Master Thesis in Sports Technology

A general marker-less motion capture approach for the Microsoft Kinect Sensor v2 using subject-specific articulated models within an iterative closest point algorithm for improved anatomical accuracy.

Mikkel Svindt Gammelgaard

Department of Health Science and Technology, Aalborg University

ARTICLE INFO

Supervisor: Michael Skipper Andersen

Group: 10201
Semester: 4 semester
Submission date: 03-06-2015

ABSTRACT

Joint kinematics obtained by the Microsoft Kinect Sensors are measured based on a simplification of the human anatomy, which don't include anatomically accurate estimations of the joint centers. Consequently the measured joint kinematics are not directly compatible with the anatomically accuracy of musculoskeletal models. Accordingly, the purpose of the present study was to propose a new marker-less tracking approach using the Microsoft Kinect v2 to obtain the 3D movement of a single leg. A subject specific articulated model of the right leg with anatomically correct joint centers was generated using the Microsoft Kinect v2 to scan the anatomical geometry of the segments. The poses of the detailed articulated model was estimated relative to a dynamic depth recording of a 90 degree hip flexion movement, using an iterative closest point algorithm to which a constrained optimization problem was defined. The proposed approach was compared to a marker-based system. Hip flexion/extension, abduction/adduction, external/internal, knee flexion/extension and ankle plantar/dorsi flexion were measured for one subject. The flexion/extension and internal/external rotation for the hip and flexion/extension for the knee showed good agreement with the marker-based results. The hip abduction/adduction and ankle flexion/extension showed poor results compared to the marker-based system.

Key words: Marker-less motion capture, articulated model, joint kinematics, pose estimation, kinematic analysis, Microsoft Kinect, iterative closest point.

Introduction

Motion capture is widely used in various fields from entertainment to biomechanical research. Within entertainment, motion capture is often used to create animations or interact with games. In biomechanical research, motion capture is used for kinematic analysis of motion. Marker based motion capture currently serve as the most common approach when performing motion analysis in the field of biomechanics, despite accommodating several limiting factors (Corazza et al. 2006; Mündermann et al. 2008). Marker-based motion capture systems depend on multiple calibrated cameras carefully placed within a laboratory setting, and the attachment of several markers upon each segment. The attachment of markers is a time-consuming process and the markers can cause discomfort for the subject. Additionally, skin movement relative to the underlying bone can result in inaccurate estimations of joint kinematics (Akbarshahi et al. 2010, Fuller et. al, 1997) . These drawbacks have created an increasing interest in maker-less motion capture systems, which have given rise to several proposed methods (Mündermann et al. 2007; Mündermann et al. 2006; Yang et al. 2014; Moeslund et al. 2006). Marker-less motion capture system are often divided into model-free and model-based approaches. Model-free approaches use structured learning by predefining image input features together with accompanying output poses to model the dependency between observed features and corresponding poses. These approaches, therefore, require training sets, which closely resemble the analyzed motion. Model-based approaches include a 3D representation of the human body with kinematically chained joints. The articulated model is fitted to the observed data, which is either 2D silhouettes or 3D representations of the human motion (Yang et al. 2014) .

Microsoft Kinect Sensors (Microsoft Corp., Redmond, WA, USA) are popular examples of marker-less motion capture devices. The device consists of a depth sensor and a color camera. The Kinect for Windows Software Development Kit (SDK) offers a model-free joint tracking algorithm to estimate joint kinematics of individuals within its field-of-view (Han et al. 2013; Zhang 2012). The tracking algorithm is based on a large training set which include various movements and body shapes (Shotton et al. 2011). The benefits of this approach is that it enables real-time estimations of joint kinematics in a computationally robust and efficient manner. Clark et al. (2015) investigated the Microsoft Kinect v2's ability to measure postural control using the joint tracking algorithm. In comparison with a marker-based system, the study found high validity for the measures of movements in the anterior-posterior direction, but not in the medial-lateral direction. Although the accuracy of the Microsoft Kinect joint estimations are sufficient for various rehabilitation purposes (Fernández-Baena et al. 2012; Mentiplay et al. 2015; Clark et al. 2015; Galna et al. 2014), it does not present the accuracy required within biomechanical research (Yang et al. 2014; Mentiplay et al. 2015; Pfister et al. 2014). Another limitation of this approach is that the segments are modelled as cylinders and does not provide an anatomically accurate representation of joints. Additionally, due to its simplistic overall representation of the human anatomy, the built-in model of the Microsoft Kinect Sensors are typically not directly compatible with more anatomically accurate musculoskeletal models such as those accompanying the AnyBody Modeling System (AnyBody Technology, Aalborg, Denmark)(Damsgaard et al. 2006) or OpenSim (Delp et al. 2007). This creates an issue for the captured motion to be used as input within musculoskeletal models, as the estimated joints within the two models do not coincide and hereby creates a discrepancy between the resulting movement of the musculoskeletal model and the original Kinect measurement (Andersen et al. 2013). Because of the Microsoft Kinect Sensors ability to perform detailed non-expensive 3D reconstruction using a rather simple setup, the device itself has may have promising technological applications for marker-less motion capture (Yang et al. 2014).

Previous studies have used a model-based approach using subject-specific 3D articulated models aligned to a visual hull captured of the movement. The alignment of the articulated model and the visual hull is either accomplished using simulated annealing or an iterative closest point algorithm. These approaches have estimated highly accurate joint kinematics which are comparable to the accuracy of maker-based systems (Mündermann et al. 2007; Corazza et al. 2010). Although marker-based motion capture is not considered a golden standard (Benoit et al. 2006; Yang et al. 2014), most validations of marker-less motion capture systems is done against marker-based systems.

Schmitz et al. (2014) utilized the Microsoft Kinect Sensor to capture the poses of a simple two segmented jig. The Kinect Fusion software (Microsoft Corp., Redmond, WA, USA) was used to build a 3D articulated surface model of the jig. An iterative closest point algorithm was used to align the articulated surface model with a depth recording of each pose of interest. The accuracy of the marker-less motion capture system was compared with both a marker-based system and an inclinometer. The inclinometer was included to serve as a ground truth because of its ability to acquire highly accurate measures of the joint angles. The results showed an agreement in accuracy within 2° or less. Since this study

was performed on a simple two segmented jig, it is difficult to know whether the same accuracy can be achieved for the human body, as the surface geometry and motions are much more complex. The aim of the present study was to propose a general marker-less model-based motion capture approach for one leg kinematics using the Microsoft Kinects Sensor. The intention behind the proposed approach was to obtain an anatomical accuracy compatible with that of musculoskeletal modelling systems like AMS and OpenSim. The general tracking approach formulated in the current article is similar to the tracking approach applied within the AMS, but reformulated to utilize depth information, rather than marker positions, as drivers. To validate the proposed approach it was compared with a marker-based system.

Method

The pipeline of the proposed marker-less motion capture approach can be split in four stages. The first stage is the data acquisition, which consists of multiple 3D scans of the segments and a dynamic depth recording of the moving segments. The 3D scan will be used to calibrate the articulated model and the dynamic recording will be used to estimate the poses of the articulated model. In the second stage, an articulated model is created based on the 3D scan. The articulated model contains each segment and its corresponding joint centers defined in local coordinates. In the third stage, each dynamic depth frame acquired from the recording is processed so that the moving segment is isolated within each frame. Within the fourth stage, a pose estimation is performed using the processed data as input. Lastly, the pose of all the segments within the articulated model are estimated for each frame using an iterative closest point algorithm. Additionally, the joint angles are calculated for each estimated pose. The presented approach was only applied for the right leg, and included the follow segments: pelvis, thigh, shank and foot. The pipeline of the approach is illustrated in Figure 1. All parts of the presented approach was written and run in MATLAB (MATLAB R2016a, The MathWorks, Inc., Natick, Massachusetts, United States).

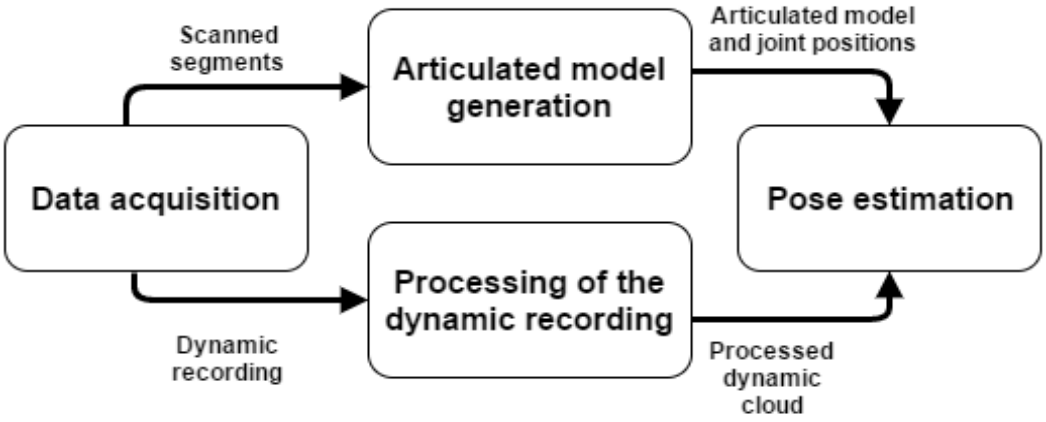

Figure 1. Pipeline of the marker-less motion capture approach. The order and output for each stage of the pipeline are illustrated.

Data Acquisition

Calibration scans

The 3D representation is obtained by scanning the segments using the Kinect Fusion software which is integrated in the Kinect for Windows Software Development Kit 2.0. Kinect Fusion enables the Microsoft Kinect Sensor v2 to be used for object scanning and 3D reconstruction (Newcombe et al. 2011). The voxel resolution of the scan was set to 384^3 voxels and the voxels per meter was set to 512. Before the scan was initiated, anatomical landmarks were marked in order to allow subsequent estimations of the joint centers following common regression equations as explained later (see Figure. 2). During the scan, the subject was asked to stand still in an upright position until the scan was complete. To acquire a sufficient detail in the representation of all segments, the scanning was divided in three separate scans. The first scan was performed from the knee to the navel. The second scan was performed from the floor to the knee. Before the third scan, the subject was asked to sit down with the leg elevated and the calf placed on a raised bench, with the knee extended. The bench was used to apply support at the calf to ensure that the subject was able to hold a steady posture of the foot. In this position, the third scan was performed from the ankle to the sole of the foot. The third scan

was performed to acquire a full representation of the foot. Examples of the scans before processing can be found in Figure. 3.

Dynamic recording

Before the dynamic depth recordings were initiated, the Microsoft Kinect Sensor v2 was placed 1.5 m away, facing the right lateral side of the subject. A 10 second recording of the background without the subject in the field view was acquired before the dynamic recordings to enable background subtraction. A MATLAB GUI was written to acquire and save the recordings. Each frame of a single recording was saved as a 512x424x3 point cloud of the scene. Due to limitations of the used hardware and the implementation of the recorder in MATLAB, the Microsoft Kinect Sensor v2 sampled at approximately 15 Hz during the recordings.

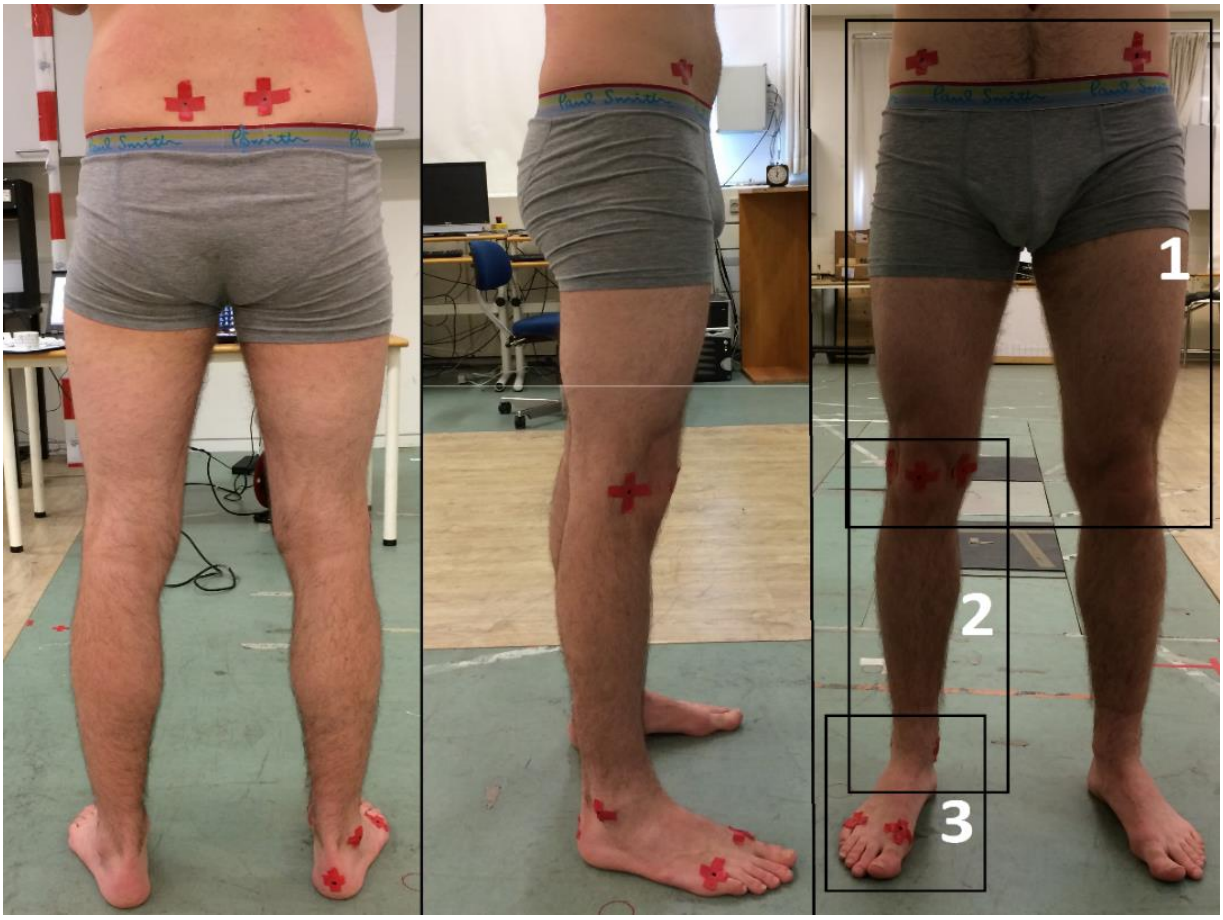

Figure 2. The subject viewed from the back (right), side (center) and front (left) in the scanned position. The anatomical landmarks was marked as red crosses with black dots. The scanned area of the first, second and third scan are illustrated

Articulated model generation

Within the 3D scanned point clouds, each of the anatomical marker positions were identified in the coordinates of the 3D scan. The marker positions was used to estimate the joint centers within each 3D scan. The hip joint center was defined using the regression equation described in Harrington et al. (2007).The knee and ankle joint centers were defined at the midpoint between the lateral and medial markers of the knee and ankle, respectively (Templeton 2002; Wu & Cavanagh 1995). The segments within the 3D scan were identified using the anatomical markers and estimated joint centers. A plane was defined for each joint to identify the transition of the connecting segments. The pelvis-to-thigh transition plane was defined by a vector from a point, which was manually selected in perineum region, passing through the hip joint center and a second vector from the midpoint between the anterior hip markers going through the midpoint of the posterior hip markers. The thigh-to-shank transition plane was defined by a vector from the lateral knee marker passing through the knee joint center and a second vector from the knee joint center passing through the hip joint center. Similarly, the shank-to-foot transition plane was defined by a vector from the lateral ankle marker passing through the ankle joint center and a second vector from the ankle joint center passing through the knee joint center. Segmentation of the 3D scan was done using the transition planes to designate each segments associated points, within the 3D scan. The designated points were then defined in the local coordinates of the associated segment along with the corresponding

joint centers. For the pelvis, thigh and shank, the origin of the coordinate systems were defined at the midpoint between the corresponding joint centers of the segment. The origin of the foot coordinate system was defined at the midpoint between the ankle joint center and the midpoint between the metatarsal markers. Points located closer than 0.05 m to the pelvis-to-thigh and thigh-to-shank transition plane were discarded to remove points near the joints. For the shank-to-foot plane only points closer than 0.03 m was discarded. The points near the joints were discarded to reduce the interference between the segments in the pose estimation when the joints were flexed or extended. Lastly, the point cloud was down-sampled by applying a uniform 3D grid, in which the position of points inside each grid was averaged. The grid size set to 0.002 m^3. The articulated model is illustrated in Figure 4a.

Figure 3. The three different 3D scans of a subject obtained using Kinect Fusion.

Processing of the dynamic recording

To isolate the moving segments within the dynamic depth recording, the background recording of the scene was used for background subtraction. The background subtraction was performed by defining a threshold for each voxel within the background scene, the thresholds were defined by subtracting one centimeter from the smallest depth distance found within all frames of the background recording. With the smallest depth distance being closest to the camera. The threshold was applied for each frame within the dynamic recording. Voxels were excluded if the depth distance of the dynamic recording was above its corresponding threshold. Additionally, a filter was applied to remove any remaining noise within the cloud. The filter applied used two input parameters, the first was a defined number of nearest neighbors' *k*, to which the average distance are calculated. The second input is a threshold for the average distance to a point's *k*-nearest neighbors, if the average distance for any point is above the threshold, it is defined as an outlier (Rusu et al. 2007). The number of nearest neighbors and threshold value was defined as 150 points and 0.01 m, respectively. Lastly, the point clouds were down-sampled using the same method as described for the articulated model. Each frame of the processed depth recording was stored as a 3D point cloud. A processed frame of a dynamic recording are illustrated in Figure 4b.

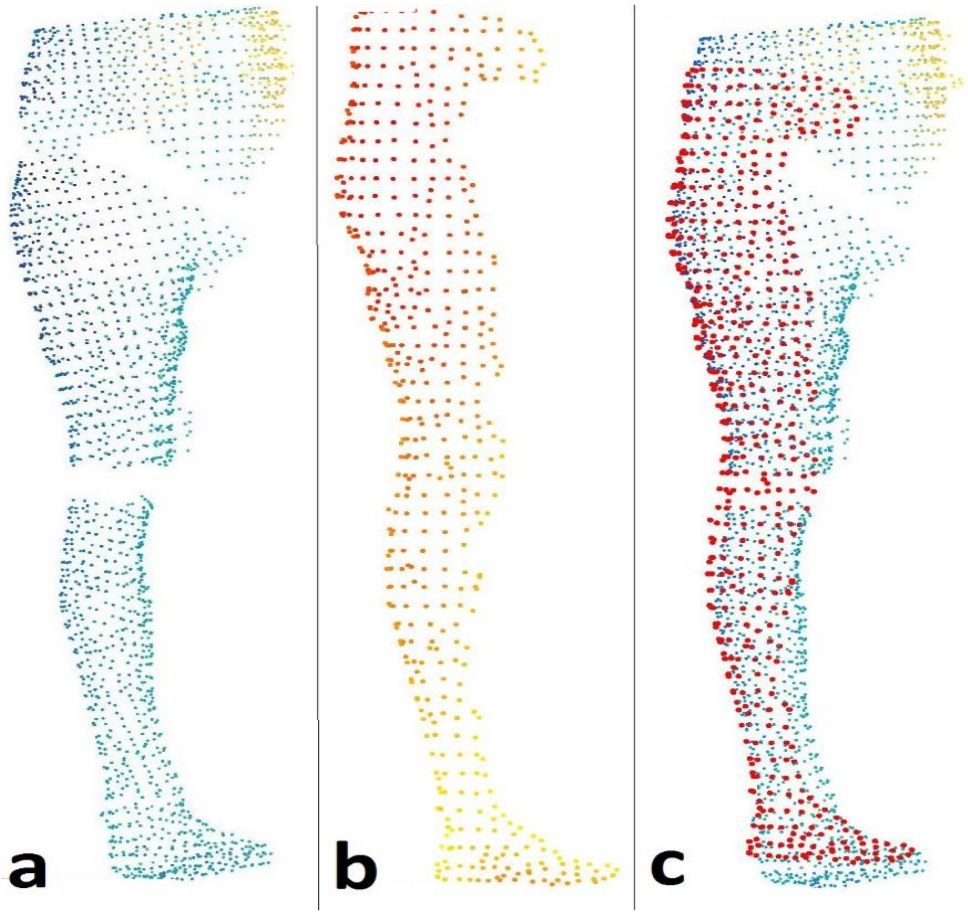

Figure 4. The articulated model of the 3D scan with points around the joints removed (a). The processed dynamic cloud (b). The pose estimation performed for the articulated model at normal stance. (c).

Pose estimation

The purpose of the kinematic analysis was to estimate the translation and rotation for all segments within the articulated model, which gave the best fit for each frame within the recording. The translation and rotation coordinates of each segment are denoted as:

$$q_j = [q_1\ q_2 \dots q_7],$$

where q_j is the vector coordinates of the *j*th segment. The first three indices of q_j contains the global position coordinate (x, y, z) and the last four contains the Euler parameters (e_0, e_1, e_2, e_3) governing the rotation of the segment. The pose of best fit for each segment was found using an iterative closest point algorithm to which a constrained optimization problem was defined. The optimization problem consisted of two equations, in which one had to be solved exactly while the other only had to be solved as well as possible (Andersen et al. 2009):

$$\min_q G\big(\Psi(q,t)\big)$$

$$s.t \quad \Phi(q,t) = 0,$$

Ψ represents the difference between each point in the dynamic cloud and its closest point within the articulated model. Φ contains the constraints defined for the Euler parameters to ensure that they have unit length (Nikravesh 1988), as well as the constraints defined for each joint. q, is the concatenated vector of each q_j. The objective function G was formulated as a least square:

$$G = \frac{1}{2}\Psi^T\Psi,$$

where Ψ is the vector difference which dimension corresponds to the number of points within the dynamic cloud. To find the closest point in the articulated model, each point M_i in the dynamic cloud was transformed into each segments

local coordinate system. Thereby defining M_i' in each segments local coordinate system before searching for its closest point in the related segment. The closest point s_j' for each segment was found by a nearest neighbor search using k-d trees (Muja & Lowe 2009). Iteratively, the search was performed for each point M_i in the dynamic cloud, thereby assigning each M_i with an associated nearest segment, denoted j. Ψ_i was calculated as:

$$\Psi_i = r_{q_j} + s_j' A_{q_j}^T - M_i,$$

where r_{q_j} and $A_{q_j}^T$ denote the global position coordinate and the rotation matrix related to the Euler parameters of q_j, respectively. s_j', denotes the closest point in the local coordinates of segment j. M_i, denotes the global coordinate of the ith point in the dynamic cloud. Each Ψ_i was weighted relative to the amount of points within the associated segment. If M_i was within its associated segments transition plane, Ψ_i would be calculated as: $\Psi_i = 0$.

Within the joint constraints the ankle and knee joints were defined as hinged joints and the hip joint was defined as a spherical joint (Andersen et al. 2009). To acquire a realistic starting guess for q, a custom MATLAB GUI was written to manually fit the pose of each segment with the initial dynamic cloud. The system was solved using the Complex optimization method (Manetsch 1990). This method uses multiple starting points for the design parameters (q), which are randomly distributed within a defined interval around the starting guess. The number of staring points were set at 56, which corresponded to two times the number of design parameters. For each iteration the worst point are replaced with a new point. The new point are defined by reflecting the worst point through the centroid of the other points. The process was continued until all points had converged within the criteria. The criteria was set to 0.005 for the position coordinates and 0.005 for the Euler parameters. To ensure that the constraints of Φ was met with certainty a weight factor of 100 was applied. The optimized translations and rotations was applied to the articulated model for each frame to acquire the dynamical orientation. The relative motion of the segments was calculated using Cardan angles (Grood & Suntay 1983) together with the ISB recommended axis conventions.

Validation

To validate the proposed marker-less motion capture approach, simultaneous recordings were performed using a marker-based system. The system consisted of eight infrared cameras from the Oqus 3+ series combined with Qualisys Track Manager v. 2.9 (Qualisys, Sweden). The marker based system was set to sample at 100 Hz. One healthy male subject (age: 26 years, height: 183 cm, weight: 76 kg) participated within the study. Before the recording was initiated 30 retroreflective markers were placed on the lower body of the subject. The subject was asked to stand in an upright position as the recordings initiated, and then slowly flex the hip to 90 degrees with the knee bend and then return to the initial position. A total of five repetitions were recorded. The three trials in which the subjects' maximum hip flexion was closest to 90 degrees were chosen for further analysis. Furthermore, one standing reference was captured using the marker-based system. Kinematic analysis of the data obtained using the marker based system was performed using the anatomical landmark model (Lund et al. 2015) in the AnyBody Modelling System v. 6.0 which uses the same calculations and conventions for the joint angles as applied within the proposed approach. This model uses the markers of the standing reference to generate a stick figure model and the joint parameters. A musculoskeletal cadaver-based model was scaled to fit subject-specific stick figure model and the joint parameters. Furthermore the kinematics of the dynamic movement was computed by applying the stick figure model to the markers. Finally, the inverse dynamic analysis was computed using the computed kinematics and subject-specific musculoskeletal model.

Data analysis

The root mean square (RMS), standard deviation of the difference and mean difference for the hip flexion/extension, abduction/adduction, internal/external rotation, knee flexion/extension and ankle plantar/dorsi flexion was calculated for the included trials. Furthermore the mean difference in peak values was calculated for all measured angles to evaluate the relative accuracy of the two systems. Additionally, for all of the measured angles the Pearson correlation coefficient (ρ) was calculated between the two systems to measure the relative dependence between the systems. All measures are calculated from a defined start and end of movement. The start and end point was defined when the knee flexion was above and below 10°, respectively. The data from both systems were resampled to contain equal number of samples. To smoothen the calculated angles measured by the marker-less system, the angles were filtered using a 2 order Butterworth filter. The cutoff frequency of the filter was set at 2.

Results

The following only include one single trial, as the proposed optimization algorithm was not able to find solutions for the other two included trials. The results presented therefore only include the one trial which were successfully analyzed.

The RMS, STD. of difference, peak difference and mean difference for the hip, knee and ankle angles are presented in Table 1. The respective ρ-values for the different angles can be found in the bottom left corner of the graphs in Figure 5 and 6. The RMS for the hip angles deviated from 1.4° to 9.4°, the abduction/adduction of the hip showed the highest RMS and the lowest RMS was found in the internal/external rotation of the hip. The STD. of difference for the hip flexion/extension, abduction/adduction and internal/external rotation were measured at 5.3°, 2.4° and 1.5°, respectively. The mean difference for the hip flexion/extension, abduction/adduction and internal/external rotation were measured at 6.1°, 3.0° and -0.5°, respectively. The peak difference for the hip flexion/extension, abduction/addiction and internal/external rotation was measured at 10.1°, 8.9° and -2.9°, respectively. The graphs in Figure 5 display strong agreement between the marker-less and marker-based systems for the hip flexion/extension and internal/external rotation angles, the values of ρ for these angles similarly display strong correlation (ρ = 0.988, ρ = 0.808). The hip abduction/adduction of the marker-less system displayed poor correlation (ρ = 0.166).

Table 1. The measured mean difference, standard deviation (STD), peak angle difference and root mean square deviation (RMSD) for the hip, knee and ankle angles for the included trials.

Joint angle	RMS (degrees)	STD. of difference (degrees)	Peak difference (degrees)	Mean difference (degrees)
Hip flexion/extension	8.5	5.3	10.1	6.1
Hip abduction/adduction	9.4	2.4	8.9	3.0
Hip internal/external rotation	1.4	1.5	-2.9	-0.5
Knee flexion/extension	3.4	3.2	-4.0	-1.0
Ankle plantar/dorsi flexion	6.8	5.3	17.5	3.7

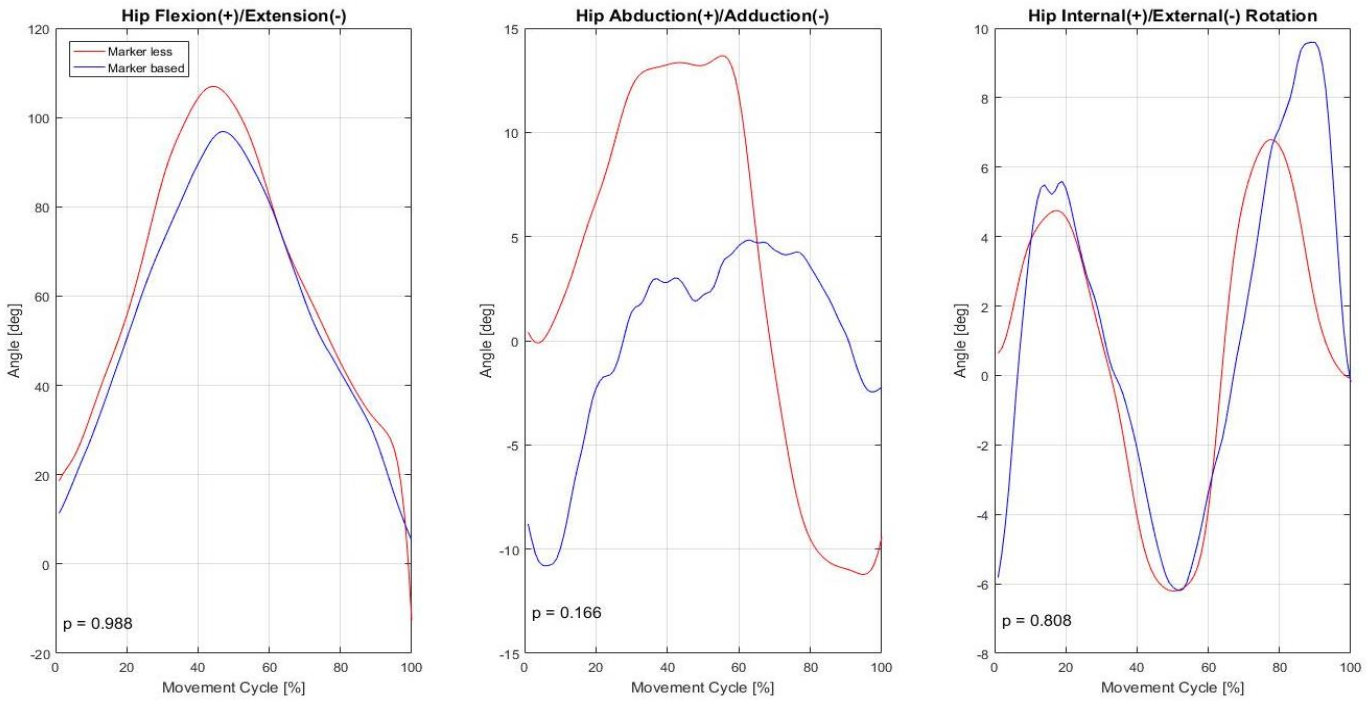

Figure 5. Hip flexion/extension (left), abduction/adduction (center) and internal/external rotation (right) for the marker-based system (blue) and the marker-less system (red). The angles are shown for 100% of the movement cycle, which is defined from above to below 10 degrees of the knee flexion. The correlation coefficient (ρ) are shown in the bottom left corner for each measure.

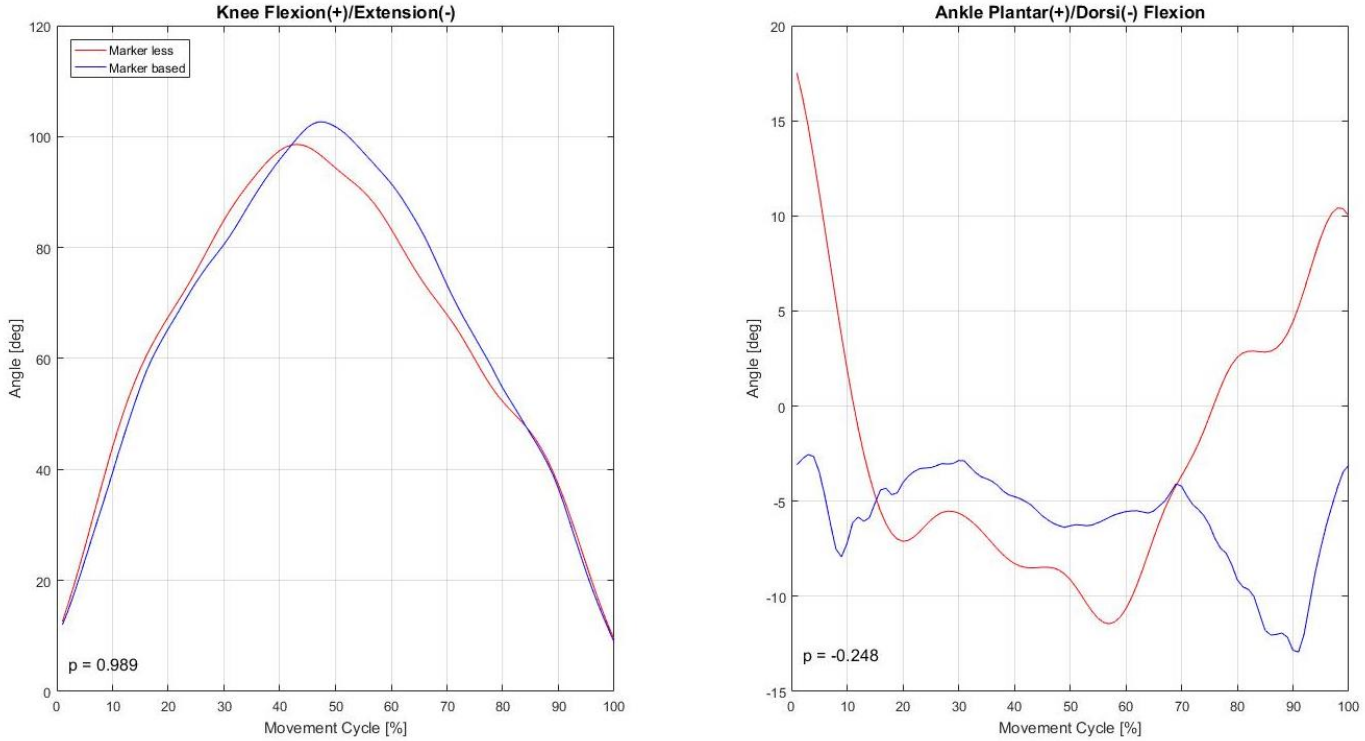

Figure 6. Knee flexion/extension (left) and ankle plantar/dorsi flexion (right) for the marker-based system (blue) and marker-less system (red). The angles are shown for 100% of the movement cycle, which is defined from above to below 10 degrees of the knee flexion. The correlation coefficient (ρ) are shown in the bottom left corner for both measures.

The knee flexion/extension displayed strong correlation for the knee flexion ($\rho = 0.989$) which is also illutstrated in Figure 6. The RMS for the knee was 3.4°. The mean difference, STD. of difference and peak difference were measured at -1.0°, 3.2° and -4.0°, respectively. The difference in ankle angles between the two systems was much higher than the knee angles as the RMS was measured at 6.8°. The mean difference, STD of difference, peak difference were measured at 3.7°, 5.3°, 17.5°, respectively. The correlation of the ankle further displayed the poor agreement between the systems ($\rho = -0.248$). This is also illustraed in the graph for the ankle displayed in Figure 6.

Discussion

The presented study sought to propose a marker-less motion capture approach using a single Microsoft Kinect v2, in which the dynamic depth recording was tracked using a model-based iterative closest point algorithm to track single leg kinematics. The presented approach was tested for a 90 degree hip flexion movement and validated against a marker-based system. Although the proposed marker-less tracking algorithm failed to analyze two of the three included trials, the approach revealed good agreement for the hip flexion/extension, internal/external rotation and knee flexion/extension. However the proposed approach was not able to accurately measure the hip abduction/adduction and ankle flexion. Although parts of the results is encouraging, the failure to analyze two out of three trials does indicate that the marker-less approach as presented in this article is in need of modifications.

Using the Kinect Fusion software to obtain the articulated model appeared to give a decent estimation of the anatomical geometry of the subject. By marking the anatomical landmarks, it was possible to estimate the joint centers in articulated model by applying the same methods as used in the AnyBody Modelling System. Previous studies measuring kinematics using the Microsoft Kinect Sensors (Andersen et al. 2013; Mentiplay et al. 2015) have used model free tracking algorithms (Shotton et al. 2011), which does not estimate the joint centers in the same way as the AnyBody Modelling System. Although, the geometry are well estimated for the articulated model, it was necessary to perform several carefully performed scans to obtain usable representations. Additionally, the hardware used in the current study was not able to scan larger areas with sufficient resolution, consequently separate scans was obtained for the pelvis-thigh and shank-foot to ensure the details of the scan was sufficient. Increasing the voxels per meter would allow a single scan to include the whole leg geometry. This was not possible using the hardware within the current study as the Kinect Fusion algorithm are very GPU intensive. A single scan would ensure that all anatomical marker positions are estimated correctly relative to the joining segments. Currently the markers around the knee are located in each of the two scans, which might inflict discrepancies between the joint center positions estimated for the knee and shank. Similar discrepancies may also occur for the shank and foot segment.

The mean difference for the flexion/extension of the hip and knee was 6.1° and -1.0°, respectively. For comparison Fernández-Baena et al. (2012) reported mean differences above 5.5° for hip flexion and 6.8° for knee flexion using the Microsoft Kinect v1 joint tracking algorithm in comparison to a marker-based system. Marker-less approaches not using the Microsoft Kinects have reported mean differences for hip and knee flexion in running and gait between -0.4° - 2.0° and 1.5° - 2.8°, respectively (Corazza et al. 2006; Sandau et al. 2014). Relative comparison to these previous studies cannot be made directly as the current study only included a single trial, but it could indicate that the proposed approach do provide accurate measures of the knee flexion/extension. Although, further investigation regarding the repeatability of the proposed approach are required, as the current results are only based on a single trial. The results for the hip abduction/adduction and ankle flexion reveal large deviations from results obtained with the marker-based system, which indicate that the system did not provide very reliable estimates of the these angles. A possible explanation for these results could be the reduced sampling rate of the Microsoft Kinect v2 within the current study. If the movement in the dynamic recording between each frame are too large, it demands a larger search interval for the optimization and it becomes increasingly difficult for the iterative closest point algorithm to find the optimal pose. A higher sample rate would decrease the relative movement between each frame in the dynamic recording, which may simplify the pose estimation process. This would of course also increase the computation time of each recording. The current computation time for each frame using the parameters described in the method is approximately 20 minutes per frame on an Intel® Core™ i5 2.60 GHz CPU. The computation time could be greatly decreased by rewriting the current approach in a programming language that is not depending on online interpretation of code such as C or Fortran.

The fact that only one of three trials was solved indicate that the optimization algorithm applied did not behave very robust relative to the obtained input. Whether this can solely be explained by the recorded input not containing sufficient information is not known and several processes in the proposed approach could have contributed to this lack of robustness. It was observed that specific frames in the dynamic recording were missing parts of the segments, which could explain why the optimization algorithm failed. These parts may incorrectly have been excluded in the dynamic processing. Additionally, positioning the Microsoft Kinect v2 sensor facing the side of the subject, might not provide sufficient information about the anatomical geometry within the dynamic recording to accurately estimate the pose of the articulated model. The dependency between the obtained anatomical geometry in the dynamic recording and the placement of the sensor relative to the subject was not investigated within the current study. Therefore, it is possible that an improved accuracy, could be achieved by placing the sensor at a different position relative to the subject. Positioning the camera to include more of the upper body in the dynamic recording, may have contributed to a better estimation of the pelvis position. This would possibly give more accurate measures of the hip angles. Another way to increase the anatomical geometry obtained in the dynamic recording would be to use multiple Microsoft Kinect v2 sensors recording simultaneously. Two Microsoft Kinect v1 sensors have previously been used for motion capture for musculoskeletal models using the iPi motion capture software (iPi Soft, Moscow, Russia) (Andersen et al., 2013), which is able to orientate the simultaneously recorded 3D point clouds to the same global coordinates. Applying multiple sensors would provide much more anatomical geometry for the segments within each frame of the dynamic recording. Using a single sensor and the current setup, each dynamic frame mainly contains information about the geometry for the lateral side of the segments, including views from the front and medial sides of the segments would give additional information regarding the segments orientation and rotations in the frontal and axial planes.

Compared to the agreement found for the flexion/extension angles for the hip and knee, the ankle showed very little agreement compared to the marker-based system. Although point-to-point difference between the articulated model and dynamic recording was weighted relative to the amount of points within the corresponding segment, the proposed pose estimation algorithm was not able to find accurate poses for the articulated foot segment. The articulated model included the whole geometry of the foot by scanning the foot in an elevated position. Generating the articulated foot using this method gave a more accurate representation of the whole foot and it also removed the floor from the articulated model. The same detail was not measured of the foot in the dynamic recording and the background subtraction process further decreased the detail because it removed the lower part of the foot near the floor as it was not able to distinguish between the foot and the floor. This creates a large difference in detail between the articulated and dynamic representations of the foot. This could explain the large discrepancies observed for the ankle flexion, as areas of the foot included in the articulated representation of foot, but not in the dynamic, could be defined as closest points thereby offering several solutions leading the pose estimation towards inaccurate translation and rotation of the foot. Both the hip abduction/adduction and the ankle flexion were largely over estimated relative to the amount of rotation measured in these planes by the marker-based system. The peak difference was measured at 8.9° for the hip abduction/adduction and 17.5° for the ankle flexion. In perspective, the angles measured by the marker based system ranged from -10 to 5 and -2.5 to -13 for the hip abduction/adduction and ankle flexion, respectively. As illustrated in Figure 5 and 6, the measured angles when the movement initiates are very different form the initial angles measured by the marker-based system.

Although, this could be explained by limited information in the dynamic recording as previously described, it could also be explained by miscalculations of the angles or incorrect definitions of the coordinate systems in the optimization algorithm. Furthermore, the initial guess for the translation and rotation of the segments are important, it is also possible that the initial guess was not accurate enough to give a correct initial orientation of the foot and the hip. If the initial solution is incorrect it is likely, that the following estimated poses would be as well.

The pose estimation rely on the parameters defined for the constrained optimization. Structured analysis of how altering the values of the optimization parameters effect the estimated translation and rotation of the articulated model may have provided more accurate results, as the current values of the parameters might not be optimal. This may also give a better understanding of the behavior of the optimization and which part of the proposed approach that needs to be changed to obtain better results. The discrepancies measured within this study clearly shows that the accuracy of the proposed approach is not compatible with that of musculoskeletal models. The current kinematic inaccuracies would only be amplified if the obtained kinematics would be used in a musculoskeletal model where not only position measurements but also velocities and accelerations are required. It is difficult to specifically predict which part of the proposed approach that limited the ability to accurately estimate the kinematics of the dynamic recording and should be investigated further.

As described in the previous sections the proposed marker-less system consisted of several limitations: (1) the dynamic recording appeared to obtain sufficient information for the thigh and shank, but not for the pelvis and foot. As previously described acquisition using multiple Microsoft Kinect v2 sensors would considerably increase the representation detail of the dynamic recording. Additionally, increasing the sampling rate from 15 Hz to the 30 Hz, which the Microsoft Kinect v2 is capable of, might also ease the pose estimation process. (2) It was also observed that some of the excluded frames was incomplete, and missing parts of the recorded segments. Whether this is caused by faults in the processing of the dynamic recording or in the acquisition stage should be investigated further, as such inconsistency within the dynamic recording greatly affect the pose estimation. (3) It is not known if the optimization parameters used to obtain the current results are optimal for finding accurate solutions. Further investigating into the behavior of the algorithm could provide more optimal values for the parameters within the algorithm, which could make it more robust in terms of finding correct solutions.

Conclusion

The study presented a new general model-based tracking approach for the Microsoft Kinect v2, using an iterative closest point algorithm with a defined constrained optimization problem. The algorithm was able to estimate the poses of the articulated model, relative to the dynamic depth recording obtained using the Microsoft Kinect v2, but only for a single recording, as the algorithm was not able to sufficiently solve the other recorded trials. The proposed approach displayed strong correlation between the marker-based and marker-less system for hip flexion/extension, internal/external rotations and knee flexion/extension. The hip internal/external rotation and knee flexion/extension also estimated the peak angles with good accuracy. However, the measured hip abduction/adduction and ankle flexion showed poor correlation and also showed large deviations in peak values relative to the amount of abduction/adduction and flexion measured by the marker-based system. Overall the results indicate that the proposed approach is in need of modifications to provide sufficient accuracy for all estimated joint angles. Further investigation is needed to address which specific parts of the current approach that should be modified to achieve better overall accuracy.

References

Akbarshahi, M. et al., 2010. Non-invasive assessment of soft-tissue artifact and its effect on knee joint kinematics during functional activity. *Journal of Biomechanics*, 43(7), pp.1292–1301.

Andersen, M.S. et al., 2013. Full-body Musculoskeletal Modeling Using Dual Microsoft Kinect Sensors and the Anybody Modeling System. *14th International Symposium on Computer Simulation in Biomechanics*, pp.1–2.

Andersen, M.S., Damsgaard, M. & Rasmussen, J., 2009. Kinematic analysis of over-determinate biomechanical systems. *Computer methods in biomechanics and biomedical engineering*, 12(4), pp.371–384.

Benoit, D.L. et al., 2006. Effect of skin movement artifact on knee kinematics during gait and cutting

motions measured in vivo. *Gait and Posture*, 24(2), pp.152–164.

Clark, R.A. et al., 2015. Reliability and concurrent validity of the Microsoft Kinect V2 for assessment of standing balance and postural control. *Gait & Posture*, 42(2), pp.210–213. Available at: http://www.sciencedirect.com/science/article/pii/S0966636215000740.

Corazza, S. et al., 2006. A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. *Annals of Biomedical Engineering*, 34(6), pp.1019–1029.

Corazza, S. et al., 2010. Markerless motion capture through visual hull, articulated ICP and subject specific model generation. *International Journal of Computer Vision*, 87(1-2), pp.156–169.

Damsgaard, M. et al., 2006. Analysis of musculoskeletal systems in the AnyBody Modeling System. *Simulation Modelling Practice and Theory*, 14(8), pp.1100–1111.

Fernández-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. *Proceedings of the 2012 4th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2012*, pp.656–661.

Galna, B. et al., 2014. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait & posture*, 39(4), pp.1062–1068.

Grood, E.S. & Suntay, W.J., 1983. A joint coordinate system for the clinical description of three-dimensional motions: applications to the knee. *Journal of biomechanical engieering*, 105, pp.136–144.

Han, J. et al., 2013. Enhanced Computer Vision with\nMicrosoft Kinect Sensor: A Review. *Ieee Transactions on Cybernetics*, 43(5), pp.1–17.

Harrington, M.E. et al., 2007. Prediction of the hip joint centre in adults, children, and patients with cerebral palsy based on magnetic resonance imaging. *Journal of Biomechanics*, 40(3), pp.595–602.

Lund, M.E. et al., 2015. Scaling of musculoskeletal models from static and dynamic trials. *International Biomechanics*, 2(1), pp.1–11. Available at: http://www.tandfonline.com/doi/full/10.1080/23335432.2014.993706#abstract.

Manetsch, T.J., 1990. Toward efficient global optimization in large dynamic systems-the\nadaptive complex method. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(1), pp.257–261.

Mentiplay, B.F. et al., 2015. Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables. *Journal of Biomechanics*, 48(10), pp.2166–2170.

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(2-3 SPEC. ISS.), pp.90–126.

Muja, M. & Lowe, D.G., 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *International Conference on Computer Vision Theory and Applications (VISAPP '09)*, pp.1–10. Available at: papers2://publication/uuid/3C5A483A-ADCA-4121-A768-8E31BB293A4D.

Mündermann, L., Corazza, S. & Andriacchi, T.P., 2007. Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Mündermann, L., Corazza, S. & Andriacchi, T.P., 2008. Markerless Motion Capture for Biomechanical Applications. *Human Motion*, (May), pp.377–398.

Mündermann, L., Corazza, S. & Andriacchi, T.P., 2006. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of neuroengineering and rehabilitation*, 3, p.6.

Newcombe, R. a et al., 2011. KinectFusion: Real-Time Dense Surface Mapping and Tracking. *IEEE International Symposium on Mixed and Augmented Reality*, pp.127–136. Available at:

http://homes.cs.washington.edu/~newcombe/papers/newcombe_etal_ismar2011.pdf.

Nikravesh, P.E., 1988. *Computer-aided Analysis of Mechanical Systems*, Upper Saddle River, NJ: Prentice-Hall International, Inc.

Pfister, A. et al., 2014. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of medical engineering & technology*, 1902(June), pp.1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24878252.

Rusu, R.B. et al., 2007. Towards 3D Object Maps for Autonomous Household Robots. *Iros '07*, pp.3191–3198. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4399309.

Sandau, M. et al., 2014. Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane. *Medical Engineering and Physics*, 36(9), pp.1168–1175.

Schmitz, A. et al., 2014. Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. *Journal of Biomechanics*, 47(2), pp.587–591.

Shotton, J. et al., 2011. Real-time human pose recognition in parts from single depth images. *Cvpr*, pp.1297–1304.

Delp,. SL et al., 2007. OpenSim: Open source to create and analyze dynamic simulations of movement. *IEEE transactions on bio-medical engineering*, 54(11), pp.1940–1950.

Templeton, A.R., 2002. Letter to the Editor. *Genetics*, 475(May), pp.473–475.

Wu, G. & Cavanagh, P.R., 1995. ISB Recommendations in the Reporting for Standardization of Kinematic Data. *Journal of Biomechanics*, 28(10), pp.1257–1261.

Yang, S.X.M. et al., 2014. Markerless motion capture systems for tracking of persons in forensic biomechanics: an overview. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2(1), pp.46–65. Available at: http://dx.doi.org/10.1080/21681163.2013.834800\nhttp://www.tandfonline.com/doi/abs/10.1080/21681163.2013.834800.

Zhang, Z., 2012. Microsoft Kinect Sensor and Its Effect. *MultiMedia, IEEE*, 19(2), pp.4–10.