Human whole body motion characterization from embedded Kinect

Consuelo Granata, Joseph Salini, Ragou Ady, Philippe Bidaud

HAL Id: hal-00944475
https://hal.sorbonne-universite.fr/hal-00944475v1

Submitted on 10 Feb 2014

Human whole body motion characterization from embedded Kinect

Consuelo Granata
ISIR, UPMC-CNRS UMR
4, place Jussieu, 75005 Paris
IMPERIAL COLLEGE OF LONDON
Exhibition Rd, London SW7 2AZ, UK
Email: granata@isir.upmc.fr

Joseph Salini, Ragou Ady
ISIR, UPMC-CNRS UMR
4, place Jussieu, 75005 Paris
Email: salini@isir.upmc.fr
Email: ady@isir.upmc.fr

Philippe Bidaud
ISIR, UPMC-CNRS UMR
4, place Jussieu, 75005 Paris
ONERA
Chemin de la Hunière, 91123 Palaiseau
Email: philippe.bidaud@isir.upmc.fr

Abstract—Non-verbal communications such as kinesthetics, or body language and posture are important codes used to establish and maintain interpersonal relationships. They can also be utilized for safe and efficient human robot interactions. A correct interpretation of the human activity through the analysis of certain spatio-temporal and dynamic parameters represent an outstanding benefit for the quality of human machine communication in general. This paper presents an effective markerless motion capture system provided by a mobile robot for sensing human activity, in non-invasive fashion. We present a physical model based method exploiting the embedded Kinect. Its performances are evaluated first comparing the results to those obtained with a precise 3D motion capture marker based system and to data obtained from a dynamic posturography platform. Then an experiment in real life conditions is performed to assess the system sensitivity to some gait disturbances.

I. INTRODUCTION

Social robots must be able to interact effectively with humans, to understand their needs and to interpret theirs orders. These challenging tasks can obviously benefit from considering the human body language (posture, gestures, etc.). Two thirds of communication between humans is non-verbal [1]. The analysis of non-verbal cues provide access to a deeper reading of emotions (e.g. boredom, interest , sadness, etc.) and even intentions of the person. Body language creates a very important part of the communication. The body then gives a certain "form" in mind. Beyond signs, being able to understand and anticipate human activity directed towards a goal is an important element to optimize, through interaction, the robot perception and motor activity. To reach this understanding, we can try to rely on how the brain generates predictions. A hypothesis on the "predictive coding" suggests that the brain continuously generates expectations on sensory input from motor activities. This hypothesis derives from the ideomotor theory [2]. It considers that actions are cognitively represented in terms of their perceived effects. This principle of "ideomotor" argues that during the execution of a particular action, a motor model (reduced order model) is automatically associated with the input of perception representing the perceptual effects of the action [3]. Perception and action planning are considered similar processes based on the characteristics of activation codes that represent external events. Having such a model must allow to design the robot functions (communication, perception, action) in a form directed by the interaction as any other form of human-machine interface.

We focus here on the analysis of human locomotor activity, that is generally based on the spatio-temporal parameters (Center of Mass (CoM) trajectory, step length, step frequency, step width, walking speed, trunk acceleration etc.) considered to be relevant for gait evaluation [4], [5], but also on the parameters characterizing the state of dynamic balance of the postural system. Typically we use the Centre of Pressure (CoP) which is by definition identical to the to the Zero Moment Point (ZMP) [6]. The CoP/ZMP is the point where the Ground Reaction Force (GRF) vector intersects the plane of Base Of Support (BOS) of the feet on the ground. This reaction force represents the resultant of the contact force distribution on the ground. In the last two decades, many technical solutions were developed to access these different parameters of human movement, e.g. treadmills, instrumented surfaces such as force platforms or particular shoes [7], [8]. To capture the kinematics movement, the most reliable results were achieved by systems such as Motion Capture (MoCap) which exploits high speed digital cameras to capture the 3D motion performed by a subject. Several systems can be used but most commonly the subject is fitted with either passive or active markers [10]. The passive markers are often covered with an infrared reflective material and then attached to the subject on predefined anatomical landmarks. The estimation of 3D human motions has been recently considered by using single or multi video cameras [9], [11] (for more details about the motion capture methods we refer to the Poppe's review [12]). The collected data are used to animate a model reproducing the anthropometric characteristics of the subject. The kinematic solver is formulated as a frame-by-frame weighted least-squares problem that minimizes the differences between the measured marker locations and the model's virtual marker locations. By introducing thus obtained joint motions, the dynamic results from the multi-body motion with the help of a forward dynamic model [13]. These technologies are designed to be used in controlled laboratory conditions (with specific environmental set-up or instrumented clothes worn by the person) and cannot be employed for activity assessment in real life environments. Gonzáles et al. [14] proposed a non-invasive (marker-less) portable system based on the combination of 2 sensors, a Wii balance board and a Kinect to compute the CoM. This system is unsuitable for the analysis of locomotion because of the use of the balance board. During locomotion, the most challenging phases for postural balancing system are beginning and end the walk, turning,
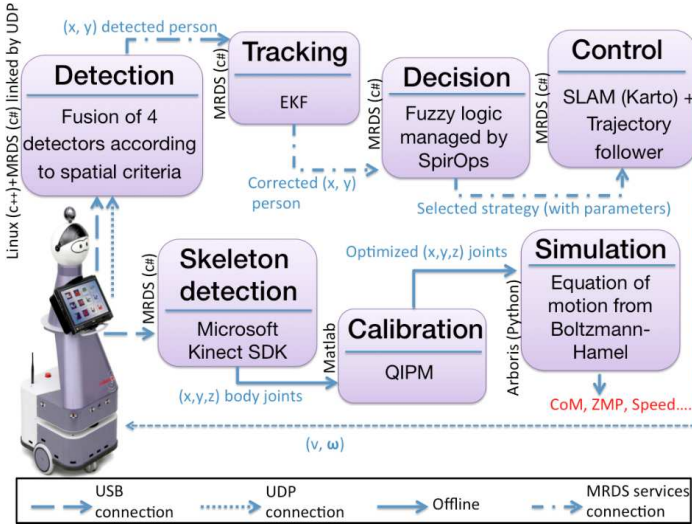
Fig. 1. Overall system architecture.

avoiding obstacles (changing the length of the pitch, changing direction, stepping over objects, etc.) and bumping into people and objects. A system able to quantify human responses to perturbations induced by the daily life environment can have a use-value. To our knowledge there is no system that can be used for an in-depth analysis of human activity (especially of posture balance and stability) in non controlled environmental conditions, in a non-invasive fashion (marker-less method) and with an accuracy comparable with marker based technologies. We approach this need with a robot which can perceive and track humans while moving and with the embedded Kinect sensor to capture in real-time multi-segmental human motion by using the Microsoft Kinect SDK.

The paper describes a system developed to analyze a person's activity by using a non-invasive (marker-less) motion capture system. A method for determining the individual's anatomical parameters and the algorithms developed for the digital animation model based on the measurement of a number of characteristic points obtained by a Kinect are described. These algorithms are validated experimentally on different exercises during which the positions of the CoP/ZMP, measured and calculated using a posturography platform, are compared. The proposed system is also assessed analyzing the gait of a subject followed by the robot. It shows strong sensitivity to some disturbances obtained manipulating artificially the lower limb joints.

II. EXPERIMENTAL SYSTEM

The robot used for the implementation is a commercial robot, Kompaï (Figure 1) specifically designed for services in private or public spaces. In order to raise the potential of this platform, we added a Microsoft Kinect and an extra laptop for computation. We developed and integrated a framework for the synthesis of smart person following behavior composed of a multimodal human detector and a fuzzy logic based decisional engine (Figure 1). For more details we refer to our previous work [15].

III. FILTERING METHOD

The Kinect sensor belongs to the class of devices known as depth cameras. These cameras interpret 3D scene information

based on projected infrared light system called light coding that generates a 3D point cloud. The Kinect sensor has an approximate resolution of $0.013\ m$ per pixel and works at $30\ Hz$. The 3D point cloud is converted into a depth map, and from a single depth map, body parts are inferred using a randomized decision forest-based approach, learned from over 1 million training examples. A mean shift algorithm is used to robustly compute 3D positions of body joints from the modes of discrete probability distributions. More details can be found in Shotton et. al ([10]). The accuracy of the joints position provided by the Kinect skeletal detection algorithm has been investigated in [16] where the inter-joint distance (the limb length) variations have been estimated as up to more than $0.1\ m$. For facing this issue an improved skeletonization solution to a space-time constraint problem is required. In previous works, the pose correction was approached in different ways: according to physical-model [17], by using Kalman-like filters [18], by adopting different kinds of regressors (random forest regressor [19], nearest neighborhood regressor [20], Gaussian process regression [21] etc.). We propose a method for pose correction that takes into account both the coherence of the anthropomorphic model of the skeleton and the spatial-temporal motion consistency. This method is composed of 3 steps (Figure 2):

Step 1: Initialization
When the person is in the Kinect field of view, his skeleton is detected and the 3D positions of all his joints are provided (Figure 6 left). During the initialization phase, we use a first set of Kinect measurements to define a reference model of the involved body. In particular, the system computes the anthropometric kinematic model of the body according to the mean distances between the consecutive joints.

Step 2: Physical model calibration
This step is required to keep the human physical skeleton consistent with the reference model (in anthropomorphic point of view). Our approach relies on a standard mathematical solution, constrained optimization. It consist in minimizing the scalar value quadratic objective function $f(X)$ while respecting a certain number of constraints $g(X)$ (1).

$$\begin{aligned} & minimize\ f(X) \\ & subject\ to\ g(X) = 0 \end{aligned} \quad (1)$$

In the objective function $f(X)$ (Equation 2), X is a vector residing in $(3*nJ)$-dimensional space (3D coordinates for nJ joints), a_i, b_i, c_i are the weights associated with the coordinates of the joint i (in our case we consider equal weights on the 3 dimensions).

$$f(X) = \sum_{i=1}^{nJ} a_i(\bar{x}_i - x_i)^2 + b_i(\bar{y}_i - y_i)^2 + c_i(\bar{z}_i - z_i)^2 \quad (2)$$

$(\bar{x}_i, \bar{y}_i, \bar{z}_i)$ and (x_i, y_i, z_i) are respectively the coordinates of the joint i provided by the skeleton detector algorithm and the coordinates that we are searching for to minimize the objective function while respecting the constraints in Equation 3.

$$\begin{aligned} g_j(X) &= d^2(J_j; J_{j+1}) - d^2(M_j; M_{j+1}) = \\ &= (x_j - x_{j+1})^2 + (y_j - y_{j+1})^2 + \\ &\quad + (z_j - z_{j+1})^2 - d^2_{j,j+1} = 0 \end{aligned} \quad (3)$$

with $j = 1,...,nJ - 1$, $(J_j; J_{j+1})$ consecutive detected joints, $(M_j; M_{j+1})$ consecutive joints in the model, and $d_{j,j+1}$ the distance between M_j and M_{j+1}.

All the constraints are equality constraints and are fixed according to the reference model (Initialization step). The optimization of the objective function is made by using the quadratic interior point method (QIPM), which is based on the improvement of initial conditions (measurements) for solving quadratic programming problems.

Step 3: Model Simulation and Parameters Analysis
Finally, the 3D trajectories of the kinect skeleton joints virtual markers are used to animate a physical model of the subject on a dynamic simulator Arboris-Python [23]. Arboris-Python is an open-source constrained multibody dynamics simulation software written in Python language. It includes a generic and easily extensible set of joints (singularity-free multi-degree of freedom joints, non-honolomic joints, etc.) for the design and modeling of tree structure mechanisms with a minimal set of state variables. It gives access to the completed mechanical properties of the system as well as to the constraints and to the controllers implemented to get the desired behavior of the virtual human. Various control algorithms have been implemented in the Arboris-Python software, from the proportional-integral-derivative controllers to the predictive model based controllers which are used for the control of the locomotion and postural balance task or the interaction tasks with adaptive impedance. The equations of motion of these multibody systems are obtained with the Boltzmann-Hamel formalism [22] from which the first-order approximation of the model is computed. The resulting equations are then integrated using a time-stepping method and a semi-implicit Euler integration scheme. In this way, it is possible to introduce and solve additional constraints, i.e. the kinematic loops, which can be either unilateral (contact) or bilateral (joint), with the help of a Gauss-Seidel algorithm [24]. For a complete description of this software you may refer to [25]. Once the model of the subject generated from a generic virtual human model (with 36 degrees of freedom) is instantiated with the kinematic data retrieved from the calibration phase and by inferring the anthropometric table Leva [26], the simulation is ran using the 3D cartesian points as target points for the selected joint axis through PID controllers. The input torque vector producing the virtual human motion in accordance with the one tracked by the Kinect are computed by solving a Linear Quadratic Program (LQP) that optimizes a set of weighted tasks (virtual joint marker trajectories and postural control) subject to equality and inequality constraints translating the physical limitations to implicitly satisfy the human motion. For the LQP problem formulation we refer to [27]. A number of parameters characterizing walking or steady state of the person can be obtained from the simulation of the virtual human (spatio-temporal gait parameters, energy, postural balance etc.). In the experiments presented hereafter we consider more particularly the CoP/ZMP. The CoP/ZMP is computed as follows:

$$\ddot{c} = J_{com}(q)\ddot{q} + \dot{J}_{com}(q,\dot{q})\dot{q}$$
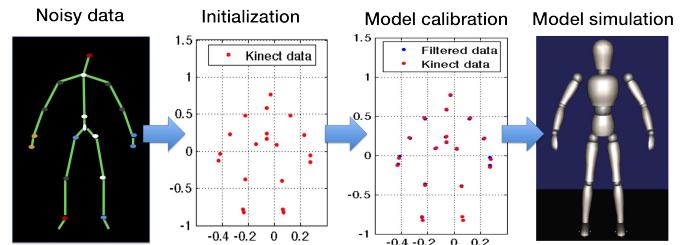$$z = c - \frac{h}{g}\ddot{c} \tag{4}$$

Fig. 2. Example of pose correction for a person in standing position. From the noisy data collected by the Kinect to the animation of the dynamic mannequin.

where J_{com} is the Jacobian of the CoM, h is its height, z and c are respectively the horizontal position vectors of the CoP/ZMP and of the CoM, and g is the gravity value. These equations directly refer to the Linear Inverted Pendulum Model (LIPM), an unstable system whose control must consider the prediction of the future states of the system. z and c would respectively represent the foot of the pendulum and the projection of the point mass onto the ground.
Others parameters such as ankles and knees trajectories are proven be distinguishing factors useful to characterize the postural balance and to detect deviations during walk.

IV. SYSTEM VALIDATION

A. Material and procedure

In order to assess the reliability of the system, a preliminary set of experiments were made in a laboratory setting (Fig. 6 right). Five healthy subjects (see Table I for physical body data) were asked to execute 3 different movements (arm movement, rocking movement and side steps) on a posturography platform while wearing 13 CodaMotion markers, to validate the consistency of our system (see Fig. 6 center for marker placement). The Kinect embedded in the robot was simultaneously used for skeleton detection. The goal of this set of experiments was to assess the CoP/ZMP trajectory provided by Arboris by comparing it with the CoP/ZMP trajectory measured with a posturography platform (ground truth). The CoP/ZMP was chosen because it's the only parameter that can be directly measured by a posturography platform. Both CodaMotion and Kinect data are collected and replayed by dynamic simulation and the resulted CoP/ZMP trajectory was recorded and compared with the posturography platform measurements. All data was filtered using a second order lowpass Butterworth filter with cutoff frequency of $10 Hz$. For Kinect data, the pose correction method described in Section III was applied.

TABLE I. MAIN PHYSICAL BODY CHARACTERISTICS OF THE SYSTEM VALIDATION EXPERIMENT PARTICIPANTS.

.

Subject	Height (m)	Weight (kg)	Gender
P1	1.66	60	F
P2	1.78	78	M
P3	1.8	71	M
P4	1.75	88	M
P5	1.86	79.5	M

B. Results

To evaluate the consistency of our system with respect to the ground truth, the CoP/ZMP trajectories computed by

Arboris replaying both the Kinect and the CodaMotion data, and the CoP/ZMP trajectories measured by the posturography platform were compared. In Figure 3, the 3 CoP/ZMP trajectories related to arm movement executed by one subject are shown. Figures 4 and 5 report the CoP/ZMP trajectories corresponding to rocking movement and to lateral steps for the same subject. The max mean error between the trajectories provided by Arboris by using only the Kinect data (with the correction pose algorithm) and the ground truth was lower than $0.08\ m$. This result proves the accuracy of the system estimating some important parameters related to human activity. Not only the overall motion was well preserved during the simulation, but also the posture assessment showed good results.
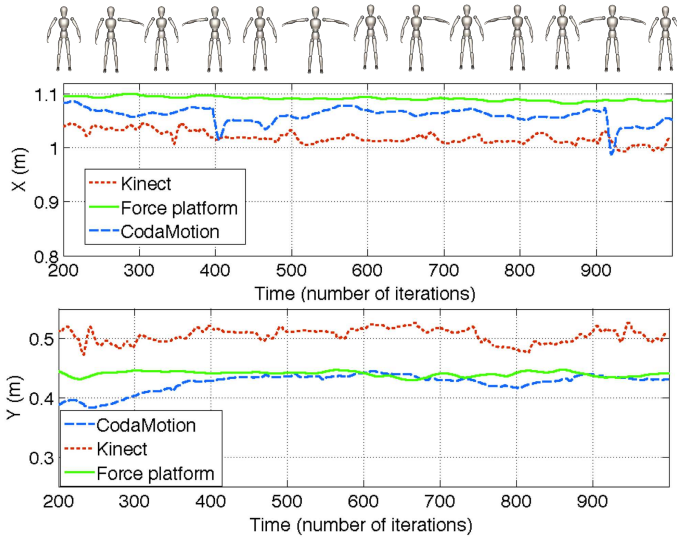

Fig. 3. CoP/ZMP trajectory results of arms movement for a subject. The trajectories issued from Kinect and CodaMotion data are calculated by Arboris.
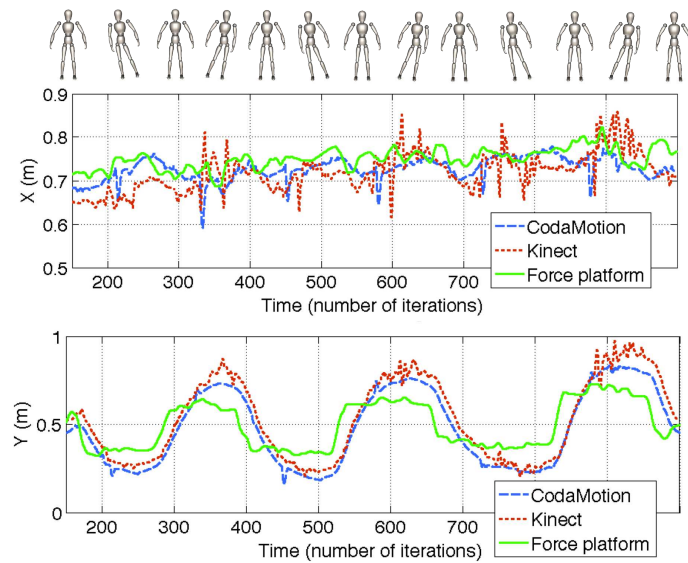

Fig. 4. CoP/ZMP trajectory results of rocking movement for a subject. The trajectories issued from Kinect and CodaMotion data are calculated by Arboris.
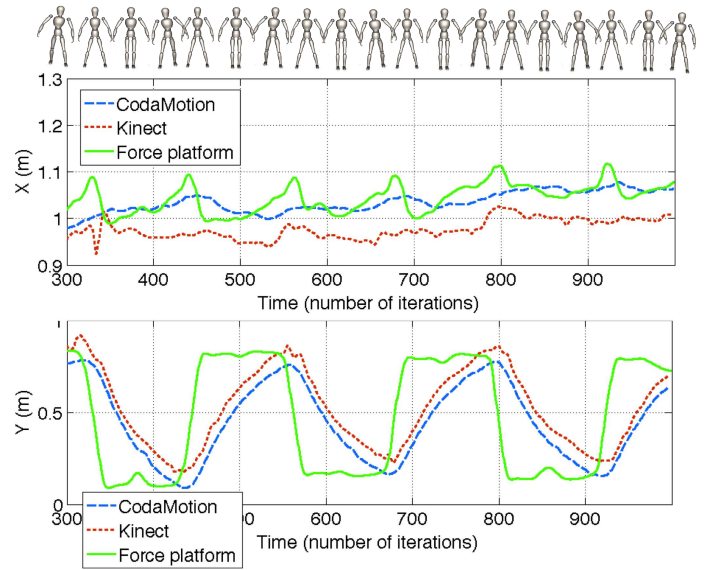

Fig. 5. CoP/ZMP trajectory results of lateral steps movement for a subject. The trajectories issued from Kinect and CodaMotion data are calculated by Arboris.
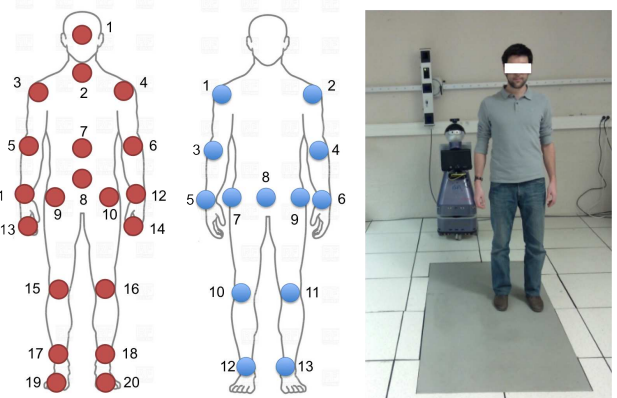

Fig. 6. The anatomical joints detected by the Microsoft Kinect SDK algorithm (left), the CodaMotion markers placement on the subjects' bodies (center) and the experimental setting (right).

V. WALKING ANALYSIS DURING PERSON FOLLOWING

A. Material and procedure

System capabilities to extract some walking spatio-temporal and dynamics parameters of a subject have been evaluated in a real life situation. For this, the robot followed a walking person and tracked him by using the on board sensors under optimal conditions. The experiments were performed with a subject walking normally, then the mobility of the lower limb joint was artificially manipulated in a controlled way. The objective of these experiments was to evaluate the sensitivity of the system to the disturbances, aiming the analysis of pathological walking activities later. Three experiments were conducted to produce permanent changes by mechanical effects on lower limb:

1) Subject walked in his comfortable walking speed (CWS) wearing his usual shoes;

Fig. 7. Mechanical disturbances on the subject: normal (left), right knee immobilization (center) and ski boot on right foot (right)
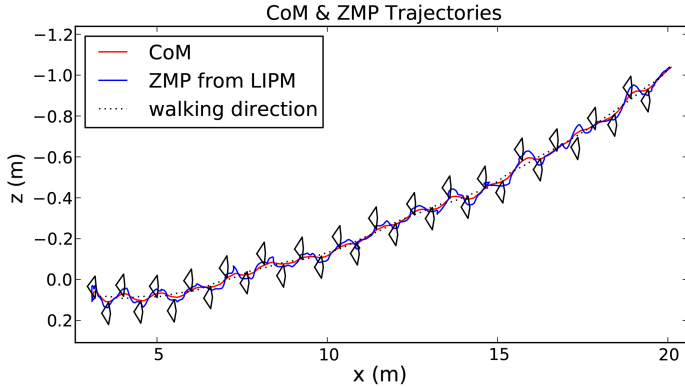

Fig. 8. CoP/ZMP, COM trajectories and feet position during the support phase of normal walking.

2) Subject walked in his CWS wearing a brace on the right leg;
3) Subject walked in his CWS wearing a ski boot on the right foot.

Experiments 1 was the reference for the normal gait to which the results of the other experiments are compared. Experiments 2 and 3 aimed to evaluate the system accuracy in detecting anomalies in the cases of loss of mobility of the knee and ankle joints, respectively. The experiments were performed in the subject's CWS to avoid the influence of velocity constraints in the deviations effects. The mechanical disturbances on the subject are shown in Figure 7.

The subject's movements were sensed by using only the Kinect embedded on the robot and running the skeleton detector algorithm provided by the Microsoft Kinect SDK. The collected data were then processed following the filtering method described in Section III.

B. Results

The CoP/ZMP trajectory during walking is an important parameter for postural balance assessment. In Figure 8 the CoP/ZMP trajectory and the feet position during the normal walk (Experiment 1) are shown. The CoP/ZMP trajectory is computing by using the LIPM.

Overall, the results presented below show that the system is accurate enough to detect both mechanically induced gait disturbances. The 2 parameters that proved key factors to label walking deviations were: the knees flexions and the ankle lateral movement. As expected, when the subject walked wearing his usual shoes, the average flexion of right and left knees were almost equal (Figure 9). Wearing the brace in the right leg, the mean flexion of the right knee during the support phase was lesser than the mean flexion of the left knee ($-0.17\ rad$ of right knee versus $-0.24\ rad$ of the left). The use of the ski boot entailed a slight flexion difference as well ($-0.23\ rad$ of right knee versus $-0.21\ rad$ of the left).

The 2 types of lower limb disturbances tested are distinguishable observing the side movement of the ankles, e.g. the movement orthogonal to the CoM/ZMP during the swing phase. As shown in Figure 10, in the case of normal gait, both movements of right and left ankles were equal. The use of the brace entailed an asymmetry of side movement. The mean amplitude of the lateral movement observed for the right ankle was bigger than the mean value observed for the left ankle ($0.07\ m$ versus $0.12\ m$). The same phenomenon was observed when the subject walked wearing the ski boot ($0.07\ m$ versus $0.13\ m$).

VI. CONCLUSION

This paper proposed a system for the analysis of human activity in everyday life contexts using an autonomous robotic system. We introduced a method for determining the individual's anatomical parameters and the algorithms developed for the digital animation model based on the measurement of a number of characteristic points obtained by the Kinect embedded in the robot. The system was assessed experimentally on different exercises during which the positions of the CoP/ZMP, calculated and measured using a posturography platform, were compared. The potential of the system was also tested in real life conditions exploiting the mobility of the robot and its ability to track persons. These experiments were performed with a subject walking normally, then the mobility of the lower limb joint was artificially manipulated in a controlled way. The system showed sensitivity to the disturbances meaning that it can be exploited to detect pathological walking activities.

REFERENCES

[1] K. Hogan and R. Stubbs. *Can't Get Through: 8 Barriers to Communication.* Pelican Publishing Company, 2003.

[2] B. Elsner and B. Hommel. Effect anticipation and action control. *Journal of experimental psychology: human perception and performance*, vol. 27, issue 1, page 229, 2001.

[3] M. Jeannerod. The cognitive neuroscience of action. Oxford, Blackwell, 1997

[4] K. E. Webster, J. E. Wittwer and J. A. Feller. Validity of the GAITRite® walkway system for the measurement of averaged and individual step parameters of gait. Gait & posture, Elsevier, vol. 22, issue 4, pages 317-321, 2005.

[5] W. Zijlstra and A. L. Hof. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait & posture, Elsevier, vol. 18, issue 2, pages 1-10, 2003.

[6] P. Sardain and G. Bessonnet. Forces acting on a biped robot. center of pressure-zero moment point. In IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 34(5), pages 630-637, 2004.

[7] Y. Brenière. Why we walk the way we do. Journal of motor behavior, vol. 28-4, 1996.
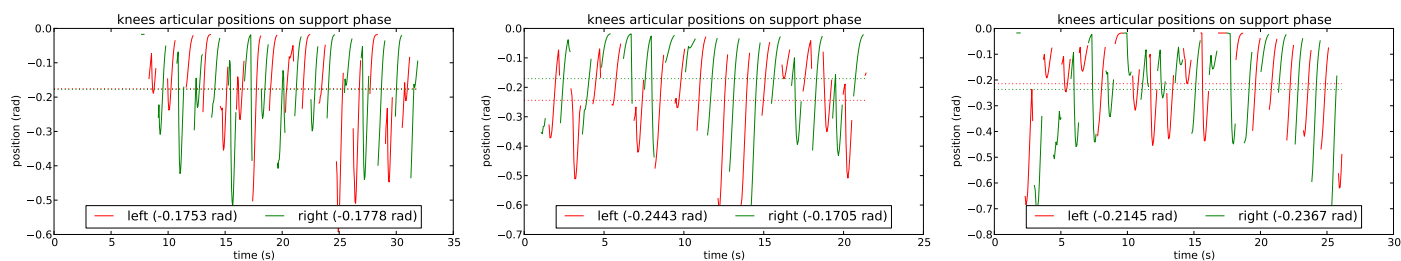
Fig. 9. Knees angular positions: normal walking (left), right knee immobilization (center) and ski boot on right foot (right)
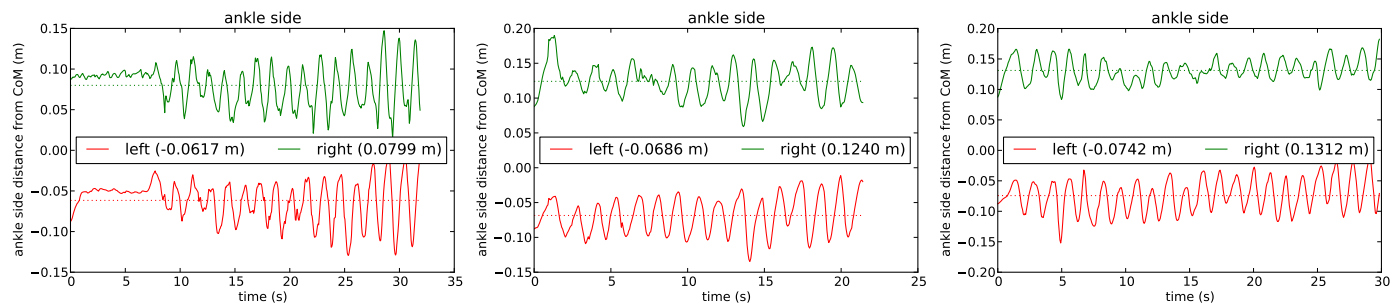
.

Fig. 10. Ankles side movement: normal walking (left), right knee immobilization (center) and ski boot on right foot (right)

.

[8] H. M. Schepers, E. van Asseldonk, J. H. Buurke and P. H. Veltink. Ambulatory estimation of center of mass displacement during walking. In IEEE Transactions on Biomedical Engineering, vol. 56-4, pages 1189-1195, 2009.

[9] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling. The International Journal of Robotics Research, vol. 22-6, pages 371-391, SAGE Publications, 2003.

[10] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A Blake, Real-time human pose recognition in parts from single depth images. In IEEE Conference on Computer Visionand Pattern Recognition (CVPR), pages 1297-1304, 2011.

[11] L. Ren, G. Shakhnarovich, J. K. Hodgins, H. Pfister and P. Viola. Learning silhouette features for control of human motion. ACM Transactions on Graphics (ToG), vol. 24-4, pages 1303-1331, 2005.

[12] R. Poppe. Vision-based human motion analysis: An overview. Computer vision and image understanding, vol. 108-1, pages 4-18, 2007.

[13] A. Sandholm, N. Pronost and D. Thalmann. MotionLab: a Matlab toolbox for extracting and processing experimental motion capture data for neuromuscular simulations. In Modelling the Physiological Human. Springer Berlin Heidelberg, pages 110-124, 2009.

[14] A. González, M. Hayashibe and P. Fraisse. Estimation of the Center of Mass with Kinect and Wii balance board. In IEEE/RSJ Int Conf on Intelligent Robots and Systems (IROS), 2012.

[15] C. Granata and P. Bidaud. A framework for the design of person following behaviors for social mobile robot. In IEEE/RSJ Int Conference on Intelligent Robots and Systems (IROS), 2012.

[16] Š. Obdržálek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, M. Pavel. Accuracy and Robustness of Kinect Pose Estimation in the Context of Coaching of Elderly Population.

[17] S. Tak and H. Ko, Physically based motion retargeting filter, US Patent App. 11/317, 844, 2005.

[18] J. Lee, S. Shin, Motion fairing, in IEEE Computer Animation '96, pages 136-143, 1996.

[19] W. Shen, K. Deng, X. Bai, T. Leyvand, B. Guo and Z. Tu, Exemplar-based human action pose correction and tagging. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[20] R. Duda, P. Hart and D. Stork, Pattern Classification and Scene Analysis 2nd ed. (1995), http://www.svms.org/classification/DuHS95.pdf retrieved March 1, 2013.

[21] B. Schölkopf, A. Smola, R. Williamson and P. Bartlett, New support vector algorithms. Neural computation 12.5, pages 1207-1245, 2000.

[22] V. Duindam and S. Stramigioli. Lagrangian dynamics of open multibody systems with generalized holonomic and nonholonomic joints. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3342-3347, 2007.

[23] S. Barthélemy, J. Salini and A. Micaelli. Arboris-Python. https://github.com/salini/arboris-pyhton

[24] T. Liu and M. Y. Wang. Computation of three dimensional rigid body dynamics of multiple contacts using time-stepping and Gauss-Seidel. In IEEE Transactions on Automation Science and Engineering method, 2005.

[25] J. Salini. Dynamic control for the task/posture coordination of humanoids: toward synthesis of complex activities. PhD thesis. Université Pierre et Marie Curie, 2012.

[26] P. de Leva, Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters, Journal of Biomechanics, vol. 29-9, pages 1223-1230, 1996.

[27] J. Salini, V. Padois and P. Bidaud. Synthesis of complex humanoid whole-body behavior: a focus on sequencing and tasks transitions. In IEEE International Conference on Robotics and Automation ICRA'11, pages 1283-1290, 2011.

[28] F. B. Horak, Clinical assessment of balance disorders, Gait & Posture 6, pages 76-84, 1997.

[29] http://www.codamotion.com/