Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2011)
A. Bargteil and M. van de Panne (Editors)

Human Motion Reconstruction from Force Sensors

Sehoon Ha Yunfei Bai C. Karen Liu

Georgia Institute of Technology

Abstract

Consumer-grade, real-time motion capture devices are becoming commonplace in every household, thanks to the recent development in depth-camera technologies. We introduce a new approach to capturing and reconstructing freeform, full-body human motion using force sensors, supplementary to existing, consumer-grade mocap systems. Our algorithm exploits the dynamic aspects of human movement, such as linear and angular momentum, to provide key information for full-body motion reconstruction. Using two pressure sensing platforms (Wii Balance Board) and a hand tracking device, we demonstrate that human motion can be largely reconstructed from ground reaction forces along with a small amount of arm movement information.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.7]: Three-Dimensional Graphics and Realism—Animation; Simulation and Modeling [I.6.8]: Types of Simulation—Animation;

1. Introduction

3D human motion capture and reconstruction came a long way from expensive, cumbersome mechanical systems to markerless, real-time devices affordable to every household. Nintendo Wii was the first immersive gaming platform with the accelerometer-based gesture recognition system. Sony Playstation Move introduced a more accurate hand tracking system using a single video camera. The success of Microsoft Kinect, a depth-camera gaming accessory sold eight millions in 60 days, demonstrated that consumer devices with capability of recognizing human movement are becoming commonplace in our daily life. Although the accuracy and quality of the motion produced by consumer-grade devices are still not on par with high-end optical motion capture systems, their capability is sufficient for everyday applications, such as video games or user interfaces for electronic devices.

Today, most widely used mocap systems employ vision-based techniques to capture 3D positions of a set of body points, followed by an inverse kinematics procedure to reconstruct joint angles. Vision-based approaches have numerous advantages over other systems, but the output quality suffers when the user is heavily occluded (including self-occlusion) or under poor lighting condition. In view of these drawbacks, we propose a new way to capture and reconstruct 3D human motion supplementary to vision-based techniques. Our approach leverages ground reaction forces from force sensors to approximate full-body pose and only uses vision devices to track the hand positions of the user. Because we largely reduce the dependency on vision devices, our technique is open to any algorithms and consumer-grade devices capable of hand tracking.

The motivation of using ground reaction force to reconstruct human motion is inspired by the previous studies of linear and angular momentum of human motion. Researchers in both computer animation and robotics have demonstrated that regulating body momenta is one of the most effective ways to control full-body human movement for a variety types of motions. Therefore, we hypothesize that body momenta provide a very important piece of clue such that only very little additional information, positions of hands and feet, is needed to reconstruct the full-body motion. We could employ a vision-based mocap system to approximate body momenta by measuring the position of a dense set of body points. However, a much more effective way to obtain body momenta is through measuring external forces applied on the user and the center of pressure of the user. With this information, we can directly derive linear and angular momentum of the full body using Newton Second Law, rather than approximating them with a set of measured body points.

To this end, we design an accessible mocap system based on two Wii Balance Boards and a hand tracking device. Our primary demonstration uses Sony Palystation Move as the hand tracking device, but we also show that other motion gaming systems, such as Nintendo Wii and Microsoft

SCA 2011, Vancouver, British Columbia, Canada, August 5 – 7, 2011.

DOI: 10.2312/SCA/SCA11/129-138

Kinect, can be incorporated with our method. Our approach leverages Wii Balance Boards to enable full-body tracking capability for Wii and Playstation systems, and improve the quality of motion for the Kinect system. During the capture session, the user simply stands on the Wii Balance Boards. If Wii or Playstation systems are used, additional Wiimotes or Move controllers need to be held in the user's hands. No special garment or markers are required to use the system. While Move and Kinect provide more accurate hand positions, an addition benefit with Wii system is that it can be operated in any location (indoors or outdoors) under any lighting condition.

Though affordable, a Wii Balance Board only measures the vertical component of the contact force and a coarse approximation of the center of pressure. Similarly, a Wiimote only provides noisy acceleration data for slow movement in the local frame of the device, rather than accurate positional data in the world frame. The main difficulty of constructing such a system arises from the inherent limitations of the low-cost sensors and numerical errors. In addition, real-time reconstruction of freeform, full-body poses based only on sparse sensor information presents great engineering challenge. Our inverse kinematics procedure based on Principle Component Analysis and Fibonocci search achieves these goals at the rate of 60 frames per second.

We demonstrate sequences of arbitrary freeform motion captured in a live setting. The goal of this work is not to compete against the vision-based mocap systems in terms of motion quality and accuracy, but rather to demonstrate that the force sensor can provide significant benefit to existing, consumer-grade mocap systems. The accuracy of the system is evaluated visually and quantitatively by comparing the results with a lab-grade optical motion capture system. Our results show that Sony Playstation Move and Wii systems can potentially track full-body motion if augmented with force sensors. We also demonstrate that, when subject to occlusion, the quality of the poses reconstructed by Kinect can be largely improved by exploiting the dynamics aspect of human movement.

2. Related Work

Vision-based motion capture systems have been widely used in applications ranging from military training to rehabilitation, to recreation. Top-of-the-line optical motion capture systems provide great accuracy and high temporal resolution but usually require the subject to wear retro-reflective markers or powered LED markers in a calibrated space with multiple cameras [Vic, Pha]. These devices produce clean 3D position of markers placed densely on the subject's body. Reconstruction of human poses usually involves marker labeling and inverse kinematics, both of which can be solved reliably and efficiently by existing techniques. Markerless motion capture systems have also been proposed and studied extensively. With careful calibrated environment, multiple video cameras can be used to capture human motion [CTMS03, CBK05, dAST*08]. Recently, markerless motion capture systems have advanced rapidly thanks to the breakthrough development in the technology of depth- cameras. With a single consumer-grade depth-camera [Mic, Pri, Can], approximated human poses can be captured and reconstructed in real-time. New applications in this extremely active and unexplored area are proposed everyday by researchers and hobbyists in public domains [Ope, Kin]. Nevertheless, the raw data collected by a single depth camera are usually ambiguous for certain human poses and sensitive to occlusion. As a result, the reconstruction of human motion highly depends on heuristics, prior knowledge, and pre-captured kinematic data. Our technology is complementary to any vision-based systems. In particular, systems using a single depth camera can benefit from our method whenever there is occlusion or ambiguity in pose reconstruction.

Based on miniature inertial sensors, such as accelerometers, gyroscopes, gravitometers, or magnetometers, inertial motion capture systems measure linear acceleration and angular velocity at different body locations by placing a dense set of inertial sensors on the subject [XSea]. To obtain the position and the orientation of the body part, we need to integrate the raw data over time which inevitably introduces significant numerical drift. A few systems combine acoustic sensors to reduce accumulation errors [VAV*07, Int], but the precision of the position information achieved by these sensors is still not on par with that of optical motion capture techniques. In spite of these limitations, Slyper and Hodgins [SH08b] used five accelerometers along with a database of pre-captured motions to reconstruct upper body motions. Shiratori and Hodgins [SH08a] demonstrated that Nintendo Wiimotes can be used to control a simulated biped character. In one of our systems, we use two Wiimotes to reconstruct arm motion. Instead of integrating the sensor data, we use a Cascade Neural Network to map the sensor input data to the most plausible arm pose.

Force platforms do not directly measure human motion, but they capture ground reaction forces generated by the body movement. Advanced platforms provide center of pressure and both vertical and shear forces [Adv]. More commonly used devices only measure the pressure pattern at the contact area but do not provide shear forces. Yin and Pai demonstrated that a pressure sensor pad, XSensor [XSeb], can be an intuitive interface to control virtual characters performing pre-captured motions [YP03]. Our system uses a much simpler and lower cost device, Wii Balance Board [Wii] and we do not require a database of pre-captured motions at runtime. Although the Wii Balance Board cannot provide horizontal force needed to compute the center of mass (COM) of the user, we observe a strong correlation between the vertical forces distribution and the horizontal movement of the COM. Our algorithm is able to exploit this relationship to reconstruct the 3D linear momentum.

Inverse kinematics (IK) is a common technique for reconstructing poses from 3D marker positions measured by optical mocap systems. Reliable and efficient methods have been proposed and adopted widely by animation and robotics applications [Wel93,ZB98,YN03]. A few IK algorithms utilize a database of pre-captured human motion to produce more natural poses [WH97,RSC01,KG04,GMHP04]. We also formulate an IK problem to reconstruct a pose from sensor data at each time step. Without dense marker positions, our IK method is still able to reconstruct plausible poses based on the linear and angular momentum constraints, and a solution space biased toward the example poses. In addition, we overcome the intensive computation of linear and angular momentum constraints by breaking a high-dimensional optimization into a sequence of one-dimensional subproblems.

The importance of linear and angular momentum for biped motion has been demonstrated in various applications from humanoid robots [KKK*03, GK04, HP08] to character animation [LP02, ALP04, KLK04, MZS09]. Much previous work focused on regulating body momenta to achieve task-level goals and maintain balance. Our work further demonstrates that body momenta provide the crucial information for kinematic pose reconstruction.

3. Overview

We design a real-time system to demonstrate that freeform, full-body human motion can be largely reconstructed from ground reaction forces along with a small amount of hand movement information. Our system uses two Wii Balance Boards to reconstruct core features of human motion including linear momentum, angular momentum and foot positions. These features enable us to approximate torso and lower body movements. The detailed arm motion, however, requires additional sensors to detect hand positions.

Our technique works with any sensor device capable of tracking hand positions. We demonstrate three stand-alone mocap systems by augmenting popular gaming input devices, PlayStation Move, Nintendo Wiimote, or Microsoft Kinect, with consumer-grade force sensors. Nintendo Wiimotes do not provide positions of hands and require additional algorithms to recover positions from accelerameter data. On the other hand, Microsoft Kinect provides 3D hand positions along with other 13 feature points on the human body, which is more than necessary for our system. In the following sections, we will describe how Wii Balance Board can enable the functionality of full-body tracking for Wii and Playstation, and improve the pose quality for Kinect. The overview of the system is described in Figure 1.

4. Force Sensors

Assuming there is no external force applied on the user other than gravity and the ground reaction force, the rate of momenta of the user can be computed by

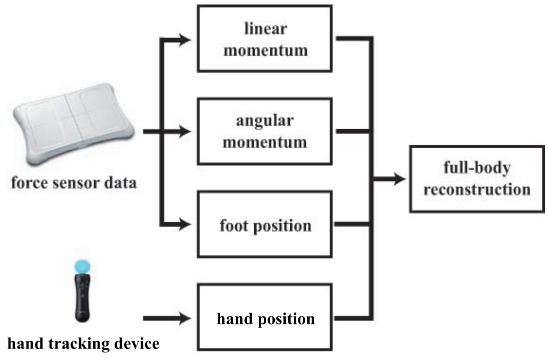

Figure 1: *System overview.*

$$\dot{\mathbf{P}} = m\mathbf{g} + \mathbf{F}$$
$$\dot{\mathbf{L}} = (\mathbf{p} - \mathbf{c}) \times \mathbf{F} \tag{1}$$

where $\mathbf{P}$ and $\mathbf{L}$ are the linear and the angular momentum respectively, $m\mathbf{g}$ is the gravitational force, $\mathbf{F}$ is the ground reaction force, $\mathbf{c}$ is the center of mass (COM), and $\mathbf{p}$ indicates the center of pressure (COP). In theory, if we can acquire accurate ground reaction forces and the COP in real-time, we can reconstruct body momenta from the current state of the user by applying Equation 1 and the numerical integration.

In practice, a 6-axis force platform capable of measuring 3D forces and the accurate COP has a steep price ranging from $5,000 to $25,000. Instead of using an expensive equipment, we hypothesize that human body moves in a highly coordinated fashion such that a simpler (and cheaper) device is sufficient for computing body momenta. To this end, we choose to use one of the most accessible pressure sensing devices, Wii Balance Board, at the cost about $80.

Wii Balance Board uses four pressure sensors at the four corners to measure the weight and the COP of the user. In contrast to high-end force platforms, each sensor at the corner of the Wii Balance Board only measures the vertical component of the contact force. The COP can be approximated by interpolating the four vertical forces, but the lack of horizontal force measurement prevents us from reconstructing body momenta using Equation 1.

Despite that the measurement from Wii Balance Board is theoretically insufficient to determine body momenta, our algorithm is able to predict the horizontal motion based on the following key observation: The aggregate vertical contact force does not provide any information about horizontal motion, but if we consider each individual vertical force measured at the corner separately, the distribution of the vertical force shows strong correlation to the horizontal motion of the COM. Our system further exploits two Wii Balance Boards, one for each foot, to measure contact force distribution in eight locations. In the rest of the section, we will describe the reconstruction of linear momentum for vertical and horizontal components separately, followed by the reconstruction of angular momentum and the foot positions.

4.1. Vertical linear momentum

Vertical linear momentum can be directly computed from contact forces measured by Wii Balance Board and numerical integration. Because our algorithm directly uses the position of the COM to reconstruct the full-body pose, we need to integrate the measured contact force twice at each time step.

$$c_y^{(t+1)} = c_y^{(t)} + \dot{c}_y^{(t)} h + \frac{F_y}{m} h^2 \tag{2}$$

We apply discrete Kalman Filter [Ras] to remove some of the noise in the input data. However, as long as the time step h is greater than zero, numerical error will accumulate and the position of the COM will eventually drift. We develop a practical solution to the issue of numerical drift.

From our empirical data, a common source of numerical drift comes from outliers in the acceleration domain. Although short lasting, one time step of outlier acceleration can cause the velocity to spike. Even if the acceleration immediately drops to zero thereafter, the velocity will remain at a high constant value and the position in turn will drift rapidly. From the empirical observation of human motion, it is a rare situation when the COM moves at a nonzero constant speed. Our analysis shows that less than 0.5% of the motion in a large dataset (four minutes) maintains a nonzero constant COM velocity more than 0.05 second. In other words, when the acceleration of the COM is zero, the velocity of the COM is very likely to be zero. Based on this observation, our algorithm gradually reduces the vertical velocity of the COM to zero, when the vertical acceleration of the COM is near zero for a short period of time.

4.2. Horizontal linear momentum

Unlike vertical linear momentum, we cannot measure the horizontal forces directly from the Wii Balance Board. We therefore employ a data-driven approach to learn a predictive model that captures the relation between the distribution of the vertical contact force and the horizontal motion of the COM (c_x and c_z).

To learn an effective model from data, we need to carefully select the features included in the input vector and the output variables. For x-axis, our input vector is defined as $\mathbf{X} = \{\mathbf{F}^{(t-1)}, \mathbf{F}^{(t)}, c_x^{(t-2)}, c_x^{(t-1)}, c_x^{(t)}, c_z^{(t)}\}$ and the output vector is $\mathbf{Y} = \{c_x^{(t+1)}\}$. For z-axis, we simply swap the subscript x and z in the input and the output vectors. The first two terms in $\mathbf{X}$ capture the change of force distribution ($\mathbf{F}$ contains vertical forces measured by eight pressure sensors) from the previous time step to the current one, while the last two terms indicate the current horizontal position of the COM. We also include two previous positions of the COM, $c_x^{(t-2)}$ and $c_x^{(t-1)}$, in the axis of interest to provide approximated velocity and acceleration of the COM.

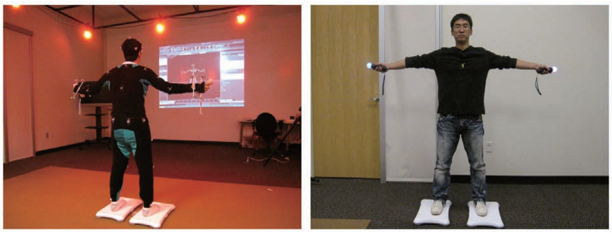

Figure 2: *Left: Acquisition of training examples. Right: T-pose for initial calibration.*

We choose a nonlinear data modeling tool, artificial neural network (ANN), capable of modeling complex relationships between inputs and outputs. To generate examples in the training set, we simultaneously capture full-body motion and vertical contact forces using a 12-camera Vicon system and two Wii Balance Boards (Figure 2). The user performs freeform motions that require change of COM horizontally for about three minutes. The COM positions in the input and output vectors of the training set can be computed from the mocap data.

Cascade ANN For training ANN, we use a cascade-correlation neural network. Cascade-correlation is a generative, feed-forward, supervised learning algorithm for artificial neural network [FL90]. It is different from standard back-propagation algorithm in that it begins with a minimal network and adds one hidden unit at a time automatically through training. Therefore, cascade ANN is useful for incremental learning and can be 10 times faster than back-propagation algorithm.

The cascade ANN starts from zero hidden neurons, and is trained until stagnation. We generate candidate hidden neurons which are connected to all input and previous hidden neurons and try to maximize the correlation between the activation of the candidate units and the residual error of the net. Candidate unit with the maximum correlation is added to the net and links between selected unit and all output units are generated. Weights to previous neurons are frozen after being added and never changed in future epochs of training. Keep adding new hidden units until overall error of the net falls below the threshold.

4.3. Angular momentum

To compute the change of the angular momentum, we modify Equation 1 for two Wii Balance Boards:

$$\dot{\mathbf{L}} = (\mathbf{p}_L - \mathbf{c}) \times \mathbf{F}_L + (\mathbf{p}_R - \mathbf{c}) \times \mathbf{F}_R \tag{3}$$

where the subscripts L and R stand for left and right foot respectively. The COP and vertical contact force are measured directly using Wii Balance Boards. The horizontal force is computed by double-differentiating the reconstructed horizontal COM.

The angular momentum at each time step $\mathbf{L}^{(t)}$ is computed via integration of $\dot{\mathbf{L}}$. Based on the same empirical rule we apply to linear momentum for numerical drift reduction, we gradually reduce $\mathbf{L}$ to zero when $\dot{\mathbf{L}}$ becomes zero for a short period of time.

4.4. Foot positions

In addition to providing momentum information, Wii Balance Boards can also provide an approximated foot position by measuring the COP. When the COP is near-static, it is a good indicator of the center location of the foot. We can then set positional constraints on the heel and the toes accordingly for the full-pose reconstruction. Unfortunately, the COP measured by the Wii Balance Board usually moves rapidly around the foot even when the foot is completely fixed on the board. This is due to the poor approximation of the COP using only four pressure sensors at the corners.

We assume that the user does not intentionally slide her feet on the Wii Balance Boards during the capture. This implies that any change of the foot position must be achieved by lifting the foot and placing it in a different location on the board. Therefore, change of foot position can be detected if a noticeable period of zero pressure on the Wii Balance Board occurs. This assumption allows us to ignore most sporadic COP movement, but it does not help us follow the user's foot positions. As soon as the foot is lifted from the Wii Balance Board, we lose all the information about the foot until it reappears at a different location some time later.

Without additional device to track the foot, it is unlikely to accurately predict the stepping length and the direction in real-time. However, as long as one foot is still on the Wii Balance Board, we still have the information of body momenta which provides an approximated location of the unconstrained leg. To generate more detailed motion for the unconstrained leg, we use a heuristic to guess the vertical position of the foot. When the COP disappears from the board, we lift the positional constraints on both heel and toes by $15cm$. When the COP reappears on the board on a new location, we gradually interpolate the foot position to that location.

5. Arm Movement

With the information provided by Wii Balance Boards, namely body momenta and foot positions, we are able to reconstruct reasonable poses with plausible lower body and torso motion. However, the arm motion, where most nuances and details of human motion lie, is difficult to be reconstructed by body momenta because arms have smaller mass comparing to the rest of the body. We describe three possible devices to obtain the additional data for arm motion reconstruction.

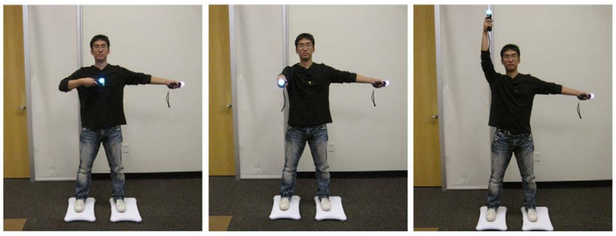

Figure 3: *We use three poses to define the origin and the three axes of the physical world for calibrating Playstation Move.*

5.1. Vision camera: Sony Playtation Move

Sony Playstation Move is best suited for the purpose of our system because it balances between accurate hand tracking and the cost of the device. The device consists of one web camera (Eye Toy, \$29) and two Move controllers (\$39 apiece). Each Move controller is equipped with an accelerometer, a gyroscope, a magnetometer, and an illuminated sphere at its head. 3D position of the sphere can be easily tracked in real-time. For the purpose of our system, we only use the information about 3D position and ignore other sensor data. The 3D position measured by a Move controller needs to be converted to the coordinate frame of the virtual world. We design a simple calibration process to match the origin and axes of real world to those of the virtual world. (Figure 3)

5.2. Inertial sensor: Nintendo Wiimote

Among all three devices, Nintendo Wiimotes (with Motion Plus) provide the least accurate information for hand tracking. Each Wiimote (\$38) is equipped with an accelerometer that measures linear acceleration and a gyroscope that measures angular velocity. Ideally, if we were able to track hand positions using inertial sensors, our system would have enough information to reconstruct full-body poses with plausible arm movement. Unfortunately, inertial sensors are designed to measure acceleration or velocity, rather than accurate positional data. Directly integrating the inertial data over time to obtain position is unreliable due to the noise and the numerical drift. In addition, a Wiimote measures the acceleration and the velocity in the local frame of the device. Using an approximated world frame to compute the hand position will further introduce errors.

Although we cannot integrate the inertial data to obtain hand positions, it is possible to learn a neural network model to directly reconstruct the arm joint angles. The main challenge is to select a set of appropriate features to include in the input vector. An exhaustive search of all possible subsets of features is a prohibitively inefficient approach. We select features based on our understanding of articulated rigid body kinematics. Consider the following two equations that describe the relation between the sensor input and the motion output:

$$\boldsymbol{\omega} = \mathbf{J}_\omega(\mathbf{q})\dot{\mathbf{q}} \tag{4}$$

$$\boldsymbol{\alpha} = \dot{\mathbf{J}}_v(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{J}_v(\mathbf{q})\ddot{\mathbf{q}} \tag{5}$$

where $\mathbf{J}_\omega$ and $\mathbf{J}_v$ are the Jacobians evaluated at the joint configuration $\mathbf{q}$. Assuming that the arm motion does not accelerate excessively (i.e. $\ddot{\mathbf{q}}$ is relatively small), given the current joint configuration of the arm $\mathbf{q}_a$, it is quite possible to train a neural network model that captures the relation between the combination of linear acceleration $\boldsymbol{\alpha}$ and angular velocity $\boldsymbol{\omega}$, and the joint velocity $\dot{\mathbf{q}}_a$ of the arm.

Guided by these kinematic relationships, we define our input vector as $\mathbf{X} = \{\mathbf{q}_a^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\omega}^{(t)}\}$, where $\mathbf{q}_a^{(t)}$ is the current arm joint configuration, and $\boldsymbol{\alpha}^{(t)}$ and $\boldsymbol{\omega}^{(t)}$ are the linear acceleration and angular velocity of the hand measured by the inertial sensor. Based on these features, we use cascade ANN to predict the joint angles of the arm in the next time step, $\mathbf{Y} = \{\mathbf{q}_a^{(t+1)}\}$.

5.3. Depth camera: Microsoft Kinect

Microsoft Kinect alone is sufficient to capture full-body motion in real-time. Using Kinect as a hand tracking device might not be an economical design choice. However, Kinect occasionally fails to estimate certain types of motion due to self occlusion of the user. We demonstrate that combining a depth camera with force sensors can improve the quality of the full-body motion reconstructed by Kinect.

Kinect consists of an infrared laser projector and a monochrome CMOS sensor (depth camera) which produces a 2D depth map in size of 640 by 480 with an 1.3mm error per pixel. Details of skeleton reconstruction algorithm of Kinect are not opened to public, but it roughly builds a model using Maximum-A-Posteriori estimation with a huge database of human poses. In terms of the implementation, we use OpenNI framework to detect feature points. OpenNI detects 15 joint positions, two of which provide the current positions of the hands.

6. Motion Reconstruction

Our system requires a simple initial calibration procedure to match the user and the character's initial poses and measure the weight of the user. The user is asked to stand on the Wii Balance Boards and assume a T-pose similar to Figure 2 for five seconds. During this procedure, we use Wii Balance Boards to measure the weight and the COP of each foot. We then set the joint angles of the character to a default T-pose with the foot positions similar to the user's. The initial COM, COP, and joint angles used in body momenta and arm pose reconstruction are taken from this initial pose of the character.

At each time step, we formulate an optimization problem to solve for a full-body pose $\mathbf{q}$ subject to reconstructed momentum $\mathbf{P}^{(t)}$ and $\mathbf{L}^{(t)}$ and measured end-effector positions $\mathbf{G}^{(t)}$.

$$\min_{\mathbf{q}} \|P(\mathbf{q}) - \mathbf{P}^{(t)}\|^2 + \|L(\mathbf{q}) - \mathbf{L}^{(t)}\|^2 + \|G(\mathbf{q}) - \mathbf{G}^{(t)}\|^2 \tag{6}$$

A real-time application usually cannot afford solving this optimization at each time step, because it is high-dimensional, nonconvex, and involving complex linear and angular momentum constraints. Furthermore, because our optimization problem is largely under-constrained, there exist many solutions but most of the resulting poses are unnatural. To handles these two issues, we introduce a much more efficient IK algorithm that leverages a generic set of recorded poses (90 second range of motion). Similar to previous methods [GMHP04, IWZL09], we use examples to represent the space of desired solutions. Our algorithm applies PCA on a generic set of poses, but unlike Ishigaki et al. [IWZL09], we break one high-dimensional optimization problem into a sequence of one-dimensional subproblems. To solve Equation 6, we perform one-dimensional Fibonacci search along each principle component in the order of its importance. By solving the variables one by one, we can achieve faster performance and more natural poses toward the principal components. One drawback of 1D Fibonacci search is that it sometimes fails to achieve accurate solution and generates footskating. In those cases, we use a conventional optimization-based IK to solve for a small IK chain from the hip to the foot. With simple Fibonacci search, our IK method runs at 160/60 Hz without/with a footskating cleaning routine respectively. For the system that uses Wiimotes, we replace the term that minimizes errors in hand positions in Equation 6 with a term that minimizes the error in arm joint angles.

7. Evaluation

We implemented and tested our system on the 3.20GHz CPU with a single core. Our system can run at real-time with 60 frames per second (fps) in the worst case. Five minutes of offline motion capture data were used to train one neural network for the COM and two for the arms. We also took 90 seconds of the motion to compute the principle components for the IK process. None of the offline data were used during online capture sessions.

We performed two different types of evaluation on our system. First, we provided live demonstrations of our system running in real-time in the accompanying video. We showed that our system is able to reconstruct freeform, full-body motion (Figure 4) on Playstation Move, Nintendo Wii, and Microsoft Kinect. Second, we showed a few comparisons to

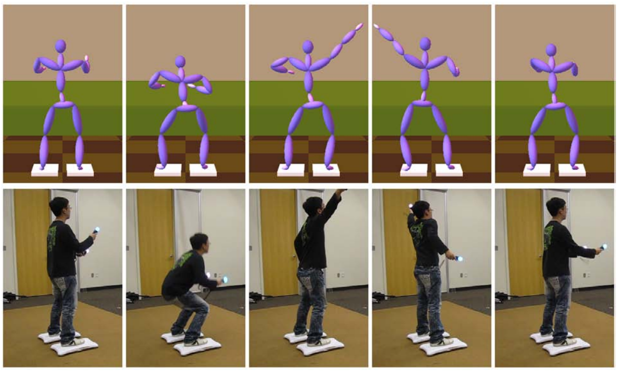

Figure 4: *A random freeform sequence.*

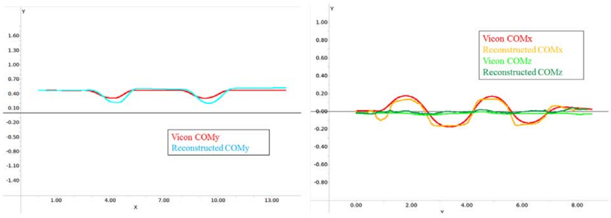

Figure 5: *Left: Reconstructed vertical COM trajectory for a squat motion compared with Vicon System. Right: Reconstructed horizontal COM trajectories for a leaning motion compared with Vicon system.*

evaluate the accuracy of the reconstructed motion features, the quality of our IK solutions, the improvement provided to Kinect, and the importance of the angular momentum. We also demonstrated that our system can be used by different users without any modification.

7.1. Accuracy of reconstructed motion features

To assess the accuracy of the motions produced by our system, we quantitatively compared the reconstructed motion

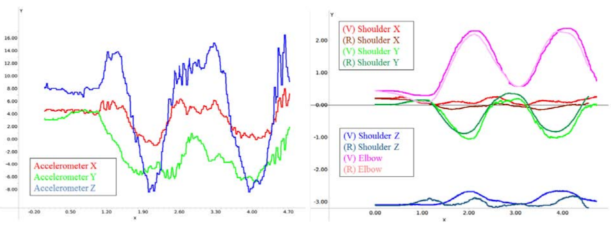

Figure 6: *Left: Linear acceleration recorded by a Wiimote. Right: Reconstructed shoulder (3-DOF) and elbow (1-DOF) trajectories compared with Vicon System.*

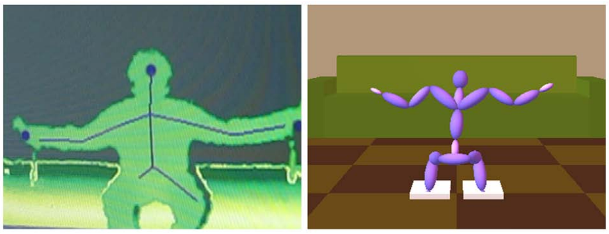

Figure 7: *Left: A squat pose reconstructed by Microsoft Kinect. Right: The same squat pose reconstructed by our system.*

features, namely momentum (by Wii Balance Boards) and arm joint angles (by Wiimotes), to those directly measured by a high-end Vicon motion capture system which consists of 12 MX40+ cameras. The user's motions were simultaneously captured by the Vicon system and our system. We showed the comparison of a squat motion, a leaning motion and a arm swing motion in Figure 5 and Figure 6 (Right). For the COM reconstruction, although our results sometimes exhibit more noise, they closely follow the motion capture trajectory for the entire sequence. The arm reconstruction is much more challenging given the large degree of noise in the input data recorded by a Wiimote (Figure 6 Left). A few parts of the reconstructed motion are not as accurate, but our results are able to capture the characteristics of the motion patterns recorded by the mocap system.

7.2. Quality of IK solutions

We provided a side-by-side comparison between the motion produced by our system (with Wii Balance Boards and Playstation Moves) and the same motion captured by Vicon MX40 12-camera system with 53 markers. If we treat the results from Vicon system as ground truth, our motion is comparable in overall quality, considering that much sparser sensory information is available in our system. Biasing the IK solution toward the subspace of example poses can largely improve the quality of the reconstructed motion. Please refer the video. However, our system occasionally generates undesired artifacts around the shoulders and fails to capture the subtle motion of the torso.

7.3. Improvement on Microsoft Kinect

Microsoft Kinect by itself can be used as a full-body pose tracking device. Occasionally, Kinect fails to produce high-quality poses due to the problem of occlusion. For example, when the user crouches down, Kinect loses track of the lower body completely. We tested a few known difficult poses for Kinect and show that force sensors can help Kinect produce reasonable poses when the depth camera does not provide reliable data (Figure 7 and the video).

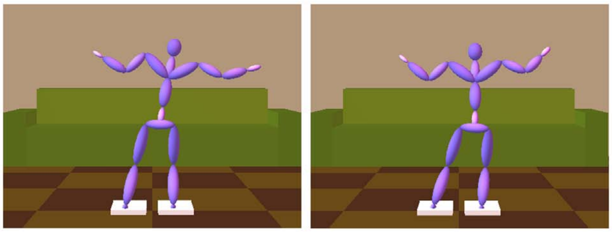

Figure 8: *Left: Pose reconstruction without angular momentum. Right: Pose reconstruction with angular momentum.*

7.4. Effect of angular momentum

The bottleneck of our system lies in the computation of angular momentum. However, considering angular momentum in pose reconstruction significantly improves the quality of the motion (Figure 8 and the video). Without the angular momentum, all body parts of the character tend to move in unison to match the desired COM. When the additional constraint on the angular momentum is considered, relative motion between body parts becomes evident due to the effect of moment of inertia.

7.5. Limitations

Our system has a few limitations due to the input sensors and reconstruction algorithms. Although we overcome many of the drawbacks, the Wii Balance Board is simply not designed for mobile usage. As a result, the user's locomotion is constrained and the feet are bound to a small area.

The accuracy in the reconstructed motion is an aspect we need to improve for applications demanding high degree of accuracy. The small numerical drift is still the source of inaccuracy and is inevitable due to the inherent limitations of force and inertial sensors.

Although our system is able to reconstruct stepping, it becomes increasingly inaccurate after the foot leaves the Wii Balance Board for a long period of time. We can approximate the leg position based on measured body momenta, but detailed leg motion is not possible to reconstruct due to the lack of sensor information. This drawback prevents us from capturing some interesting lower body motion, such as kicking or ballet dancing.

8. Conclusions

This paper introduces a real-time, 3D human motion capture system using force sensors and a hand tracking device. We are able to reconstruct freeform, full-body motion from two Wii Balance Boards in conjunction with Sony Playstation Move, Nintendo Wii, or Microsoft Kinect. Our algorithm exploits the dynamic aspects of human movement, such as linear and angular momentum, to provide key information for full-body motion reconstruction. As an alternative to vision-based, real-time mocap systems, we demonstrate that human motion can be largely reconstructed from ground reaction forces along with a small amount of arm movement information.

Besides Newtonian physics, our reconstruction depends on neural networks trained from freeform motion data. We found that cascade ANN a very effective tool to avoid overfitting when learning a complex, nonlinear relation between sensor inputs and motion outputs. However, the key to a robust neural network model lies in feature selection. By applying the prior knowledge in kinematics and dynamics of human motion, we selected features that are mostly relevant to motion reconstruction while maintaining a manageable size of input domain.

One exciting future avenue is to explore the possibility to replace force sensing platforms with a pair of force sensing shoes. The user will no longer be bound to a fixed location and will be able to locomote freely in the space or even outdoors while being motion captured. This system can potentially be a low-cost, portable mocap solution for many practical applications.

References

[Adv] AdvancedMechanicalTechnology: *Multi-Axis Force Plates, http://www.amti.biz/*. 2

[ALP04] Abe Y., Liu C. K., Popović Z.: Momentum-based parameterization of dynamic character motion. In *Eurographics/SIGGRAPH Symposium on Computer Animation* (2004), pp. 173–182. 3

[Can] CanestaTechnology: *Canesta Technology, http://canesta.com/*. 2

[CBK05] Cheung K., Baker S., Kanade T.: Shape-from-silhouette across time: Part ii: Applications to human modeling and markerless motion tracking. In *International Journal of Computer Vision* (2005). 63(3):225–245. 2

[CTMS03] Carranza J., Theobalt C., Magnor M., Seidel H.: Free-viewpoint video of human actors. In *ACM Transactions on Graphics* (2003). 22(3):569–577. 2

[dAST*08] d. Aguiar E., Stoll1 C., Theobalt C., Ahmed N., Seidel H.-P., Thrun S.: Performance capture from sparse multi-view video. In *ACM Transactions on Graphics* (2008). 27(3). 2

[FL90] Fahlman S. E., Lebiere C.: The cascade-correlation learning architecture. In *Advances in neural information processing systems 2* (1990), pp. 524–532. 4

[GK04] Goswami A., Kallem V.: Rate of change of angular momentum and balance maintenance of biped robots. In *Proc. IEEE Int'l Conf on Robotics and Automation* (2004), IEEE, pp. 3785–3790. 3

[GMHP04] Grochow K., Martin S. L., Hertzmann A., Popovic Z.: Style-based inverse kinematics. *ACM Transactions on Graphics 23*, 3 (2004), 522–531. 3, 6

[HP08] Herr H., Popovic M.: Angular momentum in human walking. In *Journal of Experimental Biology* (2008). 3

[Int] IntersenseTechnology: *Intersense Technology, http://www.intersense.com/*. 2

[IWZL09] Ishigaki S., White T. M., Zordan V. B., Liu C. K.: Performanced-based control interface for character animation. *ACM Transaction on Graphics 28*, 3 (Aug. 2009). 6

[KG04] Kovar L., Gleicher M.: Automated extraction and parameterization of motions in large data sets. *ACM Transactions on Graphics 23*, 3 (2004), 559–568. 3

[Kin] KinectHacks: *Kinect Hacks, http://kinecthacks.net/*. 2

[KKK*03] Kajita S., Kanehiro F., Kaneko K., Fujiwara K., Harada K., Yokoi K., Hirukawa H.: Resolved momentum control: humanoid motion planning based on the linear and angular momentum. In *Intelligent Robots and Systems* (2003), pp. 1644–1650. 3

[KLK04] Komura T., Leung H., Kuffner J.: Animating reactive motions for biped locomotion. In *VRST '04: Proceedings of the ACM symposium on Virtual reality software and technology* (2004), pp. 32–40. 3

[LP02] Liu C. K., Popović Z.: Synthesis of complex dynamic character motion from simple animations. *ACM Trans. on Graphics (SIGGRAPH) 21*, 3 (July 2002), 408–416. 3

[Mic] Microsoft: *Microsoft XBOX, http://www.xbox.com:80/en-US/kinect*. 2

[MZS09] Macchietto A., Zordan V., Shelton C. R.: Momentum control for balance. *ACM Trans. on Graphics (SIGGRAPH) 28*, 3 (2009), 1–8. 3

[Ope] OpenKinect: *Open Kinect, http://openkinect.org*. 2

[Pha] PhaseSpaceTechnology: *PhaseSpace Technology, http://www.phasespace.com/*. 2

[Pri] PrimesenseTechnology: *Primesense Technology, http://www.primesense.com/*. 2

[Ras] Rasco B.: *Where´s the Wiimote? Using Kalman filtering to extract accelerometer data. http://www.gamasutra.com/*. 4

[RSC01] Rose C. F., Sloan P.-P. J., Cohen M. F.: Artist-directed inverse-kinematics using radial basis function interpolation. *Computer Graphics Forum 20*, 3 (2001), 239–250. 3

[SH08a] Shiratori T., Hodgins J. K.: Accelerometer-based user interfaces for the control of a physically simulated character. *ACM Trans. on Graphics 27*, 5 (2008). 2

[SH08b] Slyper R., Hodgins J. K.: Action capture with accelerometers. In *ACM SIGGRAPH/Eurographics symposium on Computer animation* (2008), pp. 194–199. 2

[VAV*07] Vlasic D., Adelsberger R., Vannucci G., Barnwell J., Gross M., Matusik W., Popović J.: Practical motion capture in everyday surroundings. *ACM Trans. Graph. 26*, 3 (2007), 35. 2

[Vic] ViconTechnology: *Vicon Technology, http://www.vicon.com/*. 2

[Wel93] Welman C.: Inverse kinematics and geometric constraints for articulated figure manipulation. 3

[WH97] Wiley D. J., Hahn J. K.: Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications 17*, 6 (1997), 39–45. 3

[Wii] WiiBrew: *WiiBrew, http://wiibrew.org/wiki/Main_Page/*. 2

[XSea] XSenseTechnology: *XSense Technology, http://www.xsense.com*. 2

[XSeb] XSensorTechnology: *XSensor Technology Corporation, http://www.xsensor.com/*. 2

[YN03] Yamane K., Nakamura Y.: Natural motion animation through constraining and deconstraining at will. *IEEE Transactions on Visualization and Computer Graphics 9*, 3 (July 2003), 352–360. 3

[YP03] Yin K., Pai D. K.: Footsee: an interactive animation system. In *ACM SIGGRAPH/Eurographics symposium on Computer animation* (2003), pp. 329–338. 2

[ZB98] Zhao L., Badler N.: Gesticulation behaviors for virtual humans. In *Pacific Graphics* (1998), pp. 161–168. 3