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

Adso Fernández-Baena
Grup de Tecnologies Mèdia (GTM)
La Salle - Universitat Ramon Llull
Barcelona, Spain
adso@salle.url.edu

Antonio Susín
Universitat Politècnica de Catalunya
Barcelona, Spain
toni.susin@upc.edu

Xavier Lligadas
LABSID SL.
Barcelona, Spain
xavi@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

I. 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] or electromagnetic sensors [2], at present, using optical systems with reflective markers is the most common technique [3] [4]. 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 [5][6]. Several recent review articles have summarized the common shortfalls of skin based marker techniques [7] [8] [9]. Markerless motion capture [10] [11] 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 under-constrained without the spatial and temporal correspondence that tracked markers guarantee.

From the powerful game industry new devices like Kinect [12] have appeared, allowing to interact with game consoles in real time. Moreover, this new hardware is considerably cheaper than the usual complex multi-camera 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 equipment is required. The skeleton is made of 15 joints and due to its simplification it cannot 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 validated 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 consequence of our study, we have implemented Rehabtimals. Rehabtimals is a rehabilitation framework that covers all phases in physical rehabilitation cycle allowing patients recover from their injuries by playing with a serious game at their homes.

The rest of the paper is organized as follows: in section 2 we relate some previous work, section 3 describes the equipment used in our study, section 4 describes the motion capture performance and in section 5 we present the results.

II. RELATED WORK

A. Motion Capture and Rehabilitation

The interest of some videogames researchers towards use videogames in physiotherapy, occupational therapy, and psychotherapy increased last years. Many physical therapies are based on repetition to achieve a range of motion or control over a specific muscle group. That process is done without any external encouragement [13] and causes patients losing motivation during the therapy and the rehabilitation becomes slower and frustrating. By developing a serious game as a rehabilitation tool, we achieve the motivation associated to games. Low cost technologies like webcams

bring the possibility to capture patient movements in order to complement the process of rehabilitation, without losing control on the patient when they are at home [14][15]. Although, that systems are limited by their motion capture capacity. To improve capture precision, other approaches use sensors to track user movements. In [16], they used Wiimote device, another low cost peripheral from videogames industry. More recently, the appearance of Kinect causes a great improvement in low cost motion capture performance, allowing the implementation of some Kinect-based rehabilitation systems [17][18].

B. Kinect Precision

From the point of view of the Kinect's precision we can cite [19], where the authors quantify the amount of error according to the depth distance from the Kinect. The results are difficult to compare because they work with an static scene and not with motion capture data that also involve errors in the obtained skeleton. In [20] stride-to-stride gait variability is also studied with Kinect. Recently, in [18] a similar approach is adopted to implement a serious game after comparing Kinect motion capture performance against a marker-based motion capture system. Although a comparative study is conducted, their focus on joint positions rather on joint orientations without facilitating precise results.

III. EQUIPMENT DESCRIPTION

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

A. Optical Motion Capture

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). 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. After capture sessions, motion data is cleaned trying to remove noisy data and recovering missing markers. Therefore, optical motion capture data is very accurate and we will use it as reference for testing Kinect mocap accuracy. In this work we have been used MediaLab [21] facilities. MediaLab is a passive optical motion capture laboratory belonging to La Salle - Universitat Ramon Llull (Barcelona). This laboratory has 24 Vicon MX3 [22] cameras that allow a 45 m^2 of capture volume.

B. Kinect Motion Capture

Kinect device appears in November of 2010 as a entertainment device of Microsoft Xbox [23] console. It is based on software developed by Rare [24], Microsoft Game Studios affiliated company, and the technology of PrimeSense [25] cameras. Kinect RGB camera uses an 8-bit VGA resolution (640x480 pixels) while its monochrome depth sensor has a VGA resolution of 11 bits that allows 2048 sensibility levels. Kinect device has an approximate depth limitation from 0.7 to 6 meters. Horizontal angular field of view is 57° and 43° 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 [26] and Primesense's NITE (Natural Interaction Technology for End-user) [27] have been used.

IV. EXPERIMENTAL ANALYSIS

In this section we describe the technical aspects of capture sessions and which are the selected movements to study Kinect precision. Then, we show how we have processed motion data in order to compare both systems.

A. Capture Description

We have mounted a Kinect device inside the motion capture laboratory in order to record motions from both systems. As we can see in Figure 1, Kinect was placed in front of the actor at an approximate distance of 2 meters. In case of the optical motion capture, cameras cover the capture volume with multiple views.

Figure 1. Motion capture laboratory setup.

We have studied different movements from upper and lower body belonging to common physical therapies. Chosen movements are explained in Table I. Each movement is driven by a main joint which could be knee, hip or shoulder.

Movement	Treatment
Knee flexion and extension	Gain movement phase of ruptured cruciate ligament of the knee treatment.
Hip flexion and extension on the sagittal plane	Gain strength phase of ruptured cruciate ligament of the knee treatment.
Hip adduction and abduction on the coronal plane with knee extended	Gain strength phase of ruptured cruciate ligament of the knee treatment.
Shoulder flexion and extension on the sagittal plane with elbow extended	Gain movement phase of ruptured rotators of the knee treatment.
Shoulder adduction and abduction on the coronal plane with elbow extended	Gain movement phase of ruptured rotators of the knee treatment.
Shoulder horizontal adduction and abduction on the transverse plane with elbow extended	Gain movement phase of ruptured rotators of the shoulder treatment.

Table I
RECORDED MOVEMENTS.

1) Optical Motion Data: 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 2 can be seen in detail markers configuration. Motion data was recorded with a framerate of 120 and markers positions were expressed in world coordinate system.
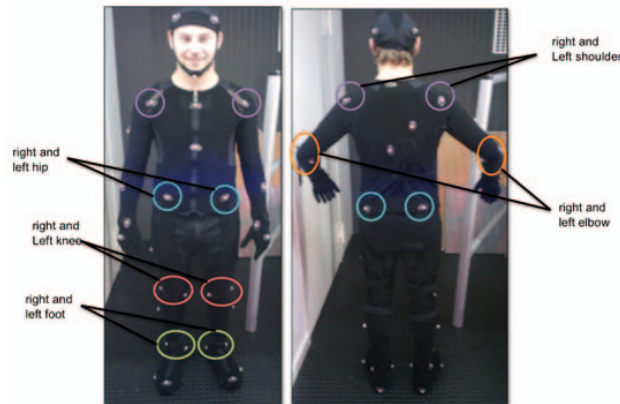

Figure 2. On top, placement of markers on the actor's body on both sides; and below, detail of the placement of markers for calculating the position of the head, shoulder, elbow, knee and foot.

2) Kinect Motion Data: We have been used joint positions from NITE user tracking algorithm rather than joint orientations. Joint specification is shown in Figure 3 which reported 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, so we do not analyze faulty values. Therefore, motion data was recorded at a framerate of 30 frames per second.
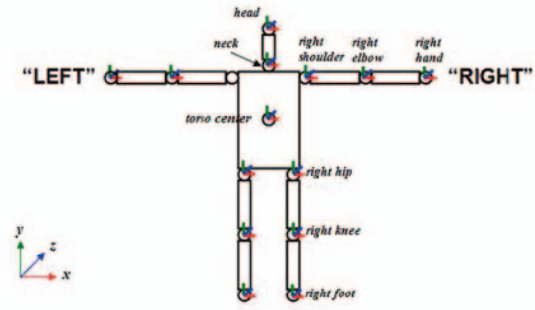

Figure 3. OpenNI joint specification.

Estimated positions by Kinect suffer some 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 low-pass filter (5 frames of time window) to eliminate local fluctuations.

B. Motion processing

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, we have time scaled data acquired by optical system (120 fps) to Kinect framerate (30 fps). Although, pairs of motions are unsynchronized because capture systems were independent manually activated. In order to fix it, we have observed that actor tends to lower the arms just after staying in t-pose. So, we have searched that moment analyzing vertical axis decrement of hands position.

Then, we have aligned the same movement from both systems in terms of coordinate system. Optical motion data is reported in its world coordinates (Figure 4 - (a)), and the same for Kinect data (Figure 4 - (d)). Both coordinate systems have the same orientation but they have exchanged axes. We have corrected it as shown in Figure 4 - (b).

Finally, we estimate each joint position that Kinect is able to track from two optical markers (see Figure 2. To achieve this is very simple since you just have to calculate the midpoint between marker pairs. In Figure 4 - (c) it can be observed the result of this step. Now the appearance of optical data is like Kinect data.

2) Rotational Data: Then, we have computed knee rotation, and hip and shoulder rotations respect to body planes. 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) because it has only one degree of
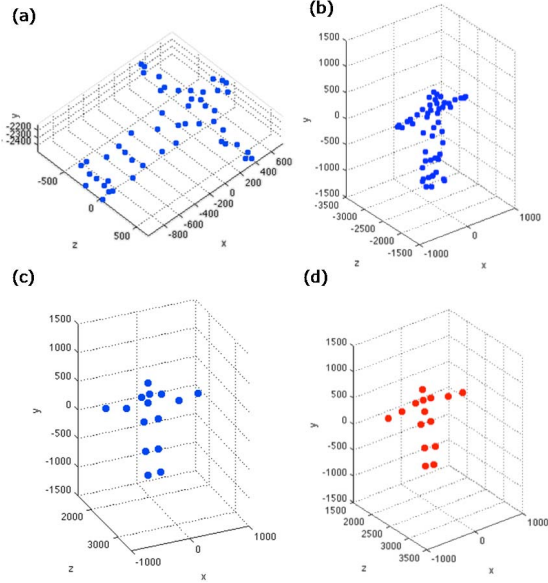
Figure 4. Steps from data correspondence. (a) Markers from optical motion capture. (b) Rotated markers from optical motion capture. (c) Joint positions from optical markers. (d) Joint positions from Kinect motion capture.

freedom. In case of other rotations, the computation consists on creating a vector from the applicable joint to its child and compute the angle between this vector and a perpendicular vector to the desired body plane.

V. 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. In this section we present the conducted comparison and the implemented application.

A. Precision study

Figure 5 shows some joint angle trajectories belonging to different motion clips. As we can see, signals from Kinect and Vicon system have an evidence correlation because they are synchronized and follow the same pattern. To measure the accuracy 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.

As we have mentioned, rotation comparisons have been done for knee, hip and shoulder joints. In the case of knee we can summarize results in Table II. All degree error are lower than 10° ranging from 6.78^o to 8.98^o. Dynamic ranges of motion are between 89° and 115°. ME is increasing when ROM is higher. It occurs because in extreme 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° 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 II
KNEE RESULTS.

In case of hip, we have compared sagittal and coronal rotation. In Table III there are hip results. Sagittal ME is around 5° and coronal is ranged from 6° to 10°. Sagittal ranges of motion are around 90° and coronal are 77° and 38°. Errors in sagittal movements are lower than the other cases, however errors in coronal plane are lower than 10°.

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 III
HIP RESULTS. MOTION 4 AND 5 CONTAIN SAGITTAL 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° to 13° in all plane rotations.

B. Rehabtimals

Results obtained allow to use Kinect for physical rehabilitations purposes, although not very accurate precision have been shown. Therefore, counting repetitions is affordable and an approximate angle value could be reported as feedback. We have implemented a serious game which is called Rehabtimals (see Figure 6). Rehabtimals is a virtual world where a patient recovers from his injury meanwhile an animal grows up and overcomes life situations. We have used NITE User Controls [27] and Panda3D game engine [28] to
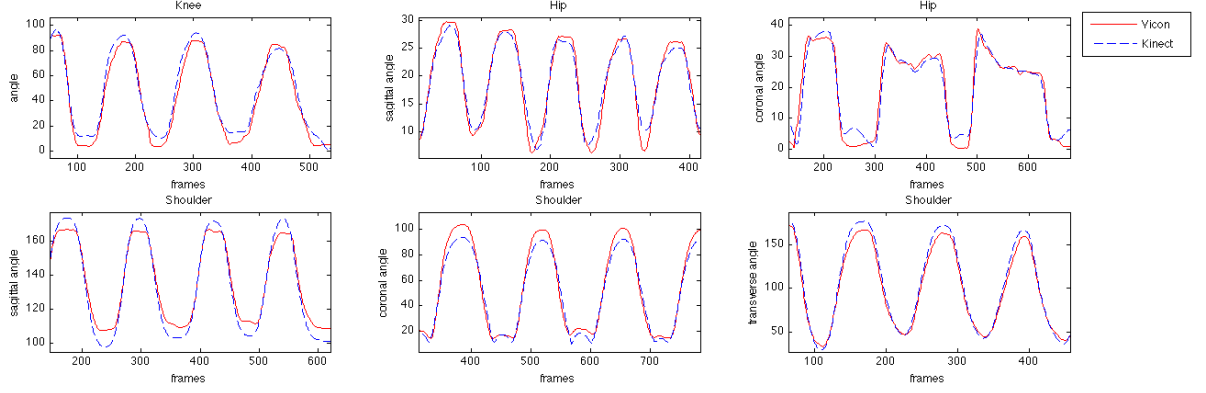
Figure 5. Joint angle trajectories. In blue color, Kinect motion data; in red color, Vicon motion data. From top to down, left to right: knee rotation, hip sagittal rotation, hip coronal rotation, shoulder sagittal rotation, shoulder coronal rotation and shoulder transverse rotation.

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 IV
SHOULDER RESULTS. FROM MOTION 8 TO 11 ARE SAGITTAL MOVEMENTS; MOTION 12 AND 13 ARE CORONAL MOVEMENTS; MOTION 14 AND 15 ARE TRANSVERSE MOVEMENTS.

implement a server-host architecture. In our first version, we cover ruptured cruciate ligament of the knee rehabilitation which is guided by a macaw (see Figure 6). We also have implemented an application for therapists which is named Rehabtimals Pro. In Rehabtimals Pro, therapist can control and manage their patients treatments by analyzing captured angle rotations, visualizing 3D motion reconstructions of patients or watching video sessions.

VI. CONCLUSIONS AND FUTURE WORK

As we have shown, the precision of the Kinect is, of course, less than the optical motion capture system, but has several other advantages: price, portability and markerless. 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

Figure 6. Capture screens from application. From top to down, left to right. Instruction screen, Level 1 screen, Level 2 screen from Rehabtimals application, and Session screen from Rehabtimals Pro application.

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.

REFERENCES

[1] Xsens. Xsens: 3D Motion Tracking. http://www.xsens.com/, May 2012.

[2] Ascension Technology Corporation. Ascension Technology Corporation. http://www.ascension-tech.com/, May 2012.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] S Corazza, L Mundermann, 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] Microsoft Xbox. Kinect. http://www.xbox.com/kinect, May 2012.

[13] Jan-Henk Annema, Mathijs Verstraete, Vero Vanden Abeele, Stef Desmet, and David Geerts. Videogames in therapy: a therapist's perspective. In *Proceedings of the 3rd International Conference on Fun and Games*, Fun and Games '10, pages 94–98, New York, NY, USA, 2010. ACM.

[14] J. W. Burke, M. D. J. McNeill, D. K. Charles, P. J. Morrow, J. H. Crosbie, and S. M. McDonough. Optimising engagement for stroke rehabilitation using serious games. *Vis. Comput.*, 25(12):1085–1099, October 2009.

[15] Mónica S. Cameirão, Sergi Bermúdez I Badia, Esther Duarte Oller, and Paul F. Verschure. The rehabilitation gaming system: a review. *Studies in health technology and informatics*, 145:65–83, 2009.

[16] Fraser Anderson, Michelle Annett, and Walter F. Bischof. Lean on Wii: Physical rehabilitation with virtual reality and Wii peripherals. *Annual Review of CyberTherapy and Telemedicine*, 8:181–184, 2010.

[17] Christian Schonauer, Thomas Pintaric, Hannes Kaufmann, Stephanie Jansen Kosterink, and Miriam Vollenbroek-Hutten. Chronic pain rehabilitation with a serious game using multimodal input. In *Virtual Rehabilitation (ICVR), 2011 International Conference on*, pages 1 –8, june 2011.

[18] Mi Zhang Sebastian Koenig Phil Requejo Noom Somboon Alexander A. Sawchuk Chien-Yen Chang, Belinda Lange and Albert A. Rizzo. Towards pervasive physical rehabilitation using microsoft kinect. In *International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, San Diego, CA, USA, May 2012.

[19] T. Dutta. Evaluation of the kinect sensor for 3-d kinematic measurement in the workplace. *Appl Ergon*, 43(4):645–9, 2012.

[20] Erik E. Stone and Marjorie Skubic. Evaluation of an inexpensive depth camera for in-home gait assessment. *JAISE*, 3(4):349–361, 2011.

[21] La Salle Universitat Ramon Llull. MediaLab. Motion Capture, RV + RA, Animation, Videogames and CAD. http://www.salleurl.edu/medialab, May 2012.

[22] Vicon. Motion Capture Systems from Vicon. http://www.vicon.com/, May 2012.

[23] Microsoft. Microsoft Xbox. http://www.xbox.com/, May 2012.

[24] Rare Ltd. Rare. http://www.rare.co.uk/, May 2012.

[25] PrimeSense Ltd. Primesense Natural Interaction. http://www.primesense.com/, May 2012.

[26] OpenNI. OpenNI Home. http://www.openni.org/, May 2012.

[27] PrimeSense Ltd. PrimeSense Natural Interaction Technology for End-user. http://www.primesense.com/nite, May 2012.

[28] Carnegie Mellon University. Panda3D Game Engine. http://www.panda3d.com/, May 2012.