Multicontact Interaction Force Sensing From Whole-Body Motion Capture

Tu-Hoa Pham [ID], Stéphane Caron, and Abderrahmane Kheddar [ID], *Senior Member, IEEE*

Abstract—We present a novel technique that unobtrusively estimates forces exerted by human participants in multicontact interaction with rigid environments. Our method uses motion capture only, thus circumventing the need to set up cumbersome force transducers at all potential contacts between the human body and the environment. This problem is particularly challenging, as the knowledge of a given motion only characterizes the resultant force, which can generally be caused by an infinity of force distributions over individual contacts. We collect and release a large-scale dataset on how humans instinctively regulate interaction forces on diverse multicontact tasks and motions. The force estimation framework we propose leverages physics-based optimization and neural networks to reconstruct force distributions that are physically realistic and compatible with real interaction force patterns. We show the effectiveness of our approach on various locomotion and multicontact scenarios.

Index Terms—Force sensing from motion capture, multicontact, neural networks, physics-based optimization, whole body.

I. INTRODUCTION

HUMAN motions result from skilled control of the physical interactions with the environment through contacts. Thus, haptic perception is a fundamental theme toward action understanding and control. The monitoring of contact forces is already widely used in various fields such as robot learning from demonstration and control [1], [2], physics-based animation [3], [4], and visual tracking [5], [6]. Measurement of contact forces is usually achieved by mounting force transducers at prefixed contact locations, making it a costly, cumbersome, and intrusive process that is difficult to use in daily settings. Mounting force transducers on the persons obstructs their natural motion and is not sustainable for daily use. In contrast, the accurate monitoring of interaction forces from motion capture alone, which can readily be achieved using consumer-grade cameras [7], [8], would enable a wide range of applications in personal robotics, human–computer interaction, and rehabilitation [9] as a new unobtrusive biosensor for the healthcare Internet of Things [10].

However, this problem is very difficult due to the indeterminacy of force distributions in multicontact. Indeed, while the knowledge of external and internal forces uniquely determines the resulting kinematics for a given articulated system, even a perfectly known motion does not suffice to fully characterize the underlying forces in multicontact. Instead, the resultant force can be distributed in infinitely many different ways on a given set of contacts. Illustrating this indeterminacy, consider a human participant standing still with both feet on the ground. Even in the elementary case of a static biped stance, the participant can exert tangential forces that compensate each other out, e.g., by pushing their feet apart. While substantial work was dedicated to the problem of force indeterminacy during gait, general contact configurations (e.g., involving hands) have been comparatively less studied in the literature (see Section II).

We address the force distribution problem in multicontact by combining the benefits of machine learning techniques and physics-based optimization to capture the variability in the way humans naturally regulate interaction forces while ensuring their physical compatibility with the observed motion.

1) We formulate an optimization problem allowing the estimation of physically valid forces either from motion observations alone or from a reference signal (see Section III).
2) We construct a novel dataset on human whole-body kinodynamics containing 2.4 h of synchronized force and motion measurements under diverse configurations of tasks, participants, and contacts (see Section IV).
3) We propose two neural network architectures allowing the prediction of contact force distributions from motion observations as well as their interactive correction by physics-based optimization (see Section V).
4) We validate our approach with ground-truth force measurements on various multicontact scenarios and assess

T.-H. Pham was with the Interactive Digital Humans Group, Montpellier Laboratory of Informatics, Robotics, and Microelectronics, UMR5506, CNRS–University of Montpellier, 34090 Montpellier, France, and with the CNRS–AIST Joint Robotics Laboratory, UMI3218/RL, Tsukuba 305-8568, Japan. He is now with IBM Research, Tokyo 103-8510, Japan (e-mail: pham.main@gmail.com).

S. Caron is with the Interactive Digital Humans Group, Montpellier Laboratory of Informatics, Robotics, and Microelectronics, UMR5506, CNRS–University of Montpellier, 34090 Montpellier, France (e-mail: stephane.caron@lirmm.fr).

A. Kheddar is with the Interactive Digital Humans Group, Montpellier Laboratory of Informatics, Robotics, and Microelectronics, UMR5506, CNRS–University of Montpellier, 34090 Montpellier, France, and also with the CNRS–AIST Joint Robotics Laboratory, UMI3218/RL, Tsukuba 305-8568, Japan (e-mail: kheddar@lirmm.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2017.2760912

the respective contributions of physics-based optimization and neural networks (see Section VI).

Finally, we discuss the limitations, applications, and future extensions of our work (see Section VII). Besides a significantly extended dataset, our current work enhances the earlier approach of [11] with: an improved formulation of the optimization problem accounting for motion measurement uncertainties, the consideration of individual contact normals in the learning features enabling more fine-grained predictions by neural network models, as well as algorithmic descriptions and extensive validation experiments that have not been presented before. To foster the research on this new topic and encourage alternative implementations, we make the whole-body kinodynamics dataset and algorithms publicly available.[1]

II. RELATED WORK

Research on human–computer interaction has resulted in multiple techniques for whole-body motion capture from markerless visual observations [7], [8], magnetic trackers [12], or wearable inertial sensors [13], [14]. Force sensors were notably used in conjunction with inertial sensors and vision to improve the motion reconstruction in [3] and [4]. Instead of physical force sensors, numerical models were used to compute *physically plausible* distributions supporting visual observations in hand-object tracking [5], [6]. The problem of estimating the *real forces* applied on the environment was tackled in the case of deformable objects [15] and conversely by considering the human body elastic [16]. In the inspiring work of [17], ground reaction forces were computed with a spring-based contact model to estimate internal joint torques during locomotion. General contact configurations are commonly addressed in simulation and robotics using constrained optimization [18], which alone may not result in the forces humans instinctively apply, as illustrated in Section III-C.

Inverse optimization approaches in kinesiology research address the force distribution indeterminacy by modeling the objective function(s) supposedly optimized by the central nervous system [19]. However, such approaches are difficult due to the redundancy of the human body and the difficulty to observe physiological parameters without invasive surgery [20]. The variability of inverse dynamics solutions with different body segment inertial parameter (BSIP) models was notably discussed in [21] and [22]. Toward this issue, Jovic *et al.* [23] introduced an optimization framework for the online estimation of robot and human BSIPs from motion and force–torque measurements. An alternative approach for BSIP reconstruction was proposed in [24], along with a data-driven approach to estimate contact forces from motion tracking between the feet and the ground.

Recent successes for the control of robot arms [25], [26] or general articulated characters [27] using neural networks illustrated their ability to account for complex model uncertainties. Neural networks were also used to resolve force indeterminacy cases during gait [28] and manipulation [29]. To account

for temporal continuity, recurrent neural networks (RNNs) [30] with long short-term memory (LSTM) [31] neurons were used in [32] and [33], still for manipulation. Whole-body interactions were first addressed using an RNN in combination with a second-order cone program (SOCP) [34] for physics-based optimization in [11]. Our current study generalizes this idea to more complex multicontact scenarios, supported by an extended dataset that is significantly more diverse in terms of contact configurations, tasks, and participants.

III. WHOLE-BODY CONTACT FORCE OPTIMIZATION

A. Equations of Motion and Friction Constraints

We consider an articulated system of rigid bodies subject to $N_{\boldsymbol{\tau}}$ internal joint torques $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^{(i)}, \dots, \boldsymbol{\tau}_{N_{\boldsymbol{\tau}}}^{(i)})^T$ and $N_{\mathbf{F}}$ external wrenches $\mathbf{F}_k = (\boldsymbol{\tau}_k, \mathbf{f}_k)^T$, with $\boldsymbol{\tau}_k$ and $\mathbf{f}_k$ being the respective external torque and force at contact k, expressed in the global frame. With the position and orientation of a chosen base link, the number of degrees of freedom is $N_{\mathrm{DoF}} = N_{\boldsymbol{\tau}} + 6$. We denote by $\mathbf{q}$, $\dot{\mathbf{q}}$, and $\ddot{\mathbf{q}}$ the respective generalized coordinates, velocity, and acceleration of the articulated system, respectively. The whole-body equations of motion can be expressed as

$$\mathbf{H}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q},\dot{\mathbf{q}}) = \begin{bmatrix} \mathbf{0}_6 \\ \boldsymbol{\tau} \end{bmatrix} + \sum_{k=1}^{N_{\mathbf{F}}} \mathbf{J}_k^T \mathbf{F}_k \qquad (1)$$

where

1) $\mathbf{H}(\mathbf{q})$ the $N_{\mathrm{DoF}} \times N_{\mathrm{DoF}}$ mass matrix;
2) $\mathbf{C}(\mathbf{q},\dot{\mathbf{q}})$ the $N_{\mathrm{DoF}} \times 1$ bias vector of the Coriolis, centrifugal forces and gravity terms;
3) $\mathbf{J}_k$ the $N_{\mathrm{DoF}} \times 6$ the kth contact Jacobian matrix;
4) $\mathbf{0}_6$ the 6×1 internal wrench directly applied at the root of the kinematic tree in case of linkage with the environment (zero for the case of the floating base).

We assume the parameters of the dynamic model to be known [23], [24]. For each contact k, we denote by $\mathbf{z}_k$ the (uniquely defined) normal vector oriented from the environment to the body, and by $\mathbf{x}_k$ and $\mathbf{y}_k$ two orthogonal vectors in the tangential plane. We thus obtain a local decomposition for each external wrench $\mathbf{F}_k$ in the contact frame $\mathcal{C}_k = (\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k)$:

$$^{\mathcal{C}_k}\mathbf{F}_k = (\boldsymbol{\tau}_k^{\mathbf{x}}, \boldsymbol{\tau}_k^{\mathbf{y}}, \boldsymbol{\tau}_k^{\mathbf{z}}, f_k^{\mathbf{x}}, f_k^{\mathbf{y}}, f_k^{\mathbf{z}})^T,$$

$$\text{with} \quad \begin{cases} \boldsymbol{\tau}_k = \boldsymbol{\tau}_k^{\mathbf{x}}\mathbf{x}_k + \boldsymbol{\tau}_k^{\mathbf{y}}\mathbf{y}_k + \boldsymbol{\tau}_k^{\mathbf{z}}\mathbf{z}_k \\ \mathbf{f}_k = f_k^{\mathbf{x}}\mathbf{x}_k + f_k^{\mathbf{y}}\mathbf{y}_k + f_k^{\mathbf{z}}\mathbf{z}_k. \end{cases} \qquad (2)$$

Having chosen $\mathbf{z}_k$ oriented toward the body, each normal force component is such that

$$f_k^{\mathbf{z}} \geq 0. \qquad (3)$$

With μ_k the friction coefficient at contact k, the tangential force is constrained by the normal component as follows:

$$\|f_k^{\mathbf{x}}\mathbf{x}_k + f_k^{\mathbf{y}}\mathbf{y}_k\|_2 \leq \mu_k f_k^{\mathbf{z}}. \qquad (4)$$

Contact torque constraints are usually obtained by discretizing the contact surface into individual contact points subject to 3-D forces only. Closed-form formulas were derived for rectangular support areas in [35]. We observed in our experiments that such

---

[1] https://github.com/jrl-umi3218/WholeBodyKinodynamics.

constraints could be violated due to motion tracking uncertainties and omitted them in this study.

B. Physics-Based Optimization

In this section, we discuss the extraction of physically plausible force distributions compatible with a given motion, characterized by generalized coordinates $\mathbf{q}$, velocities $\dot{\mathbf{q}}$, and accelerations $\ddot{\mathbf{q}}$. Such force distributions can be obtained as solutions of an SOCP of the form

$$\min \quad \mathcal{C}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{r}^T\mathbf{x}$$

$$\text{s.t.} \begin{cases} \|\mathbf{A}_j\mathbf{x} + \mathbf{b}_j\|_2 \le \mathbf{c}_j^T\mathbf{x} + \mathbf{d}_j, \quad j = 1,\ldots,m \\ \mathbf{E}\mathbf{x} \le \mathbf{f} \\ \mathbf{G}\mathbf{x} = \mathbf{h} \end{cases} \tag{5}$$

with $\mathbf{x}$ a vector of $N_{\mathbf{x}} = 6 + N_{\boldsymbol{\tau}} + 6N_{\mathbf{F}}$ force variables

$$\mathbf{x} = \left(\mathbf{F}^{\text{MEW}}, \boldsymbol{\tau}, \left({}^{\mathcal{C}_k}\mathbf{F}_k\right)_{k=1,N_{\mathbf{F}}}\right)^T. \tag{6}$$

Here, $\mathbf{F}^{\text{MEW}}$ represents a measurement error wrench (MEW) applied to the floating base of the kinematic tree. This wrench is $\mathbf{0}_6$ in the ideal case of perfect measurements and dynamic model. However, trying to enforce the strict constraint $\mathbf{F}^{\text{MEW}} = \mathbf{0}_6$ on noisy measurements and with an approximative dynamic model results in unfeasible SOCP problems. To allow for uncertainties, we relax this constraint and rather make the solver enforce it at best (i.e., minimizing $\|\mathbf{F}^{\text{MEW}}\|$), as detailed thereafter.

1) Inequality Constraints: In (5), linear inequality matrices $\mathbf{E}, \mathbf{f}$ and cone inequality matrices $\mathbf{A}, \mathbf{b}_j, \mathbf{c}_j, \mathbf{d}_j$ can directly be computed from (3) and (4), respectively.

2) Equality Constraints: We consider the whole-body equations of motion. Given an instance of $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}})$, the term $\mathbf{h}$ in (5) corresponds directly to the left-hand side of (1):

$$\mathbf{h} = \mathbf{H}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}). \tag{7}$$

$\mathbf{h}$ is a vector of N_{DoF} elements. The matrix $\mathbf{G}$ in (5) is here of size $N_{\text{DoF}} \times N_{\mathbf{x}}$ and can be decomposed using selection matrices $\mathbf{G}_{\boldsymbol{\tau}}$ and $(\mathbf{G}_{\mathbf{F}_k})_{k=1,N_{\mathbf{F}}}$ such that

$$\mathbf{G}_{\boldsymbol{\tau}}\mathbf{x} = \begin{bmatrix} \mathbf{F}^{\text{MEW}} \\ \boldsymbol{\tau} \end{bmatrix} \quad \text{and} \quad \mathbf{G}_{\mathbf{F}_k}\mathbf{x} = \mathbf{F}_k. \tag{8}$$

Note that each $\mathbf{G}_{\mathbf{F}_k}$ must incorporate the rotation matrix between the contact frame $\mathcal{C}_k$ and the world frame. We obtain

$$\mathbf{G} = \mathbf{G}_{\boldsymbol{\tau}} + \sum_{k=1}^{N_{\mathbf{F}}} \mathbf{J}_k^T\mathbf{G}_{\mathbf{F}_k}. \tag{9}$$

3) Cost Function: Having incorporated the previous constraints in the SOCP, physically plausible force distributions can be computed by minimizing a chosen cost function depending only on the optimization variables, e.g., a weighted sum of the squared L^2 norms of the optimization variables:

$$\mathcal{C}_{\alpha,\beta,\gamma}(\mathbf{x}) = \alpha\|\mathbf{F}^{\text{MEW}}\|^2 + \beta\|\boldsymbol{\tau}\|^2 + \gamma\sum_{k=1}^{N_{\mathbf{F}}}\|\mathbf{F}_k\|^2. \tag{10}$$
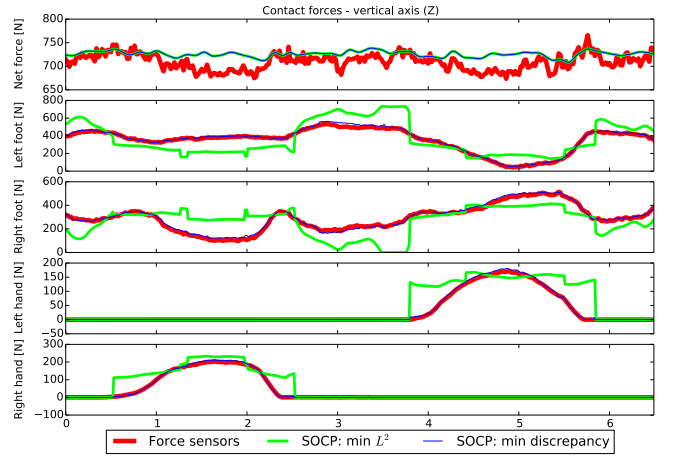

Fig. 1. Force sensor noise and uncertainties in the resultant force (top plot) can be corrected using physics-based optimization. In multicontact, directly minimizing the norm of the individual forces (green) results in forces that are physically plausible but significantly differ from real measurements (red). By minimizing the discrepancy to the latter (blue), we reconstruct forces that are both physically plausible and in agreement with natural force distributions.

In practice, it is preferable to set α greater than β and γ so that $\mathbf{F}^{\text{MEW}}$ is only used when the observed motion is otherwise unfeasible. The two other parameters β and γ can be tuned to minimize either internal joint torques or applied contact wrenches. Alternatively, when target values $\widetilde{\mathbf{F}}_k$ for the contact wrenches are available (e.g., from force–torque sensors), it is possible to extract force distributions in their vicinity that are also guaranteed to be physically plausible, by minimizing the discrepancy to the optimized wrenches in the SOCP cost function [36]:

$$\mathcal{C}_{\alpha,\beta,\gamma}^{\widetilde{\mathbf{F}}_k}(\mathbf{x}) = \alpha\|\mathbf{F}^{\text{MEW}}\|^2 + \beta\|\boldsymbol{\tau}\|^2 + \gamma\sum_{k=1}^{N_{\mathbf{F}}}\left\|\mathbf{F}_k - \widetilde{\mathbf{F}}_k\right\|^2. \tag{11}$$

Contact forces and internal joint torques occurring during gait are typically in the order of 100 N and 1 N·m, respectively [21]. In our experiments, we chose $\alpha = 10^2$, $\beta = 10^{-2}$, and $\gamma = 1$ so that $\mathbf{F}^{\text{MEW}}$ only compensates unfeasible raw motion measurements, and internal joint torques can vary as needed to prioritize matching optimized and target contact wrenches.

C. Motivating Example: Triple Contact Indeterminacy

We illustrate the crucial role played by the SOCP cost function. We consider a participant standing still next to a table and taking support on it using the right hand, then the left. We represent the vertical component of the measured forces in Fig. 1. In addition, we compute force distributions of minimal L^2 norm using (10) and minimizing the discrepancy to the sensor measurements using (11).

With the participant standing still, the equations of motion dictate that the net contact forces (top plot) should mostly oppose the participant's weight. However, individual force sensor uncertainties result in rather noisy force estimates. In contrast, all SOCP variants accurately reconstruct the net force directly from the measured kinematics. Using the cost function
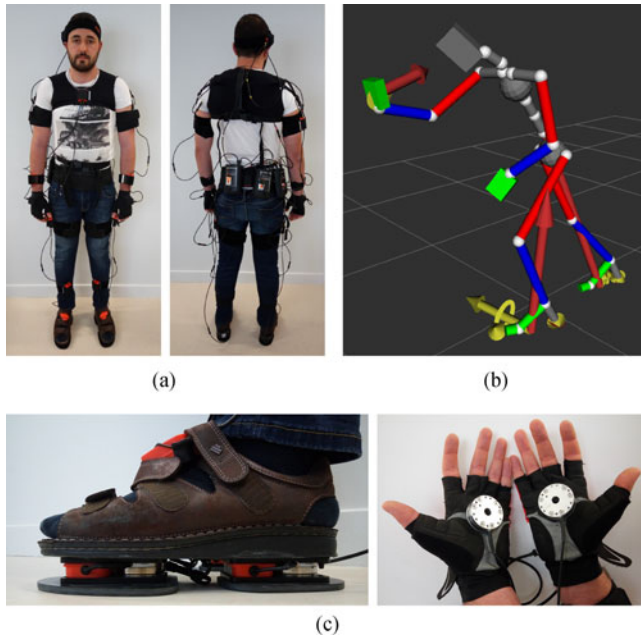
Fig. 2. Acquisition system for whole-body kinematics and contact forces. (a) Inertial motion capture system. (b) Captured motion and forces. (c) Shoes and gloves instrumented with force–torque sensors.

of (10) results in forces that are physically plausible but may greatly differ from actual measurements. Using the cost function of (11) enables the reconstruction of force distributions that are both physically plausible and in the vicinity of target forces when available. The aim of our work is to circumvent the need for force sensors. Thus, in the following, we train RNNs to predict such target force distributions directly from motion observations.

IV. WHOLE-BODY KINODYNAMICS DATASET

A. Experimental Setup

We depict our complete acquisition system in Fig. 2.

1) Whole-Body Motion: We track the whole-body motion using the Xsens MVN Awinda inertial motion capture system [13], comprised of 17 inertial measurement units (IMU) worn and strapped at specified body landmarks on the participant's body. The motion capture system is battery powered and wireless, transmitting accelerometer, gyroscope, and magnetometer measurements to the computer at 100 Hz. The motion of the human body, modeled as a 23-segment skeleton, is then readily provided in the form of the 6-DoF pose, velocity, and acceleration of each segment. We enable the dynamics analysis of Section III by converting these quantities into generalized coordinates, velocities, and accelerations $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}})$, with a kinematic tree composed of 23 segments linked by 22 spherical joints and rooted at the participant's pelvis. Measuring the participant's body measurements and weight, we compute the BSIPs using the anthropomorphic tables of [37]. Inertial motion capture systems by themselves do not provide absolute positioning and are prone to drift compared to marker-based tracking methods (e.g., Vicon). We are working toward attenuating this problem. Still,

the choice of this motion capture system was motivated by strong occlusions that are inherent to whole-body interactions with the environment and hinder vision-based motion capture systems (e.g., Vicon). In contrast, the inertial motion capture system allows us to explore various interaction scenarios in uncontrolled and cluttered environments, e.g., when crouching under a table. In the future, the system could even be employed outdoors or used in combination with a limited number of visual sensors to solve the issue of drift and absolute positioning.

2) Contact Forces: We measure the contact forces exerted by the participant onto the environment both at the feet and at the hands. Contact forces at the feet are monitored using instrumented shoes (Xsens ForceShoe). Each shoe is equipped with two force–torque sensors and two IMUs, providing contact forces measured individually at the heel and toes, and transmitted to the computer via Bluetooth at 50 Hz. We monitor contact forces exerted at the hands using two additional force–torque sensors (ATI Mini-45) attached to gloves worn by the participant during interaction experiments. The force–torque sensors are wired to dedicated acquisition cards on the computer, and measurements are also recorded at 50 Hz. Both ForceShoe and ATI sensor signals are linearly interpolated to 100 Hz, matching the motion capture sampling rate. In comparison to static force plates commonly used in gait analysis, wearable force sensors can be less accurate. Still, a major advantage of our lightweight setup is that it enables the efficient and continuous acquisition of contact forces on arbitrary contact configurations, highly dynamic motions, and relatively unrestricted movement areas. In contrast, using static force plates considerably reduce the range of possible tasks, contacts, and motions.

B. Newton–Euler Equations and Signal Synchronization

Each type of sensor used in this work (i.e., motion capture suit, force-sensing shoes, and ATI Mini-45 sensors) is individually monitored using a dedicated acquisition program. Therefore, raw measurements need to be temporally synchronized with each other before further analysis. This step is performed using the Newton–Euler equations taken at the center of mass $\mathbf{G}$ of the whole-body articulated system. For each body segment s of the 23-element set $\mathcal{S}$, we denote by m_s its mass and $\mathbf{G}_s$ its center of mass. In the global frame, we denote by $\mathbf{v}_s$ the linear velocity of $\mathbf{G}_s$ and $\mathbf{R}_s$ its orientation matrix. In the segment frame, we denote by $\boldsymbol{\omega}_s$ and $\mathbf{I}_s$ its local angular velocity and inertia tensor, respectively. With m the total mass of the articulated system and $\mathbf{G}$ its centroid, the linear momentum $\mathcal{P}$ and angular momentum $\mathcal{L}_{\mathbf{G}}$ at $\mathbf{G}$ are defined by

$$\begin{cases} \mathcal{P} = \sum_{s \in \mathcal{S}} m_s \mathbf{v}_s \\ \mathcal{L}_{\mathbf{G}} = \sum_{s \in \mathcal{S}} m_s \overrightarrow{\mathbf{G}\mathbf{G}_s} \times \mathbf{v}_s + \mathbf{R}_s \mathbf{I}_s \boldsymbol{\omega}_s. \end{cases} \tag{12}$$

With $\dot{\mathcal{L}}_{\mathbf{G}}$ and $\dot{\mathcal{P}}$ the time derivatives of the angular and linear momenta, respectively, $\mathbf{g}$ the gravity vector, and $^{\mathbf{G}}\mathbf{F}_k$ the contact wrench at contact k transformed to $\mathbf{G}$, the Newton–Euler

equations for centroidal dynamics state that

$$\begin{bmatrix} \dot{\mathcal{L}}_{\mathrm{G}} \\ \dot{\mathcal{P}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ m\mathbf{g} \end{bmatrix} + \sum_{k=1}^{N_{\mathrm{F}}} {}^{\mathrm{G}}\mathbf{F}_k. \tag{13}$$

We gather gravity, linear, and angular momenta as a centroidal wrench $\mathbf{w}^{\mathrm{G}}$ due to contact forces, taken at $\mathbf{G}$ [38]:

$$\mathbf{w}^{\mathrm{G}} = \begin{bmatrix} \dot{\mathcal{L}}_{\mathrm{G}} \\ \dot{\mathcal{P}} - m\mathbf{g} \end{bmatrix}. \tag{14}$$

With $\mathbf{P}_k$ the location of contact k, (13) becomes

$$\mathbf{w}^{\mathrm{G}} = \sum_{k=1}^{N_{\mathrm{F}}} \begin{bmatrix} \boldsymbol{\tau}_k + \overrightarrow{\mathbf{GP}_k} \times \mathbf{f}_k \\ \mathbf{f}_k \end{bmatrix}. \tag{15}$$

$\mathbf{w}^{\mathrm{G}}$ is a purely kinematic term that can be directly computed from the whole-body pose and its derivatives using (14), but also from the contact forces using (15). Thus, synchronizing motion capture and force measurements amounts to synchronizing $\mathbf{w}^{\mathrm{G}}$ estimates from kinematics and forces. For this purpose, we start each experiment by having the participant walk a few seconds and, then, take support on a table with the left and right hands, alternatively. To synchronize kinematic and ForceShoe signals, we plot the components of their respective estimates for $\mathbf{w}^{\mathrm{G}}_{\mathrm{kin}}$ and $\mathbf{w}^{\mathrm{G}}_{\mathrm{shoe}}$ during the walking phase and select by hand a constant time shift to match the two signals at best. We then compute the residual wrench $\mathbf{w}^{\mathrm{G}}_{\mathrm{res}} = \mathbf{w}^{\mathrm{G}}_{\mathrm{kin}} - \mathbf{w}^{\mathrm{G}}_{\mathrm{shoe}}$. When the participants leans on a table with one hand, $\mathbf{w}^{\mathrm{G}}_{\mathrm{res}}$ should be equal to the wrench $\mathbf{w}^{\mathrm{G}}_{\mathrm{hand}}$ measured by the corresponding force–torque sensor. Again, we find a constant time shift to match $\mathbf{w}^{\mathrm{G}}_{\mathrm{hand}}$ and $\mathbf{w}^{\mathrm{G}}_{\mathrm{res}}$ at best, thus synchronizing hand sensors with kinematic-ForceShoe signals.

Following the temporal synchronization, we perform the following signal processing. All measurements are subject to noise, e.g., from the sensors themselves or due to interferences in the transmission (both wired and wireless). We attenuate it by smoothing all signals with a Gaussian filter of kernel $\sigma = 0.05\,\mathrm{s}$. In addition, a slow-varying bias can appear in the force–torque measurements with repeated stress and battery drain. We estimate this bias through time by averaging the signals that persist when the sensors are not in contact with the environment, which should only be caused by the inertia of the moving parts attached to the sensing surface (e.g., force shoe external sole). Since the inertial motion capture system does not provide absolute positioning, we could not reliably identify the occurrence of contacts with the environment based solely on the whole-body motion observations. Therefore, we identified them by direct thresholding on the force sensor measurements. Still, this material limitation does not affect the generality of our approach and can be fully circumvented with additional visual observations (see also [39] for the retrieval of contact points without environment knowledge). Finally, we correct the remaining force sensor uncertainties by combining their measurements with the motion capture data using the SOCP approach illustrated in Section III-C. In the following, we call *ground truth* the SOCP-corrected sensor measurements (relative to the dynamic model).

C. Experiments
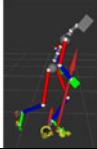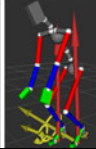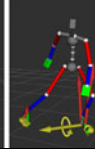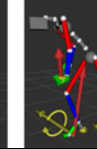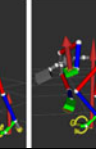
The importance of collecting ground-truth measurements on how humans naturally distribute contact forces not only during locomotion but across a variety of multicontact configurations was established in [11]. Thus, in this work, we purposefully explore a wide range of motion dynamics as well as diverse contact configurations that exhibit strong force distribution indeterminacy. The following tasks were chosen from daily activities to cover a spectrum of three features: number of contacts involved, orientation of hand contacts (when applicable), and effort required to perform the motion:

1) Walking, i.e., with always at least one foot on the ground (one contact, low effort). Straight and curved paths were considered separately;
2) Running, i.e. with at most one foot on the ground (one contact, medium effort);
3) Hopping on one foot, e.g., forward or in place (one contact, high effort);
4) Balancing the upper body while keeping both feet static, e.g., leg stretching or performing arm motions (two contacts, low effort);
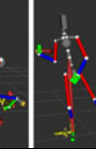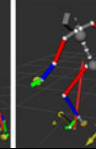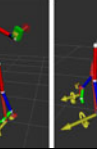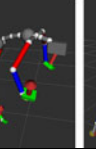5) Jumping using both feet, e.g., forward or to the side (two contacts, high effort);
6) Taking support on a table with one hand, e.g., to reach for an object further on the table (three contacts, horizontal hand contact, low effort);
7) Crouch and stand by taking support with one hand on a table, e.g., to reach for an object under the table (three contacts, horizontal hand contact, high effort);
8) Leaning against a wall with one hand (three contacts, vertical hand contact, low effort);
9) Leaning on a wall with one hand and reach forward, e.g., to look around a corner or grab an object (three contacts, vertical hand contact, high effort);
10) Taking support on a table with both hands (four contacts, horizontal hand contacts, low to high effort);
11) Leaning on a wall with both hands, e.g., to stretch or push a heavy object (four contacts, vertical hand contacts, low to high effort).

Contact is a complementarity condition involving the dual geometric and force spaces. The first two features we used to categorize our tasks (namely number and orientations of contacts) ensure coverage of the geometric part of the condition, while the last one (perceived effort) aims for coverage of the force space. Considering the two variants (straight and curved) of walking experiments separately, we thus construct a repertoire of 12 motion types, six of them involving contacts between the feet and the ground only and the six others involving both feet and hands. We illustrate this dataset in Table I.

Six volunteers, three males and three females, took part in our study. Their weights (between 45.0 kg and 86.0 kg, plus the 5.0 kg acquisition system), heights (between 1.57 m and 1.92 m), and individual body segment lengths were measured to initialize the motion capture skeletal tracking model and BSIPs following the procedure described in Section IV-A. Before each experiment, all sensors (i.e., inertial motion capture system,

TABLE I
INTERACTION CONFIGURATIONS FROM THE WHOLE-BODY KINODYNAMICS DATASET

	Feet only					Feet and hands					
Motion	Hop	Run	Jump	Walk (×2)	Balancing	Take support	Crouch	Lean	Lean and reach	Take support	Lean
Contacts	0/1F	0/1F	0/2F	1/2F	2F	1/2F-1H	1/2F-1H	1/2F-1H	1/2F-1H	1/2F-2H	1/2F-2H
Human action											
Motion-force data											

force-sensing shoes, and glove-mounted force sensors) were calibrated and reset following the manufacturers' recommended acquisition procedure to reduce the effects of measurement drift and hysteresis. We divided the 12 motion types of Table I into two sequences of six motions. Each sequence consisted in three tasks involving the hands and three tasks involving only the feet, executed in alternation for one minute each. Participants were given time between consecutive tasks to put on, or take off instrumented gloves, so locomotion tasks were not constrained by unnecessary force sensor wires. In total, each task was executed twice by each participant. For one participant, we observed force measurement errors of abnormal magnitude on the right-hand sensor and discarded the corresponding recordings from the dataset. For another participant, the motion capture system was disconnected during a hopping task. Overall, our new dataset on human whole-body kinodynamics in multicontact totals 2.4 h of synchronized motion and force measurements, classified into 12 task primitives.

V. CAPTURING HUMAN FORCE DISTRIBUTION PATTERNS

A. Learning Features

Let $\mathbf{K}$ denote a set of (input) whole-body kinematic features, and $\mathbf{D}$ a set of (output) contact force features. The desired contact force estimation mapping $\mathcal{F}$ is of the form

$$\mathbf{D} = \mathcal{F}(\mathbf{K}). \tag{16}$$

We model this mapping $\mathcal{F}$ using a neural network trained on our whole-body kinodynamics dataset. The dynamic features $\mathbf{D}$ simply correspond to the set of contact wrenches $\mathbf{F}_k$ we seek to estimate. A straightforward approach to construct the set of kinematic features $\mathbf{K}$ could be to take all the remaining parameters appearing in the whole-body equations of motion of (1), e.g., the mass $\mathbf{H}$ and bias $\mathbf{C}$ matrices, joint accelerations $\ddot{\mathbf{q}}$, and Jacobian matrices $\mathbf{J}_k$ representing the contact configuration. However, doing so would result in a particularly large number of parameters that can make the neural network training process difficult. We instead propose to construct a selection of high-level kinematic features based on the Newton–Euler equations of (13), which extract the gist of locomotory dynamics. In particular, from the formulation of (15), we take as first input features

the centroidal wrench $\mathbf{w}^{\mathbf{G}}$, which can be computed from kinematics only with (14), and the contact positions relative to the center of mass, $\overrightarrow{\mathbf{GP}_k}$. Since these quantities are expressed in the world frame, we account for translational and rotational invariances by transforming them to a reference frame $\mathcal{G}$ of origin $\mathbf{G}$ and fixed with respect to a chosen body segment (e.g., the pelvis). Walking straight to the North is thus locally equivalent to walking straight to the East. To facilitate the modeling of the mapping of (16) with a neural network, we construct $\mathbf{K}$ as a fixed-size input vector. We continuously monitor N_c potential contacting body segments over time and encode their activity with parameters $\delta_{k,i}$ such that

$$\delta_{k,i} = \begin{cases} 1, & \text{if contact } k \text{ is active at time step } i \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

In our experiments, we considered the forces applied at the heels and toes separately in both the SOCP and the neural network model, so that $N_c = 6$ including the hand palms. Finally, in addition to the contact locations, we consider their orientation through the contact normals $\mathbf{z}_k$. Denoting by $^{\mathcal{G}}\mathbf{w}^{\mathbf{G}}$, $^{\mathcal{G}}\mathbf{P}_k$, and $^{\mathcal{G}}\mathbf{z}_k$ the respective coordinates of $\mathbf{w}^{\mathbf{G}}$, $\mathbf{P}_k$, and $\mathbf{z}_k$ in the frame $\mathcal{G}$, respectively, the complete input features at time step i are

$$\mathbf{K}_i = \left({}^{\mathcal{G}}\mathbf{w}^{\mathbf{G}}_i, \left({}^{\mathcal{G}}\mathbf{P}_{k,i}, \delta_{k,i}, \mathbf{z}_{k,i}, \right)_{k=1,N_c} \right)^T. \tag{18}$$

Similarly, the output features are the target wrenches in $\mathcal{G}$:

$$\mathbf{D}_i = \left(\left({}^{\mathcal{G}}\mathbf{F}_{k,i} \right)_{k=1,N_c} \right)^T. \tag{19}$$

B. Neural Network Architecture

We model the evolution of motion and force distributions as time series using RNNs with LSTM neurons in order to account for temporal continuity between consecutive samples. In this section, we propose two neural network architectures to be used in conjunction with physics-based optimization. The first architecture, WBN$^{\mathrm{D}}$ (whole-body network, direct), directly maps the observed motion to the underlying forces

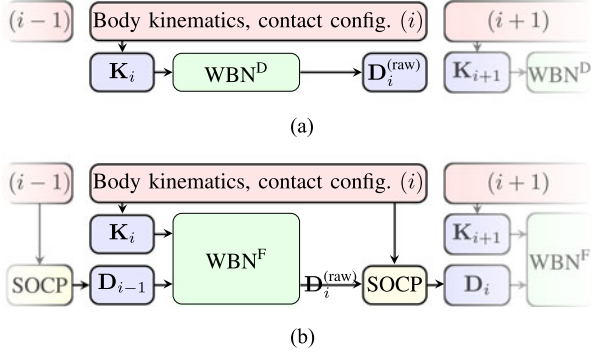$$\mathbf{D}_i = \mathrm{WBN}^{\mathrm{D}}(\mathbf{K}_i). \tag{20}$$

Fig. 3. Direct and feedback whole-body network architectures. (a) Forces are directly computed from the kinematics and contact configuration. (b) Force predictions are corrected between consecutive time steps.

Once trained, this network is used as follows. At each step, a new vector of kinematic features $\mathbf{K}_i$ is fed to WBN$^\mathrm{D}$, yielding raw output features $\mathbf{D}_i^{(\mathrm{raw})}$. Since the RNN model does not enforce the equations of motion, the corresponding forces $\mathbf{F}_{k,i}^{(\mathrm{raw})}$ may not be readily compatible with the observed motion. We compute physically plausible forces $\mathbf{F}_{k,i}$ in their vicinity using the SOCP of (5) with the cost function of (11).

Alternatively, we enable the interactive correction of RNN predictions by constructing a network WBN$^\mathrm{F}$ (whole-body network, feedback) that takes as inputs both the current kinematics and the distribution at the previous time step

$$\mathbf{D}_i = \mathrm{WBN}^\mathrm{F}\left(\mathbf{K}_i, \mathbf{D}_{i-1}\right). \tag{21}$$

When using WBN$^\mathrm{F}$ for prediction, we initialize $\mathbf{D}_0$ to the distribution of minimal L^2 norm following (10). At each time step, $\mathbf{K}_i$ and $\mathbf{D}_{i-1}$ are fed together to WBN$^\mathrm{F}$, yielding raw predictions $\mathbf{D}_i^{(\mathrm{raw})}$. By SOCP correction, we reconstruct physically accurate forces $\mathbf{F}_{k,i}$ and extract the corresponding dynamic features $\mathbf{D}_i$, used for prediction at the next time step.

We depict the two proposed architectures in Fig. 3. Note that for WBN$^\mathrm{D}$, raw predictions $\mathbf{D}_i^{(\mathrm{raw})}$ may be corrected independently from each other, enabling opportunities for parallel computing if desired. In contrast, the intertwined RNN-SOCP approach of WBN$^\mathrm{F}$ imposes a sequential prediction process.

VI. RESULTS

A. Prediction-Correction Framework

We implement the two proposed neural network architectures within the Torch7 framework [40] as two LSTM hidden layers of size 256 followed by a linear output layer of size $6N_c$, the number of output features. We partition the whole-body kinodynamics dataset into three subsets of respective size 70%, 15%, and 15% for training, validation, and testing. We train the neural networks by minimizing a mean square error regression criterion using minibatch stochastic gradient descent and dropout to avoid overfitting [41]. We estimated from the dataset that participants maintained each contact on average for 2.07 s. We thus set the length of the training batches to 2.0 s. The SOCP correction is implemented separately using the CVXOPT library

TABLE II
FORCE ESTIMATION ERRORS ON FULL TESTING SET (23 min)

	Raw	SOCP correction
Force sensors	1.6%	ground truth
SOCP min.L^2	N/A	7.0%
WBN$^\mathrm{D}$	8.3%	6.4%
WBN$^\mathrm{F}$	6.6%	5.8%

for convex optimization [42]. We run the prediction process for each task of the testing set and compute the root-mean-square errors (RMSE) between reconstructed forces and ground truth distributions. We normalize the RMSEs with the range of the normal forces measured in the testing set, $f_{\max}^{\mathbf{z}} = 1378\,\mathrm{N}$. For the sake of completeness, we also quantify the force sensor measurement uncertainties, the estimation errors for distributions computed by straightforward minimization of their L^2 norm, and prediction errors for the neural networks alone, without SOCP correction. We report the resulting normalized RMSE (NRMSE) in Table II.

Expectedly, the lowest estimation errors are attained using physical force sensors, which directly measure the applied contact wrenches. Still, this level of accuracy was obtained using costly cumbersome force sensors. Table II yields three major outcomes. First, we confirm the previous observation that physics-based optimization alone does not suffice to address the issue of force indeterminacy in multicontact, since the L^2-minimizing cost function provides the worst results of the second column. Second, even without SOCP correction (first column), the accuracy of WBN$^\mathrm{F}$ exceeds that of WBN$^\mathrm{D}$, and even that of the L^2-minimizing SOCP alone. Thus, RNNs can successfully capture interaction force patterns even without enforcing the equations of motion during training. In particular, the better performance of WBN$^\mathrm{F}$ compared to WBN$^\mathrm{D}$ shows that providing the RNN with past forces as inputs helps handle force indeterminacy, i.e., associating a given motion (unique $\mathbf{K}_i$) to multiple possible distributions (different $\mathbf{D}_i$). Third, combining RNN and SOCP yields the best results overall, improving the accuracy of WBN$^\mathrm{D}$ and WBN$^\mathrm{F}$ by 23% and 11% respectively, and that of the SOCP alone by 17%.

B. Accuracy in Multicontact Indeterminacy

The effectiveness of the SOCP to correct inaccurate force predictions is particularly visible for the hopping sequence depicted in Fig. 4. Indeed, for this motion, the presence of only one foot on the ground at each instant makes it straightforward for the SOCP to enforce that the force exerted at the only contact is exactly causing the acceleration of the centroid. We further investigate the respective contributions of RNN and SOCP by separating experiments with only feet or with both feet and hands. Since the former involves relatively large impulses (e.g., during jumping), we normalize the estimation errors of each category by the range of their respective measurements. We report the resulting NRMSEs in Table III. For both categories, combining RNN and SOCP yields significant improvements compared
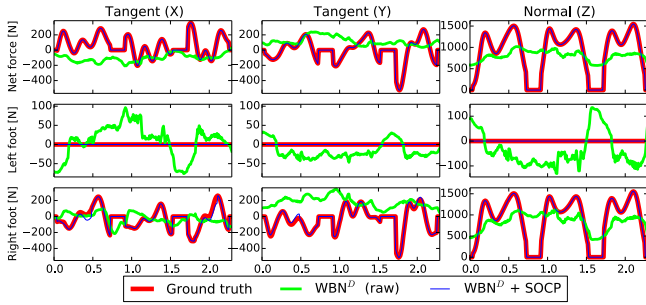
Fig. 4. SOCP correction on single contact motion (hop on the right foot).

TABLE III
ESTIMATION ERRORS BY CONTACT CONFIGURATION

| | Feet only (13 min) | | Feet + hands (10 min) | |
| | $f_{max}^{\mathbf{z}} = 1378\,\text{N}$ | | $f_{max}^{\mathbf{z}} = 750\,\text{N}$ | |
	Raw	SOCP	Raw	SOCP
Force sensors	2.1%	ground truth	1.9%	ground truth
SOCP min.L^2	N/A	8.7%	N/A	9.6%
WBN^D	9.9%	7.6%	12.2%	9.4%
WBN^F	7.5%	6.6%	10.3%	9.3%

TABLE IV
ESTIMATION ERRORS BY SEGMENT ON FEET + HAND TASKS (10 MIN)

| | Feet: $f_{max}^{\mathbf{z}} = 750\,\text{N}$ | | Hands: $f_{max}^{\mathbf{z}} = 177\,\text{N}$ | |
	Raw	SOCP	Raw	SOCP
Force sensors	2.0%	ground truth	5.7%	ground truth
SOCP min.L^2	N/A	10.8%	N/A	21.9%
WBN^D	14.2%	11.0%	14.4%	10.5%
WBN^F	12.0%	10.8%	13.2%	12.9%

to either in isolation. Importantly, the NRMSEs of all three estimation methods are larger when also considering hand contacts, which illustrates the increased multicontact indeterminacy.

Finally, we further decompose the tasks involving feet and hands and assess the estimation accuracy by body segment in Table IV. For all configurations, again, the SOCP greatly improves the accuracy of both neural network architectures. However, while for the feet, the three estimation methods yield comparable NRMSEs, the estimation errors of the SOCP alone on the hands (rightmost column) are now significantly larger than that of WBN variants. This result shows that RNNs are well suited to tackle the issue of force indeterminacy in multicontact, for which physics-based optimization can serve as a valuable complement.

We depict sample force reconstruction results for two-, three-, and four-contact motions in Fig. 5. In all cases, we confirm that the net force is reconstructed accurately by all methods, as expected. In the two-contact balancing scenario, we see that the WBN^D network fails to capture weight shifts between feet (Z component, rightmost column) and tends to predict uniform
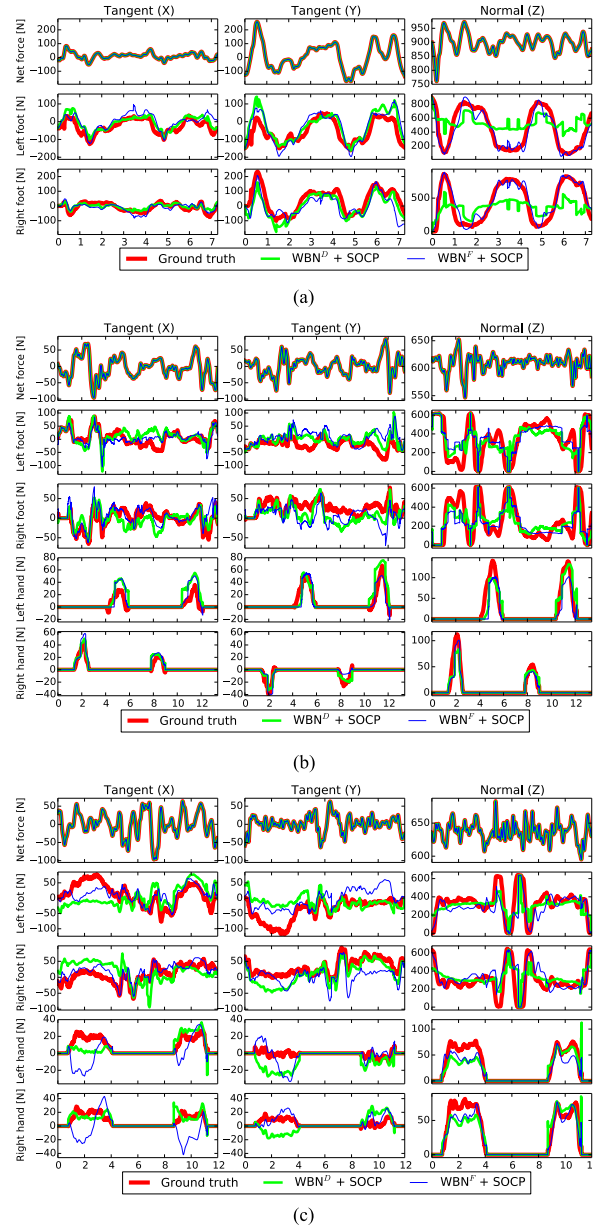

Fig. 5. Force profiles in various contact configurations. Net forces are measured in the world frame, while contact forces are reported in their respective (local) contact frames. (a) Two contacts: upper-body balancing with static feet. (b) Three contacts: taking support on a table with one hand (alternating). (c) Four contacts: leaning against a wall with two hands at the same time.

distributions, while the WBN^F network tracks them suitably, thanks to its ability to capture time-dependent variations. With more contacts, time dependence over pressure distribution becomes less significant and both networks perform reasonably.

VII. DISCUSSION AND FUTURE WORK

Our work establishes that the estimation of interaction forces, a problem that pertains to the human sense of *touch*, could be tackled through the lens of *motion capture*. The dual optimization and learning framework we propose extends the state of the art in capturing human force distribution patterns beyond gait

analysis to general multicontact configurations used to interact with the environment. This important result makes it possible to completely circumvent costly, cumbersome, and intrusive transducing technologies with any whole-body tracking system. Indeed, while we collected our (public) dataset using inertial motion capture, the RNN architectures we propose only rely on centroidal dynamics, making them agnostic with respect to the actual motion capture system employed. Meanwhile, an SOCP can be formulated for any whole-body kinematic model. As such, our framework is readily compatible with existing markerless visual tracking techniques, thus enabling novel interfaces for in-home unobtrusive force monitoring for personal robotics or rehabilitation.

In its current implementation, our work has some limitations. First, we consider the academic point-contact model, whereas, in practice, contacts are between surfaces, yielding additional complementarity conditions [35] that are, as we observed, difficult to take into account under motion-tracking uncertainties. Contacts also include a certain amount of deformation that we did not model. Assessing the contact force by a portable force sensor also affects the natural motion behavior. Instead, one could distribute force sensing devices in the experimented environment, but at the cost of many more sensing units. Considering all body limbs for contact would be presently difficult, as wearable force sensing suits do not exist in the current state of the technology. We, therefore, chose to focus on foot and hand contacts at the expense of other kinds of interaction such as shoulder or waist contacts (e.g., for seated motions).

To deal with these limitations and consider other features, our work can (and should) be extended to arbitrary contact configurations and motions. While a short-term solution could be to collect additional force and motion measurements (e.g., with force sensors at the knees and elbows), we anticipate that the increased level of instrumentation would strongly interfere with natural interaction behaviors, or even render some impossible (e.g., performing a cartwheel). Instead, our future work involves considering the distribution of contact forces as an inverse optimal control problem, i.e., finding optimization criteria privileging the forces measured in reality. Note that in the meantime, we ensured that the forces estimated by our framework would always be at least physically plausible (if not resembling human forces) by making the SOCP formulation independent of the acquired dataset. In the long term, we also plan to apply our framework to force-based robot learning from demonstration, online multicontact motion retargeting, and knowledge-based multicontact planning and control [43].

ACKNOWLEDGMENT

REFERENCES

[1] L. Rozo, P. Jiménez, and T. Carme, "A robot learning from demonstration framework to perform force-based manipulation tasks," *Intell. Serv. Robot.*, vol. 6, no. 1, pp. 33–51, 2013.

[2] J. Englsberger, P. Kozlowski, and C. Ott, "Biologically inspired dead-beat controller for bipedal running in 3D," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 989–996.

[3] S. Ha, Y. Bai, and C. K. Liu, "Human motion reconstruction from force sensors," in *Proc. ACM SIGGRAPH/Eurograph. Symp. Comput. Animation*, 2011, pp. 129–138.

[4] P. Zhang, K. Siu, J. Zhang, C. K. Liu, and J. Chai, "Leveraging depth cameras and wearable pressure sensors for full-body kinematics and dynamics capture," *ACM Trans. Graph.*, vol. 33, no. 6, 2014, Art. no. 221.

[5] N. Kyriazis and A. A. Argyros, "Physically plausible 3D scene tracking: The single actor hypothesis," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 9–16.

[6] Y. Wang *et al.*, "Video-based hand manipulation capture through composite motion control," *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. no. 43.

[7] C. Tran and M. M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.

[8] D. Michel, K. Panagiotakis, and A. A. Argyros, "Tracking the articulated motion of the human body with two RGBD cameras," *Mach. Vision Appl.*, vol. 26, no. 1, pp. 41–54, 2015.

[9] A. González, M. Hayashibe, and P. Fraisse, "Subject-specific center of mass estimation for in-home rehabilitation–Kinect-Wii board vs. Vicon-force plate," in *Converging Clinical and Engineering Research on Neurorehabilitation*. New York, NY, USA: Springer, 2013, pp. 705–709.

[10] G. Yang *et al.*, "A health-IOT platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2180–2191, Nov. 2014.

[11] T.-H. Pham, A. Bufort, S. Caron, and A. Kheddar, "Whole-body contact force sensing from motion capture," in *Proc. IEEE/SICE Int. Symp. Syst. Integr.*, 2016, pp. 58–63.

[12] S. J. Lee, Y. Motai, and H. Choi, "Tracking human motion with multi-channel interacting multiple model," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1751–1763, Aug. 2013.

[13] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors," Xsens Motion Technologies BV, Enschede, The Netherlands, Tech. Rep, 2009. [Online]. Available: https://www.xsens.com/images/stories/PDF/MVN_white_paper.pdf

[14] H. Ghasemzadeh and R. Jafari, "Physical movement monitoring using body sensor networks: A phonological approach to construct spatial decision trees," *IEEE Trans. Ind. Informat.*, vol. 7, no. 1, pp. 66–77, Feb. 2011.

[15] M. Mohammadi, T. L. Baldi, S. Scheggi, and D. Prattichizzo, "Fingertip force estimation via inertial and magnetic sensors in deformable object manipulation," in *Proc. IEEE Haptics Symp.*, 2016, pp. 284–289.

[16] Y. Zhu, C. Jiang, Y. Zhao, D. Terzopoulos, and S.-C. Zhu, "Inferring forces and learning human utilities from videos," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 3823–3833.

[17] M. A. Brubaker, L. Sigal, and D. J. Fleet, "Estimating contact dynamics," in *Proc. IEEE Int. Conf. Comput. Vision*, 2009, pp. 2389–2396.

[18] S. Nakaoka, S. Hattori, F. Kanehiro, S. Kajita, and H. Hirukawa, "Constraint-based dynamics simulator for humanoid robots with shock absorbing mechanisms," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3641–3647.

[19] V. M. Zatsiorsky, *Kinetics of Human Motion*. Champaign, IL, USA: Human Kinetics, 2002.

[20] B. I. Prilutsky and V. M. Zatsiorsky, "Optimization-based models of muscle coordination," *Exercise Sport Sci. Rev.*, vol. 30, no. 1, p. 32, 2002.

[21] G. Rao, D. Amarantini, E. Berton, and D. Favier, "Influence of body segments' parameters estimation models on inverse dynamics solutions during gait," *J. Biomech.*, vol. 39, no. 8, pp. 1531–1536, 2006.

[22] A. Muller, C. Germain, C. Pontonnier, and G. Dumont, "A comparative study of 3 body segment inertial parameters scaling rules," *Comput. Methods Biomech. Biomed. Eng.*, vol. 18, no. sup1, pp. 2010–2011, 2015.

[23] J. Jovic, A. Escande, K. Ayusawa, E. Yoshida, A. Kheddar, and G. Venture, "Humanoid and human inertia parameter identification using hierarchical optimization," *IEEE Trans. Robot.*, vol. 32, no. 3, pp. 726–735, Jun. 2016.

[24] X. Lv, J. Chai, and S. Xia, "Data-driven inverse dynamics for human motion," *ACM Trans. Graph.*, vol. 35, no. 6, Nov. 2016, Art. no. 163.

[25] C. Yang, Y. Jiang, Z. Li, W. He, and C.-Y. Su, "Neural control of bimanual robots with guaranteed global stability and motion precision," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1162–1171, Jun. 2017.

[26] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *Int. J. Robot. Res.*, 2017. [Online]. Available: https://doi.org/10.1177/0278364917710318

[27] I. Mordatch, K. Lowrey, G. Andrew, Z. Popovic, and E. V. Todorov, "Interactive control of diverse complex characters with neural networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 3132–3140.

[28] S. E. Oh, A. Choi, and J. H. Mun, "Prediction of ground reaction forces during gait based on kinematics and a neural network model," *J. Biomech.*, vol. 46, no. 14, pp. 2372–2380, 2013.

[29] T.-H. Pham, A. Kheddar, A. Qammaz, and A. A. Argyros, "Towards force sensing from vision: Observing hand-object interactions to infer manipulation forces," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 2810–2819.

[30] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[32] T.-H. Pham, N. Kyriazis, A. A. Argyros, and A. Kheddar, "Hand-object contact force estimation from markerless visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017. [Online]. Available: https://doi.org/10.1109/TPAMI.2017.2759736

[33] C. Fermüller *et al.*, "Prediction of manipulation actions," *Int. J. Comput. Vision*, pp. 1–17, 2017. [Online]. Available: https://doi.org/10.1007/s11263-017-0992-z

[34] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[35] S. Caron, Q.-C. Pham, and Y. Nakamura, "Stability of surface contacts for humanoid robots: Closed-form formulae of the contact wrench cone for rectangular support areas," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 5107–5112.

[36] Y. Nakamura, K. Yamane, Y. Fujita, and I. Suzuki, "Somatosensory computation for man-machine interface from motion-capture data and musculoskeletal human model," *IEEE Trans. Robot.*, vol. 21, no. 1, pp. 58–66, Feb. 2005.

[37] R. Dumas, L. Cheze, and J.-P. Verriest, "Adjustments to Mcconville et al. and Young et al. body segment inertial parameters," *J. Biomech.*, vol. 40, no. 3, pp. 543–553, 2007.

[38] S. Caron, Q.-C. Pham, and Y. Nakamura, "ZMP support areas for multi-contact mobility under frictional constraints," *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 67–80, Feb. 2017.

[39] S. Lengagne, Ö. Terlemez, S. Laturnus, T. Asfour, and R. Dillmann, "Retrieving contact points without environment knowledge," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2012, pp. 841–846.

[40] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A MATLAB-like environment for machine learning," in *Proc. BigLearn, NIPS Workshop*, 2011.

[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[42] M. Andersen, J. Dahl, and L. Vandenberghe, "CVXOPT: A python package for convex optimization," 2013. [Online]. Available: abel.ee.ucla.edu/cvxopt

[43] K. Bouyarmane and A. Kheddar, "Humanoid robot locomotion and manipulation step planning," *Adv. Robot.*, vol. 26, no. 10, pp. 1099–1126, 2012.

Tu-Hoa Pham received the Dipl.-Ing. SupAéro degree from Institut suprieur de l'aronautique et de l'espace, Toulouse, France, the M.Sc. degree in mathematics from Université Paul Sabatier, Toulouse, in 2013, and the Ph.D. in robotics from the Université de Montpellier, Montpellier, France, in 2016, conducted between the CNRS–AIST Joint Robotics Laboratory, Japan, and CNRS–UM LIRMM, France.

He is currently a Postdoctoral Researcher with IBM Research, Tokyo, Japan. His research interests include robot vision and learning for monitoring of human activities and learning from demonstration.
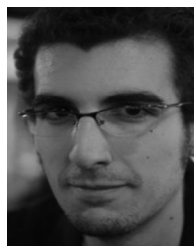
Stéphane Caron received the Ph.D. degree in mechano-informatics from the University of Tokyo, Tokyo, Japan, in 2016, with a thesis on multicontact motion planning for humanoid robots.

He is a Researcher in humanoid locomotion with LIRMM, CNRS–University of Montpellier, Montpellier, France. As an alumni of the École Normale Supérieure (ENS Paris), he stayed one year at the Technicolor Laboratory, Palo Alto, CA, USA, before joining the Nakamura Laboratory to undergo his doctoral studies. His research interests include contact interaction, numerical optimization, and predictive control, all related to the broader field of humanoid locomotion.

Abderrahmane Kheddar (M'04–SM'12) received the B.S. degree in computer science from the Institut National d'Informatique (ESI), Algiers, Algeria, in 1990, and the M.Sc. and Ph.D. degree in robotics from the University of Pierre et Marie Curie, Paris, France, in 1993 and 1997, respectively.

He is presently the Directeur de Recherche at CNRS and the Director of the CNRS–AIST Joint Robotic Laboratory, UMI3218/RL, Tsukuba, Japan. He is also leading the Interactive Digital Humans team at LIRMM, CNRS–University of Montpellier, Montpellier, France. His research interests include haptics, humanoids, and thought-based control using brain–machine interfaces. He is a founding member of the IEEE/RAS chapter on haptics and the co-Chair and founding member of the IEEE/RAS Technical Committee on model-based optimization. He is a member of the steering committee of the IEEE Brain Initiative and an Editor of the IEEE TRANSACTIONS ON ROBOTICS and within the Editorial Board of some other robotics journals. He is a founding member of the IEEE TRANSACTIONS ON HAPTICS and served in its Editorial Board during three years (2007–2010). He is a titular member of the National Academy of Technology of France and knight in the National Order of the Merit.