Biomechanical Validation of Upper-body and Lower-body Joint Movements of Kinect Motion Capture Data

Adso Fernández-Baena
La Salle - Universitat Ramon Llull
adso@salle.url.edu

Antonio Susín
Universitat Politecnica de Catalunya
toni.susin@upc.edu

Xavier Lligadas
LABSID SL.
lligadas@labsid.com

Abstract

New and powerful hardware like Kinect introduces the possibility of changing biomechanics paradigm, usually based on expensive and complex equipment. Kinect is a markerless and cheap technology recently introduced from videogame industry. In this work we conduct a comparison study of the precision in the computation of joint angles between Kinect and an optical motion capture professional system. We obtain a range of disparity that guaranties enough precision for most of the clinical rehabilitation treatments prescribed nowadays for patients. This way, an easy and cheap validation of these treatments can be obtained automatically, ensuring a better quality control process for the patient's rehabilitation.

Keywords: motion capture, markerless motion capture, depth camera

1 Introduction

Motion capture techniques are used over a very broad field of applications, ranging from digital animation for entertainment to biomechanics analysis for clinical and sport applications. Although there are other technologies, like inertial [1][2] or electromagnetics sensors [3], at present, using optical systems with reflective markers is the most common technique [4] [5] [6]. Despite their popularity, marker based methods have several limitations: usually a controlled environment is required to acquire high-quality data and the time required for marker placement can be excessive [7][8][9]. Several recent review articles have summarized the common shortfalls of skin based marker techniques [10] [11] [12]. Markerless motion capture offers an attractive solution to the problems associated with marker based methods, but the general problem of estimating the free motion of the human body is underconstrained without the spatial and temporal correspondence that tracked markers guarantee.

From the powerful game industry new devices like Kinect [13] have been appear, allowing to interact with game consoles in real time. Moreover, this new hardware is considerably cheaper than the usual complex multicamera systems. Kinect can be thought as a 3D markerless motion capture system because it gives you a simplified skeleton in real time. No especial dress or other equipament is required. The skeleton is made of 15 joints (see Figure 5) and due to its simplification it can not be used (by now) for very accurate studies. Because of that, we aim to use it when such accuracy it is not needed, like clinical rehabilitation where the correctness of a motion can be validate without been extremely precise. For these kind of applications, in this paper we consider the validation of the Kinect data in terms of joint angles when motion of the main limbs is involved. We compare these data with a professional motion capture equipment and we compute the error along the complete capture.

For the biomechanics community and clinical therapy in general, it is needed a validation of the precision of this new devices and to understand the possible appropriate applications for these cheap and portable technology. As it is shown in section 4, the obtained accuracy for the measurements of the angle joints are enough for most of the prescribed exercises in rehabilitation treatments. As a conclusion, we think that a new series of useful applications using these new technology can be developed according to our results.

The rest of the paper is organized as follows: section 2 describes the equipment used in our study, section 3 describes the motion capture performance and in section 4 we present the results.

2 Equipment description

Motion capture is the process of registering movements performed by a person, who is called actor, by some mechanisms or devices. Motion capture systems can be categorized in two main groups: markers motion capture or markerless motion capture. Capturing motion with markers implies using some kind of sensors or devices that the actor must wear to help cameras recognizing motion or to send data to a manager system for further treatment. This fact can produce some distortion in motion because actor could be uncomfortable or simply constrained for this devices. In the other hand, markerless motion capture avoid these problems because is based on computer vision algorithms that deal with images from capturing cameras. Although, this type of motion capture usually has less precision than markers one.

Markerless motion capture is based on how computer vision algorithms interprets captured images from cameras. There is a lot of literature about this type of algorithms. Moeslund has been done an extens review in [14] [15]. In addition to the algorithms is very important the type of the cameras and the set up. Markerless motion capture could be done using a camera, using stereographic cameras, 3D cameras or multiple cameras. Using only a unique camera avoid synchronization problems faced by systems with multiple cameras, although these increase precision in tracking results.

In this project, we have been compared motion capture data from Kinect against optical motion capture data. In the following points there is a description of both used systems.

2.1 Optical Motion Capture

As we have mentioned before, optical motion capture systems are the most popular and the most used in videogames and medical fields. In [16][17] there are an extensive explanation of optical motion capture process detailing all steps included. A typical optical system consists in a set of cameras from 4 to 32 and a computer that manage them. Usually, actor takes some markers that are reflective (passive) or emitters (active). Passive markers are made with reflective materials and its shape is spherical, semispherical or circular. Markers are place directly over actor skin or over an elastic suit. In passive systems, cameras are equipped with infrared LEDs diodes and light emitted is reflected by markers. In the other hand, markers in active systems are LEDs.

Cameras in these systems can capture between 30 and 2000 frames per second. At least, two cameras have to visualize one marker in order to determine its 3D position, although it is better than three or more cameras for better precision. In some occasions, who is been captured, or another actor, or some object can occlude some of the markers. When these occlusions exist, any camera can see these markers and this cause data losing. After capture sessions, motion data is cleaned trying to remove some noisy data and recovering missing markers. Because of this, optical motion capture data is very accurate and we will use it as reference for testing Kinect motion capture accuracy.

In this work we have been used MediaLab [18] facilities. MediaLab is a passive optical motion capture laboratory belonging to La Salle - Universitat Ramon Llull (Barcelona). This laboratory has 24 Vicon [19] cameras that allow a 45 m^2 of capture volum.

2.2 Kinect Motion Capture

Evolution of entertainment industry and continuous development of digital devices, it has been manufactured cameras capable of generate 3D models. This type of cameras are called depth cameras. Planar images captured by traditional cameras lose 3D information about the scenes. Depth information is very useful and important in order to visualize and perceive real world environment, so this feature permits increment the interactivity and user experience in virtual worlds. There are some type of 3D cameras on market. Nowadays, Kinect is the most popular.

Kinect device appears in November of 2010 as a entertainment device of Microsoft Xbox [20] console. It is based on software developed by Rare [21], Microsoft Game Studios affiliated company, and the technology of PrimeSense [22] cameras. These cameras interpret 3D scene information based on projected infrared light system called Light Coding. Kinect RGB camera uses a 8 bits VGA resolution (640x480 pixels) while its monocrom depth sensor has a VGA resolution of 11 bits that allows 2048 sensibility levels. Both video outputs work at 30 frames per second. Kinect device has an approximate depth limitation from 0.7 to 6 meters. Horizontal angular field of view is 57^o and 43^o vertically. Horizontal field of view has a minimum distance around 0.8 meters and 0.63 meters in vertical, so Kinect has an approximate resolution of 1.3 millimeters per pixel.

Thus, Kinect is a device capable to extract color and depth information from scenes. In order to use Kinect as a motion capture system we need an specific software connected to it. OpenNI [23] framework and Primesense's NITE (Natural Interaction Technology for End-user) [24] middleware have been used. OpenNI is a multiplatform framework that defines API's for natural interaction applications development. This framework offers a set of libraries to be implemented by devices and middleware components. Primesense's NITE middleware can work together with OpenNI and includes a set of computer vision algorithms capable of converting depth images into useful information.

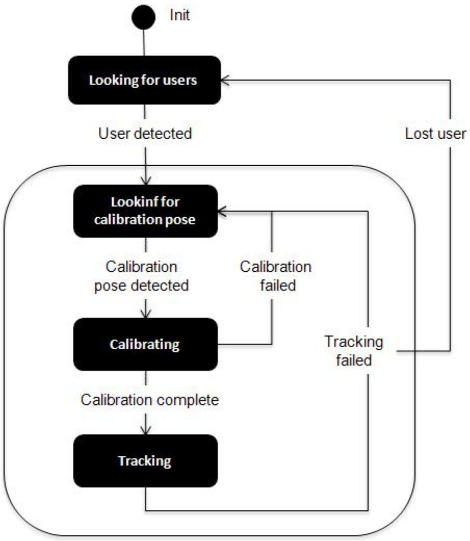

Figure 1: User segmentation and user tracking workflow.

NITE's algorithms [25] we have used are user segmentation and user tracking. In Figure 1 there is on overview of how these algorithms work together. User tracking algorithm needs user segmentation to be successful. Once the system is able to segment an user, it starts the user tracking algorithm. First, user tracking needs to calibrate the segmented user. It is mandatory that user performs a calibration pose and waits up to calibration is complete. From this moment, the system starts to report global positions of estimated joints and its global orientation. Skeleton profile is described in

Figure 5.

3 Capture Description

We want to record motions with optical motion capture and Kinect simultaneosly, and this leads to mount Kinect device inside the motion capture laboratory. As we can see in Figure 2, Kinect was mounted above a tripod placed in front of the actor. In case of the optical motion capture, it was normally set.

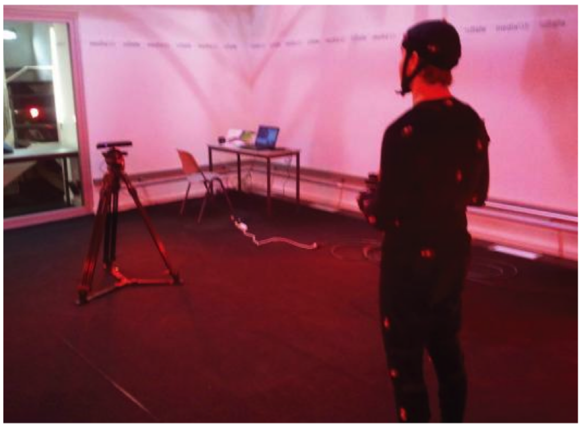

Figure 2: Actor in motion capture laboratory with a Kinect mounted in front of him.

To test Kinect accuracy, we have studied movements from upper body and lower body following all body planes (Figure 3). So, we select three joints: shoulder, hip and knee. For the shoulder we have recorded movements in all planes of the body, in case of hip, sagital and coronal rotations. To validate the movement of the knee, flexion and extension movements were recorded.

Recording motion with two independent systems requires some kind of synchronization when you start recording a movement. Additionally, each system needs some specific steps to be initialized. In case of Kinect, it is indispensable to calibrate user in order to get joint positions. So, at the beginning of the capture session, user was asked to be calibrated by Kinect system. From this point, Kinect was reporting tracking information. In the other hand, optical motion capture workflow forces actor to start every clip in t-pose for later easy data cleaning. So, we manually activate both systems when actor was in t-pose. Each system capture motion and save it in a specific way that will be explained in depth in the following sections. Temporal synchronization between pairs of motions is obviously not very accurate. In Section 4 we will explain how we have fixed synchronization problem. Finally, we have obtained the amount of data described in Table 1. We have recorded capture sessions in video too.

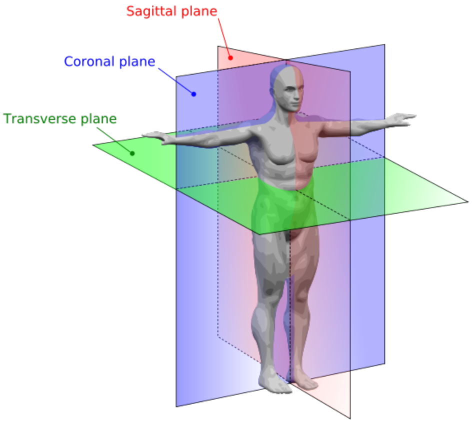

Figure 3: Body planes.

Joint	Plane of movement	Frames
Knee	Relative to parent	5998
Hip	Sagittal	3152
Hip	Coronal	1893
Shoulder	Sagittal	4261
Shoulder	Coronal	3114
Shoulder	Traversal	2018

Table 1: Recorded movements.

3.1 Optical Motion Data

As we have explained before, to capture motion using an optical system actor have to wear attached optical markers whose positions are reported. On the contrary, Kinect motion capture system reports joint positions. So, in order to compare them it is mandatory to adapt one representation to the other. We have created a markers configuration to calculate joint positions from markers. We have placed two markers in each joint that Kinect system is able to track. Markers have been placed by a therapist, trying to minimize a bad placement that will affect joint positions computing. In Figure 4 can be seen in detail markers configuration.

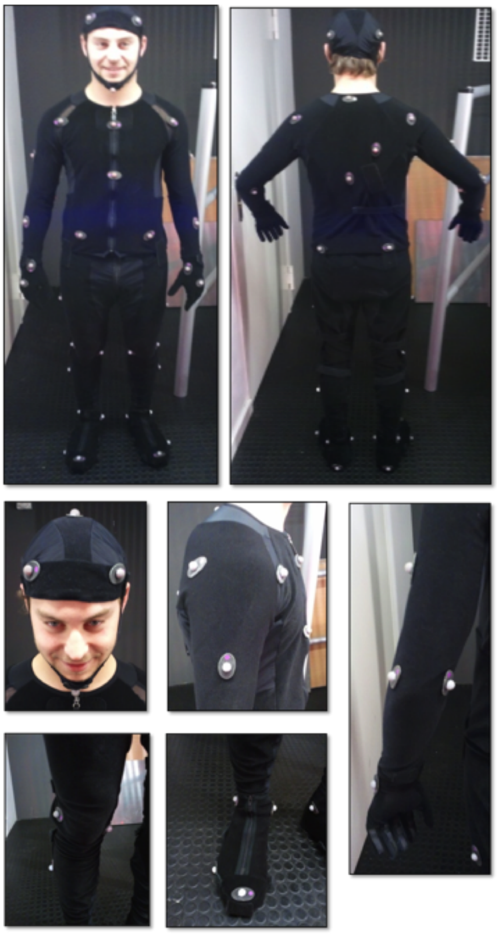

Figure 4: On top, placement of markers on the actor's body on both sides; and below, detail of the placement of markers for calculating the position of the head, shoulder, elbow, knee and foot.

Optical data was recorded with a framerate of 120 and markers positions were expressed in world coordinate system. After making the recordings, optical motion data was cleaned and post-process to eliminate mistakes in capture and recover missing markers. So, each optical capture was perfectly prepared and validated.

3.2 Kinect Motion Data

Kinect motion data depends on the joints that NITE user tracking algorithm can detect. This algorithm is able to report joints positions and global orientations. Although we want to validate joint rotations, we have been used joint positions. We made this decision because orientation data is computed from position data, so errors from position data are propagated. Therefore joint positions are more reliable. Joint specification is shown in Figure 5. Joint positions are referred to a world coordinate system with the origin placed in Kinect device. Tracking algorithm also report a confidence value for each joint. This value is a boolean parameter that indicates if joint position is reliable for the system. We have saved this information for each joint too. Apart from this, as we have mentioned in Kinect device description, Kinect can report camera data at 30 frames per second. So, Kinect motion data was recorded at this framerate.

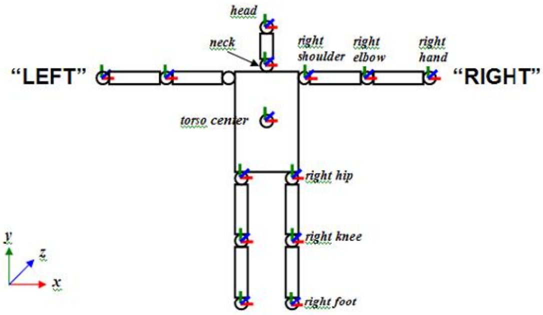

Figure 5: OpenNI joint specification.

Estimated positions by Kinect suffer a lot of noise. Tracking algorithm recompute at each frame all joint positions regardless of temporal continuity, this fact produces noise mentioned. Due to this problem we have been used a smooth lowpass filter to eliminate local fluctuations.

4 Motion Data Processing

In this section we describe how we have processed motion data from both capture systems to compare results. Data correspondence from motion data consists in the conversion of markers position data to joint position data, temporal and spatial alignment of each pair of motions. After that, we have computed joint rotations which we want to evaluate.

4.1 Data correspondence

We converted the optical motion capture data to the representation of Kinect motion data in order to enable comparison between them. First of all we have aligned the same movement from both systems in terms of coordinate system. Optical motion data is reported in its world coordinates ((a) in Figure 6), and the same for Kinect data ((d) in Figure 6). Both coordinate systems have the same orientation but they have exchanged axes. We have corrected it as shown (b) in Figure 6.

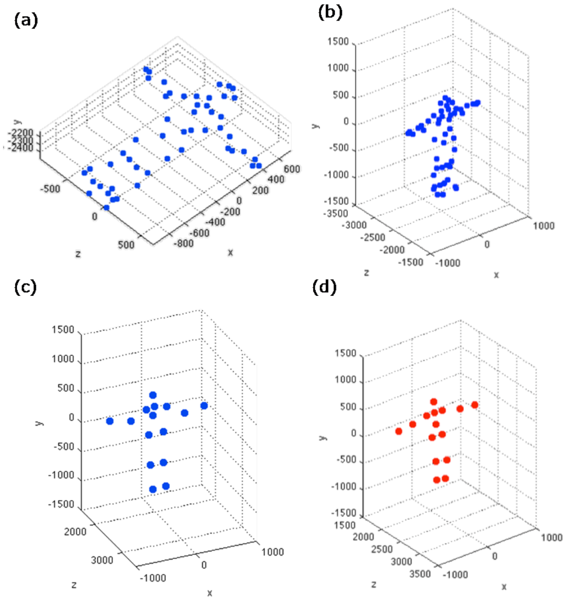

Figure 6: Steps from data correspondence. (a) Markers from optical motion capture. (b) Rotated markers from optical motion capture. (c) Joint positions from optical markers. (d) Joint positions from Kinect motion capture.

Joint correspondence consists in compute joint positions from optical motion data of the markers. As we have explained before, we have used an specific markers placement to solve this. Each joint position that Kinect is able to track is estimated from two optical markers. To achieve this is very simple since you just have to calculate the midpoint between marker pairs. In (c) of Figure 6 it can be observed the result of this step. Now the appearence of optical data is like Kinect data.

Capturing with two different systems produce some temporal differences according to technical equipment features. First, there is a difference between framerates used in both motion capture systems. In case of Kinect, a framerate of 30 frames per second is used. In the other hand, the optical motion capture that we used record at a speed of 120 frames per second. So, we have time scaled data acquired by optical system.

Both motions have the same velocity but they are still unsynchronized. At this moment we have pairs of motions that their beginnings and endings are close between them, but not exactly at same time because synchronization was manually done. In order to fix it we have searched which is the point when the actor leaves t-pose and starts performing the appropiate movement. We have observed in video recordings that actor tends to lower the arms just after staying in t-pose. So, we have detected this decreasement in vertical axis of hands position using a velocity threshold value.

Definitely we have motion pairs expressed in the same way and synchronized. To evaluate if data processing is successful, we have implemented a visualization tool for qualitative revision. In Figure 7 there is a captured screen of that application.

4.2 Rotational Data

We have 3D positions of specified joints from both motion capture systems. Our goal is to compare joint rotations instead of positions. So, we have extracted the desired rotations from position data . To compute these rotations we can distinguish the calculation of knee rotation and the rest. Knee rotation is the angle between two vectors, one from knee to foot and the other from knee to hip. This is because knee

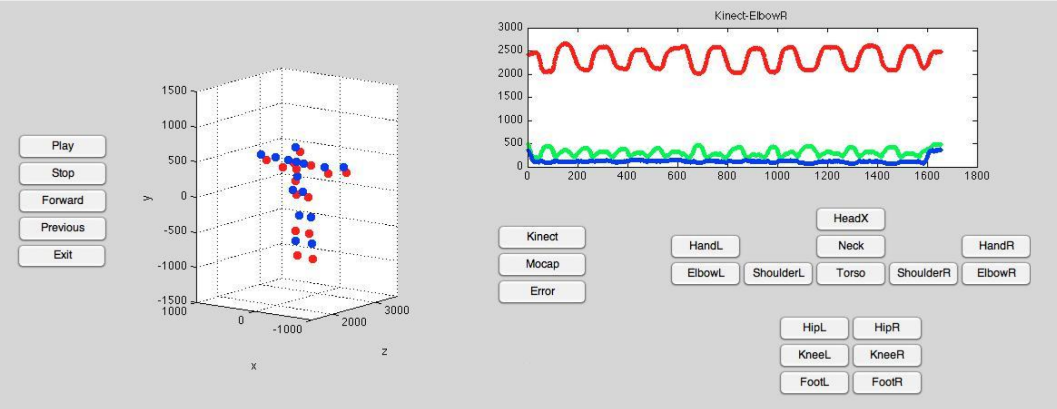

Figure 7: Visualization tool. On the left side, there are controls to manage the visualization of pairs of motions which are right next. Estimated joint positions from Kinect are coloured in blue and markers positions are in red. On the right side, we can see a graph of visualized data. The buttons below allow to select the movements data from different joints. Also we can select if we want to observe data from Kinect, from Mocap or the difference between them using the middle buttons.

has only one degree of freedom. In case of other rotations, we have computed rotation values respect to body planes (Figure 3). This calculation consists on creating a vector from joint to its child joint and compute the angle between this vector and a perpendicular vector to the desired body plane. In Figure 8 we can observe described vectors for extracting desired rotations.

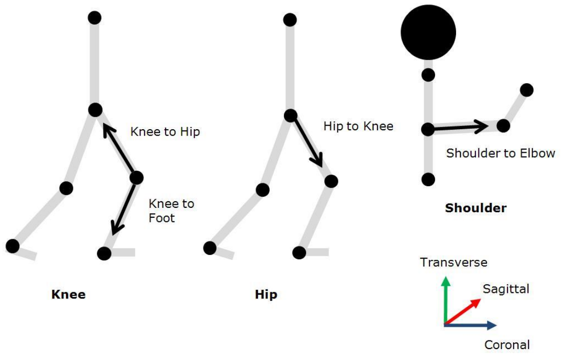

Figure 8: Extracted rotations. On the left, vectors for computing knee angle; on the middle, vector for computing hip direction; on the right, vector for computing shoulder direction; and below, perpendicular vectors to body planes.

5 Results

To evaluate the performance of Kinect as a motion capture system we have compared the reported joint rotational values from this system against data from optical motion capture. Figure 9 shows 3 joint angle trajectories from different motion clips. As we can see, signals from kinect and optical system have an evidence correlation because they are synchronized and follow the same pattern.

The aim of this comparison is to know how accurate is Kinect in terms of degrees. So, we have computed the mean error (ME) and the mean error relative to range of motion (MER) for each motion clip. MER is calculated by

$$MER_M = \frac{1}{m}\sum_{i=1}^{m}(K_i - O_i)/ROM \qquad (1)$$

where M is a motion clip, m is the frames length of motion clip M, K_i and O_i are joint angle from kinect motion capture and optical motion capture in frame i respectively, and ROM is the range of motion.

Rotation comparisons have been done for knee, hip and shoulder joints. All treatened joint values of Kinect motion data have a true

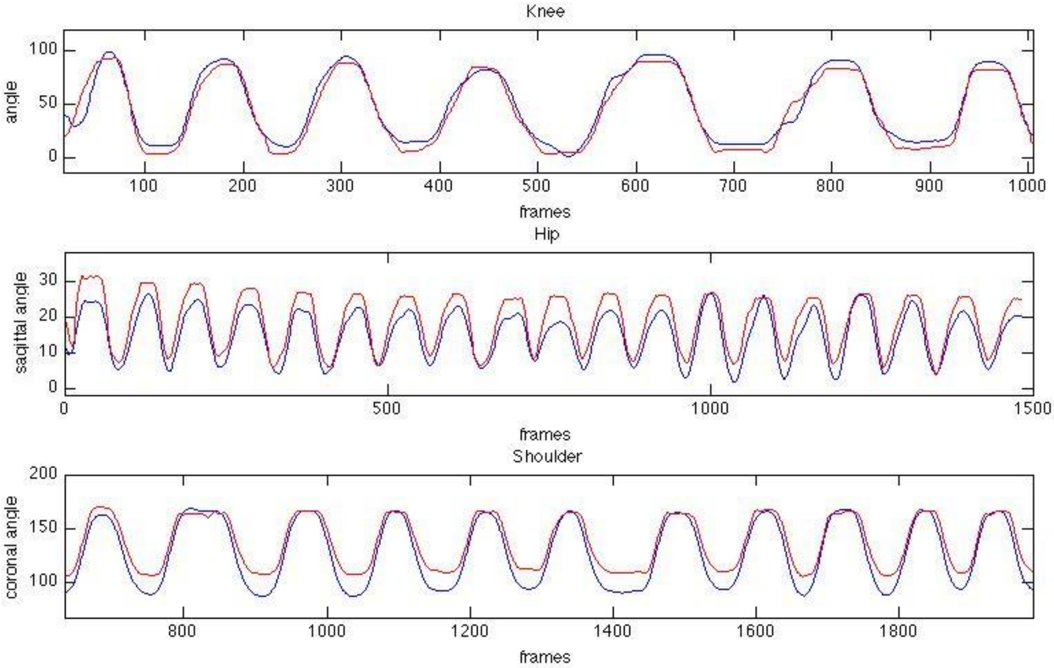

Figure 9: Joint angle trajectories. In red color, Kinect motion data; in blue color, optical motion data. From top to down, knee, hip and shoulder from different motion clips.

confidence parameter belonging to tracking algorithm. In the case of knee we can summarize results in Table 2. All degree error are lower than 10^o ranging from 6.78^o to 8.98^o. Dynamic ranges of motion are between 89^o and 115^o. ME is increasing when ROM is higher. It occurs because in extrem rotations leg is perpendicular to kinect camera doing more difficult to track hip and knee. Even in this situation would be difficult for a person to determine the pose. These results are good enough for some physical therapies based on repetitions, because nowadays therapists visually controls the range of motion and it is assumed that it has at least 10^o approximate error.

M	Frames	ROM	MER	ME
1	2112	89.29^o	0.07	6.78^o
2	2033	93.41^o	0.08	7.94^o
3	1853	115.07^o	0.07	8.98^o

Table 2: Knee results.

In case of hip, we have compared sagittal and coronal rotation. In Table 3 there are hip results where the two first motions contain the sagittal movement and the following the coronal movement. Sagittal ME is around 5^o and coronal is ranged from 6^o to 10^o. Sagittal ranges of motion are around 90^o and coronal are 77^o and 38^o. Errors in sagittal movements are lower than the other cases, however errors in coronal plane are lower than 10^o.

M	Frames	ROM	MER	ME
4	1880	89.42^o	0.06	5.53^o
5	1370	90.12^o	0.06	5.88^o
6	1272	77.07^o	0.13	9.92^o
7	1893	38.63^o	0.17	6.49^o

Table 3: Hip results. Motion 4 and 5 contain saggital movements; motion 6 and 7 are coronal movements.

Shoulder rotations are the most complete of our study because this joint has 3 degrees of freedom. In this case, we have obtained results that are varying between 7^o to 13^o in all plane rotations. The four first motions contain movements in sagittal plane; motion 12 and 13 are coronal movements and the two last are transverse movements. For this joint there is no direct relation between ROM and errors cause we have sparse results.

M	Frames	ROM	MER	ME
8	1655	125.04^o	0.06	8.02^o
9	793	139.08^o	0.05	7.19^o
10	558	128.31^o	0.08	9.75^o
11	1255	120.14^o	0.07	8.41^o
12	2302	71.39^o	0.16	11.33^o
13	812	73.14^o	0.11	8.34^o
14	1359	89.10^o	0.13	11.80^o
15	659	82.53^o	0.15	13.19^o

Table 4: Shoulder results. From motion 8 to 11 are saggital movements; motion 12 and 13 are coronal movements; motion 14 and 15 are transverse movements.

6 Conclusions and Future Work

We have presented a precision comparative study between two very different equipments. One very accurate and expensive motion capture system and the new easy and cheap device Kinect. We think our results would be very interesting for the biomechanics community and computer animation in general.

As we have shown, the precision of the Kinect is, of course, less than the optical motion capture system, but has several other advantages: prize, portability and markless. The precision ranks obtained for the main joints of the body allows as to confirm that Kinect can be a very useful technology in present rehabilitation treatments. In fact, we have developed a first application for knee rehabilitation that automatically counts repetition movements and validates the quality of such a motion.

The precision of the Kinect captures can be increased by imposing some fixed length restriction for the bones (now it can be different in each frame). One can also help the system using some incremental tracking strategy, now it is frame independent. For biomechanics applications, the human joint motion rank can also be included as a restriction of the system. These can be several future works to improve the present results obtained using Kinect. Another possibility is to work directly with the depth map information and try to get a better approximation of joint and bones positions using a retargeting method.

Acknowledgements

We are very grateful to the direction of Creation, Design and Serious Game Master Program and La Salle - Universitat Ramon Llull for the help in using their facilities. The authors would like to thank therapist Bernat Pascual for performing the movements in motion capture sessions and for placing optical markers in their correct places. Also thank Eduard Ruesga and Meritxell Aragonés for his work in recording and cleaning motion capture clips. We acknowledge David Miralles for his helpful discussions and reviewing advices. This research was partially supported by LABSID SL and TIN2010-20590-C02-01.

References

[1] Xsens. Xsens: 3D Motion Tracking. http://www.xsens.com/, Dec 2011.

[2] Animazoo. Animazoo Motion Capture Systems and Technology. http://www.animazoo.com/, Dec 2011.

[3] Ascension Technology Corporation. Ascension Technology Corporation. http://www.ascension-tech.com/, Dec 2011.

[4] A Cappozzo, F Catani, U Della Croce, and A Leardini. Position and orientation in space of bones during movement: anatomical frame definition and determination. *Clinical Biomechanics*, 10(4):171 – 178, 1995.

[5] Roy B Davis, Sylvia Ounpuu, Dennis Tyburski, and James R Gage. A gait analysis data collection and reduction technique.

Human Movement Science, 10(5):575–587, 1991.

[6] C Frigo, M Rabuffetti, D C Kerrigan, L C Deming, and A Pedotti. Functionally oriented and clinically feasible quantitative gait analysis method. *Medical and Biological Engineering and Computing*, 36(2):179–185, 1998.

[7] A Cappozzo, F Catani, A Leardini, MG Benedetti, and U Della Croce. Position and orientation in space of bones during movement: experimental artefacts. *Clinical Biomechanics*, 11(2):90 – 100, 1996.

[8] J. Fuller, L.-J. Liu, M.C. Murphy, and R.W. Mann. A comparison of lower-extremity skeletal kinematics measured using skin- and pin-mounted markers. *Human Movement Science*, 16(2-3):219 – 242, 1997. 3-D Analysis of Human Movement - II.

[9] J. A. De Guise S. Larouche Sati, M. and G. Drouin. Quantitative assessment of skin marker movement at the knee. *The Knee*, 3:121–138, 1996.

[10] S Corazza, L Mndermann, A M Chaudhari, T Demattio, C Cobelli, and T P Andriacchi. A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. *Annals of Biomedical Engineering*, 34(6):1019–1029, 2006.

[11] Ugo Della Croce, Alberto Leardini, Lorenzo Chiari, and Aurelio Cappozzo. Human movement analysis using stereophotogrammetry: Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. *Gait & Posture*, 21(2):226 – 237, 2005.

[12] Lorenzo Chiari, Ugo Della Croce, Alberto Leardini, and Aurelio Cappozzo. Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors. *Gait & Posture*, 21(2):197 – 211, 2005.

[13] Microsoft Xbox. Kinect. http://www.xbox.com/kinect, Dec 2011.

[14] Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Comput. Vis. Image Underst.*, 81:231–268, March 2001.

[15] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.*, 104:90–126, November 2006.

[16] B Bodenheimer, C Rose, S Rosenthal, and J Pella. The process of motion capture: Dealing with the data. *Computer Animation and Simulation*, pages 3–18, 1997.

[17] Gutemberg B. Guerra-filho. Optical motion capture: Theory and implementation. *Journal of Theoretical and Applied Informatics (RITA*, 12:61–89, 2005.

[18] La Salle Universitat Ramon Llull. MediaLab. Motion Capture, RV + RA, Animation, Videogames and CAD. http://www.salleurl.edu/medialab, Dec 2011.

[19] Vicon. Motion Capture Systems from Vicon. http://www.vicon.com/, Dec 2011.

[20] Microsoft. Microsoft Xbox. http://www.xbox.com/, Dec 2011.

[21] Rare Ltd. Rare. http://www.rare.co.uk/, Dec 2011.

[22] PrimeSense Ltd. Primesense Natural Interaction. http://www.primesense.com/, Dec 2011.

[23] OpenNI. OpenNI Home. http://www.openni.org/, Dec 2011.

[24] PrimeSense Ltd. PrimeSense Natural Interaction Technology for End-user. http://www.primesense.com/nite, Dec 2011.

[25] PrimeSense. Prime Sensor NITE 1.3 Algorithms notes. *PrimeSense NITE Documentation*, 2010.