Whole-Body Magnetic Resonance Imaging Enables Assessing Spatial Accuracy and Precision of Skeletal Joint Locations Inferred from Motion Capture Systems

Susan GIBLIN[1], Maeve SMITH[1], Stephen SMITH[1], Stuart O'BRIEN[1],
Jason Philip MCMORROW[2], James MEANEY[2], Friedrich WETTERLING[1]*
[1] Kitman Labs Ltd., Joyce's Walk, Dublin, Ireland; [2] Center for Advanced Medical Imaging (CAMI), St. James's Hospital & Trinity College, The University of Dublin, College Green, Dublin, Ireland

DOI: 10.15221/15.098 http://dx.doi.org/10.15221/15.098

Abstract

Motion capture systems can be used to infer skeletal joints from three-dimensional surface information for various human poses. However, to-date it remains unclear how well the estimated joint coordinates coincide with the anatomically-correct joint positions. The aim of this study was to determine the localization accuracy and localization precision of inferred joint positions using Magnetic Resonance Imaging (MRI). Color and depth information (RGB-D), and skeletal information of an athlete in static pose standing upright were acquired. A whole-body 3D tomographic scan was also recorded using a 3T MRI scanner.
The deviation of the joint location was the largest for the left upper leg (4.1cm±0.2cm) and the smallest for the lower arms (0.2cm±0.01cm). The mean surface point distance averaged 2.2cm±1.3cm (left upper leg), 1.8cm±1.3cm (left lower arm), and 1.5cm±1.0cm (right lower arm).
To our knowledge, this is the first attempt to use MRI as a gold standard to validate skeletal joint locations of a motion capture system. MRI provides a suitable means to validate skeletal joint localization for any motion capture system (markerless and marker based). However, advanced software solutions are required to validate and correct Kinect™ skeletal joint localization in the future.

Keywords: Magnetic Resonance Imaging, Kinect version 2, RGB-Depth, skeletal joint localization, CAPTURE

1. Introduction

Motion capture systems generally measure spatial surface information in order to then solve the inverse problem and infer skeletal joint locations from the prior knowledge of how surface information relates to joint positions for various human poses [1], [2]. However, to-date it remains unclear how accurately one can estimate joint coordinates relative to the anatomically-correct joint positions.
Motion capture systems can provide quantitative biomechanical measurements that link to risk of injury in professional athletes [3]. Specifically, computer vision offers a platform that can address some of the noted limitations of current movement assessment methods. For example, dynamic kinematic or kinetic information about movement is difficult to measure using ecologically valid methods. Comprehensive biomechanical and kinematical investigation has, to date, been restricted to lab-based settings. The constrained and static nature of motion analyses available for deployment by coaches in professional sports settings provide limited information about key movement processes that indicate risk of injury. More specifically, consider the prevention of the anterior cruciate ligament (ACL) among professional athletes. The ACL serves to stabilise the knee joint, preventing excessive translation of the tibia relative to the femur [4]. ACL injury/tear is one of the most severe, career debilitating injuries that occurs in high intensity intermittent sports that involve rapid changes in direction. Risk factors for ACL injury occurrence include dynamic restabilization on jump landing and coordinative dynamics during eccentric loading jump phases, equally asymmetries between knee and hip mechanics during jump movement indicate risk of ACL injury. However measures of knee, hip joint function discretely, or indeed measuring performance outcome (jump height or force production) in isolation, fails to provide optics on these essential precursors to injury (kinetic chain coordination and asymmetry). The identification of these pertinent risk factors is a challenge in sports medicine [5]. Pre-screening could determine anatomical and biomechanical parameters that put athletes at an increased risk of ACL injury [6]–[8].

*corresponding author: Dr. Friedrich Wetterling (fred@kitmanlabs.com)

Marker based motion capture systems are regarded to be the state-of-the-art to validate novel biomechanical data acquisition approaches [2], [9], [10]. However, these systems only measure marker locations on the surface and infer joint location. Hence, a technique is required that can measure both the location of the joint and the surface without any ambiguity. In this paper we propose to use whole body Magnetic Resonance Imaging (MRI) scans as a gold standard to determine the accuracy and precision of motion capture systems. Whole body MRI [11], [12] is a method which enables to measure the anatomically correct locations of the surface and joints of a human - fully non-invasively. Previously, a comprehensive framework was suggested for neurosurgical navigation where facial depth maps were registered with MRI data [13]. Yet, whole body MRI has never been applied to validate motion capture systems.
The aim of this study was to validate CAPTURE (Kitman Labs Ltd., Dublin, Ireland) - a software facilitating musculoskeletal screening of athletes with improved skeletal joint localization [14]. The use of commercially popular motion capture systems, e.g. Microsoft Kinect™, to assess movement has been previously suggested [15]. CAPTURE was developed to record, enhance, display, and analyse skeletal information about athletes for specific pre-defined movements, for example, ankle dorsiflexion (Figure 1). The estimated bone length and the distance of the estimated bone from the detected surface points was compared to the MRI scan - assumed to be the ground truth.

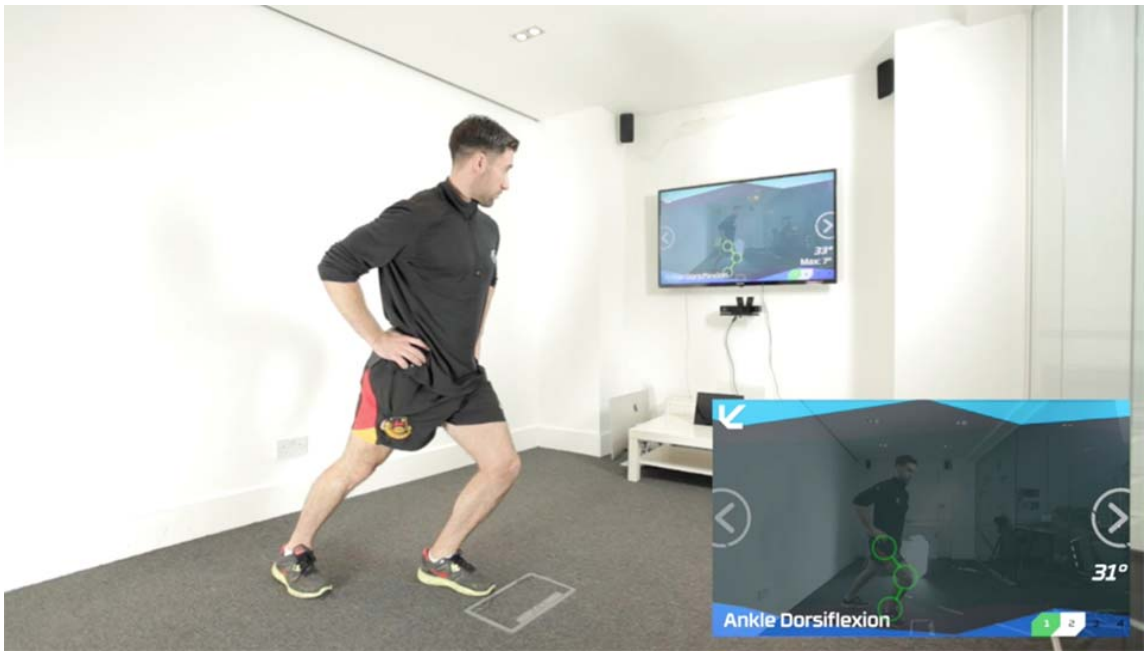

Figure 1: field test of CAPTURE during an ankle dorsiflexion measurement. The result displayed is the angle between the lower leg and a reference vector connecting the knee and the closest point on the floor.

2. Methods

2.1. Motion Capture System

The second version Kinect (Kinect for Windows v2, Microsoft Inc., Redmond, WA, USA) was used to record RGB-Depth (RGB-D) and skeletal information of an athlete in static pose standing upright (neutral stance) facing the Kinect. The depth map was converted into a 3D point cloud and extracted in conjunction with the skeleton (Figure 2) for 56 frames. The raw data was processed using self-written scripts in MATLAB (The Mathworks, Natick, MA).

2.2. MRI data acquisition and joint marking

A whole-body 3D MRI scan of the athlete was acquired using a 3T MRI scanner, and a coronal T_2-weighted Turbo Spin Echo (TSE) sequence with a voxel resolution of (1.5x1.5)mm^2 in-plane resolution and 8mm slice thickness (9mm interslice thickness). The following sequence parameter determining the imaging contrast were chosen: 270Hz pixel bandwidth, 2179ms Time of Repetition (TR), 86ms Time-to-Echo (TE), 8 segments, 256 (read direction) x 176 (phase direction) acquisition matrix per segment, an acquisition duration of 5min 50sec per segment, 8 k-space trajectories were acquired per breath hold for scans of the abdomen, the entire image stack contained 24 coronal slices (Figure 3). The final whole body image had a matrix size of 1269 rows and 320 columns.

The joint locations were marked manually in the appropriate slices to form the MRI skeleton and the outer surface of the athlete was reconstructed from the MRI scans (Figure 4). The matrix size was scaled according to the spatial pixel and slice resolution in order to arrive at a geometrically correct representation.

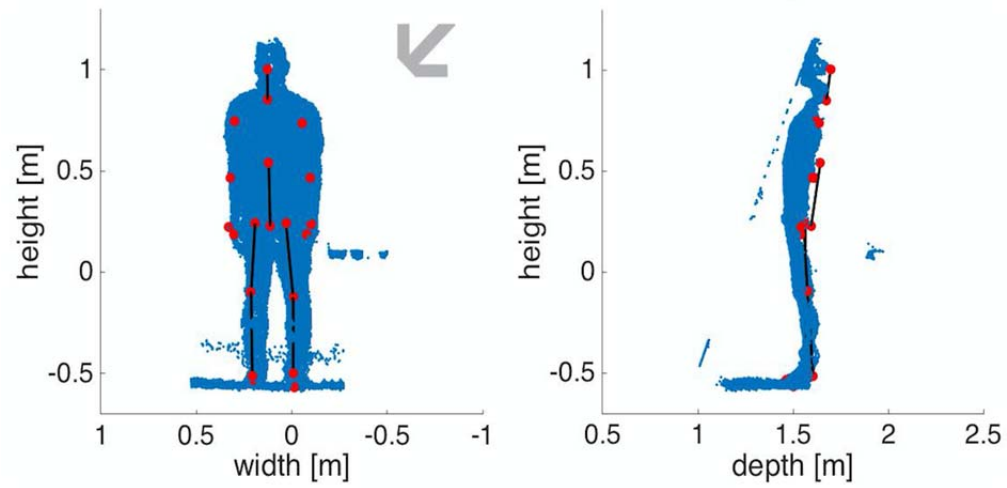

Figure 2: RGB-D point cloud (blue dots) in front (left) and side (right) view together with RGB-D skeletal points inferred from RGB-D raw data (red dots).

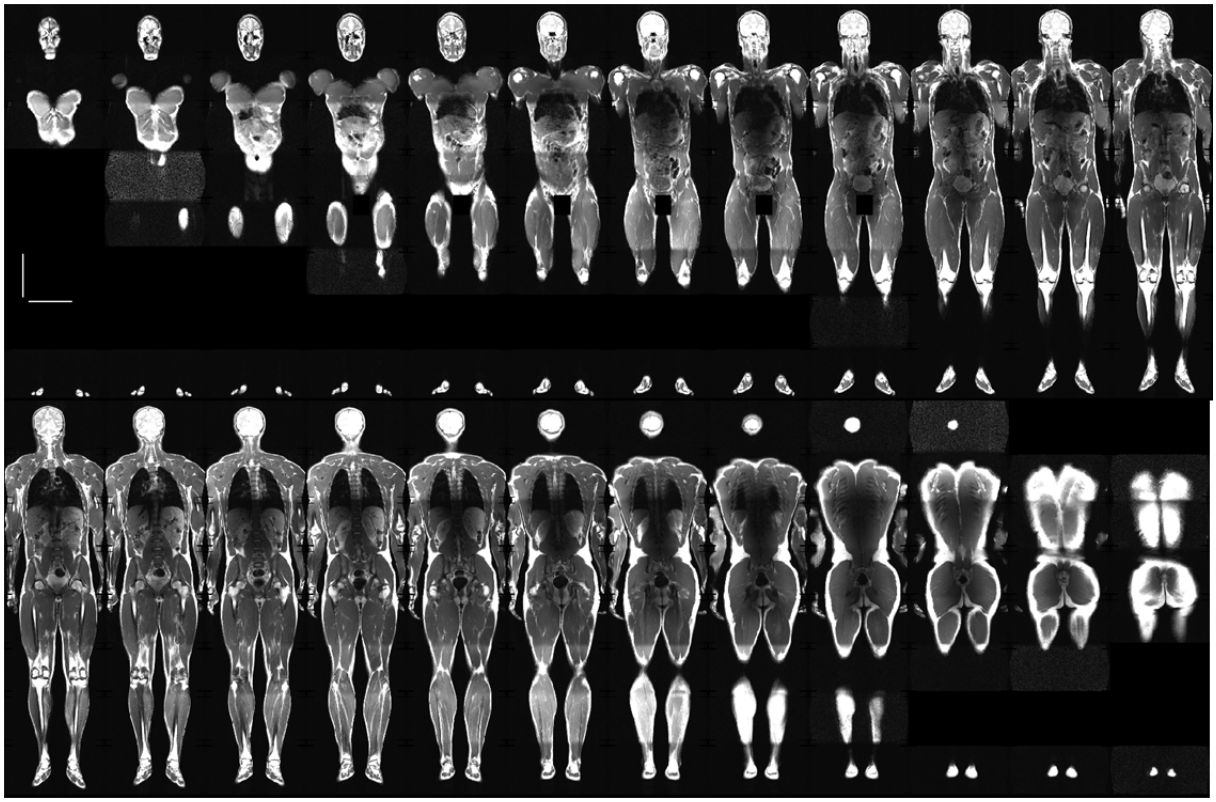

Figure 3: T_2-weighted spin-echo MRI data set of an athlete. Note the excellent resolution of the surface tissue and the skeletal structure. The vertical and horizontal scale bars in the top left corner image (top left corner) indicate a length of 20cm.

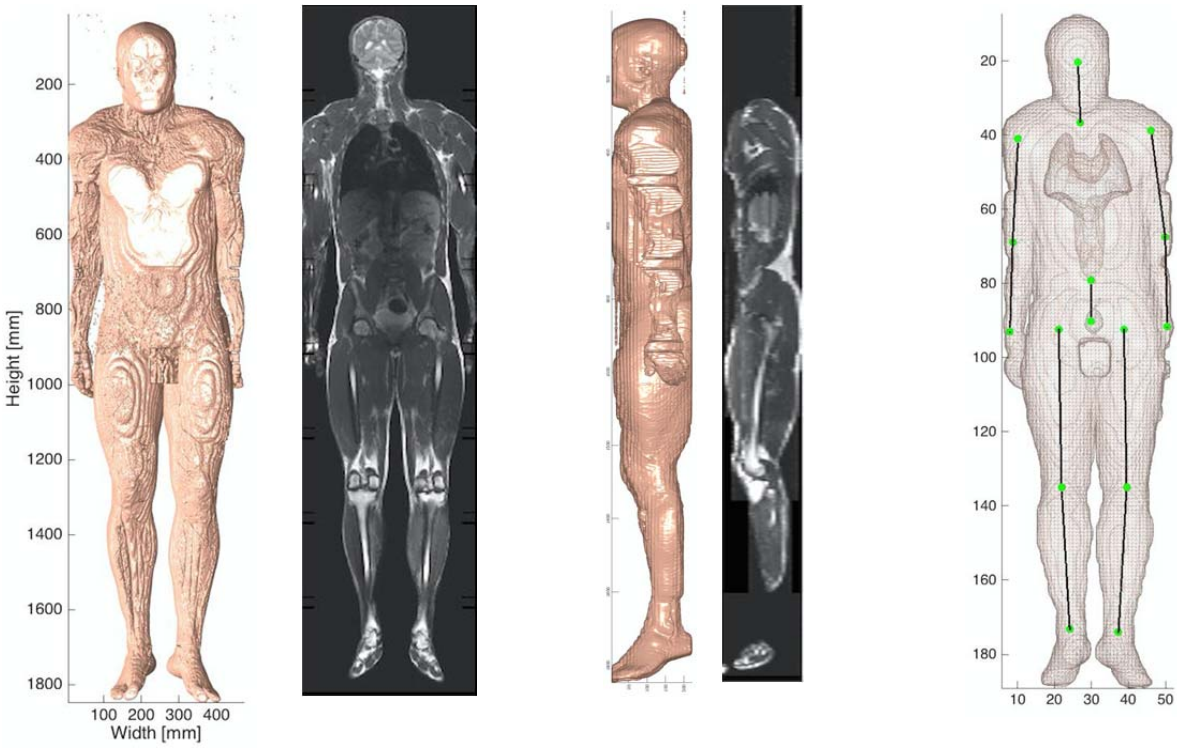

Figure 4: Surface reconstruction of athlete in front view and side view. Transparent surface rendered in 3D coordinate system with labels given in centimeters showing internally located, manually marked joint locations (green dots, right figure).

2.3. MRI and RGB-D registration

Assuming that limbs held by bones underly no deformation for different poses and hence the limb can be registered separately. The joints at each end of the reconstructed limb were used. Two more points were added as supportive markers to facilitate 3D registration. The first point was placed at half length of the bone. The second point was placed at the end point of the vector starting at the half distance point pointing backwards relative to the main plane (defined by the shoulders and hips) of the skeleton and being orthogonal to the limb. The length of this vector was defined as being half the length of the limb. A graphical example of the marker point selection for the RGB-D and MRI skeleton is shown in Figure 5.

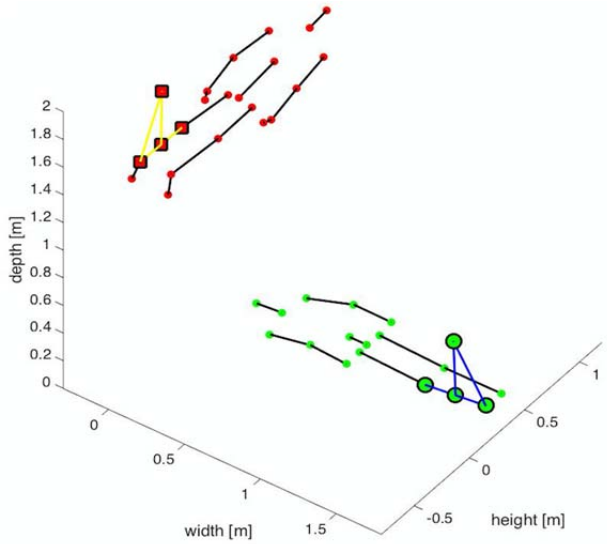

Figure 5: RGB-D skeleton (red dots, upright) and MRI skeleton (green dots, horizontal) before coregistration. Note that coregistration was processed on a joint-by-joint basis.

In a first registration step, the reference points were translated relative to the limb's center point from the RGB-D limb to the center point of the MRI limb (Figure 6, left). In a second registration step the three remaining markers were used to compute the 3D rotation necessary to rotate the RGB-D limb around the center point so that all reference markers coincide (Figure 6, right). The rigid body translation and rotation parameters were stored.
The RGB-D points closest to the limb under investigation were extracted by subdividing the limb into ten segments and selecting surface points, if they were less than 8 cm away from one of the ten segments. The extracted points were registered to the MRI space using the previously stored registration parameters specific to the limb. The nearest neighbour MRI point relative to each of the extracted RGB-D points was extracted to form the MRI surface mask (Figure 7).

2.4. Joint mismatch and surface distance measurements

After limb registration the distance of each of the two MRI joints relative to the RGB-D joint was computed as a measure of joint localization error. In a further step, each point of the RGB-D point cloud in near proximity to the limb under investigation was then transformed to the MRI space. The distance of the nearest MRI surface neighbour to each RGB-D point was computed (Figure 8). The following equation was used:

$$SurfaceDistance = \frac{1}{N}\sum_{i=1}^{N}[p_i(RGBD) - p_i(MRI)]$$

where p_i is the i^{th} point of the point cloud containing the x,y, and z coordinates. The standard deviation of the surface distance was computed.

2.5. Depth precision as a function of distance

In order to assess the localization errors and surface distance measurements it is of importance to understand the precision of the depth measurement as a function of distance from the sensor. The very same RGB-D sensor has been used to record depth frames acquired for various distances from a flat wall. The sensor was left standing in each position for up to 5secs. Ten different sensor positions were investigated. The depth variation was characterised as the difference between depth values in the same depth map pixel as measured in two consecutive frames.

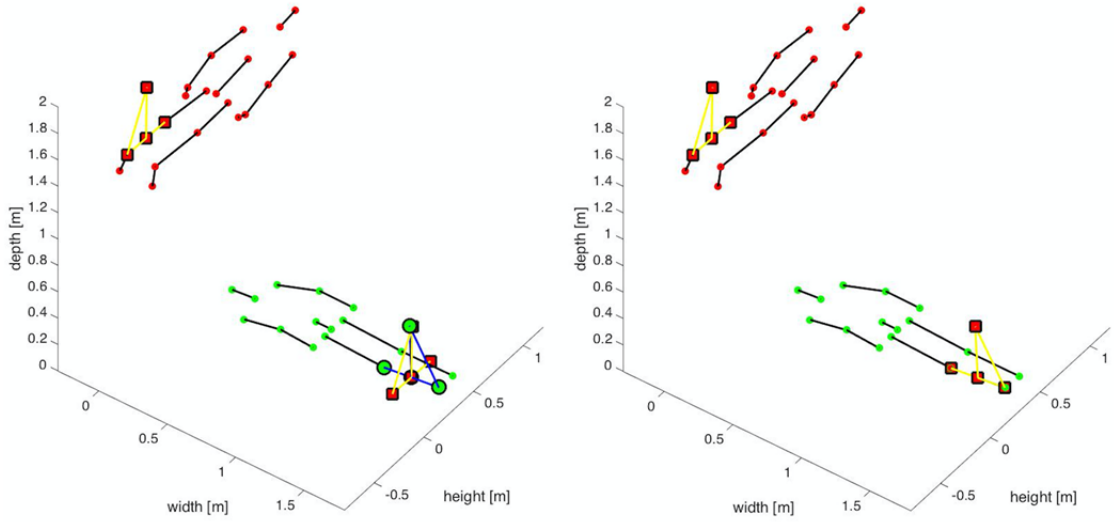

Figure 6: Translation (left) and rotation (right) of reference points from RGB-D frame (yellow line and red squares) to MRI frame (green squares). The transformation parameters were then used to align the RGB-D point cloud to the MRI point cloud.

This value was then stored in relation to the mean distance measured in this pixel. Pixels in a square region centered in the depth map frame were analysed with the square region measuring 100 pixels vertically and 100 pixels horizontally.

3. Results

An example of the extracted surface points and limbs for the upper limb bone is shown in Figure 7. The two surfaces extracted for one frame match well. The mismatch between the limb length is obvious for this most extreme case. The quantitative results for all investigated limbs are listed in Table 1.

The deviation of the joint location was the largest for the left upper leg (4.1cm±0.2cm) and the smallest for the lower arms (0.2cm±0.01cm). The mean surface point distance averaged 2.2cm±1.3cm (left upper leg), 1.8cm±1.3cm (left lower arm), and 1.5cm±1.0cm (right lower arm). The variation of the depth measurement has been assessed under real world conditions of an average gym. The result for different field of views are plotted as a function of distance in Figure 8. The outer ranges of the field of view presented with larger variations than the central part of the depth map and there was a strongly exponential and increased drastically (>10mm) with distances larger than 2.5m.

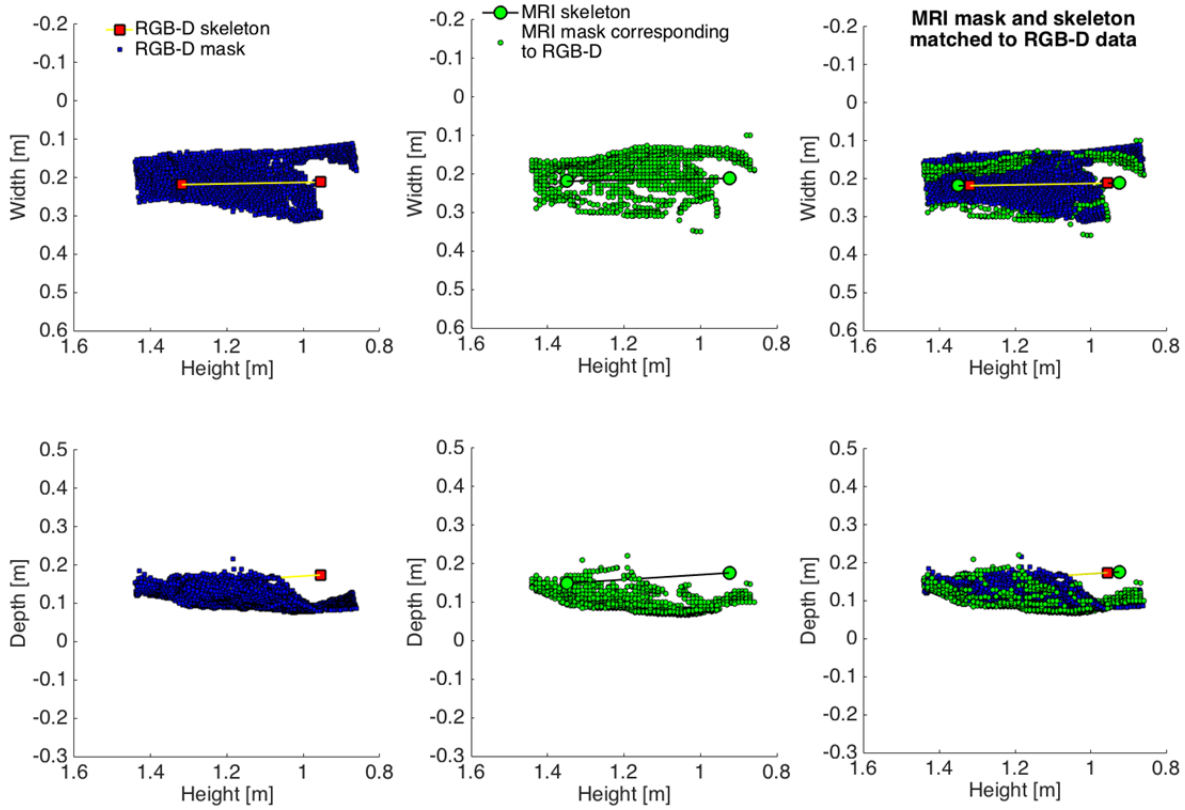

Figure 7: coregistered limb with MRI surface and RGB-D surface.

Table 1: quantitative results for 56 analysed RGB-D frames.

investigated limb	Distance of MRI *joint* from RGB-D *joint* [cm] mean ± standard deviation	Distance of MRI *surface* from RGB-D *surface* [cm] mean ± standard deviation
upper leg right	3.0 ± 0.2	2.0 ± 1.3
upper leg left	4.1 ± 0.2	2.2 ± 1.3
lower leg right	0.7 ± 0.2	1.7 ± 1.1
lower leg left	1.2 ± 0.2	1.6 ± 0.9
upper arm right	1.0 ± 0.1	1.9 ± 1.3
upper arm left	0.8 ± 0.1	1.9 ± 0.9
lower arm right	0.2 ± 0.01	1.5 ± 1.0
lower arm left	0.2 ± 0.01	1.8 ± 1.3

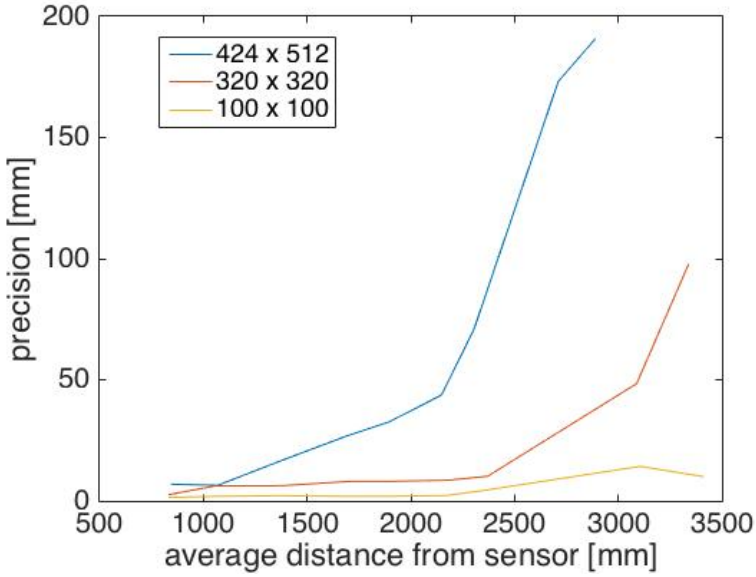

Figure 8: variation of the depth measurement for Kinect version 2 as a function of distance for different field of views as centered around the depth map's mid point.

4. Discussion

To our knowledge, this is the first attempt to use MRI as ground truth to validate skeletal joint locations of a motion capture system. While MRI is expensive, it is non-invasive and it provides equally precise localization of the joint centers and the surface tissue. Although, MRI can be prone to distortions in the outer ranges of the field of view which were especially visible at the arms, advanced adjustment methods can reduce the average distortion to 1 mm, while distortions were reported to vary between 0.7 mm and 3.7 mm without correction [16].

The large joint mismatch measured using the motion capture system in this study was 41mm for the upper leg. This deviation was expected due to the location of the hip joint being fully encapsulated in thick layers of muscular tissue making joint inference more difficult compared to other joints such as the knee, elbow, wrist, and ankle. Furthermore, the rugby player investigated presented with above average muscle mass in shoulders and thighs which may have lead to the observed joint mismatch. Others reported errors for Kinect version 2 ranging from 32mm to 207mm comparing it to a marker based system [17]. This demonstrates the need for more reliable validation methods, as suggested herein, in order to reliably assess accuracy and precision of motion capture systems in the future.

Considering the MRI surface distance from the point cloud recorded using the RGB-D sensor, the mean surface distances for points related to the upper leg was 22mm ± 13mm, while the points of the best fitting surface (the lower right arm) were on average 18mm ± 13mm apart. The standard deviation of the distance between points was measured to be within 9 to 13mm - well within the precision reported for Kinect version 2 at 2m distance from the sensor [18]. The absolute resolution precision of the sensor depends on the 70° horizontal, the 60° vertical resolution, and the matrix resolution of 512 x 424 pixel. Subsequently, the in-plane resolution limit at 2m distance from the sensor is approximately ±3.5mm. Hence, the error measured in this study most likely results from imperfections due to the depth axis. The expected variation in the depth measurement was approximately ±8mm. Lachat et al. reported a standard deviation of the depth measurement of 16mm at a distance of 3m from the sensor [19]. Yet, the deviation of the surface point locations when compared in relation to the inferred joints was between 18 to 22mm which is slightly larger than the expected depth precision of the sensor of 16mm at the investigated distance. The additional imprecision measured in this study may indeed have originated from localization errors caused by the algorithm used to infer the joint positions. The largest joint mislocalization of 40mm for the upper leg confirms this hypothesis.

Future low cost markerless motion capture systems may for instance comprise multiple cameras enabling acquisition of a full 360° view of the athlete [20] in order to improve the joint localization.

In conclusion, MRI provides a suitable means to validate skeletal joint localization for motion capture systems (markerless and marker based). By this means, CAPTURE has been validated in terms of general joint localization accuracy and precision. As a result more advanced software solutions can be developed to correct Kinect™ skeletal joint localization in the future.

References

[1] T. B. Moeslund, A. Hilton, and V. Krüger, “A survey of advances in vision-based human motion capture and analysis,” *Comput. Vis. Image Underst.*, vol. 104, no. 2–3, pp. 90–126, Nov. 2006.

[2] 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, Jul. 2014.

[3] T. E. Hewett, G. D. Myer, K. R. Ford, R. S. Heidt Jr, A. J. Colosimo, S. G. McLean, A. J. van den Bogert, M. V. Paterno, and P. Succop, “Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study,” *Am. J. Sports Med.*, vol. 33, no. 4, pp. 492–501, Apr. 2005.

[4] D. L. Butler, F. R. Noyes, and E. S. Grood, “Ligamentous restraints to anterior-posterior drawer in the human knee,” *J. Bone Joint Surg. Am.*, vol. 62, no. 2, pp. 259–270, 1980.

[5] P. Renstrom, A. Ljungqvist, E. Arendt, B. Beynnon, T. Fukubayashi, W. Garrett, T. Georgoulis, T. E. Hewett, R. Johnson, T. Krosshaug, B. Mandelbaum, L. Micheli, G. Myklebust, E. Roos, H. Roos, P. Schamasch, S. Shultz, S. Werner, E. Wojtys, and L. Engebretsen, “Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement,” *Br. J. Sports Med.*, vol. 42, no. 6, pp. 394–412, Jun. 2008.

[6] J. M. Hootman, R. Dick, and J. Agel, “Epidemiology of collegiate injuries for 15 sports: summary and recommendations for injury prevention initiatives,” *J. Athl. Train.*, vol. 42, no. 2, pp. 311–319, Apr. 2007.

[7] S. C. Landry, K. A. McKean, C. L. Hubley-Kozey, W. D. Stanish, and K. J. Deluzio, “Neuromuscular and lower limb biomechanical differences exist between male and female elite adolescent soccer players during an unanticipated run and crosscut maneuver,” *Am. J. Sports Med.*, vol. 35, no. 11, pp. 1901–1911, Nov. 2007.

[8] R. L. Lathrop, A. M. W. Chaudhari, and R. A. Siston, “Comparative assessment of bone pose estimation using Point Cluster Technique and OpenSim,” *J. Biomech. Eng.*, vol. 133, no. 11, p. 114503, Nov. 2011.

[9] R. A. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. E. Webster, L. Denehy, and A. L. Bryant, “Validity of the Microsoft Kinect for assessment of postural control,” *Gait Posture*, vol. 36, no. 3, pp. 372–377, Jul. 2012.

[10] R. A. Clark, Y.-H. Pua, A. L. Bryant, and M. A. Hunt, “Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining,” *Gait Posture*, vol. 38, no. 4, pp. 1064–1066, Sep. 2013.

[11] F. Schick, “Whole-body MRI at high field: technical limits and clinical potential,” *Eur. Radiol.*, vol. 15, no. 5, pp. 946–959, May 2005.

[12] F. Wetterling, D. M. Corteville, R. Kalayciyan, A. Rennings, S. Konstandin, A. M. Nagel, H. Stark, and L. R. Schad, “Whole body sodium MRI at 3T using an asymmetric birdcage resonator and short echo time sequence: first images of a male volunteer,” *Phys. Med. Biol.*, vol. 57, no. 14, pp. 4555–4567, Jul. 2012.

[13] L.-X. Zhang, S.-T. Zhang, H.-Z. Xie, X.-H. Zhuang, and L.-X. Gu, “Kinect-based automatic spatial registration framework for neurosurgical navigation,” *J. Shanghai Jiatong Univ.*, vol. 19, pp. 617–623, 2014.

[14] S. Smith, F. Wetterling, and S. Giblin, “PATENT: Correcting Skeletal Joint Locations,” 11-Jun-2015.

[15] S. Giblin, D. Collins, and C. Button, “Physical literacy: importance, assessment and future directions,” *Sports Med.*, vol. 44, no. 9, pp. 1177–1184, Sep. 2014.

[16] W. M. Tavares, F. Tustumi, C. da Costa Leite, L. Fernel Gamarra, E. Amaro Jr, M. J. Teixeira, and E. T. Fonoff, “An image correction protocol to reduce distortion for 3-T stereotactic MRI,” *Neurosurgery*, vol. 74, no. 1, pp. 121–6; discussion126–7, Jan. 2014.

[17] X. Xu and R. W. McGorry, “The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures,” *Appl. Ergon.*, vol. 49, pp. 47–54, Jul. 2015.

[18] K. Khoshelham and S. O. Elberink, “Accuracy and resolution of Kinect depth data for indoor mapping applications,” *Sensors* , vol. 12, no. 2, pp. 1437–1454, Feb. 2012.

[19] E. Lachat, H. Macher, M. A. Mittet, T. Landes, and P. Grussenmeyer, “First Experiences With Kinect V2 Sensor for Close Range 3D Modelling,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, 2015.

[20] K. Amplianitis, M. Adduci, and R. Reulke, “Calibration of a Multiple Stereo and Rgb-D Camera System for 3d Human Tracking,” *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, no. 1, pp. 7–14, 2014.