Biomedical Signal Processing and Control 38 (2017) 265–280

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Technical note

Performance analysis of a generalized motion capture system using microsoft kinect 2.0

Alessandro Napoli[a,*], Stephen Glass[b], Christian Ward[a], Carole Tucker[b], Iyad Obeid[a]

[a] *Department of Electrical & Computer Engineering, Temple University, Philadelphia, PA 19122, United States*
[b] *Department of Physical Therapy, Temple University, Philadelphia, PA 19140, United States*

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 4 May 2017
Accepted 4 June 2017

Keywords:
Motion capture
Kinect
Kinematics
Qualisys
Kinect-based
System reliability
Movement screening

ABSTRACT

This work presents a fine-grained analysis of the performance and limitations of the Microsoft Kinect sensor for tracking human movement in the context of biomechanical research and clinical applications. Earlier work in this field has focused on scalar summary measures or ad-hoc metrics with respect to specific movements that do not generalize well across clinical applications. In this work, the performance of the Microsoft Kinect is compared to motion tracking from a concurrently sampled professional grade Qualisys motion capture system. Subjects performed a range of clinically relevant tasks such as Sit-to-Stand and Timed Up-and-Go. Captured data included both three-dimensional joint center displacements and joint angles as recorded from both systems. Kinect performance was measured using cross correlation coefficients (CCR), root mean squared error (RMSE) relative to the Qualisys gold-standard and a new summary metric (SM) that combines both. Our results show that the Kinect-based system provides adequate performance when tracking joint center displacements in time, with overall CCR = 0.78, RMSE = 3.35 cm and SM = 1.21. On the contrary, lower accuracy was measured when tracking joint angles, with CCR = 0.58, RMSE = 24.59°, and SM = 3.76. Although performance differences for various movements and motion planes have been found, the results suggest that the Kinect is a viable tool for general biomechanical research, with specific limits on what levels of performance can be expected under various conditions.

1. Background

Capturing three-dimensional movement (or *kinematics*) is a central laboratory technique in the study of human movement. Kinematic studies have played an instrumental role in the study of joint pathology, mild traumatic brain injury, ergonomics, and athletic performance. Despite the importance of this technique, standard data acquisition methods are subject to considerable limitations. Stereophotogrammetry and electromagnetic motion tracking, for instance, require expensive, stationary equipment and time-consuming procedures for system calibration and post-processing of data. In contrast, low-cost depth sensing cameras (also known as time-of-flight or RGB-D cameras) are available off-the-shelf and may represent viable alternatives to more complex and expensive 3D camera setups. These cameras are capable of capturing RGB color images augmented with depth data at each pixel, thus providing 3D images. Such images can be used to track human motion in real-time. Among these cameras, the Microsoft Kinect™ 2.0 provides a low-cost, portable, user-friendly alternative which holds the potential to substantially increase the accessibility of kinematic data. The main advantage of using the Kinect sensor over the commercially available alternatives lies in the proprietary Microsoft algorithm that performs body and joint detection in real-time and that can be exploited using the Microsoft Software Development Kit (SDK) [1] available to .NET developers.

The performance of the Kinect™ 2.0 as a tool to evaluate kinematic variables, as compared to current standard methods, is a subject of great interest. Although several studies have been published in this area, the general trend has been to compare motion capture (MOCAP) systems based on scalar summary measures a selected metric. Examples from previous work include excursion range [2–5], mean or peak displacement [4,6–11] or timing of discrete signal events [4,5,9,11–15]. While such metrics are commonly studied in biomechanics, they do not adequately quantify the temporal structure of the signals under comparison, and are thus limited in terms of the generalizability of their results. To date, three laboratories have presented a more thorough treatment of Kinect 2.0 time series data in comparison to an existing standard. These studies still present certain limitations which restrict broader generalizability [16–19].

---

* Corresponding author.
E-mail address: a.napoli@temple.edu (A. Napoli).

http://dx.doi.org/10.1016/j.bspc.2017.06.006

The first study [19] was based on a composite signal, namely, total body center of mass, which represents a weighted sum of body segments derived from the Kinect joint center time dynamics. As a weighted sum, this output suppresses the variability inherent to the underlying time series data. Another study [17] reports measures of signal agreement measured as Intraclass Correlation Coefficients (ICC) between Kinect 2.0 sensors and an OptoTrak system (Northern Digital Inc., Waterloo, Canada). This investigation analyzed the consistency between the systems in each dimension for nearly all the joint center data natively exported from the Kinect. While this approach offers an appropriate comparison of the two systems, the results are specific to gait, a primarily sagittal plane movement pattern, and therefore represent a relatively narrow range of the human motion repertoire. Finally, a pair of studies [16,18], by a third group, quantifies Kinect signal error by its 3D L^2 norm distances from a ground truth signal as given by a 3D professional-grade MOCAP system. While they were able to classify individual data points as outliers vs. inliers based on error magnitude, information regarding the dimension and direction of signal offset is lost when using the 3D distance. As a result, this approach is not suitable for identifying systematic, direction-specific errors such as those noted by other investigators [6]. Additionally, their approach, which collapsed analyses across six tested movements, may obscure any relationships between the Kinect and ground truth signals that are specific to a given experimental condition or movement.

Considering these limitations and the expanded use of Kinect based systems in quantitative kinematic studies, a more thorough evaluation of Kinect 2.0 raw data performance as a MOCAP tool and its validation against a gold-standard 3D system is needed. The overarching goal of this research is the development of a system to collect reliable, valid kinematic data using low-cost sensors. The applications of such a technology are wide-reaching and may involve physical medicine clinics, athletics settings, and home entertainment, as well as other research domains in which kinematic data are not commonly acquired owing to prohibitive costs. The specific aim of this study was to identify limitations (and, ultimately, corrective measures) in Kinect 2.0 performance as an off-the-shelf technology for a flexible and multi-purpose MOCAP system. To that end, we have validated the Kinect against a professional three-dimensional motion capture system with 12 IR-cameras (Qualisys AB, Gothenburg, Sweden) over a range of dynamic movements and clinical tests that can be used as broad indicators of functional movement. In addition to this Kinect-vs-gold standard comparison, we present raw data from a second Kinect 2.0 sensor positioned alongside the first. These data provide an indication of reliability between Kinect 2.0 sensors. We acquired data from four healthy subjects and calculated results for point kinematics and joint angles, the latter of which are derived independently for the Kinect data using both quaternions and trigonometry applied to the joint positions.

2. Body segment orientation

The Microsoft Kinect 2.0 senses depth using an infrared camera sensor. A proprietary on-board algorithm locates bodies within the depth image and extracts parameters that describe the positions of up to six bodies in three-space in real time. For each tracked body, Kinect produces two data streams. The first is "joint location", which tracks the three-dimensional coordinates of 25 joints. The Kinect estimates three dimensional coordinates on a frame-by-frame basis using a probabilistic model that compares data from the depth image to a comprehensive database of human poses [20,21]. These measurements are in meters and are measured relative to an origin that is represented by the sensor camera itself. The second stream is "body segment orientation," in which the orientation and rotation of each segment relative to its parent, (e.g. forearm relative to upper arm) can be represented numerically by a *quaternion*. These real-time data streams are both complicated by the fact that the sampling rate varies between 5 and 30 frames per second according to instantaneous demands on the computer's processor.

A quaternion is a 4-tuple that represents the orientation and rotation of an object in three dimensions relative to some parent coordinate axis. Specifically, quaternion u can be expressed as

$$u = u_0 + u_x i + u_y j + u_z k = M\cos(\alpha) + M\boldsymbol{u}\sin(\alpha) = Me^{\boldsymbol{u}\alpha}$$

If v is some other quaternion, then v can be rotated around unit quaternion u (eg. $M=1$) by 2α radians using the following transform: $v_{rot} = uvu^*$. Although the Kinect produces a stream of "body segment orientations", these measurements must be numerically manipulated to yield clinically relevant kinematic data.

In some cases, this calculation is straightforward. For example, elbow angle can be calculated by simply calculating the angle between the quaternions of the upper arm and forearm as

$$\theta = \cos^{-1}\left(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}\right)$$

In other cases, the transformation from quaternion orientations to clinical kinematic data requires projecting body segments into the three cardinal planes (mediolateral, vertical, and anteroposterior).

Kinect quaternion mathematics are complicated by two main factors. The first is that all Kinect quaternions are defined with respect to their "parent segment" quaternion, and the second is that quaternions do not describe anatomically significant angles. Specifically, each Kinect quaternion is defined so that its y-axis points to its "child segment" quaternion, while the z-axis is normal to both the y-axis and the body segment. The x-axis is normal to both the previous axes. Since the root joint is the lower spine, all relative orientations can be re-referenced to this initial orientation by consecutive parent/child multiplication along the quaternion body chain, using Hamilton products, as follows:

$$q3_0 = q1_0 * q2_0 - q1_z * q2_z - q1_y * q2_y - q1_x * q2_x$$

$$q3_x = q1_0 * q2_x + q1_z * q2_y - q1_y * q2_z + q1_x * q2_0$$

$$q3_y = q1_0 * q2_y - q1_z * q2_x + q1_y * q2_0 + q1_x * q2_z$$

$$q3_z = q1_0 * q2_z + q1_z * q2_0 + q1_y * q2_x - q1_x * q2_y$$

where q1 and q2 are the parent and child quaternions, respectively, and q3 is the quaternion that represents the orientation of the child segment.

The second step necessary for deriving meaningful joint angles is that segment orientations expressed using quaternions must be converted into Euler angles. Specifically, the position of a limb in three-space may be considered as the result of one or more rotations in each of the cardinal planes. The values of the rotation angles and the accuracy of the conversion relative to the original quaternion depends on the order of rotation as well as the joint in question and even the movement being performed [25]. The conversion can be performed using each one of 12 possible combinations of the three axes of rotation, also known as *rotation sequences*. In this work, we chose the rotation sequences for each movement and joint that are most commonly used in biomechanics [22–24].

In addition to computing joint angles from the Kinect's quaternion stream, they can also be derived directly from the three dimensional joint locations. Specifically, the location of two joints in 3D space defines a body segment orientation. Following standard practice [26,27], the angle of each body segment is calculated relative to the normal of the floor, giving what is defined as an *absolute*