Concurrent assessment of gait kinematics using marker-based and markerless motion capture

Robert Kanko1*, Elise Laende1, Elysia Davis1, W. Scott Selbie2, Kevin J. Deluzio1

1Department of Mechanical & Materials Engineering, Queen’s University, Kingston, Canada
2Theia Markerless Inc., Kingston, Canada
*Correspondence: r.kanko@queensu.ca

Abstract
Kinematic analysis is a useful and widespread tool used in research and clinical biomechanics for the estimation of human pose and the quantification of human movement. Common marker-based optical motion capture systems are expensive, time intensive, and require highly trained operators to obtain kinematic data. Markerless motion capture systems offer an alternative method for the measurement of kinematic data with several practical benefits. This work compared the kinematics of human gait measured using a deep learning algorithm-based markerless motion capture system to those of a common marker-based motion capture system. Thirty healthy adult participants walked on a treadmill while data were simultaneously recorded using eight video cameras (markerless) and seven infrared optical motion capture cameras (marker-based). Video data were processed using markerless motion capture software, marker-based data were processed using marker-based capture software, and both sets of data were compared. The average root mean square distance (RMSD) between corresponding joints was less than 3 cm for all joints except the hip, which was 4.1 cm. Lower limb segment angles indicated pose estimates from both systems were very similar, with RMSD of less than 6° for all segment angles except those that represent rotations about the long axis of the segment. Lower limb joint angles captured similar patterns for flexion/extension at all joints, ab/adduction at the knee and hip, and toe-in/toe-out at the ankle. These findings demonstrate markerless motion capture can measure similar 3D kinematics to those from marker-based systems.

Keywords: markerless motion capture; gait analysis; kinematics; deep learning.
1. Introduction

Kinematic analysis is the measurement of the motion of rigid bodies and is an important tool in clinical and research biomechanics. While rigid body kinematics are ubiquitous in biomechanics research, the experimental and computational methods used to estimate three-dimensional (3D) segment position and orientation (pose) vary, and so do the resulting kinematics. Markerless motion capture has the potential to alleviate some of the technical and practical issues of marker-based motion analysis without sacrificing data quality by replacing physical palpation of bony landmarks with probabilistic estimation of segment pose by highly trained neural networks (Mathis et al., 2018).

In traditional marker-based motion analysis, expertise is required to identify anatomical landmarks through physical palpation. This skill necessitates knowledge of anatomy and the translation of this information to a specific individual, a process which is honed over years of experimental work. Lack of skill in this domain leads to the misrepresentation of anatomical landmarks and unrepresentative anatomical coordinate systems and subsequent joint centres, which propagate to errors in reported joint kinematics and kinetics (Della Croce et al., 1999; Holden and Stanhope, 1998; Stagni et al., 2000). This has been noted as one of the greatest sources of error in motion analysis, being greater than instrument error and similar to skin movement artefact (Della Croce et al., 2005). Systematic bias in landmarks for a given examiner may also exist which make the consolidation of different datasets and multi-center collaborations challenging (Gorton et al., 2009; Johnson et al., 2018). In the case of markerless motion analysis, landmark detection is performed by a highly trained deep neural network that applies rules consistently across individuals, thereby dissociating the tracking of human motion from the operator.

Pose estimation in both marker-based and markerless systems is achieved by using at least three noncollinear landmarks to define the pose of body segments, which is represented by a local coordinate system (LCS) relative to a global coordinate system (GCS). In a marker-based system, regression approaches based on anthropometric measurements from medical imaging or cadaveric studies (Harrington et al., 2007), or functional approaches which approximate the center of rotation between two segments (Ehrig et al., 2006; Hicks and Richards, 2005; Schwartz and Rozumalski, 2005) are used to estimate joint center positions. A markerless system estimates the probability of joint center locations directly from an annotated training set of hundreds of thousands of images. Decision rules for joint center positions are estimated in each frame of data by a highly trained neural network based only on the training the network received and the video recorded of the movement.

Reducing the reliance on the expert by removing physical markers increases the ease of collecting data. Not only does this have the potential to improve the quality of data, but also to expand the use of quantitative human movement data to areas where attaching physical markers to a participant is impractical, or an impediment to the research itself. Such instances include when a participant is not comfortable wearing minimal and skin-tight clothing, during movements where sweat reduces marker adhesion and where observation bias overwhelms the research such as behavioral studies. Markerless motion capture has the potential to measure the kinematics of human movement with fewer limitations than marker-based motion capture, including the ability to collect data outside of a laboratory environment.

The aim of this study was to compare full-body 3D gait kinematics determined by a deep neural network-based markerless motion capture software to marker-based kinematic measurements.
2. Methods

2.1 Markerless Motion Capture

*Theia3D* (Theia Markerless Inc., Kingston, ON, Canada) is a deep learning algorithm-based approach to markerless motion capture which uses deep convolutional neural networks for feature recognition (humans and human features) within 2D camera views (Kanko et al., 2020a, 2020b; Mathis and Mathis, 2020). The neural networks were trained on digital images of over 500,000 humans in the wild. In the version used for this study, 51 salient features consisting of a variety of joint locations and other identifiable surface features in the images were manually labelled by highly trained annotators and controlled for quality by a minimum of one additional expert labeller. These training images consisted of humans in a wide array of settings, clothing, and performing various activities. Deconvolutional layers are used to produce spatial probability densities for each image, representing the likelihood that an anatomical feature is in a particular location. For a given image, the network assigns high probabilities to labeled anatomical feature locations and low probabilities elsewhere. The learning that occurs during training enables the application of “rules” for identifying the learned features within a new image.

When using this markerless motion capture software, the user provides newly collected video data from multiple synchronized and calibrated video cameras that capture one or more subjects performing a physical task. The time required to collect data is largely dependent on the task of interest but can take less than five minutes for the collection of ten walking trials, for example. Two-dimensional positions of the learned features are estimated within all frames of all of the videos, which are then transformed to 3D space based on the intrinsic and extrinsic parameters of the cameras. Finally, an articulated multi-body model is scaled to fit the subject-specific landmarks positions in 3D space, and a multi-body optimization approach (inverse kinematic (IK)) is used to estimate the 3D pose of the subject throughout the physical task. By default, the lower body kinematic chain has six degrees-of-freedom (DOF) at the pelvis, and three DOF at the hip, knee, and ankle. This markerless system has been shown to accurately measure spatiotemporal gait parameters, and measure gait kinematics over multiple sessions with greater reliability than marker-based motion capture (Kanko et al., 2020a, 2020b).

2.2 Participants

A convenience sample of thirty healthy, recreationally active individuals (15 male/15 female, mean (SD) age: 23.0 (3.5) years, height: 1.76 (0.09) m, weight: 69.2 (11.4) kg) was recruited to participate in this study at the Human Mobility Research Laboratory (Kingston, ON). Participants gave written informed consent, and this study was approved by the institutional ethics committee. Exclusion criteria included having any neuromuscular or musculoskeletal impairments that could prevent their performance of walking. Participants were provided with minimal, skin-tight clothing, and wore their personal athletic shoes. Retroreflective markers were affixed bilaterally on the first, fifth, and between the second and third metatarsal heads, on the calcaneus, malleoli, tibial tuberosity, femoral epicondyles, anterior superior and posterior superior iliac spines, lateral iliac crest, suprasternal notch, C7 vertebrae, superior acromion, lateral humeral head, humeral epicondyles, radial and ulnar styloid processes, and the third metacarpus. Rigid tracking clusters (4 markers/cluster) were affixed to the shanks, thighs, and lower back, and a headband with five markers was worn.

2.3 Experimental Setup and Procedure

A marker-based camera system (seven Qualisys 3+ (Qualisys AB, Gothenburg, Sweden), 1.3 Megapixel resolution motion capture cameras) and a markerless camera system (eight Qualisys Miqus, 3
Megapixel resolution video cameras) were positioned around an instrumented treadmill (Tandem Force-Sensing Treadmill, AMTI Inc., MA). Both camera systems were connected to a single instance of Qualisys Track Manager for synchronization and to allow them to be calibrated simultaneously, resulting in a single shared global reference frame. Both systems recorded at 85 Hz.

A static calibration trial for the marker-based motion capture data was collected with the subject standing on the treadmill; no static trial is required for the markerless system. Starting at an initial speed of 1.2 m/s, participants determined a comfortable self-selected walking speed by providing feedback to researchers. Participants acclimatized to the treadmill for two minutes before ten consecutive trials of four seconds were collected simultaneously using both camera systems.

2.5 Data Analysis

Markerless motion capture video data were processed using Theia3D (v2020.3.0.962), from which 4x4 pose matrices of each body segment were exported for analysis in Visual3D (C-Motion, USA) alongside the tracked marker trajectories from the marker-based motion capture. Two skeletal models were created in Visual3D; one that tracked the markerless pose matrices, which Visual3D created automatically when data from Theia3D was loaded, and a second that tracked the marker trajectories, which was manually defined. The marker-based model was defined to have identical joint constraints as those of the Theia3D IK model. Kinematic measures were calculated using these models, time normalized over the gait cycle using force-based gait events, and analyzed further in MATLAB (MathWorks, Natick, MA). The joint position estimates, lower limb segment angles, and lower limb joint angles from both systems were compared using the difference between their measurements which were summarized as average root-mean-square (RMS) differences across all subjects. Global segment angles represent the components of the XYZ Cardan angle taken between the segment LCS and the laboratory GCS (defined as Z-Up and Y-Anterior); thus, the x-component represents rotation about the global x-axis, the z-component represents rotation about the axially aligned axis of the segment LCS, and the y-component represents rotation about the floating axis produced by the cross product of the latter with the former. Joint angles were calculated using a Cardan sequence equivalent to the Joint Coordinate System (Grood and Suntay, 1983).

3. Results

Example skeletal models from the two systems are overlaid in Figure 1A for visual comparison. The markerless skeletal model is shown overlaid on the original video data in Figure 1B. Visually, there is substantial overlap between the two models, providing face validity that the pose estimates were similar between the two systems.
Figure 1: Examples of A) the concurrent skeletal models with the global coordinate system, and B) the markerless skeletal model overlaid on original video data, with segment local coordinate systems displayed. The markerless skeletal model is shown coloured red in the image with the concurrent models, and coloured blue in the overlaid image. The global coordinate system is fixed in the treadmill, with the x-axis pointing laterally, the y-axis pointing in the direction of progression, and the z-axis pointing vertically. This subject had average joint position 3D differences at the ankle, knee, hip, shoulder, elbow, and wrist of 2.2 cm, 1.4 cm, 3.9 cm, 2.5 cm, 2.0 cm, and 1.1 cm, respectively. The joint angle RMS differences between the markerless and marker-based models for this subject were 9.9° in ankle flex/ext, 2.1° in knee flex/ext, 5.8° in hip flex/ext, 6.2° in ankle inv/ever, 3.2° in knee ab/ad, 1.9° in hip ab/ad, 8.4° in foot toe-in/-out, 12.4° in knee int/ext rotation, and 2.8° in hip int/ext rotation angles.

3.1 Joint Position Estimates

The average 3D Euclidean distance between corresponding lower limb joints throughout the gait cycle across all subjects was 2.0 cm, 2.2 cm, and 4.1 cm, for the ankle, knee, and hip, respectively. The lower limb joint position differences demonstrated some dependence on gait cycle phase, indicating that there are systematic differences in the estimates from both systems that are dependent on subject pose (Figure 2). The average 3D Euclidean distance between corresponding upper limb joints throughout the gait cycle across all subjects was 2.2 cm, 2.5 cm, and 1.0 cm at the shoulder, elbow, and wrist, respectively. The upper limb joint position estimates were less dependent on gait cycle phase than the lower limb joints, with less variable differences throughout the gait cycle (Figure 3).
Figure 2: Mean +/- SD lower limb joint position differences between both motion capture systems across the gait cycle for thirty subjects. Differences were calculated as [markerless (ML) position] - [marker-based (MB) position] and are expressed as components along each of the global coordinate system axes (row 1: medial/lateral direction; row 2: anterior/posterior direction; row 3: superior/inferior direction) and as 3D Euclidean distances (row 4). Axis labels (e.g. Lateral) indicate the position of the markerless joint relative to the marker-based joint. Root-mean-square (RMS) values are inset in the 3D Euclidean distance plots.
Figure 3: Mean ± SD upper limb joint position differences between both motion capture systems across the gait cycle for thirty subjects. Differences were calculated as [markerless (ML) position] - [marker-based (MB) position] and are expressed as components along each of the global coordinate system axes (row 1: medial/lateral direction; row 2: anterior/posterior direction; row 3: superior/inferior direction) and as 3D Euclidean distances (row 4). Axis labels (e.g. Lateral) indicate the position of the markerless joint relative to the marker-based joint. Root-mean-square (RMS) values are inset in the 3D Euclidean distance plots.

3.2 Lower Limb Global Segment Angles

The lower limb global segment angles measured using both motion capture systems and the average difference between the measurements are shown in Figure 4. The thigh and shank segments were found to have very similar global segment angles between the marker-based and markerless system for the x- and y-components, with average RMS differences of roughly 2°. The thigh and shank segment angle z-components had greater differences between the marker-based and markerless systems, of 8.4°.
and 12.1° respectively. The foot segment angle x-components capture the greatest range of motion among the lower limb segments, and very similar patterns across all subjects were measured by both systems, with an average RMS difference of 5.5°. The foot segment angle y-components (similar to inversion/eversion) had greater differences between systems with an average RMS difference of 7.0°, while the z-components (similar to toe in/toe out) had little difference with an average RMS difference of 2.8°.

Figure 4: Global segment angles of the thigh (row 1), shank (row 2), and foot (row 3) for thirty subjects measured by the markerless (orange) and marker-based (blue) motion capture systems, stacked above the average difference (markerless - marker-based) across all subjects. Segment angle x-components represent rotation about the laboratory
GCS x-axis, z-components represent rotation about the axially aligned axis of the segment LCS, and y-components represent rotation about the floating axis produced by the cross product of the latter with the former. Average RMS differences are inset in each segment angle difference plot.

3.3 Lower Limb Joint Angles

The lower limb joint angles measured using both motion capture systems are shown in Figure 5. Hip flexion/extension angles showed an offset between systems that decreased at heel-strike and toe-off, which resulted in an average RMS difference of 9.8°. The ankle flexion/extension angles showed a similar but smaller offset that resulted in an average RMS difference of 6.6°. The knee flexion/extension angles demonstrated the greatest similarity among sagittal plane angles, with an average RMS difference of 3.1°. For all three joints, the marker-based system measured greater flexion than the markerless system.

The hip and knee ab/adduction joint angles showed high agreement between the marker-based and markerless systems, which had average RMS differences of 2.7° and 3.1°, respectively. The marker-based system typically measured greater knee adduction than the markerless system, particularly during swing phase (60-100% gait cycle). The ankle inversion/eversion angles had greater differences between systems, particularly during stance phase (10-50% gait cycle); the average RMS difference between systems for this angle was 8.0°.

The hip internal/external, knee internal/external, and foot toe-in/toe-out joint angles measured by the marker-based system had greater variability across all thirty subjects, whereas the joint angle waveforms from the markerless system were more similar across subjects. The internal/external hip rotation angles measured by the markerless system were more neutral than those from the marker-based system, and the average RMS difference between systems was 6.8°. The internal/external knee joint angles differed more during early stance and late swing phases, and the average RMS difference was 11.6°. The foot toe-in/toe-out angles measured by the markerless system follow a similar pattern to those measured by the marker-based system, but with lower variability across subjects. The average RMS difference between the foot toe-in/toe-out angles was 10.6°.
**Figure 5:** Joint angles for the hip (row 1), knee (row 2), and ankle (row 3) for thirty subjects measured by the markerless (orange) and marker-based (blue) motion capture systems, stacked above the average difference (markerless - marker-based) across all subjects. Average RMS differences are inset in each joint angle difference plot.

**4. Discussion**

This study compared the kinematics of human gait measured by a marker-based motion capture system and a markerless motion capture system using treadmill gait data of thirty adult subjects, finding comparable results between the two systems.
Across subjects, the joint position estimates differed by less than 3 cm for all joints except the hip, which differed by roughly 4 cm. These differences are similar in magnitude to the errors associated with different marker-based joint center location techniques which have been reported up to 4.59 cm for functional methods and 5.08 cm for predictive regression methods (Kainz et al., 2015). The lower limb joint position differences were more sensitive to gait cycle phase compared to the upper limb joints, which may be a result of greater soft tissue artefacts for the lower limbs or less accurate markerless tracking due to occlusion by the treadmill safety bar or blurred images due to the higher speeds at which the lower limbs travel during walking.

Both systems measured similar segment poses with segment angle x-components, which capture the largest segment rotations during gait, differing by roughly 2° for the shank and thigh, and 5.5° for the foot. Shank and thigh segment angle y-components and foot segment angle z-components had similarly small differences of less than 3° between systems. The largest segment angle differences were found for the foot segment angle y-component and the shank and thigh segment angle z-components, which are angular measurements taken about the axially aligned LCS axis during gait. In combination, the relatively small rotations of these segments about these axes and the alignment of the anatomical reference frames between the segments make these movements challenging for both systems to measure, resulting in greater differences.

The lower limb joint angles captured similar patterns and ranges of motion in flexion/extension at the ankle, knee, and hip, and in ab/adduction at the knee and hip. Joint angles are measured between two adjacent body segments, and therefore any errors in segment pose are amplified in these measurements. The offset in the ankle flex/extention angles is likely a result of the foot segment angle x-component differences, which had an average RMSD of 5.5°. Similarly, the offset in the hip flex/extention angles is likely a result of an offset in the pelvis segment pose between systems, since the differences in thigh segment angle x-components were small, with an average RMSD of 2.0°. Hip and knee internal/external rotation, foot toe-in/toe-out, and ankle inversion/eversion had the most obvious differences between the marker-based and markerless systems, with respect to both the waveform patterns captured and the variability among subjects. The larger differences observed in these joint angles are both notable and challenging to interpret. There are several known factors that affect these marker-based measurements, particularly in off-sagittal plane motions, such as marker placement variation, kinematic crosstalk, soft tissue artefact, and joint position regression errors. However, there may be several unknown factors that affect markerless measurements of these motions as well. Therefore, the differences between systems cannot be interpreted as measurement errors of either system, given that the ground truth is not known and both systems likely contribute to the differences. We chose to present raw waveforms for transparency and to illustrate the differences between systems, which should not be interpreted as errors. Collectively, our results indicate the need to study these small joint movements and the sensitivities associated with their measurement more closely.

There are several potential sources of error that could affect the measurements from both systems. Marker-based kinematics are susceptible to marker placement variation, kinematic crosstalk, soft tissue artefact, and joint position regression errors. Inconsistent marker placement has been shown to contribute up to 5° of error in lower limb joint angles due to the procedure being performed by different operators and following different protocols (Gorton et al., 2009; Schwartz et al., 2004). To minimize this source of error, markers were placed by the same examiner across all subjects. It has also been shown that skin-mounted marker clusters move relative to the underlying bone during gait, with translations of up to 1.5 cm at the shank and 2.5 cm at the thigh, and rotations up to 8° (Benoit et al., 2015). These tissue artefacts...
introduce unpredictable, subject- and task-specific errors of up to $3^\circ$ in knee joint angles (Benoit et al., 2015). Finally, joint position estimation errors have been found to range from 0 cm to 4.6 cm for functional methods and from 0.7 cm to 5.1 cm for predictive regression methods, which can also contribute up to $3^\circ$ of error in lower limb joint kinematics (Kainz et al., 2015; Leboeuf et al., 2019).

Pose estimates from the markerless system may be affected by several collection environment factors. The treadmill on which subjects walked had a safety bar across the front that partially occluded the lower limbs in several camera views (Figure 1). Subjects wore minimal, skin-tight clothing and motion capture markers which were unfamiliar to the neural networks since the training dataset did not include any images of motion capture subjects or markers. Additionally, the video cameras used were necessary for spatial and temporal synchronization with the marker-based system but were not ideal for the markerless system. The video data consisted of 3 Megapixel images recorded at 85 Hz under indoor lighting conditions, producing images that were somewhat low resolution and blurry at times. More appropriate video cameras will provide higher quality images and may improve tracking. Finally, treadmill gait is known to differ from over-ground gait; however, this test compared pose estimation, not characteristics of walking, so any differences are not relevant in the context of this work.

The specific training of the neural networks used for automated pose estimation is a strength of the markerless system but is also a limitation. Any omissions or biases implicit within training sets will be propagated to error when applied to situations where the training was weak. The sensitivity of the markerless motion capture system to various factors including subject characteristics such as age, sex, ethnicity, health status, and clothing have not yet been fully tested, nor has the sensitivity to environmental factors such as lighting, the use of a treadmill, or being in a laboratory environment. These potential sensitivities will require further testing outside of this study. Finally, the subject sample used in this study consisted of young, healthy adults, and so the results may differ for samples with different characteristics.

The results presented here indicate that the overall pose of subjects is measured very similarly between systems, with small joint position differences and similar waveform patterns for most joint and segment angle measures. Given the practical benefits of markerless motion capture such as minimal subject preparation time and reduced collection environment restrictions, this technology could allow studies with truly large sample sizes that were previously not feasible. Furthermore, previous work has indicated that this system offers greater repeatability across multiple collection sessions than marker-based systems, increasing the opportunity for and potential of multi-center trials.

Conflict of Interest Statement
Scott Selbie is the President of Theia Markerless Inc. (Kingston, Ontario), the developers of Theia3D.

Acknowledgements
At the time of the study, RMK was supported by an NSERC Canadian Graduate Scholarship. We thank Human Mobility Research Laboratory members with participant recruitment, data collection, and data processing.
References


