity setting; the agreement between corresponding values is reasonable, thus enabling the use of Kinect in the timing of the motion (determination of the gait cycles). It must be borne in mind that these results have been obtained from data spanning different temporal intervals; the Vicon data cover about one-fifth (in fact, the first one-fifth) of the total time span of the Kinect data. The comparison between the Kinect and Vicon corresponding quantities thus rests on the assumption that the motion of the subject was not modified in
a systematic way during the entire length of the Kinect data acquisition, an assumption which can be partially tested after examining the constancy of the period of the gait cycle (in the Kinect-captured data). Representative waveforms, corresponding to the \( z \) coordinate of the lower-extremity nodes, are shown in Fig. 3.

5 Comparison of the Kinect and the Vicon results

The comparison of the waveforms obtained for the nodes of the extremities with the two measurement systems, as well as of those obtained for the important angles defined in Subsection 3.1, enables the assessment of the accuracy of the output of the Kinect sensor; it is assumed that the inaccuracy of the Vicon output is negligible compared to that of Kinect. (Of course, to associate the marker positions with the internal motion, i.e., with the motion of the subject’s skeletal structure, is quite another issue; we are not aware of works addressing this subject in detail.)

We first compare the LRA waveforms; the comparison of the RLD waveforms is addressed later on, at the end of the present section.

We first investigated the goodness of the association (of the waveforms obtained with the two measurement systems) for the eight node levels of the extremities (SHOULDER, ELBOW, WRIST, HAND, HIP, KNEE, ANKLE, and FOOT) and spatial directions. To this end, separately for each of the five scoring options of Section 3.3, velocity setting, and spatial direction, each node level was ranked according to the goodness of the association of the Kinect and the Vicon waveforms. The node level with the worst association was given the mark of 0, whereas the one with the best association was assigned the mark of 7. The sum of the ranking scores over all velocity settings and scoring options yielded an \( 8 \times 3 \) ‘matrix of goodness of the association’ (8 node levels for the extremities, 3 spatial directions); the minimal entry in this matrix may be 0 (which, in fact, was the overall score of the hips in the \( z \) direction!), whereas the maximal value is \( 7 \times 5 \times 5 = 175 \) (the maximal value was obtained for the hands in the \( x \) direction, a total score of 163); as these sums of scores follow the uniform distribution, total scores below \( 175/4 = 43.75 \) belong to the first quartile of the cumulative distribution, identifying the one quarter of the node levels for the extremities and spatial directions with the worst association between the Kinect and the Vicon waveforms. This test established beyond doubt that the shoulder and hip waveforms, in all three spatial directions, yielded the worst association results. From the remaining combinations of node levels for the extremities and spatial directions, the \( y \) coordinate of the knees was marginally on the good side of the first quartile of the cumulative distribution. The average value of Pearson’s correlation coefficient for the
shoulders and for the hips was found equal to about 0.187 and the rms of the distribution was equal to about 0.395; seven of the 30 scores were negative, indicating an anticorrelation between the Kinect and Vicon waveforms.

We subsequently pursued the investigation of systematic differences in the performance of the Kinect sensor regarding: a) the upper and lower parts of the subject’s body (i.e., upper- versus lower-extremity nodes) and b) the three spatial directions.

- To assess the similarity of the Kinect and the Vicon waveforms, obtained for the nodes of the extremities, one-factor ANOVA tests were performed, separately for each of the five scoring options of Subsection 3.3, on the scores obtained at each velocity setting, for all upper-extremity nodes and spatial directions, and all lower-extremity nodes and spatial directions. Assuming a significance level of 1% (the value which most statisticians associate with the outset of significance), none of these tests resulted in significant effects, at any of the velocity settings; the minimal p-value from these tests exceeded about 0.313. The exclusion of the scores for the shoulders and for the hips from these tests does not affect the results significantly; for instance, the ANOVA tests resulted in p-values exceeding about 0.269. Consequently, we conclude that no significant differences could be seen in the motion of the upper and lower extremities. We thus cannot confirm that the quality of the Kinect tracking of the lower part of the body (in relation to the upper part) deteriorates.

- We next addressed the goodness of the association of the waveforms pertaining to the three spatial directions $x$, $y$, and $z$. The corresponding ANOVA tests did not reveal significant effects at any velocity setting; the minimal p-value exceeded about 0.073. However, after the exclusion of the scores for the shoulders and for the hips from these tests, 21 (out of 25) tests yielded p-values below the significance level of 1% (median $1.75 \cdot 10^{-3}$); two additional tests indicated probable significance ($0.01 < p < 0.05$); the largest p-value was equal to 0.076. It appears that (on average) the best-matching waveforms correspond to the depth $z$, followed by the lateral direction $x$; the association of the waveforms in the $y$ direction is inferior. After the exclusion of shoulders and hips, the average values of Pearson’s correlation coefficient in the $x$, $y$, and $z$ directions were found to be: 0.845, 0.750, and 0.981, respectively. The average values of Pearson’s correlation coefficient (over all nodes and spatial directions at fixed velocity) remain reasonably close to the overall average (0.859); pronounced velocity-dependent effects have not been observed. Similar conclusions were drawn from the analysis of the results using the other four scoring options.

We will now compare the $y$ waveforms of the lower legs. This comparison is interesting for two reasons. First, the lower-leg signals are used in timing the motion; second, we intended to use these signals in order to obtain the times
(expressed as fractions of the period of the gait cycle) of the initial contact (IC) and the toe-off (TO) [9,10]; the difference of these two values is the stance fraction. Shown in Figs. 4 and 5 are the waveforms obtained from the Kinect and the Vicon outputs, separately at each velocity setting. The waveforms represent the variation of the raw signals, i.e., the y offsets of the subject’s CM have not been removed from the signals. The salient feature in the waveforms obtained with the Kinect sensor is a pronounced peak appearing around the IC; the comparison with the Vicon waveforms indicates that this peak is a Kinect-related artefact. Although it cannot influence the timing of the motion (because of its position), this artefact complicates the determination of the stance fractions, if not disabling it.

We were set on investigating the goodness of the association of the RLD waveforms, given that we intended to determine the asymmetry in the motion from the Kinect output. To this end, two-sided t-tests were performed on the score distributions between corresponding LRA and RLD waveforms, a total of 75 tests (five scoring options, three tests per scoring option, five velocity settings). Three tests were made per case: paired, homoscedastic, and unequal-variance. The p-values, obtained from the majority of these tests, were found to be small, below the significance level of 1%; the median p-value was equal to $8.12 \cdot 10^{-4}$, whereas the minimal one was equal to $3.97 \cdot 10^{-5}$. (After the exclusion of shoulders and hips, the median p-value dropped to $2.24 \cdot 10^{-4}$, whereas the minimal one to $1.39 \cdot 10^{-6}$.) The analysis showed that Pearson’s correlation coefficients of the tests on the RLD waveforms were systematically below those of the LRA waveforms, whereas all other scores were larger for the RLD waveforms, thus indicating poorer association of the Kinect and the Vicon waveforms for the RLDs compared to the LRAs. As a result, the RLD waveforms, obtained with Kinect, are not as reliable as the corresponding LRA waveforms.

We also investigated the similarity between the Kinect and the Vicon waveforms for the important angles (see Subsection 3.1). The only waveforms which match well are those for the hip and knee angles in the sagittal plane, and (to a lesser degree) those for the hip angles in the coronal plane. In the sagittal plane, the average values of Pearson’s correlation coefficients for the LRAs were equal to 0.978 and 0.953 for the hip and knee angles, respectively. In the coronal plane, the average value of Pearson’s correlation coefficient for the LRAs was equal to 0.835 for the hip angles. Pronounced velocity-dependence effects were not observed. The corresponding RLD waveforms do not much well; in the sagittal plane, the average values of Pearson’s correlation coefficients were equal to 0.379 and 0.558; the agreement for the hip angles in the coronal plane was found better (Pearson’s correlation coefficient came out around 0.800). The average value of Pearson’s correlation coefficients for the LRAs in the case of the knee angle in 3D was 0.950; for the corresponding RLDs, 0.522.
Large values of Pearson’s correlation coefficient must not be taken as indicating ‘almost identical’ waveforms\(^4\); they simply suggest that the tested quantities are linearly related, thus ensuring that results of similar quality may be obtained from the analysis of the data acquired within each of the tested measurement systems. To demonstrate this, we show in Figs. 6 and 7 the Kinect and the Vicon waveforms for the left- and right-knee angles in 3D, at all velocity settings\(^5\). The dedicated analysis of these data shows that, though the overall averages of Pearson’s correlation coefficient for the left- and right-knee angles come out to be large (equal to 0.941 and 0.952, respectively), Kinect systematically underestimates the knee angle in 3D by about 25%.

We now advance our argument regarding the systematic underestimation of the knee angle by the Kinect sensor. As long as the knee is extended, the Kinect sensor associates the knee node with the physical joint reasonably well. As the knee is flexed, the Kinect sensor is prone to moving the knee node upward, along the femur; the dedicated data analysis showed that this displacement may be as large as about 10 cm. A similar effect has been observed for the ankles, which also ‘move upward’, along the tibia, as the knee is flexed. The combined effect of these two displacements results in the systematic underestimation of the knee angle by Kinect. Additionally, the spurious peak, observed in the Kinect data in Figs. 4 and 5, originates in the ankle shift along the tibia as the knee is flexed from the extended position.

We finally address the comparison of the RoMs obtained from the Kinect and the Vicon waveforms. It might be argued that one could simply use in a study the RoMs, rather than the waveforms, as representative of the motion of each node. Of course, given that each waveform is essentially replaced by one number, the information content in the RoMs is drastically reduced compared to that of the waveforms. Shown in Fig. 8 is a scatter plot of the Kinect and Vicon RoMs of the LRA waveforms. The ideal relation between these two quantities should be a straight line with slope equal to 1, yet a sizeable scattering of the values is observed; Pearson’s correlation coefficient between the two RoM arrays came out equal to 0.982. Shown in Fig. 9 is a scatter plot of the Kinect and the Vicon RoMs of the RLD waveforms. The ideal relation between these two quantities should also be a straight line with slope equal to 1; however, the scattering of the values is substantial and Pearson’s correlation coefficient dropped to 0.616. The exclusion of the shoulder and hip nodes has a minor impact on these results. Again, we draw the conclusion that the RLD

\(^4\) On the other hand, small values of Pearson’s correlation coefficient cannot be misinterpreted; they are indicative of a mismatch.

\(^5\) Included in Figs. 4-7 are also the results obtained at 9 km/h, in spite of the exclusion of these data in the main tests. Given that the two detached markers had been placed on the upper part of the subject’s body, their detachment has no impact on the evaluation of the quantities shown in Figs. 4-7, which pertain to the lower part of the subject’s body.
waveforms, obtained with Kinect, are less reliable than the corresponding LRA waveforms.

It is worth mentioning that, as we are interested in capturing the motion of the lower legs (i.e., of the ankle and foot nodes) of the subject, the Kinect sensor was placed at about 82 cm above the ground or about 62 cm above the treadmill floor; placing the sensor higher leads to significantly more lost signals of the lower leg (the ankle and foot nodes are not tracked), as the lower leg is not visible by the sensor during a sizeable fraction of the period of the gait cycle, shortly after the TO. As, according to Microsoft, the Kinect sensor performs better if placed between 2 ft (60.96 cm) and 6 ft (182.88 cm) off the floor, the use of the sensor in the present study is in accordance with the specifications.

6 Discussion and conclusions

The present study investigated the possibility of using the original Microsoft Kinect\textsuperscript{TM} (hereafter, simply ‘Kinect’)\textsuperscript{[1]} sensor in the analysis of human motion. The output of the sensor has been compared to that of a marker-based system (MBS), which was assumed to be accurate. All data were analysed using the same software application; this was enabled after the MBS output was transformed into Kinect format using reasonable associations between the Kinect nodes and the MBS markers; owing to the removal of the constant offsets in the data analysis (see Subsection 3.2.3), the exact matching between the Kinect nodes and the locations of the MBS markers is inessential.

The output of the data analysis comprises waveforms, representing the variation within the gait cycle of (some of the) quantities which may be used in order to characterise motion, i.e., of the coordinates of extremities in the three spatial directions and of the important angles in the sagittal and coronal planes (see Subsection 3.1). We attempted the validation of the waveforms obtained from the analysis of the data acquired with the Kinect sensor on the basis of those obtained using the Vicon system as the reference solution. The similarity of corresponding waveforms was judged on the basis of five scoring options (see Subsection 3.3).

The experimentation involved one male subject, walking and running on a commercially-available treadmill. Five velocity settings were used: walking data were acquired at 5 km/h; running data were acquired at 8, 10, 11, and 12 km/h. In our processing of the data, the motion of the subject was split in two parts: one relating to the subject’s centre-of-mass (CM), the other to the motion of the subject’s body parts with respect to that CM. The CM of the present paper should rather be considered to be one reference point moving
synchronously with the physical CM. The coordinates of the Kinect-obtained CM were extracted from seven main-body nodes.

Regarding the validation of the Kinect output, the main conclusions of the present work are as follows.

- Regarding the performance of the Kinect sensor in the motion of the upper and lower extremities, no significant effects could be detected in the output.
- The Kinect and the Vicon waveforms for the shoulders and the hips do not match well.
- After excluding the shoulders and the hips, significant effects in the performance of the Kinect sensor in the three spatial directions were observed, equally affecting all velocities; the best-matching waveforms are those involving the depth $z$, whereas the worst association involves the vertical direction $y$.
- The results on the similarity of the Kinect and the Vicon waveforms for the hip and knee angles (sagittal plane) were found satisfactory as far as the existence of a linear relation between the corresponding quantities is concerned. On the other hand, it is clear that Kinect systematically underestimates the knee angle by as much as 25%, see Figs. 6 and 7.
- Spurious effects were seen in the Kinect raw $y$ coordinate of the lower leg (ankle), see Figs. 4 and 5.
- We advanced an argument aiming at the explanation of the systematic underestimation of the knee angle by Kinect, as well as of the appearance of the spurious peak in the Kinect raw $y$ coordinate of the lower leg; both effects originate in Kinect’s tendency to displace the knee and the ankle nodes ‘upwards’ (along the femur and the tibia, respectively) as the knee is flexed from its extended position.
- The Kinect-extracted waveforms, used to investigate the asymmetry in the motion of corresponding left and part nodes, are less reliable than the corresponding average waveforms (for the same left and right nodes). Therefore, it seems difficult to use Kinect in order to investigate the asymmetry in motion data.

The analysis of the validation data suggests that the original Kinect sensor cannot be easily employed in motion analysis. The systematic underestimation of the knee angle, coupled with the artefacts observed in the lower leg, are serious. Last but not least, the Kinect sensor does not provide reliable information regarding the asymmetry in the motion of the subject. The determination and application of corrections, needed in order to suppress these artefacts, comprises an interesting research subject. It would also be interesting to investigate whether the new Kinect sensor (‘Kinect for Windows v2’) resolves (some of) the problems associated with the original version of the sensor.
Finally, one might pose the question of whether results obtained from one subject are sufficient in enabling reliable conclusions. The answer is simple. Had we obtained *consistent* results from the two measurement systems, it would have made sense to ‘pursue statistics’; on the contrary, serious differences were found. Even in the unlikely case that, whichever the reason, Kinect failed for the specific subject we used in our experimentation (and that it does not fail for the general user), the involvement of the sensor in motion analysis would not be possible as each such use should have to be separately validated; therefore, the use of an MBS in motion analysis, the replacement of which comprised the subject of this study, appears to be indispensable at the present time.

**Note added in proof**

After the submission of the present work for reviewing, the authors discovered that a paper on the same subject had recently appeared [24]. Using similar methodology, the authors of that study came to the same main conclusion regarding the use of the Microsoft Kinect™ sensor in the analysis of human motion.

**Conflict of interest statement**

The authors certify that, regarding the material of the present paper, they have no affiliations with or involvement in any organisation or entity with financial or non-financial interest.

**Acknowledgements**

We are indebted to the Institute of Physiotherapy (School of Health Professions, ZHAW) for allowing us to use the Movement Laboratory (Bewegungslabor) and for providing technical assistance in the experimentation. We also acknowledge discussions with C. Bauer, E. Graf, B. Heinlein, I. Kokemor, R. Kuster, and M. Regniet. The present work was funded by the School of Engineering of Zurich University of Applied Sciences (Zürcher Hochschule für angewandte Wissenschaften, ZHAW), Winterthur, Switzerland, project No. 9710.5.14.5.0177 (internal-funding grant).

Fig. 2 has been produced with CaRMetal, a dynamic geometry free software (GNU-GPL license), originally developed by R. Grothmann and recently under E. Hakenholz [25].
References

[1] http://www.xbox.com/en-GB/Kinect/, http://www.microsoft.com/en-us/Kinectforwindows/

[2] R.A. Clark et al., Validity of the Microsoft Kinect for assessment of postural control, Gait Posture 36 (2012) 372-377.

[3] A. Schmitz, Mao Ye, R. Shapiro, Ruigang Yang, B. Noehren, Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system, J. Biomech. 47 (2014) 587-591.

[4] B. Galna et al., Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease, Gait Posture 39 (2014) 1062-1068.

[5] B. Bonnechère et al., Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry, Gait Posture 39 (2014) 593-598.

[6] G.A. Cavagna, M. Kaneko, Mechanical work and efficiency in level walking and running, J. Physiol. 268 (1977) 467-481.

[7] P.R. Cavanagh, M.A. Lafortune, Ground reaction forces in distance running, J. Biomech. 13 (1980) 397-406.

[8] M.A. Cairns, R.G. Burdett, J.C. Pisciotta, S.R. Simon, A biomechanical analysis of racewalking gait, Med. Sci. Sports Exercise 18 (1986) 446-453.

[9] S. Öunpuu, The Biomechanics of walking and running, Clin. Sport Med. 13 (1994) 843-863.

[10] T.F. Novacheck, The biomechanics of running, Gait Posture 7 (1998) 77-95.

[11] A.G. Schache, K.L. Bennell, P.D. Blanch, T.V. Wrigley, The coordinated movement of the lumbo-pelvic-hip complex during running: a literature review, Gait Posture 10 (1999) 30-47.

[12] J. Shotton et al., Real-time human pose recognition in parts from single depth images, Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, Issue Date: 20-25 June 2011.

[13] http://www.vicon.com/System/TSeries

[14] http://www.idmil.org/mocap/Plug-in-Gait+Marker+Placement.pdf, http://wweb.uta.edu/faculty/ricard/Classes/KINE-5350/PIGManualver1.pdf

[15] R.B. Davis, S. Öunpuu, D. Tyburski, J.R. Gage, A gait analysis data collection and reduction technique, Hum. Movement Sci. 10 (1991) 575-587.

[16] J. Daniels, Daniels' running formula, Third Edition, Human Kinetics (2013), pp. 26-28.
[17] M. Mongiardini, M.H. Ray, M. Anghileri, Development of a software for the comparison of curves during the verification and validation of numerical models, 7th European LS-DYNA Conference, Salzburg, Austria, May 14-15, 2009.

[18] M.P. Kadaba et al., Repeatability of kinematic, kinetic, and electromyographic data in normal adult gait, J. Orthop. Res. 7 (1989) 849-860.

[19] J.L. McGinley, R. Baker, R. Wolfe, M.E. Morris, The reliability of three-dimensional kinematic gait measurements: a systematic review, Gait Posture 29 (2009) 360-369.

[20] P. Garofalo et al., Inter-operator reliability and prediction bands of a novel protocol to measure the coordinated movements of shoulder-girdle and humerus in clinical settings, Med. Biol. Eng. Comput. 47 (2009) 475-486.

[21] A. Ferrari, A.G. Cutti, A. Cappello, A new formulation of the coefficient of multiple correlation to assess the similarity of waveforms measured synchronously by different motion analysis protocols, Gait Posture 31 (2010) 540-542.

[22] H. Theil, Economic Forecasts and Policy, Second Edition, North-Holland, Amsterdam (1961).

[23] A. Ferrari et al., First in vivo assessment of “Outwalk”: a novel protocol for clinical gait analysis based on inertial and magnetic sensors, Med. Biol. Eng. Comput. 48 (2010) 1-15.

[24] 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.

[25] http://db-maths.nuxit.net/CaRMetal/index_en.html
Table 1

The notation for the marker positions according to the ‘Plug-in Gait’ placement scheme [14].

| Marker number | Marker-position identifier | Placement                                      |
|---------------|----------------------------|------------------------------------------------|
| 1             | LFHD                       | left front head                                |
| 2             | RFHD                       | right front head                               |
| 3             | LBHD                       | left back head                                 |
| 4             | RBHD                       | right back head                                |
| 5             | C7                         | 7\textsuperscript{th} cervical vertebrae       |
| 6             | T10                        | 10\textsuperscript{th} thoracic vertebrae      |
| 7             | CLAV                       | clavicle                                       |
| 8             | STRN                       | sternum                                        |
| 9             | RBAK                       | right back (middle of the right scapula)       |
| 10            | LSHO                       | left shoulder                                  |
| 11            | LUPA                       | left upper arm                                 |
| 12            | LELB                       | left elbow                                     |
| 13            | LFRA                       | left forearm                                   |
| 14            | LWRA                       | left wrist A                                   |
| 15            | LWRB                       | left wrist B                                   |
| 16            | LFIN                       | left fingers (second metacarpal head, dorsum)  |
| 17            | RSHO                       | right shoulder                                 |
| 18            | RUPA                       | right upper arm                                |
| 19            | RELB                       | right elbow                                    |
| 20            | RFRA                       | right forearm                                  |
| 21            | RWRA                       | right wrist A                                  |
| 22            | RWRB                       | right wrist B                                  |
| 23            | RFIN                       | right fingers (second metacarpal head, dorsum) |
| 24            | LASI                       | left anterior superior iliac spine            |
| 25            | RASI                       | right anterior superior iliac spine           |
| 26            | LPSI                       | left posterior superior iliac spine           |
| 27            | RPSI                       | right posterior superior iliac spine          |
| 28            | LTHI                       | left thigh                                     |
Table 1 continued

| Marker number | Marker-position identifier | Placement                              |
|---------------|----------------------------|----------------------------------------|
| 29            | LKNE                       | left knee                              |
| 30            | LTIB                       | left tibia                             |
| 31            | LANK                       | left ankle                             |
| 32            | LHEE                       | left heel, on the calcaneus            |
| 33            | LTOE                       | left toes, second metatarsal head      |
| 34            | RTHI                       | right thigh                            |
| 35            | RKNE                       | right knee                             |
| 36            | RTIB                       | right tibia                            |
| 37            | RANK                       | right ankle                            |
| 38            | RHEE                       | right heel, on the calcaneus           |
| 39            | RTOE                       | right toes, second metatarsal head     |

Table 2

The average values of the period of the gait cycle $T$, of the cadence $C$, and of the stride length $L$ at the five velocity settings used in the data acquisition (see Section 4), separately for the Kinect and the Vicon systems.

|          | 5 km/h      | 8 km/h      | 10 km/h     | 11 km/h     | 12 km/h     |
|----------|-------------|-------------|-------------|-------------|-------------|
| **Kinect** |             |             |             |             |             |
| $T$ (s)  | 1.08702(88) | 0.8658(23)  | 0.8370(14)  | 0.8047(15)  | 0.7981(15)  |
| $C$ (steps/min) | 110.394(89) | 138.60(36)  | 143.36(25)  | 149.13(28)  | 150.35(28)  |
| $L$ (m)  | 1.5097(12)  | 1.9239(51)  | 2.3251(40)  | 2.4587(46)  | 2.6604(49)  |

|          | 5 km/h      | 8 km/h      | 10 km/h     | 11 km/h     | 12 km/h     |
|----------|-------------|-------------|-------------|-------------|-------------|
| **Vicon** |             |             |             |             |             |
| $T$ (s)  | 1.0916(11)  | 0.85262(96) | 0.8391(35)  | 0.80548(89) | 0.8028(29)  |
| $C$ (steps/min) | 109.93(11)  | 140.74(16)  | 143.01(59)  | 148.98(16)  | 149.47(53)  |
| $L$ (m)  | 1.5162(15)  | 1.8947(21)  | 2.331(10)   | 2.4612(27)  | 2.676(10)   |
Fig. 2. The 20 nodes of the original Microsoft Kinect™ sensor.
Fig. 3. Representative motion of the lower extremities (relative to the subject’s centre of mass, as defined from the Kinect data, see Subsection 3.2) in the \( z \) direction (depth, positive away from Kinect, see Fig. 1). The waveforms correspond to the Kinect data at 12 km/h (see Section 4). The quantity \( f \) is the fraction of the period of the gait cycle.
Fig. 4. The waveforms for the raw $y$ coordinate of the left lower leg (ankle), separately at each velocity setting, obtained from the Kinect and the Vicon output (see Section 4). The quantity $f$ is the fraction of the period of the gait cycle.
Fig. 5. The waveforms for the raw $y$ coordinate of the right lower leg (ankle), separately at each velocity setting, obtained from the Kinect and the Vicon output (see Section 4). The quantity $f$ is the fraction of the period of the gait cycle.
Fig. 6. The waveforms for the left-knee angle in 3D, separately at each velocity setting, obtained from the Kinect and the Vicon output (see Section 4). The quantity $f$ is the fraction of the period of the gait cycle.
Fig. 7. The waveforms for the right-knee angle in 3D, separately at each velocity setting, obtained from the Kinect and the Vicon output (see Section 4). The quantity $f$ is the fraction of the period of the gait cycle.
Fig. 8. The Kinect ranges of motion (RoMs) plotted versus the Vicon RoMs, obtained from the left/right average (LRA) waveforms of the extremities at all velocity settings used in the data acquisition (see Section 4).
Fig. 9. The Kinect ranges of motion (RoMs) plotted versus the Vicon RoMs obtained from the ‘right-minus-left’ difference (RLD) waveforms of the extremities at all velocity settings used in the data acquisition (see Section 4).