Precise Extraction of Anatomical Segment Orientations from the Kinect One Sensor

Nuno Filipe Inácio Vaz Matias*
nuno.vaz.matias@gmail.com
Thesis to obtain the Master of Science Degree in Biomedical Engineering
Supervisors: Professor Miguel Tavares da Silva* and Daniel Simões Lopes, PhD**
*Instituto Superior Técnico, Universidade de Lisboa
**Visualization and Intelligent Multimodal Interfaces, INESC-ID Lisboa
October 2016

Abstract — The recent advances for the Kinect One, a low-cost marker-less motion capture sensor, and its human motion tracking algorithm have led to the development of applications on the most diversified fields such as clinical rehabilitation, biomechanical analysis and entertainment. The tracking algorithm estimates efficiently the three-dimensional position of twenty-five anatomical joints which all together constitute a human skeleton. However, for the simulation of a human biomechanical model, the orientations it provides to guide each anatomical segment are poorly implemented and the documentation available is inconsiderable. For the purpose of obtaining a better estimation for the anatomical segments' orientations, six techniques based on vector orthogonalization were implemented. To validate the orientations estimated by each method, twenty-eight subjects performed a set of ten movements that were recorded by the Kinect One sensor and by a marker-based motion capture system that is considered the reference. Discrepancies between orientations estimated by the marker-based system and by each technique used on the marker-less system were graphically compared and were evaluated using the Pearson's correlation coefficient. The results obtained show that the six techniques implemented estimated with good to very good correlation the orientation around one axis of rotation while for the orientations around the other axes of rotation, the performance was different between the different techniques and movements. These findings encourage the development of a parametrized model to predict the optimal orientation of each anatomical segment taking into account the position of the human joints estimated by the Kinect One sensor.

Keywords — Motion analysis; Skeleton tracking; Kinect One; Vector orthogonalization; Kinematics; Euler angles.

I. Introduction

Human motion tracking, the process of recording spatiotemporal information of the human movement, presents nowadays a huge importance in the progress of several domains. On the rehabilitation field, for a patient who suffered a stroke or another pathology that have partially incapacitated his motor function, the continuously monitoring of the patient's movement reveals to be essential in order to rectify his motions and restore his independence for every daily activity [1]. The tracking of human movement is also extensively implemented on the analysis of the sports performance of top-levels athletes in order to improve their results and on the analysis of gait patterns for clinical purposes [2].

To achieve the best outcomes on the previous applications it is required a motion tracking system to be the most accurate as possible. Currently, the one that provides better precision and accuracy is the marker-based visual system [1]. In this system, several cameras are used to track reflective markers placed on the body of the person. However, this motion capture system has some drawbacks: it is expensive [3]; it takes too much time to set all the markers on the subject [3]; the markers may not be placed in the correct position [4] or they may move due to the displacements caused by the soft tissue [5]; it lacks portability which makes it impossible, for example, to perform physical therapy at the patient's home [6].

An alternative that is being explored by the researchers is the implementation of marker-less visual systems, such as the Kinect One sensor, in which a depth image is created and combined with a color camera to estimate the position of the centers of several human joints [7]. This device is much less expensive than the marker-based system, it does not require the placement of markers on the body and it is much easier to carry it to different places. Several studies have been performed in order to investigate the accuracy and reliability of the Kinect sensor on the estimation of the joints' position and the kinematic parameters that result from those [8-11]. The results found seem to be promising for the implementation of this sensor on the different applications.

However, the Kinect sensor presents one major drawback which is the estimation of the joint's orientation from the joints' position [11, 12]. The software produced for the sensor has a built-in algorithm that provides those orientations but, due to their difficult interpretation and lack of documentation about how to use them, they are not reliable. Although several investigators had analyzed the joint angle formed between two anatomical segments (for example, the angle between the forearm and the arm for elbow flexion motion [3]) they did not look for the three possible rotations that some joints can perform. This drawback leads to a poor representation of the human body model which may induce misleading results on the

computations of several kinematic parameters. Besides, this problem would also affect the visual feedback of the human body motions which is vital on several domains that need motor learning such as rehabilitation.

The main objective of this work is to study the effectiveness of the implementation of different vector orthogonalization methods for the estimation of the relative orientations of the anatomical body segments using the Kinect One sensor. Vector orthogonalization corresponds to a linear algebraic problem whose purpose is to find a set of linear independent orthogonal vectors, which form a basis for a given vector space. To assess the validation of the techniques proposed, the orientation data estimated from the Kinect One sensor is compared with the orientation data estimated by a standard marker-based system. Within this objective, it is intended to obtain anatomical rotations of the virtual model that correspond more faithfully to the real rotations performed by the human body.

II. METHODOLOGY

A. *Biomechanical Model*

For the formulation of a biomechanical problem and its analysis, a biomechanical model representative of the physical system under study needs to be created. In this work, that model corresponds to an open-loop multibody system composed of a set of rigid bodies, each completely specified by six independent coordinates that define their position and orientation. The physical system to be considered corresponds to the full human body composed of eleven major anatomical segments: the head, the chest, the abdomen and the both arms, forearms, thighs and legs. It was decided to formulate this physical system in order to be similar to the human skeleton tracked by the marker-less Kinect sensor, as shown in Figure 1 (a). The biomechanical model composed of 11 rigid bodies is illustrated in Figure 1 (b). Each anatomical segment will correspond to one rigid body and each rigid body will be represented by a reference frame with its origin located at the center of mass and its basis vectors defining the orientation.

The reference frames located on each anatomical segment can relate to the articular movements that occur on an anatomical plane around a given axis of rotation perpendicular to that plane. The human movements mainly occur on three anatomical reference planes (sagittal, frontal and transverse) and around the respective orthogonal anatomical axis of reference (mediolateral, anteroposterior and longitudinal). The axes of the reference frame of each anatomical segment tend to be aligned with the anatomical axes of reference and so it is possible to relate the rotation around a given axis with the human movements. For example, in Figure 2 is depicted the movement of knee flexion that occurs on the sagittal plane around the mediolateral axis. When the leg rotates, the reference frame with respect to the rigid body representative of the leg suffers a rotation around the axis ξ_B (which is the axis aligned with the mediolateral axis). The reference frame of the thigh remains static during that movement and it is possible to estimate the orientations of the leg during the movement by comparing the reference frame of both rigid bodies. Note that, in this work, the global reference frame is designated by xyz [13] and the local reference frame is designated by $\xi\eta\zeta$ [14].

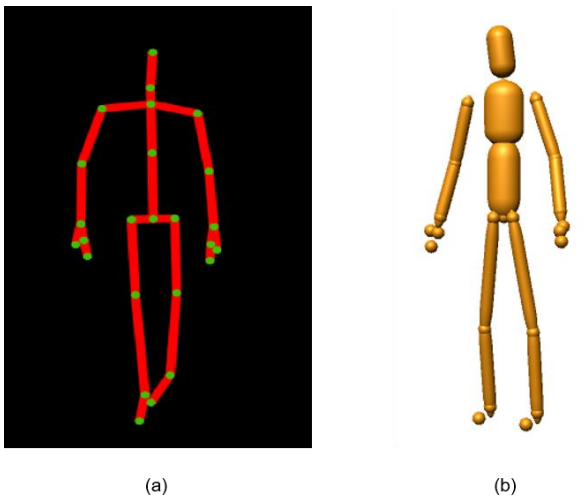

Figure 1 - Skeleton tracked by the Kinect One sensor (a). Multibody system composed of 11 rigid bodies (b).

The basis vectors of the local reference frame of each rigid body correspond directly to the rigid body's orientations in relation to the global reference frame since the local reference frame is estimated using information from the global reference frame. More specifically, the local reference frames are estimated through several points on the data acquisition space which are measured on coordinates from the global reference frame. There is one global reference frame for each motion capture technology used. For the marker-based system, these points will correspond to several markers that need to be positioned strategically on the body to cover all the anatomical segments, as shown in Figure 3. For the marker-less system, those points are estimates of the subject's joint centers computed automatically by the associated software, as shown by the green points in Figure 1 (a).

When the local reference frames of every rigid body of the model are reconstructed, then the biomechanical model is fully defined as depicted in Figure 4. The reconstruction of those reference frames had to be adapted from the ISB guidelines for joint coordinates systems [15, 16] since the skeleton estimated by the marker-less system is more limited on the number of points. The axes aligned with the anteroposterior axis, the longitudinal axis and the mediolateral axis are illustrated, respectively, by red, green and blue, either for the local reference frames either for the global reference frames of both systems.

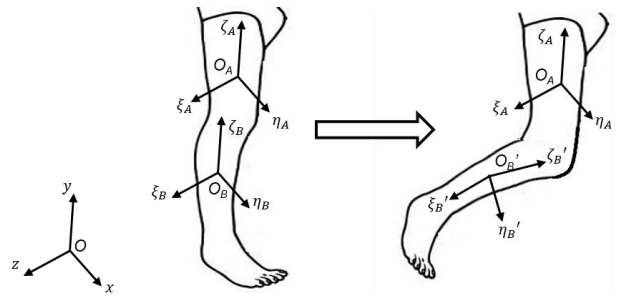

Figure 2 - Representation of a knee flexion movement and its effect on the rigid bodies' reference frame. During the knee flexion movement, the leg's reference frame follows rotates around the $\boldsymbol{\xi_B}$-axis leading to a new orientation. The thigh's reference frame remains static during the movement.

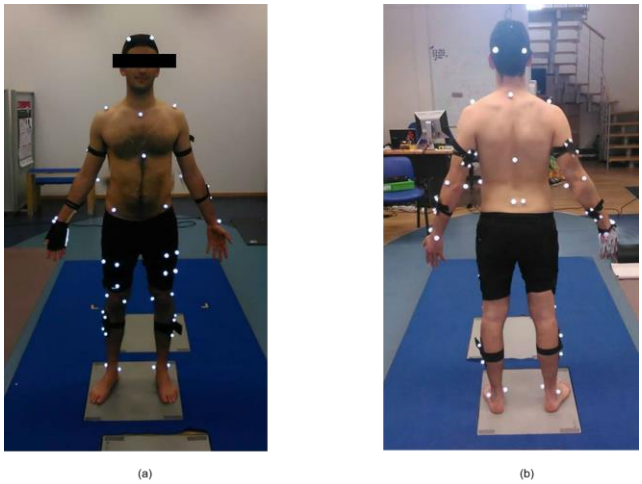

Figure 3 - Photography of the marker-set with all the anatomical and tracking markers on the front (a) and on the back (b) for the marker-based motion capture system.

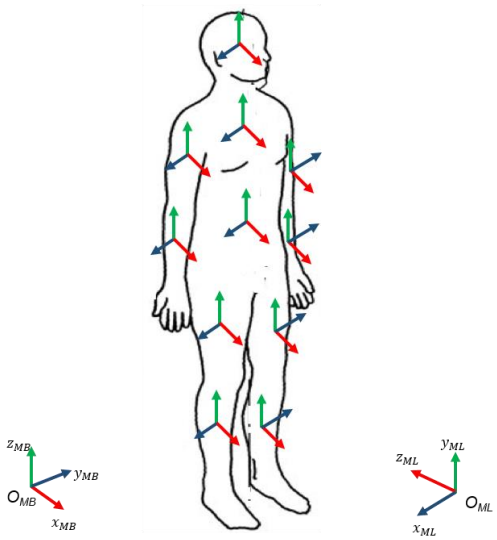

Figure 4 - Representation of the full body with the local reference frames of each rigid body shown in the respective anatomical segments and the global reference frames depicted for each motion capture system.

The computations to reconstruct the local reference frames of each rigid body are different between the marker-based system and the marker-less system. In Figure 5 is shown the necessary points and vectors to reconstruct the local reference frame of the right thigh for the marker-based system. With the three-dimensional (3-D) data of the positions of each marker (black dots) it is possible to estimate the positions of the hip and knee joints (white dots). The origin of the local reference frame is given by the mid-position between the points *RHJC* and *RKNE* since they are the only two points of the given rigid body,

$$\mathbf{r}_{O_A}^{MB} = \frac{\mathbf{r}_{RKNE}^{MB} + \mathbf{r}_{RHJC}^{MB}}{2} \tag{1}$$

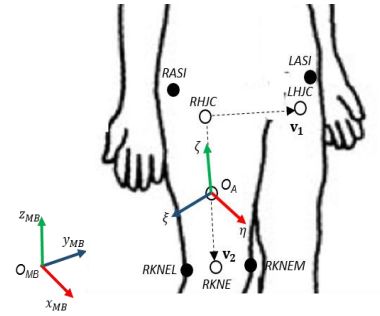

Figure 5 - Local reference frame of the thigh on the marker-based system. The markers are illustrated by black dots while the points estimated from those by white dots.

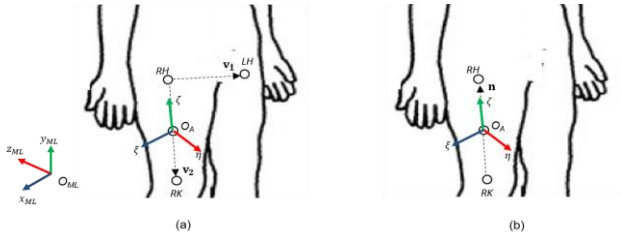

Figure 6 - Local reference frame of the right thigh on the marker-less system. It can be reconstructed with three points by the TP technique (a) or with two points with one of the other techniques (b).

where $\mathbf{r}_{RKNE}^{MB}$ and $\mathbf{r}_{RHJC}^{MB}$ corresponds to the position vector of the points *RKNE* and *RHJC* in relation with the global reference frame. In order to form an orthogonal reference frame, the axes of the local reference frame are computed by a series of cross-products and with the aid of two auxiliary vectors $\mathbf{v_1}$ and $\mathbf{v_2}$,

$$\zeta = -\mathbf{v_2} \tag{2}$$

$$\eta = \mathbf{v_1} \times \xi \tag{3}$$

$$\xi = \eta \times \zeta \tag{4}$$

In Figure 6 is represented the local reference frames estimated with the marker-based system for one of the vector orthogonalization techniques which requires three points (a) and for the other five techniques which require only two points (b). The technique that needs three points is called Three Points (TP) and it makes use of two auxiliary vectors similar to the marker-based system. Besides, the computations to estimate the origin of the local reference frame and its basis vectors are also identical to the Equations (1)-(4) but with the points *RH*, *LH* and *RK*. The remaining five vector orthogonalization techniques are called Householder (HH), Eberly (EB), Square Plate (SP), Spherical (SF) and Projection Matrix (PM) [17]. These techniques receive a normalized input vector **n**, which is given as the difference between the points *RH* and *RK*, and estimate the remaining basis vectors of the local reference frame by giving as output two orthogonal vectors, **t** and **b**, to the input vector.

For the remaining rigid bodies of the biomechanical model, the process to estimate their local reference frames is similar to the thigh's rigid body, either for the marker-based and marker-

less systems. The only differences are the points to choose and the order some computations are performed.

B. Data Acquisition

This study recruited twenty-eight healthy adults (25 ± 9 years old, 170 ± 9 cm height, 61 ± 9 kg weight, 13 women) to participate on the biomechanical analysis. All subjects needed to write the informed consent provided to them in order to get permission for their participation with a clear understanding of the implications and consequences of the experiment. The acquisition was conducted at the Lisbon Biomechanics Laboratory on Instituto Superior Técnico [18] with equipment provided by the laboratory and by INESC-ID [19].

Prior to the motion analysis, the subject was requested to wear appropriate clothing, that is, a lycra cap and lycra shorts, and personal information of age, height and weight was registered for each subject. Ten different elementary movements were performed: shoulder flexion/hyperextension, shoulder abduction/adduction, shoulder transversal abduction/adduction, shoulder medial/lateral rotation, elbow flexion, forearm pronation/supination, hip flexion/hyperextension, hip abduction/adduction, knee flexion and hip medial/lateral rotation. For each type of movement, five repetitions were performed by the subject in order to get a subject's representative movement. On each repetition, the subject started from an adapted pose of the anatomical reference position and finished on the same position it started.

Data was collected, simultaneously, using two different motion capture systems: marker-based and marker-less systems.

The marker-based system tracked the subject's movement by using the Qualisys motion capture system [20], whose software, Qualisys Track Manager (QTM), reconstruct the 3-D points from a set of two-dimensional (2-D) projected images at a sampling frequency of 100 Hz. This system needs several infra-red (IR) cameras which emits radiation that is reflected on several reflective markers and captured by the cameras. The 3-D reconstruction of each marker is then performed by the algorithm of Direct Linear Transformation (DLT). Those reflective markers can be seen placed in the body in Figure 3.

The disposition of those reflective markers were based on a modified version of the Cleveland Clinic Foundation's marker set [21], in which plastic fixtures that aggregate three or four markers called clusters are used to track the anatomical segment. The markers placed on the clusters are called tracking markers while the remaining markers are called anatomical markers. In this work, the anatomical markers are used to reconstruct the local reference frames of the biomechanical model during a static acquisition. During the movement, the local reference frames are estimated using the data from the tracking markers. This can be performed by finding a relation between the anatomical markers and the tracking markers on the static acquisition. A cluster's local reference frame based on the tracking markers can be defined as shown in Figure 7 [22]. Then it is possible to express those markers' position with respect to the local reference frame of the segment and, thereby, find a relation between the anatomical reference frame and the cluster reference frame. This relation is invariant since the distance of the anatomical markers to the respective cluster

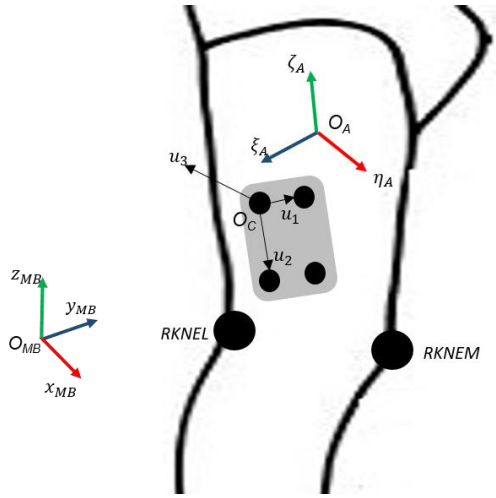

Figure 7 - Representation of the right thigh with two anatomical markers *RKNEL* and *RKNEM* and one cluster with four tracking markers. The reference frame associated with the cluster is centered at O_C and has three axis u_1, u_2 and u_3.

remains constant. In Figure 7, the anatomical markers of *RKNEM* and *RKNEL* on the knee can be related to the reference frame *C* of the cluster located on the right thigh. When the thigh moves, the anatomical markers and the cluster follows the movement by keeping a constant distance between them. This process have the advantage of minimizing the skin-effect errors that may occur on the bony landmarks where the anatomical markers are placed [23].

The marker-less system tracked the subject's movement by estimating the joints' positions of the subject with the Kinect One motion capture device [7] whose operation is based on the time-of-flight (TOF) method. This device is depicted in Figure 8.

Systems use this method to measure the time that a given object or a given signal take to travel some distance. The motion capture device used in this thesis is constituted by a TOF camera (512x424 pixels resolution), an RGB camera (1920x1080 pixels resolution) and an array of four microphones. That TOF camera emits IR radiation, thought a built-in IR emitter, which is reflected on some target and is collected back by a camera matrix of IR sensors on the camera. By measuring on each pixel the time that the light took on that process it is possible to calculate the distance the targets are to the camera and consequently obtain a depth image [24]. The Kinect One sensor can acquire the information at a sampling frequency of 30 Hz.

With the Software Development Kit (SDK) developed by Microsoft for the Kinect sensor, on which is possible to create several applications using the device, the positions of 25 joints can be estimated using the body tracking algorithm [25]. On this algorithm, those joints are inferred using a machine learning algorithm called randomized decision forest which was learned using a training data set with more than one million samples. The joints estimated with this algorithm form a human skeleton as can be seen in Figure 1 (a). Besides, in contrary to the marker-based system, this marker-less device does not require any prior calibration since the skeleton model originated by the

Figure 8 - Kinect One sensor used for the motion capture acquisition without markers [7].

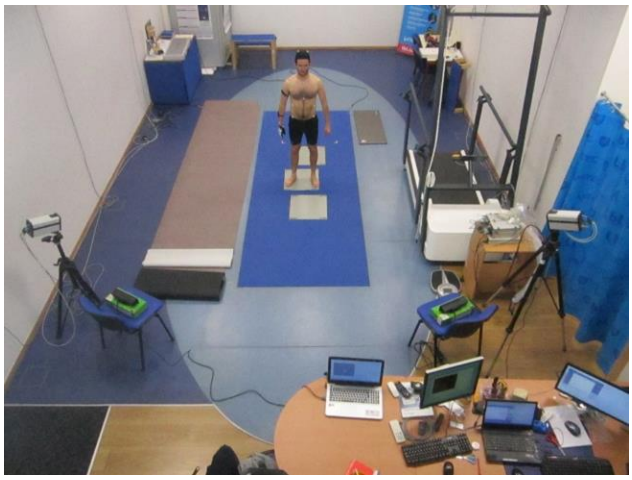

Figure 9 - Photography of the workplace for the experiment. The subject is placed on the middle gray plate on the floor. Two Kinect sensors are placed, above a chair, on the right and left sides of the subject. Two of the fourteen IR cameras are also depicted on a tripod.

algorithm is automatically detected whenever the subject is in the camera field-of-view.

The accuracy of how the sensor predicts the joints' positions is sensitive to the position and orientation of the camera regarding the location of the subject [26]. For example, self-occlusion of some body parts by other parts could lead to a poor skeleton's model estimation by the camera. Therefore, one solution to solve that is to use more than one camera in the workspace to capture the subject from a different angle. In this thesis, two Kinect sensors were positioned at 2.5 m in front of the subject and at 0.7 m above the ground as suggested in [27].

Upon the subject's arrival and after wearing the appropriate clothing, he is requested to place on the middle of the laboratory where 14 IR cameras tracked 62 reflective markers (32 tracking markers + 30 anatomical markers) and two Kinect sensors tracked the joints' positions of the subject, as depicted in Figure 9. Four computers were needed to obtain all the acquisition data since it is not currently possible to connect more than one Kinect to a single computer and it was not possible to acquire the marker-less and marker-based data on the same computer due to expensive computational cost. In one of the computers, an interface built in Unity software [28] was created using C# language, in order to receive and store the data from each Kinect sensor and to keep monitoring the acquisitions in order to

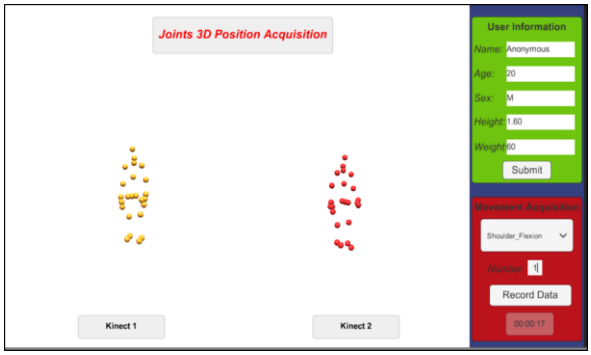

Figure 10 - Interface developed on Unity for the gathering of the information provided by the Kinect sensors.

guarantee that there were no errors during the experiment. That interface is depicted in Figure 10.

C. Trajectory Data Processing

Before estimating the orientations for the both motion capture systems, the data acquired of the 3-D positions of the markers and joints needs to be processed. For that purpose, computational methods were implemented mainly using several scripts created on the software MATLAB [29].

1) Data Storage

On the marker-based system, the movements acquired on a *.qtm* file which is a format that can be open by the QTM software and allows the user to observe all the markers tracked during the acquisition. Two tasks must be performed for each file: model implementation and trajectory reconstruction. The first consists on the creation and application of a model on each measurement, for each subject, in order to label the markers that come unidentified. The second corresponds to the connection of unidentified trajectories that, either by occlusion either by the proximity of two markers, were split but, in effect, corresponds to the same marker.

On the marker-less system, there is no need of pre-processing steps since the data from the Kinect sensors is directly saved as *.txt* file from the Unity interface.

Each participant performed five repetitions for each of the ten movements plus a static acquisition, providing then a total of 51 acquisition files for each subject and for each technology. Since two Kinect sensors were used for the marker-less system, putting together with the marker-based system, a total of 153 acquisition files arises for each subject. The information of each of those files is opened on MATLAB and stored in a matrix for a better handling of the data on the processing steps. For each technology i, each line of the matrix will represent an acquisition frame k while the columns will represent the coordinates of the markers measured by the marker-based system or the coordinates of the joints estimated by the marker-less system,

$$\mathbf{y}_k^i = \left[\left(\mathbf{j}_{k,1}^i\right)^T \ \left(\mathbf{j}_{k,2}^i\right)^T \dots \left(\mathbf{j}_{k,M}^i\right)^T \right] \qquad (5)$$

where M is the total number of markers (marker-based system) or estimated joints (marker-less system), and $\mathbf{j}_{k,m}^{i}$ corresponds to the 3-D coordinates of each m-th point measured on the global reference frame of the i-th device,

$$\mathbf{j}_{k,m}^{i} = \left[x_{k,m}^{i} \;\; y_{k,m}^{i} \;\; z_{k,m}^{i}\right]^{T} \tag{6}$$

Therefore, each acquisition file is represented by a matrix of the form

$$\begin{bmatrix} \mathbf{y}_1^i \\ \mathbf{y}_2^i \\ \vdots \\ \mathbf{y}_N^i \end{bmatrix} = \begin{bmatrix} \left(\mathbf{j}_{1,1}^{i}\right)^T & \left(\mathbf{j}_{1,2}^{i}\right)^T \ldots \left(\mathbf{j}_{1,M}^{i}\right)^T \\ \left(\mathbf{j}_{2,1}^{i}\right)^T & \left(\mathbf{j}_{2,2}^{i}\right)^T \ldots \left(\mathbf{j}_{2,M}^{i}\right)^T \\ & \vdots \\ \left(\mathbf{j}_{N,1}^{i}\right)^T & \left(\mathbf{j}_{N,2}^{i}\right)^T \ldots \left(\mathbf{j}_{N,M}^{i}\right)^T \end{bmatrix} \tag{7}$$

where N corresponds to the total number of frames of the acquisition. The conjunct of all the matrixes for every subject constitute the input for the following data processing steps.

2) Interpolation

The data acquired by the marker-based system needs to be interpolated in order to fill gaps that may be present on the trajectories. Even after the trajectory reconstruction step, the marker may not could be detected for a given set of frames which automatically maps the trajectory's values on those frames to zero. After the detection of the gaps on each trajectory curve, cubic spline interpolation was performed to fit a smooth curve between the edges of the gaps.

Interpolation on the joints' positions estimated by the Kinect sensors was performed but with a different reason in relation to the marker-based system. The operating sampling frequency of the Kinect was 30 Hz while for the cameras on the marker-based system were 100 Hz. In order for the signals acquired on both system have the same frequency, which is necessary for further processing steps, resampling of the Kinect signals with a cubic spline interpolation was performed [11].

3) Filtering

Due to the use of several electronic devices and different reconstruction and digitalization algorithms, the data that is directly acquired in the laboratory presents several unwanted oscillations which correspond to noise introduced on the trajectories [22]. This noise is usually characterized by high frequencies of the signal while the low frequencies of the signal corresponds to the information that matters. Therefore, to eliminate those high frequencies and keep the low frequencies of the signal, a low-pass filter must be implemented. This low-pass filter is characterized by a cut-off frequency, f_c, which sets a boundary: frequencies of the signal lower than f_c are not attenuated while frequencies of the signal higher than f_c are attenuated. This value must not be neither too high, which guarantees that the signal is preserved but the noise may not be removed, neither too low, which assures that most part of the noise is attenuated but the signal may be slightly distorted.

The trajectory data was filtered by a wavelet filter with Daubechies wavelets and levels of decomposition of 6 and 3, respectively, for the marker-less and the marker-based systems. A higher level of decomposition means that the frequency of the signal was decomposed into high and low frequencies more times in relation to a lower level of decomposition which, in other terms, means the applying of a lower cut-off frequency. The choice of the level of decomposition for each motion capture system was by, systematically, decompose the signal one level and compare the reconstruction signal with the original signal to verify if the noise was attenuated and the main part of the signal was preserved.

4) Synchronization

To compare the information from two or more different devices it is essential that the data from one device is synchronized with the data from the others devices. This should be achieved during the acquisition process by initializing the data collection, on both systems, at the same time. However, sometimes this is not possible to perform since the data is being acquired by different software on different computers and a phase shift of several seconds may occur, in the same trial, between the data collected from different technologies.

In this work, the data was synchronized by an observation method in which it was detected the moments a given movement starts and returns to its initial position. By performing this step on both data of the marker-based and marker-less system it was guaranteed that the data of both systems became time aligned.

5) Feature Scaling

During the acquisition, the subject was being tracked by the IR cameras of the marker-based system and two Kinect sensors. Therefore, the same anatomical joint was defined on three different global reference frames which are differently located and oriented. Since the coordinates measured for that joint will present different scales for each system, data normalization on the amplitude should be performed in order to standardize the range of the trajectories. This normalization was performed with an algorithm called feature scaling which rescales each signal to the range [0 1].

6) Cross-Correlation

Cross-correlation is a method to measure the degree of similarity between two signals. In this work, this method was implemented to choose the best trajectory signal between both Kinect sensors.

For each movement trial, the trajectories of a given joint estimated by each Kinect were cross-correlated with the trajectories of the markers of the cluster attached to the moving anatomical segment. The maximum value of cross-correlation was taken and compared between both sensors. The Kinect sensor whose maximum value of cross-correlation was higher

was chosen to proceed for the further computations, for that movement and for that joint, while the other is discarded.

Note that to perform the cross-correlation it is necessary that the signals from the marker-less and the marker-based systems have the same sampling frequency, thus the resampling of the marker-less trajectories on the beginning of the data processing being crucial here.

7) Temporal Normalization

The signals from the Kinect sensor and the marker-based system do not have the same length, therefore they are not comparable. For that purpose, resample by spline interpolation of all the data was performed in order for every trajectory's length be the same.

D. 3-D Orientations Estimation

The estimation of the orientations of a given anatomical segment in relation with other segment is relatively different between the marker-based and the marker-less systems. The former needs an additional step which is the static calibration.

1) Static Calibration

In this step, the anatomical reference frame is reconstructed using the anatomical markers and the relation between those markers and the cluster attached to the respective segment. For the case of the leg and thigh, as depicted in Figure 11, the basis vectors of the local reference frame of each segment are computed with similar expressions to Equations (2)-(4). Regarding the cluster on each segment, the position vector of the k-th tracking marker, defined on the global reference frame, is given by

$$\mathbf{r}_k^{MB} = \mathbf{R}_i^{MB}\,\mathbf{r}_k^i + \mathbf{r}_{O_i}^{MB} \tag{8}$$

where $\mathbf{r}_k^i$ corresponds to the position vector of the k-th tracking marker on the $i \in \{A, B\}$ local reference frame, $\mathbf{r}_{O_i}^{MB}$ corresponds to the position vector of the origin of the $i \in \{A, B\}$ local reference frame on the global reference frame and $\mathbf{R}_i^{MB}$ is the rotation matrix that represents the rotation of the $i \in \{A, B\}$ local reference frame in relation with the global reference frame. That rotation matrix gives directly the orientation of the local reference frame of the segment in relation with the global reference frame, and is given by its three basis vectors,

$$\mathbf{R}_i^{MB} = [\eta_i \quad \zeta_i \quad \xi_i] \tag{9}$$

By computing the difference between the position vectors of Equation (8) for a given set of tracking markers on the same cluster and, with a series of cross-products, it is possible to define the cluster's reference frame in relation to the global reference frame as,

$$\mathbf{U}^{MB} = [\mathbf{u}_1^{MB} \quad \mathbf{u}_2^{MB} \quad \mathbf{u}_3^{MB}] \tag{10}$$

For every acquisition frame, the matrices in Equations (9) and (10) are computed since they depend only on values defined on the global reference frame. With those expressions, it is possible to obtain an invariant matrix that represents the basis vectors of the cluster's reference frame but defined on the i-th local reference frame,

$$\mathbf{U}^i = [\mathbf{u}_1^i \quad \mathbf{u}_2^i \quad \mathbf{u}_3^i] = (\mathbf{R}_i^{MB})^{-1}\mathbf{U}^{MB} \tag{11}$$

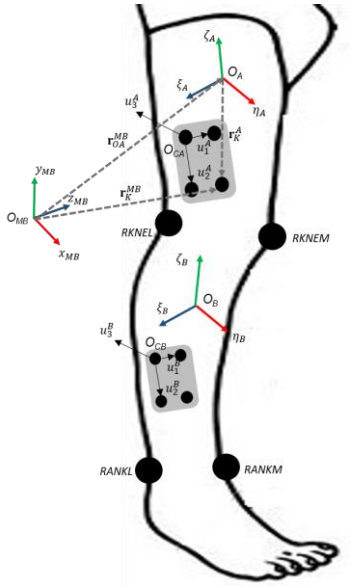

Figure 11 - Representation of the local reference frames for the anatomical segments of the thigh, leg and clusters. The dash lines represent the position vectors regarding one tracking marker *K* of the thigh.

2) Movement Analysis

For every movement trial, to compute the orientation of one anatomical segment relative to another segment, it is necessary to compute the local reference frame for each rigid body's representation of those segments. On the marker-less system, this step is fairly straightforward since the basis vectors for each local reference frame are given by one of the six techniques of vector orthogonalization. The orientations are expressed by the means of rotation matrices similar to the one in Equation (9).

On the other hand, for the marker-based system, the process to get those orientations is different since it makes use of the invariant matrix, $\mathbf{U}^i$, computed on the static calibration step in Equation (11). For every acquisition frame, it is possible to compute the reconstruction matrix described on the global reference frame, $\mathbf{U}^{MB}$, since it is known the positions of the tracking markers on the global reference frame obtained from the acquisition. Then, it is possible to estimate the orientations of a given anatomical segment, in the form of a rotation matrix, with the expression,

$$\mathbf{R}_i^{MB} = \mathbf{U}^{MB}(\mathbf{U}^i)^{-1} \tag{12}$$

The orientation of one anatomical segment in relation with other segment can be estimated, for every acquisition frame, by a relation between the rotations matrices defined for each rigid body. For the case of two rigid bodies *A* and *B*, such as in the

case of Figure 11, the orientation of the rigid body *B* in relation with the orientation of the rigid body *A* is given by,

$$\mathbf{R}_B^A = \left(\mathbf{R}_A^j\right)^{-1}\mathbf{R}_B^j \quad (13)$$

where j corresponds to the marker-based or marker-less capture systems, $j \in \{MB, ML\}$.

The rotation matrix in Equation (13) is converted to a sequence of three Euler angles which corresponds to a sequence of three successive rotations around each axis of reference, in order to evaluate easily the orientations performed for each movement.

III. Results

With the purpose to find which vector orthogonalization technique estimates better the orientations for a determined movement, the three inter-segment rotations obtained for a given rigid body on all the participants are compared between the marker-less system and the marker-based system.

In this extended abstract, the analysis of only one movement is presented. The discussion for the other movements is available on the full document this abstract is based on. Since it was presented how to estimate the orientations of the leg in relation with the thigh, the knee flexion movement is discussed. For this movement, only the orientations with respect to the leg are shown since it is the only moving anatomical segment during the motion.

From the data obtained on the methodology, three representative orientations are estimated for the marker-based system and for each technique used on the marker-less system. Those orientations are obtained by computing, initially, the mean orientation between the 5 trials acquired for each subject and, then, by taking the mean of the mean orientations for all the subjects that participated in the experiment.

In Figure 12 is depicted the three orientations of the anatomical segment of the leg in relation with the segment of the thigh estimated by the marker-based system. The orientation around ξ-axis presents a much higher amplitude in relation with the other two orientations. This outcome is expected since the knee flexion movement occurs mainly on the sagittal anatomical plane whose anatomical reference axis orthogonal to it, the mediolateral axis, is aligned with the ξ-axis. Since the flexion of the knee corresponds to the rotation of the leg towards the posterior direction, then the rotation around the ξ-axis will be in the negative direction during flexion of the leg, which occurs until approximately half of the movement, and then in the positive direction during extension of the leg. Note that the curve does not start at zero degrees, which is when the subject stands at the anatomical reference position and the axis of the different reference frames should be aligned. This offset can be due to the wrong marker placement on the skin of the subject which leads to the computation of mistaken axes for the reference frames.

The orientations estimated with the six vector orthogonalization algorithms on the marker-less system are depicted in Figure 13 against the orientations of the marker-based (depicted on gray). Likewise the standard system, the orientation corresponding to the main rotation is around the ξ-axis. For all the techniques implemented, the pattern of that orientation is similarly to the standard one with an initial decrease in the angular values followed by an increase. Most of those curves begin at the same value of the standard and not at zero as it should be, which implies that the offset is not due to the marker placement but the way the biomechanical is defined. Furthermore, it is possible to observe a big difference between the peak's value of the standard curve and the curves of all the techniques implemented. The reason for this outcome may be due to the poor estimation of the lower body joints' center by the algorithm of the Kinect sensor [30]. A different reason that can be also applied for the knee flexion movement is the possibility of occlusion of the ankle joint by an anatomical segment.

For the orientations around the ζ-axis and η-axis, the kinematic patterns of the marker-less curves do not correspond to the same patterns of the marker-based curve, with most of them revealing a negative relation in relation with the standard ones. This may be due to the combination of several reasons. First, the poor estimation of the lower body joints' center as mentioned before leads to a poor reconstruction of the reference frame. Second, five of the six techniques uses only two points to obtain the axes of reference which induces a lack of precision for the estimation of two of the three axes. During the

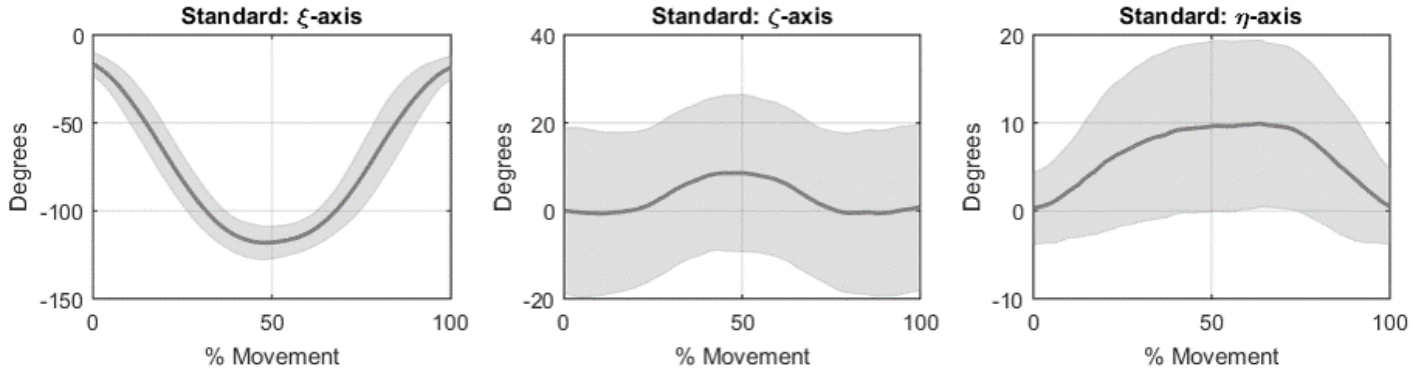

Figure 12 - Orientation curves for the knee flexion movement on the marker-based system on the right leg segment. Each column represents the orientation around each axis of the reference frame ± 1SD (n=28).

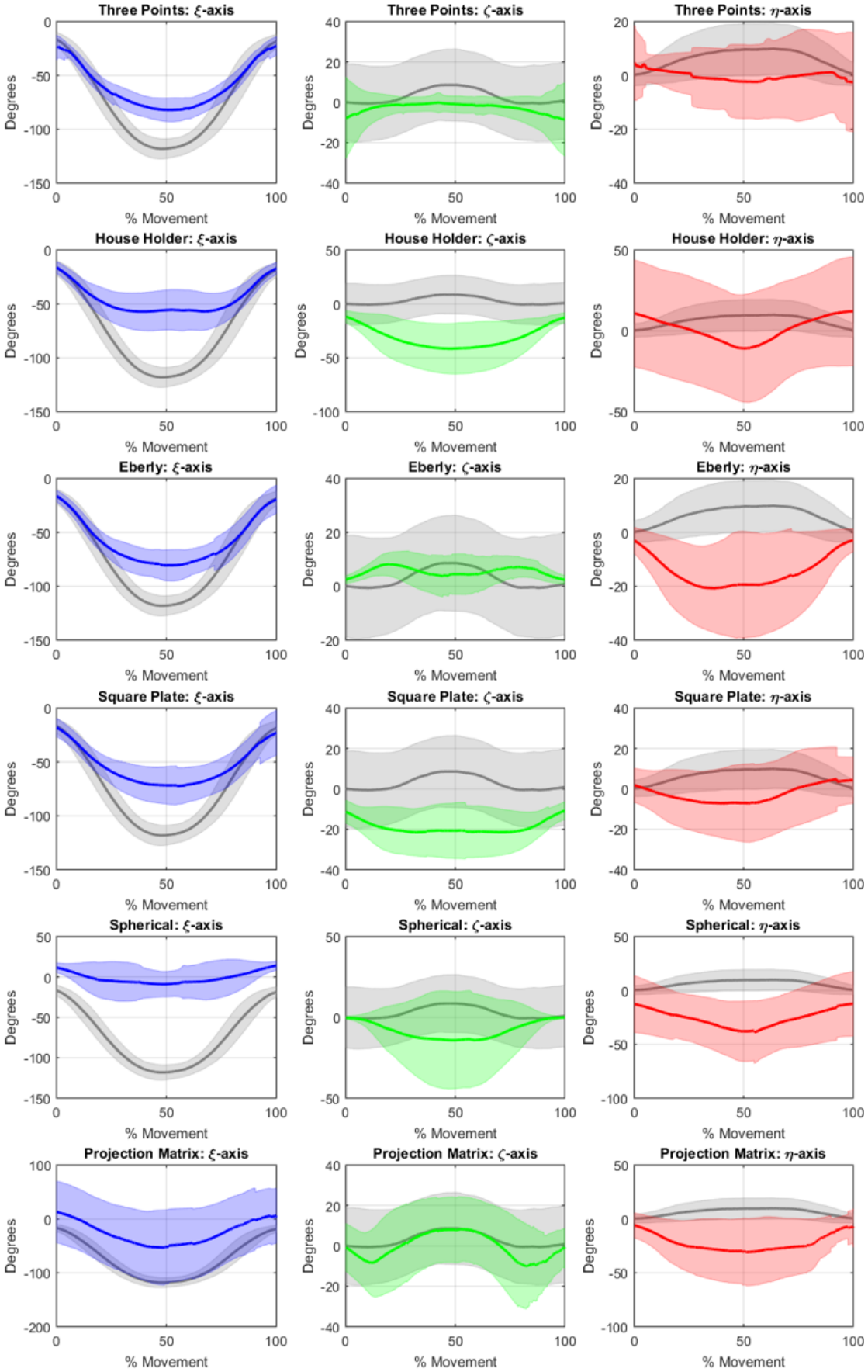

Figure 13 - Orientation curves for the knee flexion movement for each technique applied on the marker-less system ± 1SD . Each column represents the orientation around each axis of the reference frame. Each line represents one of the six techniques used (n=28).

Table 1 - Pearson's correlation coefficient for the knee flexion movement for each technique.

	Three Points		HouseHolder		Eberly		Square Plate		Spherical		Projection Matrix	
Orientation	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
ξ-axis	0.9356	0.051	0.786	0.2979	0.9251	0.0762	0.8911	0.1772	0.5898	0.495	0.8681	0.1616
ζ-axis	0.2221	0.486	-0.3756	0.6031	-0.198	0.5562	-0.244	0.5255	-0.277	0.7312	0.4637	0.4512
η-axis	-0.121	0.591	-0.3919	0.5662	-0.438	0.6423	-0.28	0.6263	-0.439	0.6352	-0.477	0.6871

processing of the orientations' results, although the axes seemed to be pointing in the correct directions they must have rotated in the opposite direction around the ζ and η axes. Third, the tracking algorithm of the Kinect presents inability to determine another point that could originate an orthogonal axis to the main axis of rotation and reducing, subsequently, the uncertainty about the directions the basis vectors would be aligned.

In Table 1 it is presented the Pearson's correlation coefficients for each technique where it is possible to verify that the correlations for the rotations around the ξ-axis are relatively high while most of the correlations for the rotations around the other two axes are low and negative.

IV. Conclusions

The results obtained showed that the vector orthogonalization techniques estimate with relatively good accuracy one of the orientations around one axis of rotation and that the techniques estimate differently the anatomical segments' orientations providing then some versatility for their use on the motion of human movements. However, this work presented limitations and drawbacks that require attention in order to improve further researches.

A biomechanical model of 11 rigid bodies was implemented to estimate the orientations on the two different motion capture systems and was based on the skeleton estimated by the Kinect sensor. It is not known how the biomechanical model implemented could affect the results obtained. Through the discussion of the results it was found that the offsets shown at some orientation curves could be caused by the way the biomechanical model was defined but future works that should focus on the comparison of different models between the two vision-based systems should be performed.

The orientations estimated by the Kinect sensor are directly dependent on the values of the joints' position obtained from the acquisitions. The accuracy of the Kinect sensor on estimating those positions is affected when a given anatomical segment occludes a given anatomical joint or some segment is out of the field-of-view of the Kinect. In this work, two sensors were placed in different locations to be able to track the user from different angles of vision and the cross-correlation method was used to choose the Kinect whose data was more similar to the standard data. It is not known if this is an ideal method to process the data from both sensors but it is recommended the implementation, for future works, of other processing methods such as data fusion [31] to verify if the results improve.

Most of the vector orthogonalization methods implemented on this work present two or three different expressions [17] which are evaluated according to the conditions that are verified. This presented a drawback for the evaluation of the kinematic patterns of the orientations since, when a given orthogonalization method swaps between conditions, a gap is formed on the kinematic pattern which leads to a sudden rotation of the anatomical segment. Although a solution was implemented to solve this problem, a more robust correction should be performed in order to obtain smoother orientations especially for techniques with more than two conditions.

Another limitation found on the application of the vector orthogonalization techniques was the reconstruction of the initial state of the reference frame for different subjects. It is expected that the axes of the reference frame to be aligned with the anatomical axes of reference. However, during the computational processing, it was noted that for some subjects and for the same orthogonalization technique, the orientations curves were reversed in relation to the expected kinematic pattern. The reason for this could be due to the average repeatability of the Kinect to estimate the joints' position leading to different values that would alter the initial condition the technique is evaluated. A numerical analysis of the values at which this problem occurs should be taken into account for future developments and a calibration expression should be formulated to correct the reference frames.

Nevertheless, the results obtained are promising for the improvement of the reproduction of virtual anatomical segments orientations closer to the real anatomical orientations. The results present the orientations estimated by six different vector orthogonalization methods and it was concluded that they present different performances depending on the type of movement or anatomical segment. As future developments, it is suggested the development of a parametrized model using the information from the results. By taking a machine learning algorithm, a biomechanical model which takes as input the estimated joints' position and predicts the best orientations values for each anatomical segment and movement could be trained.

Kinect One constitutes a marker-less motion capture system, whose low cost, portability and open-source software development tool, opens a wide range of possibilities for the development of real-world applications. Therefore, it is imperative to understand and improve its precision on the estimation of the joints' position and body segments' orientation in order to obtain more faithfully virtual representations of the real human motion.

V. References

[1] Zhou, H., & Hu, H. (2008). Human motion tracking for rehabilitation—A survey. Biomedical Signal Processing and Control, 3(1), 1-18.

[2] Moeslund, T. B., Hilton, A., & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 104(2), 90-126.

[3] Bonnechère, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L., Moiseev, F., Sholukha, V., Cornelis, J., Rooze, M., & Jan, S. V. S. (2014). Validity and reliability of the Kinect within functional assessment activities: comparison with standard stereophotogrammetry. Gait & Posture, 39(1), 593-598.

[4] Della Croce, U., Leardini, A., Chiari, L., & Cappozzo, A. (2005). Human movement analysis using stereophotogrammetry: Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. Gait & Posture, 21(2), 226-237.

[5] Leardini, A., Chiari, L., Della Croce, U., & Cappozzo, A. (2005). Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation. Gait & Posture, 21(2), 212-225.

[6] Best, R., & Begg, R. (2006). Overview of movement analysis and gait features. Computational intelligence for movement sciences: neural networks and other emerging techniques, 1, 1-69.

[7] Microsoft: Marker-less Sensor Kinect One and Microsoft SDK [Internet] [cited 2016 October 17]. Available from: https://www.microsoft.com/

[8] Obdržálek, Š., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., & Pavel, M. (2012). Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1188-1193.

[9] Kurillo, G., Chen, A., Bajcsy, R., & Han, J. J. (2013). Evaluation of upper extremity reachable workspace using Kinect camera. Technology and Health Care, 21(6), 641-656.

[10] Fern'ndez-Baena, A., Susín, A., & Lligadas, X. (2012). Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference, 656-661.

[11] Clark, R. A., Pua, Y. H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. Gait & Posture, 36(3), 372-377.

[12] Huber, M. E., Seitz, A. L., Leeser, M., & Sternad, D. (2015). Validity and reliability of Kinect skeleton for measuring shoulder joint angles: a feasibility study. Physiotherapy, 101(4), 389-393.

[13] Wu, G., & Cavanagh, P.R. (1995). ISB recommendations for standardization in the reporting of kinematic data. Journal of Biomechanics, 28(10), 1257-61.

[14] Nikravesh, Parviz E. (1988). Computer-aided analysis of mechanical systems. Englewood Cliffs.

[15] Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., D'Lima, D. D., Cristofolini, L., Witte, H., & Schmid, O. (2002). ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine. Journal of Biomechanics, 35(4), 543-548.

[16] Wu, G., Van der Helm, F. C., Veeger, H. D., Makhsous, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A. R., McQuade, K., Wang, X., & Werner, F. W. (2005). ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand. Journal of Biomechanics, 38(5), 981-992.

[17] Lopes, Daniel S., Silva, M. T., & Ambrósio, J. A. (2013). Tangent vectors to a 3-D surface normal: A geometric tool to find orthogonal vectors based on the Householder transformation. Computer-Aided Design, 45(3), 683-694.

[18] LBL: Lisbon Biomechanics Laboratory [Internet] [cited 2016 October 17]. Available from: http://lbl.tecnico.ulisboa.pt/index.htm

[19] INESC-ID: Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento [Internet] [cited 2016 October 17]. Available from: https://www.inesc-id.pt/

[20] Qualisys: company that provides technology for motion capture systems [Internet] [cited 2016 October 17]. Available from: http://www.qualisys.com/

[21] Kirtley, C. (2006). Clinical gait analysis: theory and practice. Elsevier Health Sciences.

[22] Winter, D. A. (2009). Biomechanics and Motor Control of Human Movement (4th Ed.). John Wiley & Sons.

[23] Manal, K., McClay, I., Stanhope, S., Richards, J., & Galinat, B. (2000). Comparison of surface mounted markers and attachment methods in estimating tibial rotations during walking: an in vivo study. Gait & Posture, 11(1), 38-45.

[24] Corti, A., Giancola, S., Mainetti, G., & Sala, R. (2016). A metrological characterization of the Kinect V2 time-of-flight camera. Robotics and Autonomous Systems, 75, 584-594.

[25] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011). Real-time human pose recognition in parts from single depth images. Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition Computer, 1297-1304.

[26] Plantard, P., Auvinet, E., Pierres, A. S. L., & Multon, F. (2015). Pose Estimation with a Kinect for Ergonomic Studies: Evaluation of the Accuracy Using a Virtual Mannequin. Sensors, 15(1), 1785-1803.

[27] Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. Applied Ergonomics, 43(4), 645-649.

[28] Unity Technologies: game engine for the development of video games [Internet] [cited 2016 October 17] Available from: https://unity3-D.com/

[29] MATLAB® - The Language of Technical Computing [Internet] [cited 2016 October 17]. Available from: http://www.mathworks.com/products/matlab/

[30] Skals, S. L., Rasmussen, K. P., Bendtsen, K. M., & Andersen, M. S. (2014). Validation of musculoskeletal models driven by dual Microsoft Kinect Sensor data. Proceedings of the 13th International Symposium on 3-D Analysis of Human Movement (3-D-ahm), 347-350.

[31] Moon, S., Park, Y., Ko, D. W., & Suh, I. H. (2016). Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering. International Journal of Advanced Robotic Systems, 13.