An Evaluation of a 3D Multimodal Marker-Less Motion Analysis System

Thiago Braga Rodrigues
Faculty of Engineering & Informatics
Athlone Institute of Technology
Athlone, Ireland
t.brodrigues@research.ait.ie

Ciarán Ó Catháin
Faculty of Science & Health
Athlone Institute of Technology
Athlone, Ireland
ciaranocathain@ait.ie

Declan Devine
Materials Research Institute
Athlone Institute of Technology
Athlone, Ireland
ddevine@ait.ie

Kieran Moran
Health and Human Performance
Dublin City University
Dublin, Ireland
kieran.moran@dcu.ie

Noel E O'Connor
Insight Centre for Data Analytics
Dublin City University
Dublin, Ireland
noel.oconnor@dcu.ie

Niall Murray
Faculty of Engineering and Informatics
Athlone Institute of Technology
Athlone, Ireland
nmurray@research.ait.ie

ABSTRACT

Motion analysis is a technique used by clinicians (among many others) that quantifies human movement by using camera-based systems. Marker-based motion analysis systems have been used across a variety of application domains, from Interactive 3D Tele-Immersion (i3DTI) environments to the diagnosis of neuromuscular and musculoskeletal diseases. Although such analysis is performed in several laboratories in many countries, numerous issues exist: (1) the high cost of precise motion capture systems; (2) scarcity of qualified personnel to operate them; (3) expertise required to interpret their results; (4) space requirements to install and store these systems; (5) complexity in terms of measurement protocol required for such systems; (6) limited availability; (7) and in some situations the use of markers means they are unsuitability for certain clinical use cases (e.g. for patients recovering from orthopaedic surgery). In this paper, we present, from a system perspective, an alternative, cheaper, and more accessible system for motion analysis. The ultimate aim is to use the output of this multimodal marker-less system as part of an immersive multimedia gait re-education tool. In real-time, it will advise the user on their gait performance (as well as potentially providing accurate clinical data to clinicians). With the initial focus on the capture system, we have developed and evaluated a novel multimodal system which integrates Multiple Microsoft Kinects (which employ RGB-D cameras) with multiple Shimmer Inertial Measurement Unit (IMU) sensors. We have compared this system with the VICON system (the gold standard in motion capture). Our marker-less motion capture system combines data from 4 skeletons generating 3D and complete 360 degrees in view skeleton. The system combines unit quaternions from each Kinect joint with quaternions from 4 inertial measurement units to promote integration. We used our system to measure 3D points of 12 joints from the Kinect fused skeleton and flexion-extension angles of the knee and hip in a walking trial in 8 participants with 8-10 trials per participant. The analysis found component similarity of 0.97 for knee angles and 0.98 for hip angles. These results show that our system, through combination of Multi Kinect system and Shimmer IMUs, offers a cheaper, sufficiently accurate and more accessible human motion analysis system.

CCS CONCEPTS

• **Information systems → Multimedia information systems**

• **Computing methodologies → Motion Capture**

• **Computing methodologies → 3D imaging**

KEYWORDS

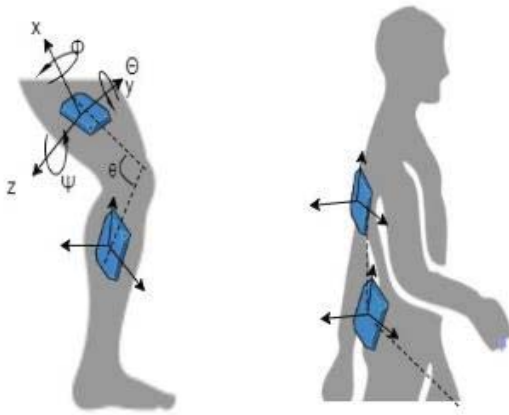Motion capture, marker-less, inertial sensors, 3D Model, Multimodal data fusion, GAIT Re-education, Immersive Multimedia

Figure 1. IMU leg sensor placement. The sensors were placed at mid-points of chest, sacrum, thigh, and tibia.

Figure 2. Multimodal module diagram with server, Kinects, and inertial sensors.

1 Introduction

Motion Capture, or MoCap, is a term used to describe the process of digitalizing a motion, transposing it into a digital format [1]. MoCap technologies have been used in military, sports, medical, robotics, cinema, game creation, and in health, to evaluate a user's performance [2]. There are two main types of MoCap technologies: marker-based motion capture and marker-less-based MoCap. Marker-based technology involves the recording of actions and movements of participants, with reference to markers attached on the body at specific pre-determined anatomical landmarks. Whilst well known marker-based MoCap systems such as the VICON system [3] provide highly accurate data, some of the aforementioned issues exist. Although VICON systems [3] have excellent precision, the cost of equipment can be inhibitive, more than 100000 thousand euros [11]. There are some situations where high-performance systems like the VICON are not required. The high cost and specific installation make it difficult to be used in clinical applications for example. Also, the need of reflective markers can be an obstacle and invasive for human analysis in elders and persons recovering from orthopaedic surgeries. As a result, research and industry have looked at marker-less MoCap technologies. Technologies such as the Microsoft Kinect [4] (which uses RGB-D cameras) can capture 3D skeleton data using anatomic landmarks. Such systems also potentially provide a better user experience, with no movement restriction and faster set up. Thus, marker-less tracking technologies like the Kinect are an attractive alternative.

However, the skeletal tracking of a single Kinect has some limitations: (i) it was designed to track the users facing the sensor (frontal views); (ii) it cannot discriminate between the frontal and rear views (e.g. even if the user is facing the opposite direction, the sensor still assumes a frontal view); (iii) its tracking frequently fails due to occlusions (e.g. the arms are occluded by other body parts); (iv) it can tolerate rotations of a user around the vertical axis only up to certain limits; (v) the area within which it can track users is quite limited [5]. Crucially, in terms of model accuracy, different
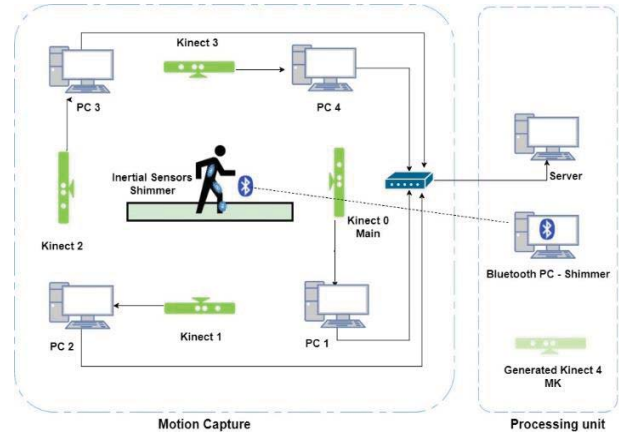
joints have variations in accuracies e.g. lower joints are less accurate compared to upper joints [6]. To have a 360° of tracking area, increasing the field of view and increase precision in skeleton detection, multiples Kinects can be employed for a wider detection range. Thereafter skeletons from all Kinects can be aligned, synchronized, and a full 360° fused skeleton can be generated using matrix transformations and quaternion fusion [7]. However, even with Multi Kinect systems, there are some accuracy issues in particular application domains (e.g. Gait analysis, interaction with virtual reality environments in tasks where high levels of dexterity are required etc.).

To address some of the short comings of a full 360° fused skeleton, the proposed system augments the Multi-Kinect system with an Inertial Measurement Unit (IMU) system. An IMU is an electronic device capable of measuring different parameters of a body like acceleration, force, angular rate, and magnetic field around the object by combining several sensors like accelerometers, gyroscope, and magnetometers [8][9]. To perform biomechanical measurements, such as range of motion, it is necessary to incorporate more than one IMU so that angular evaluation is possible. The sensors are positioned at the midpoint of the limbs to facilitate this evaluation as shown in the scheme in Fig.1 [10]. Hence, between the multi-Kinect and the IMUs, we propose a multimodal marker-less motion analysis system. The aim of this paper is to evaluate if the novel multimodal system which combines marker-less 3D motion capture with low cost inertial sensing provides an accurate method of human motion capture.

2 Related Work

Research on human movement analysis using Motion Capture techniques has grown substantially in recent years [10][11] with existing systems varying in: the number and configuration of cameras; whether or not they use markers; the representation of the captured data; processing algorithms; and application [12].

Marker-based optical systems are expensive and require high resolution cameras and bespoke software for data processing [13].
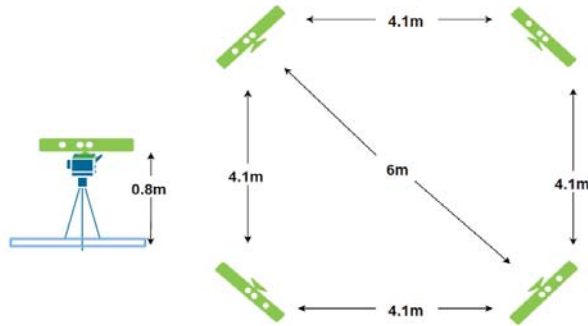
Figure 3. Kinect arrangement setup.

They are the system of choice for motion analysis [14] due to their precision but some drawbacks include high cost, complex set up and the occlusion of markers (post-processing is required to fill gaps). As a result, Marker-less systems such as the Microsoft Kinect are employed to estimate 3D human body motion analysis. Otte et al. [15], investigated the accuracy of a single Kinect for clinical measurement and reported less than 5cm error for all joints when compared with the VICON system. Schlagenhauf et al. [16], proposed a Dual-Kinect tracking system. They compared their system with the VICON and reported absolute error not exceeding 20mm. Geerse et al. [17], proposed a multi-Kinect assessment for quantitative gait against marker-based MoCap system to extract full body kinematics (it did not include inertial sensors) showing spatiotemporal Gait mean of 0.88 compared with VICON. The results of this study concluded that a Multi-Kinect system could provide sufficiently accurate motion analysis. However, their system was not capable tracking full 360º motion and did not consider the view from the back. Such related works have demonstrated the occlusion limitation of a single Kinect and potential of multiple Kinect systems as a viable alternative for motion analysis.

To compliment the capability of single Kinect systems, inertial sensors have been included to capture relevant and additional motion capture data. The work of Bó et al. in [18] and [19] are the closest to our work found in the literature. These authors employed a single Kinect integrated with 6 DOF (Degrees of Freedom) inertial sensors for rehabilitation purposes. They found out that the integration is possible but further improvement in calibration and more degrees of freedom from inertial sensors was required (Our system applies a 9DOF inertial sensor and Multiples Kinect). [19] has also used a single Kinect and inertial sensors to build a system to help post-stroke patients showing application in a clinical setting. However, no integration is described. Other authors fused a single Kinect for low-cost skeleton tracking and refined gesture recognition [20][21] achieving an overall recognition rate of 93%.

Considering these works and to the best of the author's knowledge; the novelties of the system proposed here lies in the fact that: (i) no works in the literature have integrated inertial sensors with a fused Kinect skeleton from multiple Kinect by the combination of 3D points (x, y, z) and quaternion orientation (w, i,

j, k) in a 360º view, (ii) it proposes a real-time wireless synchronization and streaming protocol for multiple IMU's. This is important as it supports easy set up and is low cost (the final cost of the system is less than 10% of the cost of a full VICON system); provides 3D and kinematic data with 9 degrees of freedom; enables fully body reconstruction; provides accurate join angles; is marker-less, motion capture to be used in any environment. To demonstrate the utility of our system, we report a comparison with the VICON.

In the next section, we introduce our multimodal system and discuss the results of system testing. The proposed system framework, signal processing, and experimental protocol is also described.

3   Experimental and Computational Details

The developed multimodal system captures and combines metadata from 4 MS Kinect cameras, and data from 4 Shimmer wearable inertial sensor units. The multimodal system architecture is composed of two main modules: The Multi Kinect module, and the IMU module as shown in Fig. 2.

3.1   Multi Kinect Module

Being aware of the limitations of a single Kinect; such as occluded joints and limited area of movement (only front view), a module containing 4 MS Kinects was developed as per Fig. 2. Each Kinect was powered by its own computer (4 quadcore Intel Core i7, 16GB DDR3 RAM, 3.2Ghz and Graphics processing unit) and was connected to the master server which process the data from each Kinect and generates the fused 3D skeleton. Each Kinect captured a skeleton from one view perspective. The angle range of the Kinect is 57.5° horizontal and 43.5º vertical [5]. The multi-Kinect fusion enables full human body motion capture in 4 views. The multi Kinect system contains 3 components: input, processing, and output. The input component consisted of 4 skeletons (one for each Kinect). The processing component was responsible for synchronization, calibration, noise reduction, and skeleton fusion. The output component returned the original 4 skeleton data and a fused 360° skeleton. For calibration purposes, all Kinects were kept at same height (0.8 meters), and the distance between Kinects in a square arrangement is 4.1 meters. The diagonal distance is 6m. This arrangement is kept for all experiments to provide consistent data (Fig. 3).

To use skeleton data of 4 Kinects, the coordinate transformation between each Kinect into a "master" Kinect was performed. For this transformation, we accepted the frontal Kinect $(K0_{xyz})$ as the "master" Kinect and all other Kinects referred to that 3D system. Using the coordinate transformation relationship as per eqn. 3, one Kinect's skeleton coordinate system was transformed into the second Kinect's skeleton coordinate system. As a result, the skeleton coordinates in both Kinect skeleton were represented in the same coordinate data system.

To discover the coordinate transformation matrix, we adopted a closed-form solution using unit quaternions to get a 4×4 transformation matrix [22].

$$M_{AB} = \begin{pmatrix} R_{[0][0]} & R_{[0][1]} & R_{[0][2]} & Tx \\ R_{[1][0]} & R_{[1][1]} & R_{[1][2]} & Ty \\ R_{[2][0]} & R_{[2][1]} & R_{[2][2]} & Tz \\ 0 & 0 & 0 & 1 \end{pmatrix}$$ eqn. 1

Considering:

a) M_{AB} – 4x4 transformation matrix to change one local Kinect A into a global coordinate B.

b) $R_{[m][n]}$ – The 3x3 rotation matrix

c) T – The 3x1 translation vector

To discover the transformation matrix of each Kinect (local coordinate system) into the K0 global coordinate system, we applied eqn. 2 as per:

$$Bi = sR * Ai + T$$ eqn. 2

or

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = s \begin{pmatrix} R_{[0][0]} & R_{[0][1]} & R_{[0][2]} \\ R_{[1][0]} & R_{[1][1]} & R_{[1][2]} \\ R_{[2][0]} & R_{[2][1]} & R_{[2][2]} \end{pmatrix} * \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} Tx \\ Ty \\ Tz \end{pmatrix}$$ eqn. 3

a) Bi – 3xN matrix representing the unit quaternion of the Kinect global 3D point

b) Ai – 3xN matrix representing the unit quaternion of the Kinect local 3D point

c) s – Scale factor if needed (default 1)

To obtain the transformation matrix and calibrate one Kinect skeleton to K0 skeleton, at least four joints must be detected by the two calibrating Kinect at the same time. We assumed that the captured skeleton data was reliable when the person was standing in front of the sensor and two Kinects can track all 20 joints at the same time in a static trial. To get the more accurate transformation matrix, 120 frames of reliable skeleton data were captured. The sum of the 20 joints coordinate difference values between calibrated Kinect and KO was calculated as per eqn. 4.

$$CDV_i = \sum_{j=0}^{20} \left(\begin{bmatrix} A_j.x \\ A_j.y \\ A_j.z \\ 1 \end{bmatrix} - M_{AB} * \begin{bmatrix} B.x \\ B_j.y \\ B.z \\ 1 \end{bmatrix} \right), (i = 0, \dots, 119)$$

eqn. 4

By comparing 120 sums of the coordinate difference values, the transformation matrix with minimum coordinate difference sum was chosen. Note that because sampling frequency of the multi Kinect system is approximately 35Hz, an oversampling was performed to synchronize the multimodal system This synchronization was achieved by the Kinect server. The server ensured the packets for each Kinect arrived at the simultaneously to ensure synchronization.

3.2 Inertial Measurement Unit

The IMU module contained 4 IMU Shimmer sensors. A real-time wireless synchronization and streaming protocol for multiple IMU was developed. The developed protocol fused in real-time accelerometer, gyroscope, and magnetometer data and generated the quaternion orientation. The data was synchronized with the computer CPU clock ensuring no data is lost.

A MATLAB script for the multi IMU screaming was developed to perform the following capture protocol: sampling frequency of all sensors was defined to be 51.2Hz; internal configuration of each IMU; synchronization between the sensors; start/stop IMU system data capture. More specifically in terms of internal configuration of each IMU, 10 streams of data were captured: 3D acceleration from accelerometer (Acc_{xyz}), 3D angular velocity from gyroscope ($Gyro_{xyz}$), 3D magnetic field from a magnetometer (Mag_{xyz}), and a timestamp [23]. As discussed later in this section, the Acc_{xyz}, $Gyro_{xyz}$, Mag_{xyz}, were fused to provide quaternion representation. The Shimmer IMU has sufficient internal memory to store sessions but each of the sensors has its own time clock. Hence, synchronization of all 4 IMU during data capture is required. An algorithm was designed and implemented to achieve synchronization. Pseudo code for this algorithm is provided in Algorithm 1.

To represent orientation of a rigid body or frame coordinates in 3D space, a quaternion representation was employed. This complex number representation can define any spatial rotation around a fixed point or coordinate system. A quaternion $q = [q_0\ q_1\ q_2\ q_3]$, can be used to get an angle θ about a fixed Euler axis [7][22] and eqn. 5. To get the angle between two joints with Shimmer, quaternion matrixes were obtained by the fusion of the 3 Shimmer internal modules (Acc_{xyz}, $Gyro_{xyz}$, Mag_{xyz},) using a Madgwick-based orientation filter [24].

The quaternion generated by the function can represent spatial rotation of each shimmer and represents angle in each axis. Having each Euler angle, it is then possible to reference one Shimmer to another and get the angle between two sensors. The angle between two IMU was used as part of the walking evaluation during experiments. The integration of the multi Kinect skeleton and IMU Modules was achieved by combining unit quaternions from 2 Kinect joints K_{xyz} and quaternions from the IMU located in the mid-point of those 2 joints $q = [q_0\ q_1\ q_2\ q_3]$ by rotating the quaternion q around vector v directing the two Kinect joints (v = K1 - K2) (eqn. 6 and Fig. 4). After getting the angles of IMU and Kinect, the angles are merged, and a combined output is generated.

$$\begin{bmatrix} \phi \\ \theta \\ \Psi \end{bmatrix} = \begin{bmatrix} \arctan \frac{2(q_0\ q_1 + q_2\ q_3)}{1 - 2(q_1^2 + q_2^2)} \\ \arcsin(2(q_0\ q_2 - q_3\ q_1)) \\ \arctan \frac{2(q_0\ q_3 + q_1\ q_2)}{1 - 2(q_2^2 + q_3^2)} \end{bmatrix}$$ eqn. 5

ϕ – rotation about the x axis
θ – rotation about the y axis
Ψ – rotation about the z axis

Algorithm 1: Real-time Multi Shimmer Streamer

---

function MultiShimmer(comPorts, jointNames, captureTime)
Input: 4 com Ports (one for each IMU), joint names, and capture time
Output: The .csv files containing each IMU data
1:　　**if** all sensors are connected through BT protocol **then**
2:　　　　Define Shimmer Handle Class instance;
3:　　　　Define sample rate;
4:　　　　Set internal board to 9DOF;
5:　　　　Enable Shimmer internal sensors (Acc, Gyro, Mag);
6:　　　　Synchronize sensor clock with PC
7:　　　　**if** IMUs are ready to capture **then**
8:　　　　　Start assessment and capture
9:　　　　　Audio alert
10:　　　　　**while** elapsedTime < captureTime **do**
11:　　　　　　Write data in CVS file
12:　　　　　　IMU quaternion sensor fusion
13:　　　　　**end while**
14:　　　　　Stop assessment and capture
15:　　　　　Audio alert
14:　　　　　Write the percentage of received packets to detect any lost information
15:　　　　**end**
16:　　　　Disconnect Shimmers
21:　　**end**
22: **end**

[Algorithm 1. The function receives COM ports, joint names and capture time and generates synchronized and calibrated sensor data. The function also generates quaternions for Euler angle calculation]

$$v' = \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix} = \begin{bmatrix} (1-2q_2^2-2q_3^2) & 2(q_1q_2+q_0q_3) & 2(q_1q_3-q_0q_2) \\ 2(q_1q_2-q_0q_3) & (1-2q_1^2-2q_3^2) & 2(q_2q_3+q_0q_1) \\ 2(q_1q_3+q_0q_2) & 2(q_2q_3-q_0q_1) & (1-2q_1^2-2q_2^2) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad \textbf{eqn. 6}$$

3.3　Experimental protocol

The experimental protocol adheres to the approach taken in numerous related works in the literature [25][26][27] and included a number of steps:

a)　Participants recruitment;
b)　Information sheet and consent form;
c)　Joint measurements;
d)　Plug-in Gait marker placement [28];
e)　8 to 12 Trials per participant (1 calibration and walks);

8 healthy participants completed 12 trials each, providing a total of 96 individual datasets. During each trial, motion data was captured using our multimodal system and VICON simultaneously, hence allowing direct comparison.

Before the experiment, an information sheet was given to each participant to explain the experiment, purposes of the project, and data confidentiality. The participant was also required to sign a consent form. As part of the set-up stage of the experiment, reflective markers were placed on the body following the Plug-In-Gait methodology from VICON [28]. Each marker placement and measurement took approximately 45 minutes. After mark placement, joint measurements were taken to give additional data of each participant. This measurement was taken in the following body segments: arms, legs, height, hip, and shoulders. Finally, the
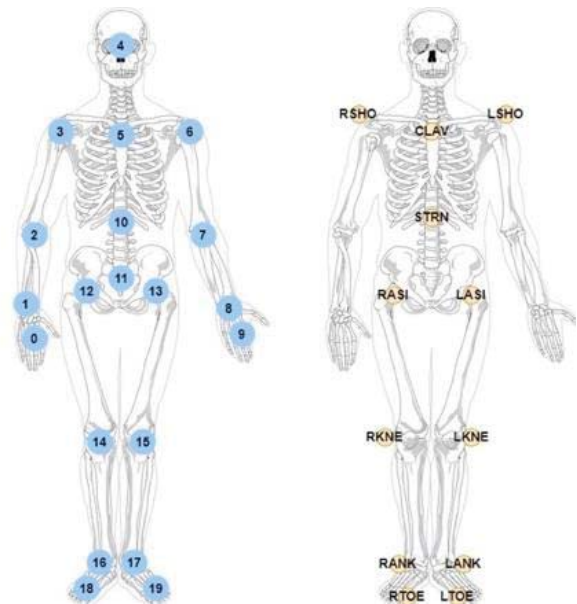

Figure 4 – (a) Kinect joint index, (b) VICON Plug-in-Gait marker placement

4 Shimmer IMUs were attached to the participant's body located at mid-points of chest, sacrum, thigh, and tibia (Fig. 1). Before data collection, static trials were performed to calibrate the Multi Kinect system, the IMUs, and VICON systems.
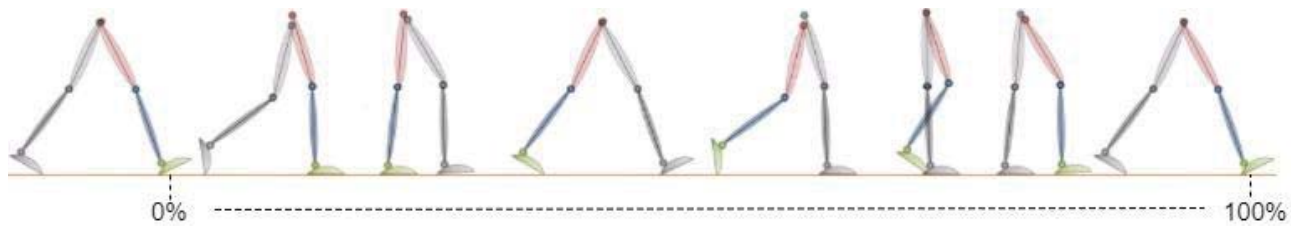
Figure 5 – Gait cycle representation

3.4 Data and signal processing

3 distinct datasets were captured for each user in each trial: Multi Kinect, Shimmer, and VICON. The Multi Kinect dataset of 5 skeletons (4 single Kinect Skeletons and 1 fused Kinect Skeleton) was stored in a .csv file and each skeleton were composed of 20 joint points. Data from Shimmer IMU was stored in a matrix format as described above. The VICON dataset, like Kinect, was in 3D position format, and was captured (VICON$_{xyz}$) on a per reflective marker basis. During each trial, some of the markers were occluded (a known problem with VICON). Hence, post the test, each VICON trial needed to be processed separately, frame per frame, to ensure all gaps were filled using spline fill and pattern fill gap filling operations [28].

To compare VICON and each Kinect three tasks needed to be completed: (1) select Kinect joints that can be related to VICON reflective markers (e.g. same body segment; right arm, right leg, hip; (2) change VICON local coordinate system into Kinect global coordinate system; (3) synchronize both systems with an external event. For (3), a jump in a force plate was used and as such each event captured by the force plate generated a trigger for the systems.

As per Fig. 4, 12 joints were selected via the Plugin GAIT and were compared with respective Kinect joints. These were: right shoulder (3_RSHO), shoulder centre (5_CLAV), left shoulder (6_LSHO), spine (10_STRN), right hip (12_RASI), left hip (13_LASI), right knee (14_RKNE), left knee (15_LKNE), right ankle (16_RANK), left ankle (17_LANK), right foot (18_RTOE), and left foot (19_LTOE); all of which are important for GAIT variable extraction and analysis. Fig. 6 shows that there is a difference between VICON markers and Kinect joint. This difference is explained because each marker is attached onto skin whilst Kinect joints are inferred in the anatomic position for the joint. However, this difference is not considered for joint signal comparison, and it was filtered during calibration procedure.

It is possible to develop comparison methods using Cross-correlation for each single Kinect and Multi Kinect with VICON. The Cross-correlation is the similarity measurement of two signals with the displacement of one relative to other. This correlation has many uses like pattern recognition and signal displacement [29]. Considering each joint point in space (Kinect$_{xyz}$ and VICON$_{xyz}$) as 3 distinct signals, the correlation of the signal can be evaluated. For the multimodal system, Principal Component Analysis (PCA) was applied. PCA is the statistic method to orthogonally convert a set
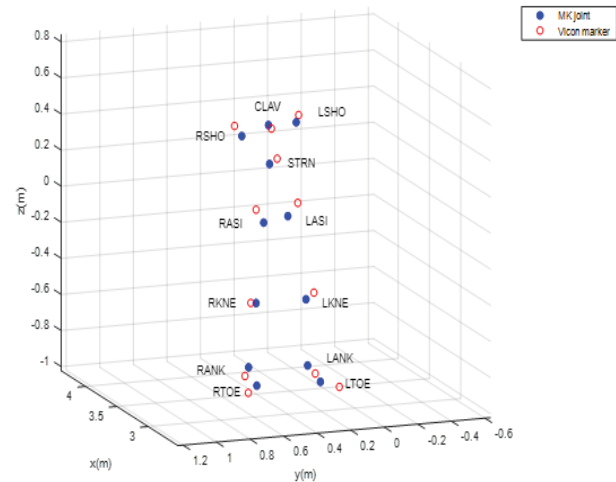

Figure 6 – 3D VICON and Multi Kinect points placement

Table 1. Joint location difference in cm: VICON and Multi Kinect comparison.

Joints/differences in cm					
RSHO	CLAV	LSHO	STRN	RASI	LASI
4.06	8.70	0.15	1.12	1.81	0.98
RKNE	LKNE	RANK	LANK	RTOE	LTOE
4.53	2.58	0.33	2.31	3.92	3.92

of observations of possibly correlated variables into linear observations called Principal Components [30].

A single trial was achieved when the participant completed a full gait cycle 0-100% (Fig. 5). This cycle happens when the participant step on the ground (heel strike), removes the heel stepping with other foot (initial swing), and step on the ground with the same foot in heel strike (terminal swing) [31]. The three streams of data (Kinects, Shimmer, and VICON) were normalized and synchronized showing a gait cycle.
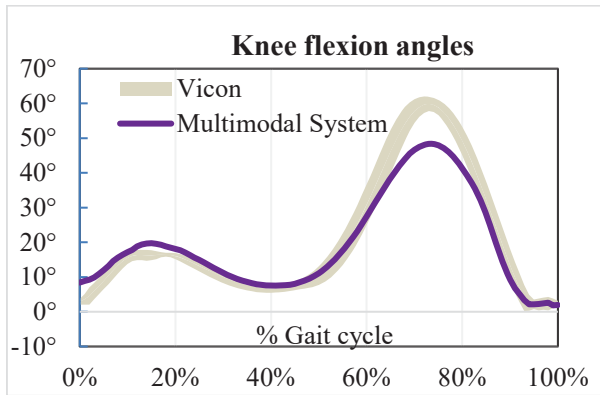
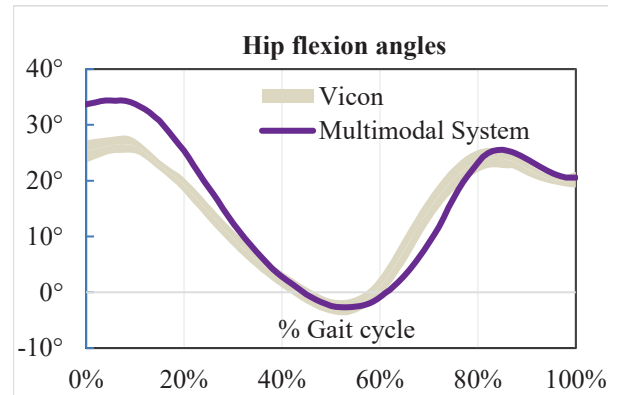**Figure 7. VICON and Multimodal System knee flexion angles.**

Figure 8. VICON and Multimodal System hip flexion angles.

4   Results and Discussion

To evaluate the Multimodal System, we performed analysis on the cross-correlation of VICON and each Kinect, and principal components analysis of VICON with the multimodal system. The results are presented in the following sections.

4.1   Comparing VICON vs multi Kinect: difference in 3-dimensional space per joint and Cross-correlation

Table 1. shows difference in 3-dimensional space comparison between Kinect system joint locations and the VICON system. The mean difference between joint locations was found to be 3.10 cm for all joints in all trials. As expected and stated by the author in [6], some joints were more accurate than others. For lower extremity joints, the results did not surpass 4cm. This 3D signal difference did not influence the joint angles and variables that were captured and analyzed for knee and hip flexion angle. To further investigate the results on Table 1, we have done cross-correlation analysis on VICON and Kinect signals. This analysis can measure the similarity of two signals over time. The results of this cross-correlation are presented in Table 2.

The Cross-correlation between VICON and Kinect shows that the frontal Kinect (K_0) and the Multi Kinect (MK) have better results for cross-correlation 0.86 and 0.85 than the other combinations as expected. The results for VICON and Kinects can be explained because the walking motion was performed towards the K0 and the sensor was programmed in a manner such that the best capture occurs when user faces the device. However, a single Kinect is not enough for a full 360° capture. Considering that the Multi Kinect skeleton is formed by the fusion of lateral and posterior Kinects (K_1, K_2, and K_3), the Multi Kinect could maintain a stable joint correlation (0.85). The results for cross-correlation for Kinects 1, 2, and 3 have demonstrated that the correlation of each Kinect compared against VICON depends on which Kinect the participant is facing. This problem is overcome with a 3D multi Kinect in use.

Table 2. Cross-correlation per joint

Joint	K0	K1	K2	K3	MK
RSHO	0.98	0.58	-0.39	0.44	0.98
CLAV	0.74	0.41	-0.34	0.26	0.74
LSHO	0.98	0.60	-0.44	0.45	0.98
STRN	0.97	0.51	-0.37	0.34	0.97
RASI	0.95	0.36	-0.26	0.18	0.95
LASI	0.94	0.39	-0.29	0.16	0.94
RKNE	0.82	0.35	-0.11	0.09	0.85
LKNE	0.80	0.08	-0.14	0.12	0.81
RANK	0.82	0.36	-0.19	0.47	0.76
LANK	0.89	0.21	-0.11	0.34	0.84
RTOE	0.60	0.28	-0.18	0.39	0.64
LTOE	0.80	0.19	-0.09	0.33	0.78
Mean	0.86	0.36	-0.24	0.30	0.85

4.2   VICON vs (multi Kinect + shimmer): Knee and Hip angles

Hip and knee angles were evaluated as these joints are weight-bearing joints and most susceptible to require bone surgical interventions [32]. To evaluate angles from all data sets, two approaches were used: first, having three distinct points in space and second: from the quaternion function and Euler angles from two IMU. The angle between 3 points is equivalent of the angle between two vectors defined by same 3 points (eqn.7-9). Considering 3 points P1, P2, and P3 and vectors u and v:

$$u = P1 - P2 \text{ and } v = P3 - P2 \qquad \textbf{eqn. 7}$$

$$\cos\theta = \frac{u.v}{|u||v|} \qquad \textbf{eqn. 8}$$

$$\theta = arcos\left(\frac{u.v}{|u||v|}\right) \qquad \textbf{eqn. 9}$$

Fig. 7 presents a comparison of the VICON and the Multimodal System for knee flexion angles in a gait cycle. Fig. 8 shows the same comparison for hip flexion angles. The main areas of

Table 3. PCA of flexion angles – Component table

System	Knee Flexion		Hip Flexion	
Component:	1	2	1	2
Vicon	1	0.01	1	0.19
Kinect 0	0.97	0.00	0.98	0.10
Kinect 1	-0.49	0.06	-0.55	0.45
Kinect 2	-0.36	0.08	-0.37	-0.10
Kinect 3	0.71	-0.27	0.33	0.85
Multi Kinect	0.80	0.24	0.97	-0.19
Shimmer	0.96	0.12	0.99	-0.03
MM System	**0.97**	0.19	**0.98**	-0.02

difference are highlighted in red: where the results exceed one standard deviation from VICON. From the figures presented, the x axis shows the percentages of gait from 0 to 100%. The initial phase 0% happens when the heel contacts the ground and 100% when the same heel contacts the ground. The y axis represents flexion angles of hip and knee. The grey curve reflects one standard deviation from Vicon. From these graphs, we can see visual representation of the gait cycle graphs of Shimmer IMU, Multimodal System and VICON.

To ensure we considered similarities between the VICON and the other systems (Kinect frontal "K_0", Multi Kinect, Shimmer, and Multimodal System), statistical analysis was performed. Principal Component Analysis (PCA) was employed as it can extract the main components of variance and correlation. KMO & Bartlett's Test of Sphericity [33] was employed as it is a measure of sampling adequacy of the data for component analysis for knee and hip angles. The KMO value must be above 0.6. and the sigma value less than 0.05%. The KMO value for knee angles was 0.714 and 0.657 for hip angles. Sigma values found of 0.02% and 0.03%. These values assure PCA suits the data we analysed.

PCA in knee and hip flexion. PCA was employed to determine (via variance), how many components could be used in the analysis. The Total Eigenvalue must be above 1 in order to have a valid component. Eigenvalues of two main components of knee flexion groups were 5.34 and 1.4 and only components 1 and 2 were used. Component 1 explained 69.68% of variance, and component 2 explained 15.89%. Table 3 shows similarity of systems based on component 1 and 2 for the knee flexion group. We can state that two groups are similar when they have component analysis close to 1. Because of that, we have found out similarity in: VICON with Kinect 0, Shimmer, and Multimodal System (0.97, 0.96 and 0.97, respectively). Eigeinvalues of two main components of hip flexion groups are 6.17 and 1.04. Components 1 and 2 are used. Component 1 explained 7.14% of variance, and component 2 explained 13.05%. Table 3 also shows similarity of systems based on component 1 and 2. Similarity found in: VICON with Multimodal System (0.98). These results indicate that only a single Kinect or an IMU would be enough to represent angles, but they do not represent the combined system. A single Kinect does not represent the full 3D body in 360º and an IMU does not give 3D points in space.

5 Conclusion

This work presented a novel multimodal motion capture system, that combined a multiple Kinects module and inertial sensors module. The system performance was compared with the gold standard VICON system. This comparison included comparing VICON signals against all possible multi Kinect 3D Kinect skeleton composed of 4 physical Kinects and 1 generated 360º multi Kinect skeleton. We have also compared VICON flexion-extension angles with our multi Kinect system and multimodal system (multi Kinect + Shimmer IMU). The results have presented analysis and discussion on signals differences between the proposed system and VICON; as well as angles estimation differences from inertial sensors integration. Our analysis has demonstrated the utility of our multimodal system (inclusive of its limitations). Based on this, many potential use cases of the Multimodal system can be proposed. The system can promote angle estimation from the IMUs and position in space of multi Kinect.

The proposed system is cheaper; easy to be set up; show clear and easily interpretable results; marker-less; supports 360 degrees of motion analysis; is easily portable and does not require large set up space or environment. Future work will include the use of the Multimodal System for Gait variable analysis, build a Machine Learning model on captured dataset to detect patterns for the "normal" Gait, and the use the model in real-time as an immersive multimedia haptic/Augmented Reality feedback tool for Gait re-education.

ACKNOWLEDGMENTS

REFERENCES

[1]   Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. 2002. Multimodal human discourse: gesture and speech. ACM Trans. Comput.-Hum. Interact. 9, 3 (September 2002), 171-193. DOI=http://dx.doi.org/10.1145/568513.568514

[2]   Peng-zhan Chen, Jie Li, Man Luo, and Nian-hua Zhu. 2015. Real-Time human motion capture driven by a wireless sensor network. Int. J. Comput. Games Technol. 2015, Article 4 (January 2015), 1 pages. DOI: https://doi.org/10.1155/2015/695874

[3]   Vicon. Motion Capture Systems. Retrieved March 10, 2017 from https://www.vicon.com/

[4]   Kinect for Xbox One | Xbox. Retrieved March 10, 2017 from http://www.xbox.com/en-US/xbox-one/accessories/kinect

[5]   Microsoft Kinect Sensor Evaluation. Retrieved March 10, 2017 from https://ntrs.nasa.gov/search.jsp?R=20110022972

[6]   Simon Choppin, Ben Lane, and Jonathan Wheat. 2014. The accuracy of the Microsoft Kinect in joint angle measurement. Sports Technology 7, 1-2 (March 2014), 98–105. DOI:http://dx.doi.org/10.1080/19346182.2014.968165

[7]   Quaternion from rotation matrix. [four-parameter representation of coordinate transformation matrix]. Retrieved March 10, 2017 from https://ntrs.nasa.gov/search.jsp?R=19780048191

[8]   Volker Kempe. Preface. Inertial MEMS, xiii-xiv. DOI:http://dx.doi.org/10.1017/cbo9780511933899.001 Ian Editor (Ed.). 2008.

[9]   Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, Nazirah M. Khairi, and Vijayabaskar Kasi. 2013. Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications. International Journal of Signal Processing Systems (2013), 256–262. DOI:http://dx.doi.org/10.12720/ijsps.1.2.256-262

[10] Gazihan Alankus, Amanda Lazar, Matt May, and Caitlin Kelleher. 2010. Towards customizable games for stroke rehabilitation. Proceedings of the 28th international conference on Human factors in computing systems – CHI 10 (2010). DOI:http://dx.doi.org/10.1145/1753326.1753649

[11] Patric Eichelberger et al. 2016. Analysis of accuracy in optical motion capture – A protocol for laboratory setup evaluation. Journal of Biomechanics 49, 10 (2016), 2085–2088. DOI:http://dx.doi.org/10.1016/j.jbiomech.2016.05.007

[12] T.p. Andriacchi, L. Muendermann, S. Corazza, and A. Chaudhari. 2006. A new era in the capture of human movement; markerless capture of human movement. Journal of Biomechanics 39 (2006). DOI:http://dx.doi.org/10.1016/s0021-9290(06)83680-8

[13] Szykman, A. and Gois, J. (2014). Evaluating the Feasibility of Low Cost Motion Capture Systems. In: XIII SBGames. Pp.1133-1136.

[14] Alexandra Pfister, Alexandre M. West, Shaw Bronner, and Jack Adam Noah. 2014. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. Journal of Medical Engineering & Technology 38, 5 (2014), 274–280. DOI:http://dx.doi.org/10.3109/03091902.2014.909540

[15] Otte K, Kayser B, Mansow-Model S, Verrel J, Paul F, Brandt AU, et al. (2016) Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Function. PloS ONE 11 (11): e0166532. Doi:10.1371/journal.pone.0166532

[16] Schlagenhauf F, Sahoo P P, Singhose W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. Robot Autom Eng J. 2017; 1(2): 555558.

[17] Daphne J. Geerse, Bert H. Coolen, and Melvyn Roerdink. 2015. Kinematic Validation of a Multi-Kinect v2 Instrumented 10-Meter Walkway for Quantitative Gait Assessments. Plos One 10, 10 (2015). DOI:http://dx.doi.org/10.1371/journal.pone.0139913

[18] Antônio P.L Bó, Mitsuhiro Hayashibe and Philippe Poignet. 2011. Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2011). DOI:http://dx.doi.org/10.1109/iembs.2011.6090940

[19] Hossein Mousavi Hondori, Maryam Khadem, and Cristina Videira Lopes. 2012. Monitoring Intake Gestures using Sensor Fusion (Microsoft Kinect and Inertial Sensors) for Smart Home Tele-Rehab Setting. 1st Annual IEEE Healthcare Innovation Conference of the IEEE EMBS(2012).

[20] Shimin Feng and R. Murray-Smith. 2014. Fusing Kinect Sensor and Inertial Sensors with Multi-rate Kalman Filter. IET Conference on Data Fusion & Target Tracking 2014: Algorithms and Applications (2014). DOI:http://dx.doi.org/10.1049/cp.2014.0527

[21] Kui Liu, Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. 2014. Fusion of Inertial and Depth Sensor Data for Robust Hand Gesture Recognition. IEEE Sensors Journal 14, 6 (2014), 1898–1903. DOI:http://dx.doi.org/10.1109/jsen.2014.2306094

[22] Berthold K.P. Horn. 1987. Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America A 4, 4 (January 1987), 629. DOI:http://dx.doi.org/10.1364/josaa.4.000629

[23] Shimmer Documentation Support. Retrieved March 10, 2017 from http://www.shimmersensing.com/support/wireless-sensor-networks-documentation/

[24] Sebastian O. H. Madgwick, Andrew J. L. Harrison and Ravi Vaidyanathan. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. 2011 IEEE International Conference on Rehabilitation Robotics (2011). DOI:http://dx.doi.org/10.1109/icorr.2011.5975346

[25] Felix Stief, Harald Böhm, Katja Michel, Ansgar Schwirtz, and Leonhard Döderlein. 2013. Reliability and Accuracy in Three-Dimensional Gait Analysis: a Comparison of Two Lower Body Protocols. Journal of Applied Biomechanics 29, 1 (2013), 105–111. DOI:http://dx.doi.org/10.1123/jab.29.1.105

[26] Alberto Ferrari, Maria Grazia Benedetti, Esteban Pavan, and Carlo Frigo. 2007. Quantitative comparison of five current protocols in gait ... (November 2007). Retrieved April 8, 2018. DOI: :10.1016/j.gaitpost.2007.11.009

[27] Hans Kainz and others. 2017. Reliability of four models for clinical gait analysis. Gait & Posture 54 (2017), 325–331. DOI:http://dx.doi.org/10.1016/j.gaitpost.2017.04.001

[28] Vicon Plug-in-Gait. Downloads. Retrieved March 10, 2017 from https://www.vicon.com/downloads/documentation/plug-in-gait-product-guide

[29] Richard D. Keane and Ronald J. Adrian. 1993. Theory of cross-correlation analysis of PIV images. Fluid Mechanics and Its Applications Flow Visualization and Image Analysis (1993), 1–25. DOI:http://dx.doi.org/10.1007/978-94-011-2690-8_1

[30] Hervé Abdi and Lynne J. Williams. 2010. Principal component analysis. WIREs Comput. Stat. 2, 4 (July 2010), 433-459. DOI: https://doi.org/10.1002/wics.101

[31] Streifeneder. The eight phases of human gait cycle. https://www.streifeneder.com/downloads/o.p./400w43_e_poster_gangphasen_druck.pdf

[32] Chaganti R-K, Lane N-E. (2011) Risk factors for incident osteoarthritis of the hip and knee Curr Rev Musculoskelet Med (2011) 4: 99.

[33] KMO and Bartllet's Test. Retrieved March 18, 2017 from https://www.ibm.com/support/knowledgecenter/en/SSLVMB_sub/spss/tutorials/fac_telco_kmo_01.html