# Kinematic and Kinetic Validation of an Improved Depth Camera Motion Assessment System Using Rigid Bodies

Robert Peter Matthew [ID], Sarah Seko [ID], Ruzena Bajcsy, and Jeffrey Lotz [ID]

***Abstract*—The study of joint kinematics and dynamics has broad clinical applications, including the identification of pathological motions or compensation strategies and the analysis of dynamic stability. High-end motion capture systems, however, are expensive and require dedicated camera spaces with lengthy setup and data processing commitments. Depth cameras, such as the Microsoft Kinect, provide an inexpensive, marker-free alternative at the sacrifice of joint-position accuracy. In this work, we present a fast framework for adding biomechanical constraints to the joint estimates provided by a depth camera system. We also present a new model for the lower lumbar joint angle. We validate key joint position, angle, and velocity measurements against a gold standard active motion-capture system on ten healthy subjects performing sit to stand (STS). Our method showed significant improvement in mean absolute error and intraclass correlation coefficients for the recovered joint angles and position-based metrics. These improvements suggest that depth cameras can provide an accurate and clinically viable method of rapidly assessing the kinematics and kinetics of the STS action, providing data for further analysis using biomechanical or machine learning methods.***

***Index Terms*—Depth camera, rigid-body model, kinematics, sit-to-stand, biomechanics, lower lumbar angle.**

## I. INTRODUCTION

**M**USCULOSKELETAL disorders of the spine and knee lead to approximately 39 million visits to clinical care facilities each year in the United States [1]. Despite the prevalence of these conditions, there remains a lack of scalable, accessible, and quantitative assessments for whole body biomechanics in clinic. The current clinical gold standard for documenting functional spine impairment is the measurement of Cobb angles in flexion and extension [2], [3], or the Sagittal Vertical Axis (SVA) from radiographs [4]. Such radiographs are inexpensive and offer a precise measurement of vertebral range of motion, but they only assess static postures. During daily functional activities such as sit-to-stand (STS), the strategy used to stand can vary [5]–[7], potentially changing the loads experienced by the joints. This results in both inconsistencies in patient care throughout the recovery process and challenges in understanding the relationship between static observations and functional abilities.

Full-body motion analysis can provide insight into pathological motions and compensation strategies. This analysis is performed in biomechanics labs using gold-standard techniques such as motion capture, force platforms, and surface electromyography. This data can be processed using full-body biomechanics software such as Anybody [8] or OpenSIM [9]. While these systems are a staple in obtaining high resolution kinematic, force, and muscular measurements, their application to regular clinical practice is limited by the time required to setup these measurements, the cost of the equipment, required expertise, and the need for a dedicated motion-capture space.

This has resulted in a dichotomy in analysis, with patients assessed with static measures focused at a particular body segment, while biomechanical labs are able to track and analyse the dynamic motion of the whole-body. Some researchers have explored the used of specialised wearable sensing systems for tracking spine function. Marras developed an exoskeletal tracking system for the lumbar spine to identify motions during occupational tasks, and to identify differences in individuals with low back pain [10], [11]. This system was shown to provide a quantitative kinematic measure of dysfunction based on a specific set of flexion tasks. Taylor and Consmüller developed a system for non-invasive back measurement using flexible strain gauges to measure the curvature of the spine [12], [13]. This system was shown to provide a reliable quantitative assessment of spine shape and range of motion when compared to X-ray. While these systems have been shown to provide good estimates of spine motion and can discriminate between pain and asymptomatic subjects, as they only track spine motion, they are not able to assess changes in full-body motion.
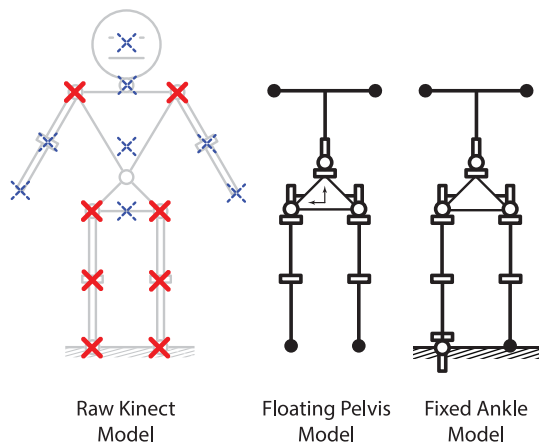
Fig. 1. The three skeletal models used in this paper. Left: Raw Kinect Skeleton. The joint centres obtained from the Kinect are shown as crosses. The markers which are not used in this work are shown in the dashed blue. A cartoon outline of a subject is shown for reference. Centre: Model I: Floating Rigid-body Model. The pelvis is defined as the base link, with three serial chain branches. The sequence of revolute joints are shown as cylinders. Right: Model II: Fixed-ankle Rigid-body Model. The right ankle is fixed to the ground and used as the base link. The revolute joint sequence and segment lengths are the same as in the centre figure.

Depth cameras such as the Microsoft Kinect have been used as a marker-less method for assessing function. Unlike the prior motion capture strategies, no hardware (markers, sensors etc.) needs to be attached to the subject. This allows for rapid testing and simplifies clinical deployment. One of the disadvantages of the use of depth cameras is the method used to identify subject landmarks. As no markers are placed on the subject, the location of a subject's joint centres (Fig. 1) relies on machine learning to label the pixels corresponding to each body segment. The intersection between body segments is then taken to be the estimated joint location [14]. This form of joint centre data from a depth camera is not unique to the Kinect; alternative depth camera sensors (Orbec, Intel RealSense, VicoVR, Depthsense, PMD, SIC), as well as skeletal tracking systems (Nuitrack, OpenNI) are commercially available. As there is no underlying rigid-body model, the estimated joint centres may be biologically inconsistent. This can lead to errors at the ankle, knee, and hip which complicate the use of depth sensors for later analysis [15]. Researchers have found that retro-reflective markers could be used to supplement the recovery process [16]. The addition of these markers adds to the experiment setup time, and sensitivity to the accuracy of marker placement.

An important distinction between this work and the work performed in the computer vision community is the underlying assumptions and goals of the final system. We develop a tool for rapid clinical assessment by applying a biomechanically realistic model to impose constraints on unconstrained estimates of joint position for a controlled task and environment. In contrast, the problem tackled by a number of these other works are the estimation of human poses across a wide range of tasks while being robust to real-world situations and environments [17].

Two approaches are generally taken when performing pose estimation: creating a skeletal model with a prior on the

associated surface geometry, or the generation of a direct map between camera inputs and pose using machine learning. Pavlakos [18] uses convolutional neural networks to estimate the likelihood that a voxel contains a joint. This method resulted in an average 3D joint error of 9.6 cm for the Human3.6M sitting down motion and an average marker reconstruction error of 5 cm. This outperforms a number of other deep learning methods [19], [20] yet still highlights the inherent challenges in joint estimation, particularly in self-occluding tasks such as sitting. This is consistent with the work by Mehta [21] who adopted a similar approach at the pixel level providing a real time (30 fps) system, but with a mean joint position error of 14–15 cm for the sit-down task. While these methods offer a promising method for versatile estimation of human motion, the current joint estimation error is high relative to the surface fitting methods.

Surface fitting methods usually use a simplified approximation of human shape, consisting of scaled cylinders or ellipsoids that are adjusted to a subjects body morphology. This simplified model is then used to estimate pose by relating these volumes to camera depth data. Recent advances have involved the use of Gaussian models to approximate body shape [22], with Ding [23] developing a method that can estimate joint centre position at 20 fps with an associated position error of 3.5 cm. Shuai [24] used spherical harmonic decomposition rather than Gaussians to track subjects with multiple depth cameras. The resulting model exhibited low marker re-projection error, though this error increased in actions with self occlusion such as sitting. Zhang [14] used a full-body skinned mesh model in conjunction with multiple depth cameras and force sensing shoes to estimate kinematic and dynamic state. The resulting system was slow, but accurate with a mean joint error of 3.8 cm at 6 fps. Unfortunately no results were published for any sitting or standing actions, but the authors do state the the system performance did decrease on self-occluding activities. The lower errors and potential for these methods to run in real time suggests that these methods may be suitable for clinical use, but the need for initial calibration of the shape model by performing an explicit calibration motion [14], [22], [23] or through manual labelling [24] detracts from their use. Similarly, the use of multiple cameras suggests a requirement of a dedicated motion capture space where the system can be setup and left undisturbed between sessions.

### A. Contributions

This paper assesses the feasibility of using a single depth camera as a clinical assessment tool for whole-body kinematic and kinetic assessment. As such, this paper prioritises: *1) Accurate anatomical joint centre locations and joint angles* which are needed for clinical assessment and future dynamic/musculoskeletal modelling. *2) Fast computation time* to allow for immediate review by the clinician. *3) Ease of use* by non-specialists in a clinical environment to perform a rapid motion assessment.

To these aims, we present a simple, fast method for taking any pre-estimated joint centre locations, automatically scaling skeletal parameters based on the subject height and recovering kinematic and kinetic measures from the biomechanical model.

This system provides accurate, reproducible, and consistent estimates of anatomical joint centre locations, with a mean joint position error of 2.63 cm. An additional estimate of L5/S1 location is added to the kinematic model allowing for assessment of the lower back, an important site of analysis in clinical and occupational health scenarios. The proposed system is used as a post-processing step on the raw Kinect 2 skeleton, with the mean computation speed of 524 frames per second. This suggests this method can be incorporated into many existing real-time methods without a significant drop in frame-rate. Only a single RGB-D camera is used, allowing for deployment clinical space without the need of a dedicated, calibrated motion capture space. The extraction of kinematic states is performed only requiring the user to specify the subject's height, without any manual model tuning, or joint labelling. The entire time to setup the camera, coach the subject to perform the STS action, data collection, and kinematic recovery takes under 1 minute.

## II. MODELLING FRAMEWORK

Rigid-body models are commonly used in biomechanics research to estimate joint kinematics and loading [25]–[29]. The mathematical formulations for the kinematics, kinetics, and dynamics of these systems can be taken from the robotics literature [30], providing a versatile method for analysing arbitrary rigid-body systems. In this work, we present and evaluate two rigid-body models:

1) *Floating rigid-body model*: Constrained body segment lengths. This form of model is typically used in motion analysis, with no environmental constraints.
2) *Fixed-ankle rigid-body model*: Constrained body segment lengths and angle-ground contact. As the ankles do not move in the sit-to-stand action, a kinematic constraint on ankle position can be used to determine the effect on the recovered kinematic and kinetic measures.

The models are driven by the raw Kinect shoulder, hip, knee, and ankle joint centre positions.

### A. Model I: Floating Rigid-Body Model

*1) Model Structure:* The human body is commonly modelled as a *floating* tree system, consisting of a pelvic base-link with serial chains that terminate at the head, hands, and feet [31]. In this work, we study the kinematics of the lower limbs and trunk during STS, neglecting the motion of the arms. We consider a 3D rigid-body model with six segments: (left and right) lower leg, (left and right) upper leg, pelvis, and torso (Fig. 1). The corresponding joint centres are at the ankle, knee, hip, and lower-lumbar joints. The knee joint is modelled as a cylindrical joint. The ankles, hips, and lower-lumbar (L5/S1) joint are modelled as spherical joints with three successive rotations. The order of these rotations is based on relations to common range of motion measures [32].

Segment lengths are determined by recommended height-scaled, sex-specific, allometric relations [27]. These relations provide estimated link length for the upper and lower leg, ($l_{UL}$ and $l_{LL}$), shoulder width and hip width, ($w_S$ and $w_P$), the length
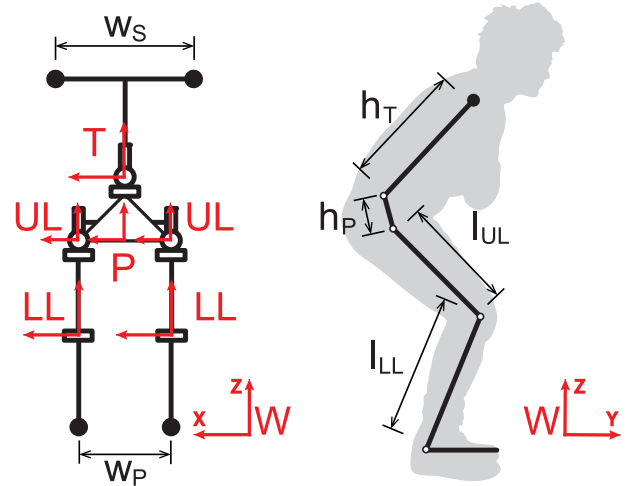


Fig. 2. The frame labelling and key lengths for the floating pelvis model. Left: Frontal plane view showing shoulder and pelvic width $w_S$, $w_P$, coordinate frames for the torso $T$, pelvis $P$, the upper and lower legs ($UL$, $LL$), and the world frame $W$. Axes are aligned with the Z axis lying along the primary axis of the segment. Right: Sagittal plane view showing torso and pelvis heights $h_T$ and $h_P$, and the upper and lower leg lengths $l_{UL}$, $l_{LL}$.

between the midpoint of the hip centres and L5/S1 ($h_P$) and the length between L5/S1 and midpoint of the shoulder centres ($h_T$).

*2) Kinematic Formulation:* The following mathematical formulation utilizes relative coordinate frame transformations to relate the observed Kinect joint centre positions to corresponding joint angles in the rigid-body model. Local coordinate frames are defined in Fig. 2. The origin of the base pelvic frame ($P$) is located at the midpoint between the hip joint centres. All other frame origins are located at the joint centre with $Z$ axis pointing along the segment length in the sagittal plane. We represent the position and orientation of the pelvic frame as an $X, Y, Z$ translation ($\boldsymbol{t}_P$), and three sequential rotations about the $X, Y,$ and $Z$ axes ($\theta_{PX}, \theta_{PY}, \theta_{PZ}$). This is formulated as the homogeneous transformation between World and Pelvic frames $\boldsymbol{g}_{W,P}$:

$$\boldsymbol{g}_{W,P}\left(\boldsymbol{t}_P, \theta_{PX}, \theta_{PY}, \theta_{PZ}\right) = \begin{bmatrix} \boldsymbol{R}_X \boldsymbol{R}_Y \boldsymbol{R}_Z & \boldsymbol{t}_P \\ \boldsymbol{0} & 1 \end{bmatrix} \quad (1)$$

where $\boldsymbol{R}_X$, $\boldsymbol{R}_Y$, and $\boldsymbol{R}_Z$ are the standard rotation matrices about the $X, Y,$ and $Z$ axes, respectively.

The relative transformations between each adjacent segment are defined in the same notation. For example, the transformation between the Pelvis and Torso ($T$) frames, $\boldsymbol{g}_{P,T}$ can be written:

$$\boldsymbol{g}_{P,T}\left(h_P, \theta_{TX}, \theta_{TY}, \theta_{TZ}\right) = \begin{bmatrix} \boldsymbol{R}_X \boldsymbol{R}_Y \boldsymbol{R}_Z & \begin{bmatrix} 0 \\ 0 \\ h_P \end{bmatrix} \\ \boldsymbol{0} & 1 \end{bmatrix} \quad (2)$$

using the coordinate frames and segment lengths defined in Fig. 2.

These homogeneous pose matrices are used to estimate the World frame locations of the left and right shoulder centres from

their local positions and relative frame transformations:

$$\begin{bmatrix} \boldsymbol{p}_{sho\_left} & \boldsymbol{p}_{sho\_right} \\ 1 & 1 \end{bmatrix} = \boldsymbol{g}_{W,P}\boldsymbol{g}_{P,T} \begin{bmatrix} \begin{bmatrix} -w_S \\ 0 \\ h_T \\ 1 \end{bmatrix} & \begin{bmatrix} +w_S \\ 0 \\ h_T \\ 1 \end{bmatrix} \end{bmatrix} \tag{3}$$

This process can be repeated for each joint centre to create the observation model for all joints:

$$\begin{bmatrix} \boldsymbol{p}_{sho\_left} \\ \boldsymbol{p}_{sho\_right} \\ \boldsymbol{p}_{hip\_left} \\ \boldsymbol{p}_{hip\_right} \\ \boldsymbol{p}_{knee\_left} \\ \boldsymbol{p}_{knee\_right} \\ \boldsymbol{p}_{ankle\_left} \\ \boldsymbol{p}_{ankle\_right} \end{bmatrix} = h_{obs}\left(\boldsymbol{\eta}, \boldsymbol{X}\right) \tag{4}$$

where $\boldsymbol{\eta}$ are the model parameters:

$$\boldsymbol{\eta} = \begin{bmatrix} h_P, h_T, l_{UL}, l_{LL}, w_S, w_P \end{bmatrix} \tag{5}$$

and $\boldsymbol{X}_I \in \mathbb{R}^{17}$ is the state vector containing the corresponding translations and rotations:

$$\boldsymbol{X}_I = \begin{matrix} [\boldsymbol{t}_P, \boldsymbol{\theta}_P, \boldsymbol{\theta}_T, \boldsymbol{\theta}_{UL\_left}, \boldsymbol{\theta}_{UL\_right}, \\ \boldsymbol{\theta}_{LL\_left}, \boldsymbol{\theta}_{LL\_right}]^T \end{matrix} \tag{6}$$

### B. Model II: Rigid-Body Model With Fixed-Ankle

Our second rigid-body model introduces an additional constraint by fixing the position of the ankle joint centres. The raw joint centres from the depth camera are not constrained by the ground plane, allowing the ankle to phase through the floor or hover above the floor while the person is standing. During STS, we assume the position of the ankle remains fixed and can be constrained at a fixed position throughout the motion. To implement this constraint, we select the base link to be one of the feet and fix this position to the ground. The mathematical formulation of the observation model is similar to Section II-A, with the model starting at one foot and moving up the leg, before branching at the pelvis into the torso and second leg branches.

The state vector $\boldsymbol{X}_{II} \in \mathbb{R}^{14}$ for the fixed ankle model has three fewer states when compared to the floating pelvis model, with the addition of the ipsilateral ankle rotation $\boldsymbol{\theta}_{A\_ipsi} \in \mathbb{R}^3$ and the removal of the pelvis translation and orientation ($\mathbb{R}^6$):

$$\boldsymbol{X}_{II} = \begin{matrix} [\boldsymbol{\theta}_{A\_ipsi}, \theta_{LL\_ipsi}, \boldsymbol{\theta}_{UL\_ipsi}, \boldsymbol{\theta}_{UL\_contra}, \\ \theta_{LL\_contra}, \boldsymbol{\theta}_T]^T \end{matrix} \tag{7}$$

where the subscripts *ipsi* and *contra* refer to the ipsilateral and contralateral sides to the base foot.

In our model, the ankle joint centre is fixed at a position based on the observed motion. The $X$ and $Z$ coordinates are taken to be the mean observed position throughout the motion. The $Y$ coordinate is fixed to be equal to the mean anterior-posterior position of the knee at a standing posture.
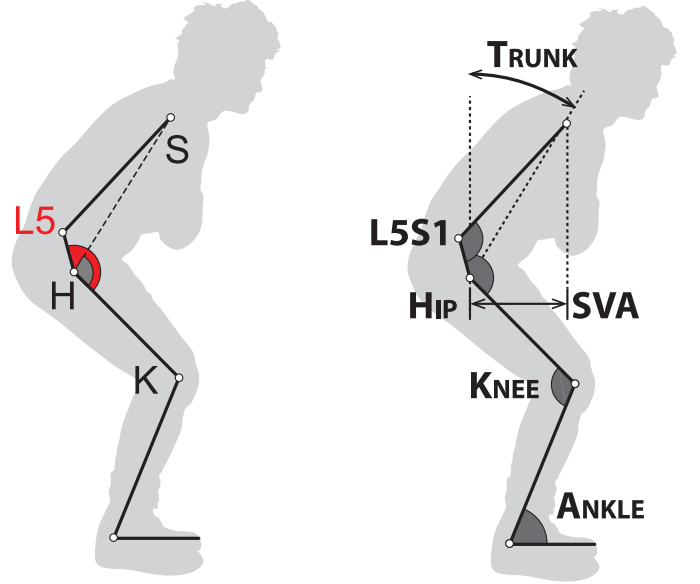


Fig. 3. Left: Model used for L5S1 angle estimation. The sagittal location of the shoulder, L5S1, hip, and knee are shown. The included angle at the hip is used to predict the angle at L5S1. Right: sagittal plane model. Joint centres and angle definitions are shown. The definition of the Sagittal Vertical Axis (SVA) metric is also shown.

### C. Inverse Kinematics

The kinematic recovery process allows for the estimation of joint angles from observations of joint position. We use two methods of kinematic recovery are: Non-linear Least Squares (NLS) and Unscented Kalman Filtering (UKF).

*1) Non-Linear Least Squares (NLS):* The error between the observed joint centres $\boldsymbol{q}$ and the expected joint centres $h_{obs}\left(\boldsymbol{\eta}, \boldsymbol{X}\right)$ is minimised for each frame $k$:

$$\min_{\boldsymbol{X}_k} \|\boldsymbol{q}_k - h_{obs}\left(\boldsymbol{\eta}, \boldsymbol{X}_k\right)\|_2^2 \tag{8}$$

*2) Unscented Kalman Filtering (UKF):* While the NLS method allows for the estimation of the state at each frame, it does not enforce any relationship between sequential states. The UKF balances inaccuracies in measurement with an estimate of the change in state between two successive states [33], [34]. Using the notation for Kalman filters, every observed joint centre at frame $k$ can be written:

$$\boldsymbol{q}_k = h_{obs}\left(\boldsymbol{\eta}, \bar{\boldsymbol{X}}_k\right) + \boldsymbol{v}_k \tag{9}$$

where $\boldsymbol{v}_k$ is a model of the sensor noise which is taken to be white noise: $\boldsymbol{v}_k \sim \mathcal{N}\left(0, \boldsymbol{R}_k\right)$, $\boldsymbol{R}_k$ is the covariance matrix of the Kinect, and the state $\bar{\boldsymbol{X}}_k$ is the true state that underlies each observation.

This observation model is combined with a process model $f_{proc}$ which relates previous estimates of the the true state $\bar{\boldsymbol{X}}_{k-1}$ to the current true state:

$$\bar{\boldsymbol{X}}_k = f_{proc}\left(\bar{\boldsymbol{X}}_{k-1}\right) + \boldsymbol{w}_k \tag{10}$$

where $\boldsymbol{w}_k$ is a model of the process noise which is taken to be white noise: $\boldsymbol{w}_k \sim \mathcal{N}\left(0, \boldsymbol{Q}_k\right)$. To set limits on the variation of each of the states between samples, the process covariance
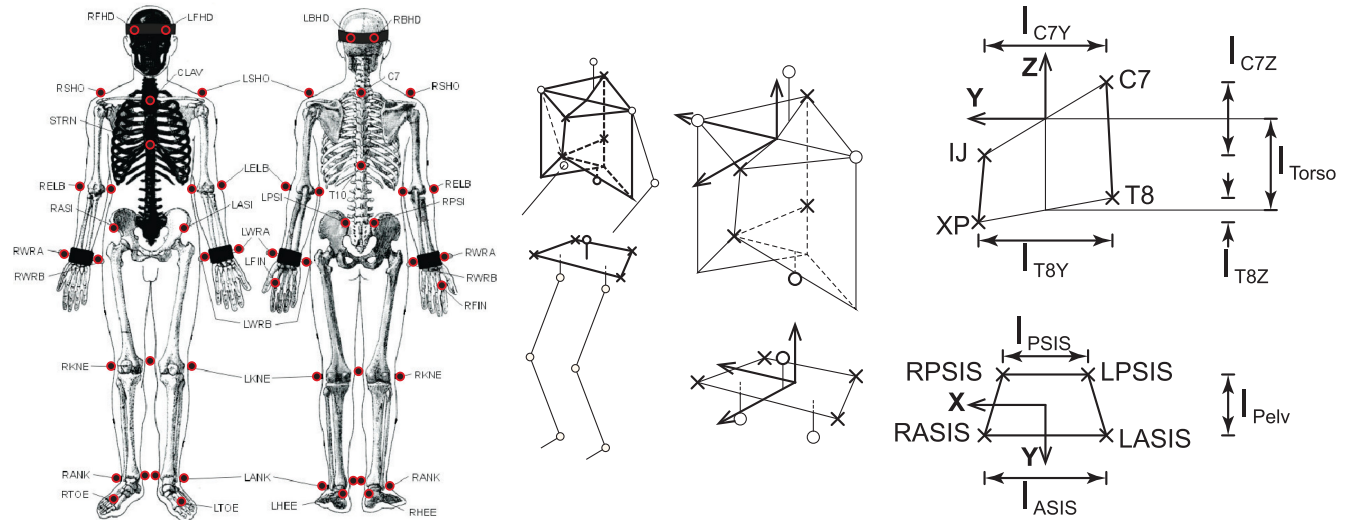
Fig. 4. Left: Cartoon of the motion capture marker protocol used in this work. Markers (red) are shown superimposed on the standard Plug-in-Gait model. Centre: Rigid body frames used for the validation model. Torso and pelvis frames are highlighted, with markers shown as crosses, and joint centres shown as circles. Right: Segment-marker definitions used for NLS recovery. The sagittal view of the torso frame, and caudal view of the pelvis frame are shown. Torso markers were located at the Incisura jugularis sternalis (IJ), Xiphoid Process (XP), and at the C7 and T8 spinous processes which were found during standing. Pelvis markers were located at the right and left Anterior Superior Iliac Spines (ASIS) and Posterior Superior Iliac Spines (PSIS).

$Q$ is fixed to be the expected change due to the velocities $\boldsymbol{\sigma}_V$. This allows the process covariance to be written explicitly as the diagonal matrix:

$$Q = \Delta t^2 diag\left(\boldsymbol{\sigma}_V^2\right) \tag{11}$$

where $\Delta t$ is the time between samples, and the process model as the identity matrix.

### D. Planarisation

The recovered 3D kinematic data is planarised for analysis of the sagittal kinematics. A plane is fit to the motion of the Kinect joint centres and the data is projected onto the plane. For the symmetric joint centres (ankles, knees, hips, shoulders), the mean of the sagittal plane positions is taken.

### E. Lower Lumbar Joint (L5/S1) Estimation

The raw Kinect skeleton provides a single joint centre along the spine. We found the position of this joint centre to be inconsistent between subjects and within single trials. Due to this unreliability and lack of relation to an anatomical landmark, we disregard the mid-spine marker in our kinematic analysis and consider an alternate method for determining a joint between the hip and shoulders in the sagittal plane.

From marker-based motion capture data, the position of the lower lumbar joint, located at L5/S1, can be estimated from pelvic landmarks. An allometric model for the position of L5/S1 in a pelvic frame is presented in Reed *et al.* [35] Unfortunately, the pelvic orientation is not observable from the Kinect data, so we cannot apply this method.

A model for lumbosacral orientation using knee flexion and trunk inclination is presented by Anderson *et al.* [36]. In that work, a quadratic model was trained on four subjects in multiple static lifting postures. This model was not assessed on any test data. Using active motion capture data, we tested the

Anderson model against the marker-based Reed method. We found that the model did not accurately predict the sacral orientation during STS.

In this work, we present a new regression model for $KHL5$, the angle formed by the knee, hip, and L5/S1 joints, driven by $KHS$, the angle formed by the knees, hips, and shoulders (Fig. 3, joints present in the Kinect data). This model assumes that coordination between the hip and L5S1 joints follows a predictable pattern across subjects.

The model is trained using marker-based motion capture data (protocol detailed in Section III-B). We define the pelvic frame by anterior and posterior superior iliac spine (ASIS and PSIS) markers shown in Fig. 4. The location of the L5S1 joint centre is based on the model presented in Reed, in which the L5/S1 joint centre in given a frame defined by the ASIS and pubic symphysis (PS) landmarks. The PS landmark is not possible to mark on a clothed subject or easily observable from motion capture data. Using dry pelvis data from Reynolds *et al.* [37], we re-derived the position of the L5/S1 joint in the ASIS-PSIS pelvic frame:

$$L5S1 = \begin{bmatrix} 0 & 0 & .11PW \end{bmatrix}^T \tag{12}$$

where the pelvic width $PW$ is the distance between the left and right ASIS landmarks. From the observed L5/S1 joint center, we compute the joint angles in the sagittal plane. A linear model was fit to the data:

$$KHL5 = \alpha SHK + \beta \tag{13}$$

with derived model parameters $\alpha = 0.82$, $\beta = 0.54$ and r-squared 0.9784. This model was trained on four subjects performing STS and tested against six subjects. Each subject performed STS three times. The fitted model and test data are shown in Fig. 5. The mean absolute error (MAE) between the predicted and observed $L5S1$ angles was $3.63 \pm 1.72$ degrees for the test set $4.21 \pm 2.73$ degrees for the training set.
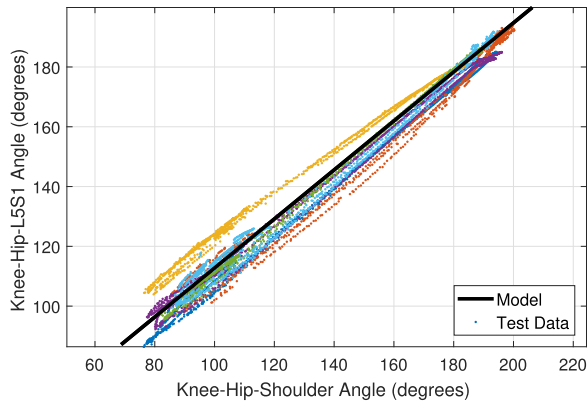
Fig. 5.  Linear regression for $L5S1$, the angle formed by the knee, hip, and L5S1, joints, from $KHS$, the angle formed by the knee, hip, and shoulder joints. The model is shown in solid black. Data from the test set of 6 subjects is shown in dotted color.

From Eq. (13) and an allometrically scaled pelvic height, $h_p$, we can express the position of the L5S1 joint centre. In our recovery framework, this L5S1 model is used after the kinematic recovery and planarisation steps are performed.

## III. EXPERIMENTAL VALIDATION

The modelling and kinematic recovery methods introduced in Section II were tested experimentally and validated against marker-based motion capture data on non-clinical subjects.

### A. Experimental Protocol

Ten subjects (3F/7M, age: 30.9 ± 9.6, height: 1.76 ± 0.12 m, mass: 67.4 ± 11.2 kg) were recruited under informed consent (UCSF IRB 16-21015). Subjects wore close fitting exercise clothing (sports bra, exercise shorts). The chair height was adjusted so that the subject's thighs were parallel to the ground, and their knees directly above their ankles during natural sitting. Subjects were asked to perform STS with their arms folded across their chest, hands touching the opposite elbow. The standing action was otherwise non-coached, with subjects performing the action naturally. Three trials, each consisting of three STS, were recorded for each subject.

### B. Active Motion Capture Model

An 8-camera active motion capture system was used in this study to provide a ground-truth estimate of position and orientation of each body segment. Motion data of STS was simultaneously recorded from the Kinect and the motion capture system. The Kinect camera was located 2.5 meters directly in-front of the subject. The Kinect joint centres were streamed at 30 Hz and saved with a UNIX timestamp onto a desktop computer. Each trial consisted of 883 ± 87 frames of Kinect depth data, and 14224 ± 1548 frames of Phasespace data for three successive stand-sit-stand motions (around 30 seconds). The Kinect and motion capture systems were time synchronised using a network time protocol server.

Thirty-two LED markers (Phasespace, San Leandro, CA) were recorded at 480 Hz with an associated UNIX timestamp.

Kinematic recovery was performed offline in MATLAB. The markers were placed onto the subjects skin using adhesive Velcro based on the Plug-in-Gait markers set [38] (Fig. 5). Additional markers were placed on the medial elbow, knee, and ankle positions to allow for estimates of joint centre from the medio-lateral marker pairs. In cases where the subject's shorts or sports bra obscured the ASIS, PSIS, or XP landmarks, a clip was used to secure the marker to the clothes band at the desired landmark.

In addition to the STS protocol, a dataset was collected for identifying the functional joint centres for each segment using the Recap2 protocol [39]. Subjects were asked to move each joint through its full range of motion three times, starting with the wrists, elbows, and shoulders, before moving the ankles, knees, and hips. The Recap2 protocol was only used to find the functional centres for the ground truth motion capture model.

NLS (Section II-C1) was used to recover the instantaneous position and orientation of each limb segment in 3D coordinates. Each limb segment was recovered independently without any modelling of the connection between connected limbs.

The rigid-body models used for each each segment are shown in Fig. 5. The coordinate system is based on Wu [40], with the exception of the pelvis segment where the origin is located at the midpoint of the ASIS and PSIS markers. The labelling of the coordinate axes were also modified to simplify plotting and analysis in MATLAB. NLS was used to estimate the marker positions in the local coordinate frame for each subject.

The joint centres for the ground-truth model were recovered using functional methods (hip and shoulder), and marker-based methods (ankle, knee, and L5/S1). Geometric sphere fitting for the hip was chosen based on the recommendation by the ISB [40] and as all subjects were able to move sufficiently [41]. The inter-malleolar point was selected for the ankles from Wu [40], the inter-epicondyle point for the knee [42], and L5/S1 from the allometric model described in Section II-E. The recovered joint-centres were planarised and the relative angles were determined at each frame.

### C. Data Analysis

All data processing was performed on previously stored Kinect 2 data on an Intel i7-5820K processor, with 32 GB of RAM running Windows 7 Enterprise. Each trial of three stand-sit-stand actions consisted of roughly 880 frames and was post-processed at 524 ± 140 fps. A graphics card was not used to aid computation.

The joint angles recovered from each method were filtered and numerically differentiated to obtain joint velocity estimates. A first-order, low-pass Butterworth filter at 5 Hz was applied to the active motion capture and both rigid-body Kinect models [43], [44]. The raw kinect data was filtered more heavily, using a first-order low-pass Butterworth filter at 2 Hz. This was to account for significant noise in the raw joint angles leading to unrealistic velocity estimates.

We compute the horizontal distance between the shoulder joint and hip joint centres at each frame as well as its velocity. This is a surrogate for the Sagittal Vertical Axis (SVA), a metric for spinal alignment, measured by static radiographs as the distance between C7 and L5S1 [4]. We also compare recovered

TABLE I

INTER-RATER ASSESSMENTS FOR ESTIMATED JOINT POSITION. MEAN ABSOLUTE ERRORS, CONCORDANCE AND INTERCLASS CORRELATION COEFFICIENTS FOR EACH SKELETAL JOINT. MEAN ABSOLUTE ERRORS ARE STATED AS MEAN WITH THE STANDARD DEVIATION IN PARENTHESIS. ANKLE CORRELATION COEFFICIENTS FOR THE FIXED-ANKLE MODEL ARE OMITTED AS THE ANKLE POSITION IN THIS MODEL DOES NOT VARY WITH TIME

| Measure | Model | Ankle Position | | Knee Position | | Hip Position | | Shoulder Position | |
|---|---|---|---|---|---|---|---|---|---|
| | | X | Y | X | Y | X | Y | X | Y |
| MAE (cm) | Raw | 8.66 ± 3.17 | 2.01 ± 1.69 | 1.96 ± 1.69 | 3.57 ± 2.38 | 1.93 ± 1.95 | 4.07 ± 2.44 | 1.98 ± 1.38 | 4.05 ± 2.04 |
| | Floating | 8.31 ± 3.21 | 1.95 ± 1.59 | 1.88 ± 1.26 | 1.46 ± 1.42 | 2.40 ± 1.69 | 2.36 ± 1.28 | 1.71 ± 1.57 | 2.93 ± 2.11 |
| | Fixed | 1.62 ± 0.07 | 1.97 ± 1.29 | 1.30 ± 1.08 | 0.78 ± 0.62 | 1.98 ± 1.36 | 2.07 ± 1.27 | 1.58 ± 1.28 | 3.08 ± 1.97 |
| CCC | Raw | 0.24 | 0.03 | 0.95 | 0.62 | 0.98 | 0.95 | 0.98 | 0.97 |
| | Floating | 0.26 | 0.04 | 0.96 | 0.88 | 0.98 | 0.98 | 0.98 | 0.98 |
| | Fixed | - | - | 0.97 | 0.97 | 0.99 | 0.99 | 0.99 | 0.98 |
| ICC(2,1) | Raw | 0.24 | 0.03 | 0.95 | 0.62 | 0.98 | 0.95 | 0.98 | 0.97 |
| | Floating | 0.26 | 0.04 | 0.96 | 0.88 | 0.98 | 0.98 | 0.98 | 0.98 |
| | Fixed | - | - | 0.97 | 0.97 | 0.99 | 0.99 | 0.99 | 0.98 |
| ICC(3,1) | Raw | 0.72 | 0.05 | 0.95 | 0.62 | 0.99 | 0.99 | 0.99 | 0.99 |
| | Floating | 0.73 | 0.05 | 0.96 | 0.91 | 0.99 | 0.99 | 0.99 | 0.99 |
| | Fixed | - | - | 0.97 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 |

TABLE II

INTER-RATER ASSESSMENTS FOR ANGLE AND POSITIONS MEASURES. MEAN ABSOLUTE ERRORS, CONCORDANCE AND INTERCLASS CORRELATION COEFFICIENTS FOR SELECTED TRAJECTORY MEASURES. MEAN ABSOLUTE ERRORS ARE STATED AS MEAN WITH THE STANDARD DEVIATION IN PARENTHESIS. ASTERISKS INDICATE THAT THE ABSOLUTE ERRORS WERE SIGNIFICANTLY DIFFERENT FROM THE RAW KINECT MODEL AT THE 5% (*) AND 1% (**) SIGNIFICANCE LEVELS. CORRELATION COEFFICIENTS ARE GIVEN WITH THE 95% CONFIDENCE INTERVAL IN PARENTHESIS

| Measure | Model | Ankle Angle (deg) | Knee Angle (deg) | Hip Angle (deg) | L5/S1 Angle (deg) | Trunk Angle (deg) | SVA Position (cm) |
|---|---|---|---|---|---|---|---|
| MAE | Raw | 12.6 (4.9) | 9.4 (6.4) | 11.1 (7.7) | - | 3.9 (2.4) | 3.4 (2.2) |
| | Floating | 11.5 (5.0)** | 9.0 (5.5)** | 5.6 (4.7)** | 6.4 (4.4) | 3.6 (2.4)** | 2.6 (1.7)** |
| | Fixed | 4.2 (2.6)** | 5.2 (3.5)** | 5.8 (5.4)** | 6.1 (4.1) | 3.5 (2.4)** | 2.6 (1.7)** |
| CCC | Raw | 0.43 (0.42 - 0.44) | 0.97 (0.96 - 0.97) | 0.95 (0.95 - 0.95) | - | 0.93 (0.93 - 0.94) | 0.93 (0.93 - 0.93) |
| | Floating | 0.43 (0.42 - 0.44) | 0.97 (0.97 -0.97) | 0.98 (0.98 - 0.98) | 0.79 (0.78 - 0.80) | 0.95 (0.95 - 0.95) | 0.96 (0.96 - 0.96) |
| | Fixed | 0.82 (0.81 - 0.83) | 0.99 (0.99 - 0.99) | 0.98 0.97 - 0.98) | 0.81 (0.80 - 0.81) | 0.95 (0.95 - 0.95) | 0.96 (0.96 - 0.96) |
| ICC(2,1) | Raw | 0.43 (0.00 - 0.77) | 0.97 (0.76 - 0.99) | 0.95 (0.91 - 0.97) | - | 0.93 (0.71 - 0.97) | 0.93 (0.75 - 0.97) |
| | Floating | 0.43 (0.00 - 0.76) | 0.97 (0.46 - 0.99) | 0.98 (0.96 - 0.99) | 0.79 (0.71 - 0.84) | 0.95 (0.71 - 0.98) | 0.96 (0.89 - 0.98) |
| | Fixed | 0.82 (0.78 - 0.85) | 0.99 (0.99 - 0.99) | 0.98 (0.97 - 0.98) | 0.81 (0.75 - 0.85) | 0.95 (0.70 - 0.98) | 0.96 (0.88 - 0.98) |
| ICC(3,1) | Raw | 0.85 (0.84 - 0.86) | 0.98 (0.98 - 0.98) | 0.96 (0.96 - 0.96) | - | 0.96 (0.96 - 0.96) | 0.96 (0.95 - 0.96) |
| | Floating | 0.83 (0.82 - 0.84) | 0.99 (0.99 - 0.99) | 0.98 (0.98 - 0.98) | 0.81 (0.80-0.82) | 0.97 (0.97 - 0.97) | 0.97 (0.97 - 0.97) |
| | Fixed | 0.83 (0.83 - 0.84) | 0.99 (0.99 - 0.99) | 0.98 (0.98 - 0.98) | 0.82 (0.81 - 0.83) | 0.97 (0.97 - 0.97) | 0.97 (0.97 - 0.97) |

peak values for several metrics during STS: flexion and extension velocities of the torso, torso inclination angle, and SVA.

We consider each combination of sensor and model (raw Kinect, floating rigid-body Kinect, fixed-ankle rigid-body Kinect, and active motion capture) to be a different *rater*, allowing for the use of inter-rater reliability assessment methods. Three statistical measures were used to analyse the performance of the raw and rigid-body Kinect models against the active motion capture ground truth:

*1) Mean Absolute Error (MAE):* Identifies the raw position or velocity error between methods.

*2) Lin's Concordance Correlation Coefficient (CCC):* Assesses inter-rater reliability between methods [45].

*3) Inter-class Correlations Coefficients (ICC):* Identifies the absolute agreement (ICC(2, 1)) and relative consistency (ICC(3, 1)) between methods [46]–[48]. ICC values were interpreted as poor ($< 0.4$), fair (0.4–0.59), good (0.6–0.74), and excellent ($\geq 0.75$) based on the treatment by [49], [50].

## IV. RESULTS

MAE, CCC, and ICC statistics are given for joint center positions (Table I), joint trajectories (Table II), velocity trajectories (Table III), and selected peak metrics (Table IV). A representative motion capture trace is shown in Fig. 6.

Both rigid-body Kinect models (floating and fixed-ankle) achieved significantly lower MAE than the raw Kinect for all joint angle and position measures (Tables I, II). In comparison to the floating model, the fixed-ankle model had significantly less error in the ankle and knee positions and angles, and comparable error in all other measures.

Higher CCC and ICC values indicate greater reliability, relative consistency, and absolute agreement. For the position measures, the fixed-ankle model has higher CCC and ICC values than the raw Kinect in all cases. The floating pelvis model was better than the raw Kinect model, but has poor performance in recovering the ankle positions and angle.

The MAE of the velocity trajectories shown in Table III are comparable between the raw, floating, and fixed-ankle models. This similarity in performance was also seen in the CCC and ICC values, with the knee, hip, trunk and SVA velocities showing high agreement and repeatability for all methods. The estimated ankle velocities had lower ICC and CC values across the methods, but showed an improvement using the fixed-ankle model. The recovery of the angles and angular velocities at L5/S1 were consistently worse than that of the other joints, classified as good-excellent instead of excellent.

The peak measures in Table IV show the performance of the different methods in extracting candidate performance metrics

TABLE III
INTER-RATER ASSESSMENTS FOR ANGULAR AND LINEAR VELOCITY MEASURES MEAN ABSOLUTE ERRORS, CONCORDANCE AND INTERCLASS CORRELATION COEFFICIENTS FOR SELECTED TRAJECTORY MEASURES. MEAN ABSOLUTE ERRORS ARE STATED AS MEAN WITH THE STANDARD DEVIATION IN PARENTHESIS. ASTERISKS INDICATE THAT THE ABSOLUTE ERRORS WERE SIGNIFICANTLY DIFFERENT FROM THE RAW KINECT MODEL AT THE 5% (*) AND 1% (**) SIGNIFICANCE LEVELS. CORRELATION COEFFICIENTS ARE GIVEN WITH THE 95% CONFIDENCE INTERVAL IN PARENTHESIS

| Measure | Model | Ankle Angular Velocity (deg/s) | Knee Angular Velocity (deg/s) | Hip Angular Velocity (deg/s) | L5/S1 Angular Velocity (deg/s) | Trunk Angular Velocity (deg/s) | SVA Velocity (cm/s) |
|---|---|---|---|---|---|---|---|
| MAE | Raw | 7.5 (7.3) | 13.0 (14.0) | 16.7 (17.0) | - | 8.0 (7.9) | 7.2 (7.0) |
| | Floating | 7.9 (9.0)* | 11.3 (12.1)** | 12.0 (12.6)** | 9.9 (10.1) | 8.1 (9.4) | 6.8 (6.7)** |
| | Fixed | 6.5 (7.2)** | 13.5 (15.0) | 13.7 (13.8)** | 10.1 (10.0) | 8.8 (9.0)** | 7.1 (6.4) |
| CCC | Raw | 0.80 (0.79 - 0.81) | 0.93 (0.93 - 0.93) | 0.93 (0.93 - 0.93) | - | 0.95 (0.95 - 0.95) | 0.95 (0.94 - 0.95) |
| | Floating | 0.73 (0.72 - 0.75) | 0.94 (0.94 - 0.95) | 0.95 (0.95 - 0.96) | 0.69 (0.68 - 0.70) | 0.95 (0.95 - 0.95) | 0.95 (0.95 - 0.96) |
| | Fixed | 0.87 (0.86 - 0.87) | 0.93 (0.92 - 0.93) | 0.94 (0.94 - 0.95) | 0.70 (0.69 -0.72) | 0.95 (0.95 - 0.95) | 0.95 (0.95 - 0.96)) |
| ICC(2,1) | Raw | 0.80 (0.79 - 0.81) | 0.93 (0.93 - 0.93) | 0.93 (0.93 - 0.93) | - | 0.95 (0.95 - 0.95) | 0.95 (0.94 - 0.95) |
| | Floating | 0.73 (0.72 - 0.75) | 0.94 (0.94 - 0.95) | 0.95 (0.95 - 0.96) | 0.69 (0.68 - 0.71) | 0.95 (0.95 - 0.95) | 0.95 (0.95 - 0.96) |
| | Fixed | 0.87 (0.86 - 0.87) | 0.93 (0.92 - 0.93) | 0.94 (0.94 - 0.95) | 0.70 (0.68 - 0.72) | 0.95 (0.95 - 0.95) | 0.95 (0.95 - 0.96) |
| ICC(3,1) | Raw | 0.80 (0.79 0.81) | 0.93 (0.93 - 0.93) | 0.94 (0.94 - 0.94) | - | 0.95 (0.95 - 0.95) | 0.95 (0.94 - 0.95) |
| | Floating | 0.73 (0.72 - 0.75) | 0.94 (0.94 - 0.95) | 0.95 (0.95 - 0.96) | 0.69 (0.68 - 0.71) | 0.95 (0.95 - 0.95) | 0.95 (0.95 - 0.96) |
| | Fixed | 0.87 (0.86 - 0.88) | 0.93 (0.92 - 0.93) | 0.95 (0.94 - 0.95) | 0.70 (0.69 - 0.72) | 0.95 (0.95 - 0.95) | 0.95 (0.95 - 0.96) |

TABLE IV
INTER-RATER ASSESSMENTS FOR PEAK MEASURES. THE ANGLE AT L5S1 IS MODELLED USING THE AUTHOR'S PROPOSED METHOD. MEAN ABSOLUTE ERRORS, CONCORDANCE AND INTERCLASS CORRELATION COEFFICIENTS FOR SELECTED PEAK MEASURES. MEAN ABSOLUTE ERRORS ARE STATED AS MEAN WITH THE STANDARD DEVIATION IN PARENTHESIS. ASTERISKS INDICATE THAT THE ABSOLUTE ERRORS WERE SIGNIFICANTLY DIFFERENT FROM THE RAW KINECT MODEL AT THE 5% (*) AND 1% (**) SIGNIFICANCE LEVELS. CORRELATION COEFFICIENTS ARE GIVEN WITH THE 95% CONFIDENCE INTERVAL IN PARENTHESIS

| Measure | Model | Peak Trunk Flexion Angle (deg) | Peak SVA Position (cm) | Peak Trunk Flexion Angular Velocity (deg/s) | Peak Trunk Extension Angular Velocity (deg/s) |
|---|---|---|---|---|---|
| Typical Value | Mocap | 61.8 (4.9) | 24.1 (3.6) | 73.3 (10.0) | 55.4 (10.8) |
| | Raw | 60.4 (3.5) | 25.1 (2.9) | 58.8 (11.0) | 50.5 (7.2) |
| | Floating | 57.6 (6.2) | 24.5 (4.1) | 76.3 (17.0) | 65.1 (15.4) |
| | Fixed | 57.2 (6.2) | 24.9 (4.0) | 75.7 (16.3) | 67.7 (14.1) |
| MAE | Raw | 3.1 (2.5) | 2.9 (2.4) | 16.7 (9.4) | 6.1 (5.3) |
| | Floating | 4.2 (3.1)** | 1.9 (1.6)** | 11.1 (9.9)** | 8.8 (12.7)** |
| | Fixed | 4.6 (3.1)** | 1.9 (1.7)** | 8.2 (11.4)** | 12.6 (7.5)** |
| CCC | Raw | 0.58 (0.44 - 0.69) | 0.35 (0.17 - 0.51) | 0.15 (0.04 - 0.25) | 0.66 (0.56 - 0.74) |
| | Floating | 0.65 (0.55 - 0.73) | 0.79 (0.70 -0.85) | 0.40 (0.24 - 0.53) | 0.51 (0.38 - 0.61) |
| | Fixed | 0.63 (0.54 - 0.71) | 0.78 (0.69 - 0.85) | 0.47 (0.31 - 0.59) | 0.54 (0.43 - 0.63) |
| ICC(2,1) | Raw | 0.58 (0.40 - 0.71) | 0.36 (0.16 - 0.52) | 0.15 (-0.07 - 0.36) | 0.66 (0.27 - 0.82) |
| | Floating | 0.65 (-0.03 - 0.86) | 0.79 (0.70 - 0.86) | 0.40 (0.21 - 0.56) | 0.51 (0.08 - 0.73) |
| | Fixed | 0.64 (-0.01 - 0.87) | 0.78 (0.68 - 0.86) | 0.47 (0.29 - 0.62) | 0.54 (-0.09 - 0.82) |
| ICC(3,1) | Raw | 0.61 (0.46 - 0.73) | 0.96 (0.95 - 0.96) | 0.28 (0.09 - 0.46) | 0.75 (0.64 - 0.83) |
| | Floating | 0.83 (0.75 - 0.88) | 0.97 (0.97 - 0.97) | 0.41 (0.22 - 0.57) | 0.64 (0.50 - 0.75) |
| | Fixed | 0.84 (0.77 - 0.89) | 0.80 (0.71 - 0.86) | 0.47 (0.30 - 0.62) | 0.79 (0.70 - 0.86) |

from the trajectory data. The MAE for the floating and fixed-ankle models are significantly lower than the raw Kinect for the peak SVA and peak flexion velocity, but significantly higher for the flexion angle, and the extension angular velocity. The CCC and ICC values for the fixed-ankle model show fair-excellent performance, compared to the raw Kinect which had poor performance measuring the peak SVA and trunk flexion velocity. The peak trunk flexion angle and extension velocity were found to be consistent (ICC(3, 1): excellent), but with lower agreement (ICC(2, 1): good/fair).

## V. DISCUSSION

The introduction of segment length constraints in the floating rigid-body model resulted in significant improvement in all joint position and angle measures. The fixed-ankle model, which combined segment length constraints with an ankle contact constraint, further improved the recovery of the ankle and knee joint angles. Accurate lower-limb recovery is essential for performing whole-body dynamic analysis. This model had excellent estimates of joint position and velocity trajectories when compared to the gold-standard motion capture. The peak metrics associated with the position data were found to provide good to excellent agreement and consistency. These improvements are also seen peak metrics obtained from the floating-pelvis model, though the MAE, CCC, and ICC values for the for the ankle, knee, and hip are comparable or worse than the fixed-ankle model. In contrast, the raw Kinect had higher MAE in recovered joint angles, notably at the ankle, knee, and hip, and poor-fair agreement for the peak position metrics.

The mean joint position error of 2.63 cm (in the sagittal plane) is substantially lower than those seen in more generalised camera methods. In comparison to the joint errors on the Human 3.6 m dataset presented by Mehta [21], the lowest mean joint position error for the sit-down action was found to be 10.4 cm reported by Pavlakos [18]. The approach proposed by Shuai [24] for the used of multiple depth cameras resulted in marker residuals of approximately 3.5 cm for the sit down then stand
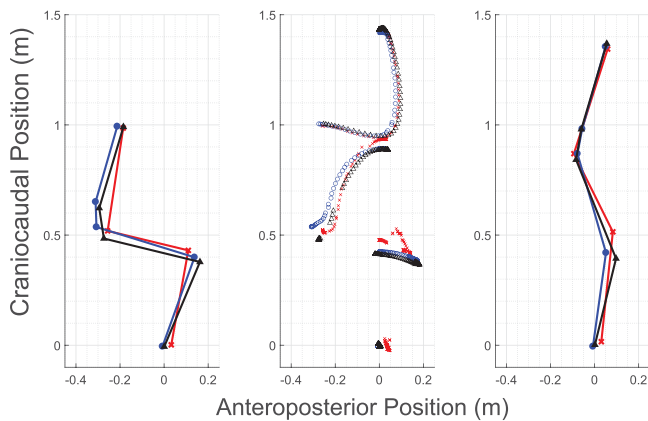
Fig. 6. Sagittal views of a representative subject performing the STS action. Left: Subject in the seated position. Black triangles outline the subject joint position from the active motion capture system. Red crosses show the raw Kinect 2 joint positions. Blue circles show the estimates from the proposed method. Middle: Motion traces for the entire STS action. L5S1 joints for the Phasespace and proposed models are hidden for clarity. Right: Subject in the near standing position. Note that the raw Kinect 2 skeleton does not provide an estimate for the L5S1 joint centre.

up action (MHAD action 9 [51]), but requires the use of three synchronised Kinect 2 cameras.

The trunk angle trajectory and peak trunk flexion angle were found to be consistent (ICC(3, 1): 0.95, 0.84) for the fixed-ankle model, but with lower absolute agreements (ICC(2, 1): 0.95, 0.64). This indicates that there may be a consistent offset between the motion-capture and the Kinect models. The MAE for these values suggest that there is an approximately 4 degree offset in the trunk angles between the active motion capture model and the fixed-ankle model. This offset may be based on the difference in the location of the shoulder centre between the motion capture and the Kinect models. The motion capture model defines the shoulder as the functional centre of the arm which is then fixed in the torso frame. This functional centre was estimated from the subject performing arm *windmills* in the sagittal plane. This motion also includes motion of the scapula, translating the location of the gleno-humeral centre (GHC). The estimated shoulder centre is likely to be in the centre of that space. During the sit to stand action, the subject's arms were placed across their chest, with their hands touching the opposite elbows. This action protracts the scapula, moving the GHC anteriorly. The Kinect however is estimating the location of the GHC for each frame accounting for this new location for the GHC.

The error in angular velocities were comparable across the raw Kinect and rigid-body methods. To obtain the velocities, the raw Kinect position data was heavily filtered. There is a notable trade-off in the performance of the models between the joint angle and angular velocity MAEs. In particular, the angular velocities of the floating-pelvis model outperforms the fixed-ankle model for all joints other than the ankle. As the floating-model is not constrained the model is able to respond rapidly to changes in observed position, resulting in lower velocity errors. In contrast, these rapid changes are moderated by the constraints imposed by the fixed-ankle model.

The improved position accuracy from the models, combined with accurate angular velocities suggests the suitability of using the fixed-ankle model for further dynamic analysis. This is not possible using the original raw Kinect data due to the inaccurate joint centre positions and corresponding angle errors.

It is important to note that this study was conducted with the person directly in front of the depth camera. From a previous study [52], the authors did find that there is an increase in estimated joint centre position when the camera is set at increasing angles from the subject. These joint centre errors were found to be higher in the lower limbs (+1.5 cm) in both standing and sitting activities, with higher error seen in the limb distal to the camera. While this issue could arise in clinic, especially in cases where the system may be rapidly set-up, the 30 degree offset used in the study is larger than a reasonably expected set-up error. Furthermore, the addition of an constraints to the ankle may improve recovery performance in the lower limbs.

## VI. CONCLUSION

This work presents a framework for improving kinematic recovery from depth-camera data through the use of rigid-body modelling. We validated the performance of our proposed method and raw Kinect data through comparison against a ground-truth active motion capture system. The use of a rigid-body model and contact constraints significantly improves the accuracy of joint angles measured by a depth camera. This framework enhances the utility of a depth-camera for quantitative motion analysis. Accurate kinematic and kinetic measurements allow for expansion to dynamic analysis of joint torques. The proposed system has low cost, space, and time requirements and can be easily deployed in clinic, with the total time to setup, collect, and process the motion data taking less than a minute.

While the validation analysis was performed on sagittal plane measures during STS, the modelling framework and kinematic recovery are performed in 3D. This framework will be extensible to actions outside of the plane, but further validation must be performed on the 3D recovery.

## REFERENCES

[1] W. T. Crow and D. R. Willis, "Estimating cost of care for patients with acute low back pain: A retrospective review of patient records," *J. Amer. Osteopathic Assoc.*, vol. 109, no. 4, pp. 229–233, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19369510

[2] A. J. Cardin and C. Hadida, "Evaluation of lumbar intersegmental range of motion using flexion-extension radiographs of asymptomatic versus low back pain adults," *J. Can. Chiropractic Assoc.*, vol. 38, no. 2, pp. 83–89, 1994.

[3] B. Cakir, M. Richter, W. Käfer, M. Wieser, W. Puhl, and R. Schmidt, "Evaluation of lumbar spine motion with dynamic X-ray–a reliability analysis," *Spine*, vol. 31, no. 11, pp. 1258–1264, 2006.

[4] B. J. Van Royen *et al.*, "Accuracy of the sagittal vertical axis in a standing lateral radiograph as a measurement of balance in spinal deformities," *Eur. Spine J.*, vol. 7, no. 5, pp. 408–412, 1998.

[5] S. S. Coghlin and B. J. McFadyen, "Transfer strategies used to rise from a chair in normal and low back pain subjects," *Clin. Biomechanics*, vol. 9, no. 2, pp. 85–92, 1994.

[6] E. Papa and A. Cappozzo, "Sit-to-stand motor strategies investigated in able-bodied young and elderly subjects," *J. Biomechanics*, vol. 33, no. 9, pp. 1113–1122, 2000.

[7] S. L. Jones, S. M. Henry, C. C. Raasch, J. R. Hitt, and J. Y. Bunn, "Individuals with non-specific low back pain use a trunk stiffening strategy to maintain upright posture," *J. Electromyography Kinesiology*, vol. 22, no. 1, pp. 13–20, 2012.

[8] J. Rasmussen, V. Vondrak, M. Damsgaard, M. D. Zee, and Z. Dostal, "The anybody project—computer analysis of the human body," in *Proc. Biomechanics Man*, 2002, pp. 270–274.

[9] S. L. Delp *et al.*, "OpenSim: Open source to create and analyze dynamic simulations of movement," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 11, pp. 1940–1950, Nov. 2007.

[10] W. S. Marras, F. A. Fathallah, R. J. Miller, S. W. Davis, and G. A. Mirka, "Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics," *Int. J. Ind. Ergonom.*, vol. 9, no. 1, pp. 75–87, 1992.

[11] S. A. Ferguson, W. S. Marras, and P. Gupta, "Longitudinal quantitative measures of the natural course of low back pain recovery," *Spine*, vol. 25, no. 15, pp. 1950–1956, 2000.

[12] W. R. Taylor, T. Consmüller, and A. Rohlmann, "A novel system for the dynamic assessment of back shape," *Med. Eng. Phys.*, vol. 32, no. 9, pp. 1080–1083, 2010.

[13] T. Consmüller, A. Rohlmann, D. Weinland, C. Druschel, G. N. Duda, and W. R. Taylor, "Comparative evaluation of a novel measurement tool to assess lumbar spine posture and range of motion," *Eur. Spine J.*, vol. 21, no. 11, pp. 2170–2180, 2012.

[14] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[15] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis," *J. Med. Eng. Technol.*, vol. 38, no. 5, pp. 274–280, 2014.

[16] J. K. Nichols, M. P. Sena, J. L. Hu, O. M. OReilly, B. T. Feeley, and J. C. Lotz, "A Kinect-based movement assessment system: Marker position comparison to Vicon," *Comput. Methods Biomechanics Biomed. Eng.*, vol. 20, no. 12, pp. 1289–1298, 2017. [Online]. Available: https://doi.org/10.1080/10255842.2017.1340464

[17] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014.

[18] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3D human pose," in *Proc. 30th IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1263–1272.

[19] S. Park, J. Hwang, and N. Kwak, "3D human pose estimation using convolutional neural networks with 2D pose information," in *Proc. Eur. Conf. Comput. Vision Workshop*, 2016, pp. 156–169.

[20] X. Zhou, X. Sun, W. Zhang, S. Liang, and Y. Wei, "Deep kinematic pose regression," in *Lecture Notes Comput. Sci. (including subseries Lecture Notes Artif. Intell. Lecture Notes Bioinf.)*, vol. 9915 LNCS, pp. 186–201, 2016.

[21] D. Mehta *et al.*, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, 2017.

[22] C. Stoll, N. Hasler, J. Gall, H. P. Seidel, and C. Theobalt, "Fast articulated motion tracking using a sums of Gaussians body model," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 951–958.

[23] M. Ding and G. Fan, "Articulated and generalized Gaussian kernel correlation for human pose estimation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 776–789, Feb. 2016.

[24] L. Shuai, C. Li, X. Guo, B. Prabhakaran, and J. Chai, "Motion capture with ellipsoidal skeleton using multiple depth cameras," *IEEE Trans. Visualization Comput. Graph.*, vol. 23, no. 2, pp. 1085–1098, Feb. 2017.

[25] NASA, "Anthropometric Source Book Volume I: Anthropometry for Designers," vol. 1, no. 2, pp. 1–606, 1978.

[26] D. B. Chaffin, "Development of computerized human static strength simulation model for job design," *Human Factors Ergonom. Manuf.*, vol. 7, no. 4, pp. 305–322, 1997.

[27] D. B. Chaffin, G. B. J. Andersson, and B. J. Martin, *Occupational Biomechanics*. New York, USA: Wiley, 1999.

[28] S. L. Delp and J. P. Loan, "A graphics-based software system to develop and analyze models of musculoskeletal structures," *Comput. Biol. Med.*, vol. 25, no. 1, pp. 21–34, 1995.

[29] M. A. Rajaee, N. Arjmand, A. Shirazi-Adl, A. Plamondon, and H. Schmidt, "Comparative evaluation of six quantitative lifting tools to estimate spine loads during static activities," *Appl. Ergonom.*, vol. 48, pp. 22–32, 2015.

[30] K. M. Lynch and F. C. Park, *Modern Robotics: Mechanics, Planning, and Control*. Cambridge, U.K.: Cambridge University Press, 2017.

[31] K. Ayusawa, G. Venture, and Y. Nakamura, "Identification of humanoid robots dynamics using floating-base motion dynamics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 2854–2859.

[32] K. Buckup, J. Buckup, and H. P. Paessler, *Clinical Tests for the Musculoskeletal System: Examinations-Signs-Phenomena*. New York, USA: TPS, 2016.

[33] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Proc. SPIE*, 1997, pp. 182–193.

[34] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," *Proc. IEEE Adaptive Syst. Signal Process., Commun., Control Symp.*, 2000, pp. 153–158.

[35] M. Reed, M. A. Manary, and L. W. Schneider, "Methods for measuring and representing automobile occupant posture," Society of Automotive Engineering, Warrendale, PA, USA, Tech. Rep. 1999-01-0959, 1999. [Online]. Available: http://papers.sae.org/1999-01-0959/

[36] C. K. Anderson, D. B. Chaffin, and G. D. Herrin, "A study of lumbosacral orientation under varied static loads," *Spine*, vol. 11, pp. 456–462, 1986.

[37] H. Reynolds, C. Snow, and J. Young, "Spatial Geometry of the Human Pelvis," Federal Aviation Administration, Rep. AAC-119-81-5, Washington, DC, USA, 1981.

[38] Vicon, *The Standard Vicon Full-Body Model (Plug-In Gait) Marker Placement Scheme*. [Online]. Available: http://www.idmil.org/mocap/Plug-in-Gait+Marker+Placement.pdf

[39] *Recap2 User's Guide*, Phasespace, San Leandro, CA, USA, 2010.

[40] G. Wu *et al.*, "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion Part I: Ankle, hip, and spine," *J. Biomechanics*, vol. 35, no. 4, pp. 543–548, 2002.

[41] H. Kainz, M. Hajek, L. Modenese, D. J. Saxby, D. G. Lloyd, and C. P. Carty, "Reliability of functional and predictive methods to estimate the hip joint centre in human motion analysis in healthy adults," *Gait Posture*, vol. 53, pp. 179–184, 2017.

[42] T. F. Besier, D. L. Sturnieks, J. A. Alderson, and D. G. Lloyd, "Repeatability of gait data using a functional hip joint centre and a mean helical knee axis," *J. Biomechanics*, vol. 36, no. 8, pp. 1159–1168, 2003.

[43] M. K. Mak, O. Levin, J. Mizrahi, and C. W. Hui-Chan, "Joint torques during sit-to-stand in healthy subjects and people with Parkinson's disease," *Clin. Biomechanics*, vol. 18, no. 3, pp. 197–206, 2003.

[44] J. Musić, R. Kamnik, and M. Munih, "Model based inertial sensing of human body motion kinematics in sit-to-stand movement," *Simul. Model. Pract. Theory*, vol. 16, no. 8, pp. 933–944, 2008.

[45] L. I. Lin, G. McBride, J. Bland, and D. Altman, "A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient," *NIWA Client Rep.*, vol. 45, no. 1, pp. 307–310, 2005.

[46] P. E. Shrout and J. L. Fleiss, "Intraclass correlations-uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.

[47] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological Methods*, vol. 1, no. 1, pp. 30–46, 1996.

[48] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jcm.2016.02.012

[49] D. V. Cicchetti, "Guidlines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology," *Psychological Assessment*, vol. 6, no. 4, pp. 284–290, 1994.

[50] M. Eltoukhy, C. Kuenze, J. Oh, and J. Signorile, "Validation of static and dynamic balance assessment using microsoft kinect for young and elderly populations," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 147–153, Jan. 2018.

[51] F. Ofli, R. Chaudhry, and G. Kurillo, *Berkeley Multimodal Human Action Database (MHAD)*. [Online]. Available: http://tele-immersion.citris-uc.org/berkeley_mhad/

[52] Q. Wang, G. Kurillo, F. Ofli, and R. Bajcsy, "Evaluation of pose tracking accuracy in the first and second generations of microsoft Kinect," in *Proc. - IEEE Int. Conf. Healthcare Inform.*, 2015, pp. 380–389.