Physics-Based Models for Human Gait Analysis

Petrissa Zell, Bastian Wandt, and Bodo Rosenhahn

Abstract

This chapter deals with fundamental methods as well as current research on physics-based human gait analysis. We present valuable concepts that allow efficient modeling of the kinematics and the dynamics of the human body. The resulting physical model can be included in an optimization-based framework. In this context, we show how forward dynamics optimization can be used to determine the producing forces of gait patterns.

To present a current subject of research, we provide a description of a 2D physics-based statistical model for human gait analysis that exploits parameter learning to estimate unobservable joint torques and external forces directly from motion input. The robustness of this algorithm with respect to occluded joint trajectories is shown in a short experiment. Furthermore, we present a method that uses the former techniques for video-based gait analysis by combining them with a nonrigid structure from motion approach. To examine the applicability of this method, a brief evaluation of the performance regarding joint torque and ground reaction force estimation is provided.

Keywords

Computer vision • Human motion analysis • Physics-based simulation • Forward dynamics optimization • Data-driven regression • 3D motion reconstruction • Video-based force estimation

Contents

P. Zell (✉) • B. Wandt • B. Rosenhahn
Institut für Informationsverarbeitung, Leibniz Universität Hannover, Hannover, Germany
e-mail: zell@tnt.uni-hannover.de; wandt@tnt.uni-hannover.de; rosenhahn@tnt.uni-hannover.de

B. Müller, S.I. Wolf (eds.), *Handbook of Human Motion*,
DOI 10.1007/978-3-319-30808-1_164-1

Introduction

The human locomotor system is a complex construction consisting of the skeleton, the nervous system, muscles, tendons, and ligaments. Its functionality is a basic human need and the focus of numerous biomechanical studies. In the course of these studies, researchers require a measure to quantify healthiness of movement. A mathematically straightforward approach to describing the action of the neuromuscular system is to compute the net joint moments acting on the segments through inverse dynamics analysis. These net moments form a first approximation to the analysis of stress at the skeletal joint and are the foundation for classifying human motion. Let us consider an isolated joint in the human body. It might be effected by the pressure of neighboring bone segments, muscular exertion, and the strain of tendons and ligaments. In order to facilitate research, all of these forces are summarized in one vector, the joint torque.

Based on inverse dynamics, researchers develop and evaluate rehabilitation techniques, e.g., prosthetics alignment (Schmalz et al. 2002) and patient-specific gait modification (Fregly et al. 2007). Others investigate interdependencies in the locomotor system and their effect, such as the influence of abnormal hip mechanics on knee injury (Powers 2010). Many of these studies focus on analyzing gait patterns, as the basic form of human movement. A well-balanced walking style is an essential feature of a healthy locomotor system.

Unfortunately joint torques are not directly measurable but have to be assessed by interpreting their external effect, that is, the ground reaction force (GRF), which is exerted by the feet on the ground. The clinical standard to estimate joint torques is to inversely calculate them from the GRF vector, the geometry of the skeleton, and the recorded motion of the body. For this purpose, the subject's gait pattern has to be recorded by a motion capture (MoCap) system, and the GRF has to be measured with force plates, which restricts this method to a laboratory setup.

For more flexibility, researchers use the profound knowledge introduced by physical models to support gait analysis. These models yield a comprehensive description and control over the correlation between acting forces and resulting motion. Based on a physical model, joint torques can be estimated in the context of an optimization problem. There exist several established optimization formulations for this particular problem, namely, forward, inverse, and predictive dynamics optimization. A comparison of these approaches and a detailed description of forward dynamics optimization can be found in section "Optimization-Based Methods."

Optimization-based torque estimation entails several challenges: A representation of the entire body rather than just isolated segments (e.g., only one leg) is often required, and model properties such as the inertial parameters and the formulation of foot-ground contact model have to be defined with care. Furthermore, optimization approaches are usually connected with high computational cost and the necessity for considerable stabilization of the dynamic system. These issues might be addressed by the use of sophisticated optimization algorithms and refined constraints, but researchers also propose an alternative route, circumventing the optimization problem and instead relying on machine learning techniques. In such a framework, the connection between a motion pattern and the underlying forces is learned on the basis of a training set of motion sequences and associated optimization results. This way, the gait pattern of a new subject can be assessed using the knowledge gained in the training phase, i.e., unobservable joint torques are directly inferred from the motion data.

With the ever-increasing amount of easily accessible video data, the demand for motion analysis based on image sequences is inevitable. In this context, a combination of 3D motion reconstruction from 2D landmarks (nonrigid structure from motion) and physics-based motion analysis is sought for and determines the direction of research in this field.

This chapter is structured in the following way: First of all, we present state-of-the-art methods for physics-based modeling and motion analysis. Then, we describe fundamental concepts to formulate the kinematics and dynamics of physical models and introduce optimization-based methods for the estimation of effective forces. On the basis of this methodology, we present two approaches and related experimental results in more detail: a two-dimensional statistical model and a video-based framework for human gait analysis. Both approaches are data-driven, i.e., the respective algorithm learns parameter correlations on a training set and uses this knowledge to directly infer underlying forces from input motion data. Finally, we point to possible future research directions in this field.

State of the Art

A wide range of state-of-the-art methods related to human motion simulation and analysis rely on physics-based modeling as a tool to control movement or to gain valuable insides into the dynamics of human motion. In the field of computer graphics, physical models enable researchers to synthesize realistic looking human motions (Fang and Pollard 2003; Liu et al. 2005; Safonova et al. 2004; Sok et al. 2007; Wei et al. 2011; Zordan et al. 2005). The naturalness of a generated motion is either ensured by the closeness to motion capture (MoCap) data or by an efficiency measure. The latter case is based on the assumption that humans perform low energetic movement. Both approaches can be applied in an optimization framework. General issues of physics-based optimization are computational expense concerning calculation of objectives, constraints, and the associated derivatives and a high-dimensional search space, which complicates the optimization process. Fang and

Pollard (2003) introduce efficient physical constraints and objective functions for a fast optimization-based character animation. The authors formulate constraints and objectives in such a way that first derivatives can be computed analytically in linear time. The reduction of a high-dimensional parameter space to an appropriate subspace is addressed in Safonova et al. (2004). In this work, the authors use MoCap data to transfer the optimization problem to a low-dimensional subspace that includes the desired motion characteristics, exclusively.

While the minimization of an energy function to simulate movement corresponds to the overall tendency of humans to avoid energy expenditure, this technique fails to catch subject-specific motion traits and is not suitable for expressive motions, such as dancing or acting. In those cases, an approach based on MoCap data that provides more extensive information is sensible. Following this observation, Liu et al. (2005) propose a physics-based dynamical model that incorporates specific features of a persons' motion style, e.g., the tendency to strain certain muscles more than others. These characteristics are learned on a training set of MoCap sequences using an introduced method called nonlinear inverse optimization.

A physical background is also suitable to support statistical approaches by providing necessary constraints on the utilized model, e.g., (Wei et al. 2011) combines statistical motion priors with physical constraints in a probabilistic framework. The authors use a maximum a posteriori approach to synthesize a wide range of physically realistic motions and motion interactions.

Apart from these graphic applications, physics-based models can be found, facilitating typical computer vision tasks, like robust person tracking and 3D pose estimation (Bhat et al. 2002; Brubaker and Fleet 2008; Vondrak et al. 2008; Wren and Pentland 1998). Brubaker and Fleet (2008) use a simple planar model to estimate biomechanical characteristics of gait and combine it with a 3D kinematic model for monocular tracking. Due to the underlying physics, the model is able to handle large occlusions. While this approach solely focuses on walking motions, a related work by Vondrak et al. (2008) considers a wide range of motion types. The authors introduce a full-body 3D physical prior that integrates the corresponding dynamics into a Bayesian filtering framework. In addition to the motion dynamics, the algorithm is able to model ground contact and environment interaction.

The previously described works concentrate on human pose tracking with physics-based constraints but do not analyze the resulting force patterns. This biomechanical objective, i.e., the estimation of inner joint torques and external contact forces, is referred to as motion analysis and has been extensively treated in various fields (Blajer et al. 2007; Brubaker et al. 2009; Johnson and Ballard 2014; Stelzer and von Stryk 2006; Xiang et al. 2010; Zell and Rosenhahn 2015). In the following, we will list a selection of recent works.

Brubaker et al. (2009) use an articulated body model to infer joint torques and contact dynamics from motion data. They accelerate the optimization procedure by introducing additional root forces and effectively decoupling the problem at different time frames. This way an optimization step does not include the integration of EOM. A similar goal is pursued by Xiang et al. (2010). The authors introduce predictive dynamics as an approach for human motion simulation. Recently, researchers tried to

learn a direct mapping from a motion parametrization (joint angles) to the acting forces (joint torques) on the basis of MoCap data: Johnson and Ballard (2014) investigate sparse coding for inverse dynamics regression and Zell and Rosenhahn (2015) introduces a two dimensional statistical model for human gait analysis.

The transfer of motion analysis to a 2D setting, i.e., a video-based framework, is a relatively novel subject of research (Brubaker et al. 2009). By combining physics-based modeling with standard nonrigid structure-from-motion techniques (e.g., (Bregler et al. 2000)), it is possible to infer the unobservable physical parameters from monocular image sequences. State-of-the-art human pose estimation methods rely either on anthropometric constraints (Akhter and Black 2015; Ramakrishna et al. 2012; Wang et al. 2014) or on temporal bone length constancy (Wandt et al. 2015, 2016). These assumptions are relatively weak compared to a kinematic model. Therefore, the next step is a combination of pose estimation with a physics-based model of humans. This will not only improve the 3D reconstruction but also allow for the estimation of joint torques from an image sequence. An example implementation of a joint model for pose estimation and physical parameter regression is described in section "Combining Physical Models with 3D Reconstruction."

Models and Methodology

Kinematics and Dynamics of Physical Models

The groundwork of physics-based human motion analysis is a physical model of the human body. The literature offers a large amount of different models with varying complexity from which many have been developed in the scope of robotics applications. A frequently used type are mass-spring models that consist of bone segments and connected by joints which are provided with torsional springs to create joint torques. Beyond that, muscular skeletal models are applied (especially in biomechanical research) for a more realistic presentation of the human locomotor system, modeling the interaction between bone segments, joints, and an elaborate structure of muscles. Since the methods described in this chapter are more suitable for simple skeletal models, we focus on the portrayal of mass-spring models in this section.

The modeled skeleton comprises a number of bone segments with associated inertial properties (mass and inertial tensor) and connecting torsional springs for each joint degree of freedom (DOF). Examples of a 2D and a 3D model can be seen in Fig. 1. An essential part of these models is the kinematic chain that provides the linkage between all segments and predefines the space of possible motions. In the following part, we show how a kinematic chain can be defined using the Denavit-Hartenberg notation (Steinparz 1985), well-known from robotics.

Denavit-Hartenberg Convention

In 1955, Jacques Denavit and Richard Hartenberg introduced a convention to characterize a kinematic chain by a minimal set of four parameters for each interlink

transformation. A link refers to a rotational or translational degree of freedom. In this presentation, coordinate frames are attached to every link of the chain, and the Denavit-Hartenberg parameters specify the transformation between successive link coordinate systems.

Let $\boldsymbol{H}_j^{j-1}$ be the transformation matrix from link $j-1$ to link j and $[\theta_j,d_j,\alpha_j,a_j]$ the constituting parameters. Then the following operations are performed consecutively: A rotation of the coordinate system around the z_{j-1}-axis by the angle θ_j, a translation along the z_{j-1}-axis by the distance d_j, a second rotation around the new x_j-axis by the angle α_j, and a final translation along the x_j-axis by the distance a_j. The corresponding transformation matrix $\boldsymbol{H}_j^{j-1}$ is composed of the total rotation $\boldsymbol{R}_j^{j-1}$ and total translation $\boldsymbol{T}_j^{j-1}$, with

$$
\begin{aligned}
H_j^{j-1} &= \left(\begin{array}{ccc|c} & R_j^{j-1} & & T_j^{j-1} \\ \hline 0 & 0 & 0 & 1 \end{array}\right) \\
&= \left(\begin{array}{ccc|c} \cos\theta_j & -\sin\theta_j\cos\alpha_j & \sin\theta_j\sin\alpha_j & r_j\cos\theta_j \\ \sin\theta_j & \cos\theta_j\cos\alpha_j & -\cos\theta_j\sin\alpha_j & r_j\sin\theta_j \\ 0 & \sin\alpha_j & \cos\alpha_j & d_j \\ \hline 0 & 0 & 0 & 1 \end{array}\right).
\end{aligned} \tag{1}
$$

When considering a kinematic chain consisting of several links, the position and orientation of every link are determined by its predecessor transformations. Therefore, the transformation from the coordinate system at link j to the world coordinate frame L_0 is obtained by the product of all transformations in the sub-chain from the root link to the respective link:

$$
\boldsymbol{H}_j^0 = \prod_{i=1}^{j} \boldsymbol{H}_i^{i-1} = \begin{pmatrix} \boldsymbol{R}_i^0 & \boldsymbol{T}_j^0 \\ 0^T & 1 \end{pmatrix} \tag{2}
$$

The linear coordinates $\boldsymbol{p}_{c_i}^0$ of the center of mass (COM) of segment i can be derived from the total translation and the total rotation of the coordinate system attached to link j with respect to L_0:

$$
\boldsymbol{p}_{c_i}^0 = \boldsymbol{T}_j^0 + \boldsymbol{R}_j^0 \boldsymbol{p}_{c_i}^j, \tag{3}
$$

where $\boldsymbol{p}_{c_i}^j$ are the coordinates of the considered segment COM in the coordinate system of the preceding link.

TMT Method

The dynamics of the considered mass-spring model are specified by a set of equations of motion (EOM), which we will formulate using the TMT method (Schwab and Delhaes 2009). This approach combines advantages of both the Newton-Euler method and the Lagrange-Euler formulation by basically transforming from constrained to unconstrained dynamics. In the Newton-Euler approach, the EOM have to be extended to differential algebraic equations by additional constraints that model the interconnection (kinematic chain) between the model coordinates. This often leads to poor accuracy of solutions provided by numerical solvers. In contrast to that, the Lagrange-Euler approach deals with a minimum set of independent generalized coordinates, and the EOM are formulated based on an energetic view point. For this purpose, all arising energies have to be identified and their derivatives have to be calculated analytically, which might be difficult in the case of complex, large DOF models.

To exploit the benefits of both methods, the TMT method uses a force approach, but incorporates the kinematic constraints in a transformation $\boldsymbol{T}$ from dependent coordinates $\boldsymbol{x}$ (segment COM position and orientation) to independent generalized coordinates $\boldsymbol{q}$ (joint angles):

$$\boldsymbol{x} = \boldsymbol{T}(\boldsymbol{q}) \tag{4}$$

Derivation supplies the corresponding linear and angular velocities and accelerations:

$$\dot{\boldsymbol{x}} = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}} \tag{5}$$

$$\ddot{\boldsymbol{x}} = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\ddot{\boldsymbol{q}} + \frac{\partial}{\partial \boldsymbol{q}}\left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}\right)\dot{\boldsymbol{q}} \tag{6}$$

The second term in Eq. (6) is referred to as convective acceleration $\boldsymbol{G} = \frac{\partial}{\partial \boldsymbol{q}}\left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}\right)\dot{\boldsymbol{q}}$.

Furthermore the relation

$$\delta \boldsymbol{x} = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\delta \boldsymbol{q} \tag{7}$$

between virtual displacements of the coordinates is valid and can be inserted into D'Alembert's principle of virtual work:

$$\delta \boldsymbol{x}(\boldsymbol{F} - \boldsymbol{M}\ddot{\boldsymbol{x}}) = 0 \tag{8}$$

$$\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\delta \boldsymbol{q}\left(\boldsymbol{F} - \boldsymbol{M}\left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\ddot{\boldsymbol{q}} + \boldsymbol{G}\right)\right) = 0 \tag{9}$$

Here $\boldsymbol{F}$ summarizes all applied forces and torques and $\boldsymbol{M}$ is the system's inertia matrix. Because of the independence of $\delta\boldsymbol{q}$ and the validity of virtual displacements larger than zero, we can write Eq. (9) as

$$\left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\right)^T \left(\boldsymbol{F} - \boldsymbol{M}\left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\ddot{\boldsymbol{q}} + \boldsymbol{G}\right)\right) = 0 \tag{10}$$

Rearranging of Eq. (10) yields the EOM, formulated in independent generalized coordinates:

$$\left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\right)^T \boldsymbol{M}\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\ddot{\boldsymbol{q}} = \left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\right)^T (\boldsymbol{F} - \boldsymbol{M}\boldsymbol{G}) \tag{11}$$

This equation can be simplified by introducing the Jacobian $\boldsymbol{J} = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}$ and the generalized inertia matrix $\boldsymbol{\mathcal{M}} = \left(\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}\right)^T \boldsymbol{M}\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{q}}$, so that

$$\boldsymbol{\mathcal{M}}\ddot{\boldsymbol{q}} = J^T(\boldsymbol{F} - \boldsymbol{M}\boldsymbol{G}). \tag{12}$$

Formulating the Equations of Motion

For a clearer insight into the described methods, we will now provide an outline of the derivation of EOM following Eq. (12) and using the 3D mass-spring model depicted in Fig. 1 on the right-hand side. The model has 23 joint DOF and 6 DOF for the global position and orientation of the root joint, resulting in a total of 29 DOF. Therefore, we will receive a set of 29 coupled EOM.

The kinematic interconnection between the DOF $\boldsymbol{q}$, i.e., between the generalized coordinates, is defined via the Denavit-Hartenberg transformation matrices. In this context, the coordinates $\boldsymbol{q}$ appear either as a translation $\boldsymbol{a}$ or $\boldsymbol{d}$ (prismatic joints) or as a rotation α or θ (revolute joints). The Jacobian $\boldsymbol{J}$ of the according coordinate transformation $\boldsymbol{T}(\boldsymbol{q})$ from generalized coordinates to Cartesian coordinates (position and orientation of segment COM) is composed of a linear and a rotational part:

$$\dot{\boldsymbol{x}} = \begin{pmatrix} J^{\text{lin}} \\ J^{\text{rot}} \end{pmatrix} \dot{\boldsymbol{q}} \tag{13}$$

Instead of formulating $\boldsymbol{T}(\boldsymbol{q})$ and calculating every differential $\frac{\partial T_i}{\partial q_j}$ to determine the Jacobian, we choose a geometric approach. In the Denavit-Hartenberg convention, the rotation and translation axis associated with a DOF q_j is always equal to the z-axis of the respective coordinate frame L_{j-1}. This leads to the following equations for prismatic joints (Spong et al. 2005):

$$\begin{aligned} J_{ij}^{\text{lin}} &= \boldsymbol{z}_{j-1}^0 \\ J_{ij}^{\text{rot}} &= 0, \end{aligned} \tag{14}$$

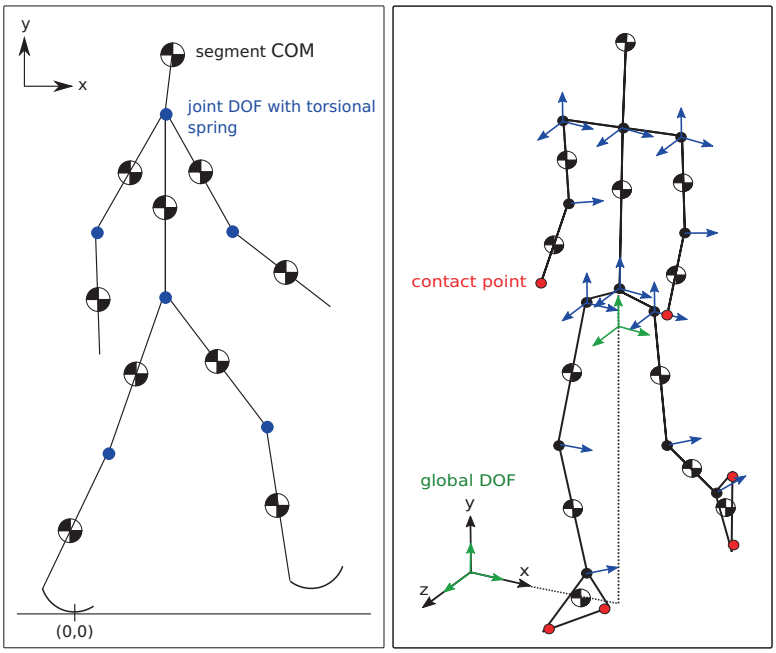

Fig. 1 Physical mass-spring-models in 2D (left) and 3D (right). Joint degrees of freedom (DOF) are depicted in blue and global DOF in green. Each joint degree of freedom is equipped with a torsional spring and each segment is assigned with a mass and a tensor of inertia.

and for revolute joints:

$$\begin{aligned} J_{ij}^{\text{lin}} &= \boldsymbol{z}_{j-1}^{0} \times \left(\boldsymbol{p}_{c_i}^{0} - \boldsymbol{T}_{j-1}^{0} \right) \\ J_{ij}^{\text{rot}} &= \boldsymbol{z}_{j-1}^{0}, \end{aligned} \tag{15}$$

with $\boldsymbol{T}_{j-1}^{0}$ and $\boldsymbol{p}_{c_i}^{0}$ calculated according to Eqs. (2) and (3), respectively.

Apart from the Jacobian, we also need an expression for the convective acceleration $\boldsymbol{G}$, which can be derived in a similar fashion with

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}^{\text{lin}} \\ \boldsymbol{G}^{\text{rot}} \end{pmatrix} \dot{\boldsymbol{q}}. \tag{16}$$

The corresponding sectional matrices for prismatic joints are

$$\begin{aligned} \boldsymbol{G}_{ij}^{\text{lin}} &= 0 \\ \boldsymbol{G}_{ij}^{\text{rot}} &= 0, \end{aligned} \tag{17}$$

and for revolute joints, we get

$$\begin{aligned} \boldsymbol{G}_{ij}^{\text{lin}} &= \sum_{k=0}^{i-1} \boldsymbol{z}_{\min\{j-1,k\}}^{0} \times \left(\boldsymbol{z}_{\max\{j-1,k\}}^{0} \times \left(\boldsymbol{p}_{c_i}^{0} - \boldsymbol{T}_{\max\{j-1,k\}}^{0} \right) \right) \dot{q}_k \\ \boldsymbol{G}_{ij}^{\text{rot}} &= \boldsymbol{z}_{j-1}^{0} \times \sum_{k=\text{j}}^{i-1} \boldsymbol{z}_{k}^{0} \dot{q}_k. \end{aligned} \tag{18}$$

The Jacobian together with the convective acceleration describes the kinematic of our model. To characterize the dynamics of the system, i.e., formulate the EOM, we need further properties, more precisely, the inertia matrix $\boldsymbol{M}$ and the active forces and torques $\boldsymbol{F}$ (cf. Eq. (12)).

The inertia matrix is composed of mass values and inertia tensors for all model segments. Let $\boldsymbol{M}^{c_i}$ and $\boldsymbol{I}^{c_i}$ be diagonal matrices containing masses and moments of inertia in the segment coordinate frames, i.e., the values for the inertial properties correspond to the rotation axes in the local frames attached to the individual segment COM. Then the inertia matrix related to the coordinates $\boldsymbol{x}$ in the global frame L_0 is

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}^{c_i} & \\ & \boldsymbol{R}_{c_i}^{0} \boldsymbol{I}^{c_i} \boldsymbol{R}_{c_i}^{0^T} \end{pmatrix}, \tag{19}$$

with the rotation $\boldsymbol{R}_{c_i}^{0}$ given by Eq. (2). Finally, the generalized inertia matrix required for the EOM is obtained by $\boldsymbol{\mathcal{M}} = \boldsymbol{J}^T \boldsymbol{M} \boldsymbol{J}$.

The motion of the 3D mass-spring model is driven by spring torques and outer contact forces. The spring torques τ_j are applied at every joint DOF of the model (represented by blue arrows in Fig. 1). A commonly used parametrization is

$$\tau_j = -\kappa_j \left(q_j - q_j^{(0)} \right) - d_j \dot{q}_j, \tag{20}$$

with a linear resetting torque depending on the spring constant κ_j and the resting angle $q_j^{(0)}$ and an additional damping term proportional to the angular velocity of the joint DOF with damping constant d_j (Brubaker and Fleet 2008). To simulate increasingly dynamic and natural-looking motions, additional nonlinear terms might be added, e.g., Zell and Rosenhahn (2015):

$$\begin{aligned} \tau_j &= -\kappa_j \left(q_j - q_j^{(0)}(\boldsymbol{q}) \right) - d_j \dot{q}_j, \\ q_j^{(0)}(\boldsymbol{q}) &= q_j^{(0)} + \sum_{k=1}^{4} c_k x_{CoM}^{k}(\boldsymbol{q}). \end{aligned} \tag{21}$$

Here, the resting angle is characterized by a fourth-order polynomial of the position of the whole body COM. It is noteworthy that this representation is only sensible for motions like walking or running that exhibit an even forward movement of the COM. The single joint torques are summarized in the vector $\boldsymbol{\tau}$ with $\tau_j = 0$ for $j \leq 6$ (six global DOF that are not actuated by a joint torque).

When the model is in contact with its environment, e.g., standing on the ground or lifting objects, it is effected by contact forces, denoted by $\boldsymbol{F}_c$ in the following. For the majority of human motion types, our feet have to touch the ground and so-called ground reaction forces (GRF) act on the sole of the foot. These forces mostly accelerate the body in vertical direction, balancing the gravitational force. They act as a braking force when the foot contacts the ground and a propulsive force at push off from the ground. Furthermore, a horizontal friction force, depending on the properties of the touching materials and the vertical force, prevents a sliding of the feet on the ground. There are several possibilities to model the ground contact: The 2D model illustrated in Fig. 1 is equipped with circular feet that enables a rolling motion of the feet on the ground during gait. The GRF is concealed in a model constraint, by defining the contact point as the root of the kinematic chain and effectively setting its vertical coordinate to zero (Zell and Rosenhahn 2015). This approach is not advisable for a 3D model with a significantly larger DOF, since it would result in intolerably long kinematic chains that entail high simulation uncertainties.

As an alternative, Brubaker et al. (2009) suggest a spring contact force that accelerates points on the sole of the foot, when approaching the ground plane. The force is modulated by two sigmoid functions, which on the one hand ensure that a contact point $\boldsymbol{p}$ is only effected when it is very close to the ground surface S and on the other hand prevent an acceleration toward the ground. The force is defined by the following equations:

$$\begin{aligned} \boldsymbol{F}_c(\boldsymbol{p},\dot{\boldsymbol{p}},\boldsymbol{\theta}_S) &= h(-60d_S(\boldsymbol{p}))h(5n_c(\boldsymbol{p}))[n_c(\boldsymbol{p})\boldsymbol{n}_S(\boldsymbol{p}) + \boldsymbol{t}_c(\boldsymbol{p})], \\ n_c(\boldsymbol{p}) &= -\kappa_N(d_S(\boldsymbol{p}) - 1) - \delta_N\dot{\boldsymbol{p}}^T\boldsymbol{n}_Y(\boldsymbol{p}), \\ \boldsymbol{t}_c(\boldsymbol{p}) &= -\delta_T\left(\dot{\boldsymbol{p}} - \left(\boldsymbol{n}_S(\boldsymbol{p})^T\dot{\boldsymbol{p}}\right)\boldsymbol{n}_S(\boldsymbol{p})\right), \end{aligned} \tag{22}$$

with the sigmoid $h(x) = \frac{1}{2}(1 + \tanh(x))$. The total force is composed of the normal force $n_c\boldsymbol{n}_S$ and the tangential attenuation $\boldsymbol{t}_c$, where n_c denotes the magnitude and $\boldsymbol{n}_S$ is the surface normal. The sigmoidal modulation depends on the shortest distance $d(\boldsymbol{p})$ between the point $\boldsymbol{p}$ and the ground surface S. Finally, model parameters are included in the vector $\boldsymbol{\theta}_S$, namely, the position and orientation of the ground plane, the spring stiffness κ_N, the normal damping constant δ_N, and the tangential damping constant δ_T.

The last force term we are still missing for a complete description of the 3D model with our EOM is the gravitational force:

$$\boldsymbol{F}_g = \boldsymbol{M}\boldsymbol{g}, \tag{23}$$

where $\boldsymbol{g}$ is the gravitational acceleration with respect to the dependent coordinates $\boldsymbol{x}$. Since the acceleration $g = -9.81\ \mathrm{m/s^2}$ only affects the linear coordinates of the segment COM in vertical direction, all other components are set to zero.

Finally, we insert all discussed terms into Eq. (12) and receive the model EOM:

$$\boldsymbol{\mathcal{M}}\ddot{\boldsymbol{q}} = J^T\left(\boldsymbol{\tau} + \boldsymbol{F}_c + \boldsymbol{F}_g - \boldsymbol{M}\boldsymbol{G}\right). \tag{24}$$

This equation determines the dynamics of our model; in other words, it predicts the development originating from an initial configuration of joint angles and angular velocities. The acting forces and therefore the accelerations of the generalized coordinates are completely defined by the model parameters $\boldsymbol{\theta}$ and the current state of the model $[\boldsymbol{q},\dot{\boldsymbol{q}}]^T$. The model parameters include all spring and contact parameters, e.g., stiffness and damping constants, resting angles and the position and orientation of the contact plane.

Optimization-Based Methods

On the basis of the previously described physical models, we can estimate the underlying forces and torques of motion data in an optimization framework. The corresponding approaches found in the literature can be divided into three different categories, inverse dynamics, forward dynamics, and predictive dynamics.

The inverse dynamics approach considers the motion parameters (e.g., joint angles $\boldsymbol{q}$) as optimization variables. A suitable objective function, e.g., the distance to a target motion or a human performance measure, is minimized and joint torques are calculated inversely. This way, the expensive integration of EOM is avoided and the optimization problem is easily controlled, i.e., no sophisticated constraints are necessary. A disadvantage of this method is the indirect derivation of forces that is affected by the characteristic inaccuracy of joint trajectories and their accelerations.

In contrast to that, forward dynamics treats the forces and joint torques as design variables that generate a motion through integration of EOM. As a result, the optimization problem induces large computational cost and accurate boundaries are crucial for the convergence of the optimization algorithm. Advantages of this approach are the directly optimized forces and a natural conclusion of motion parameters via EOM. Inverse as well as forward dynamics depends on the accuracy of the inertial properties to achieve a sound torque estimation.

A rather novel approach, called predictive dynamics, was introduced by Xiang et al. (2010). Both the joint torques and the model states are optimization parameters and the EOM are treated as equality constraints. The method is closely related to direct collocation methods (Stelzer and von Stryk 2006). It is computationally efficient since no integration of EOM is required but because of the large number of design variables an optimization algorithm suitable for large-scale problems that has to be used.

Optimization Problem Setup

In this section, we will focus on forward dynamics optimization, since it is based on the natural relation between forces and motion and provides a convenient frame to introduce optimization principles. We will formulate the problem and present appropriate objective functions, regularization, and constraint terms. Let us consider a problem of the form

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \{f(\boldsymbol{\theta})\} \\ &\text{s.t.} g_i(\boldsymbol{\theta}) \leq 0, \;\; i = 1, \ldots, m \\ &\quad h_i(\boldsymbol{\theta}) = 0, \;\; i = 1, \ldots, n, \end{aligned} \tag{25}$$

where $\boldsymbol{\theta}$ are model parameters, f denotes the objective function, and g_i and h_i are inequality and equality constraints, respectively. The forward dynamics step, i.e., the integration of the EOM, is represented by the function $\mathcal{D}$ and results in the temporal development of the model states

$$[\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)]^T_{\text{mod}} = \mathcal{D}(t, \boldsymbol{\theta}, \boldsymbol{q}(0), \dot{\boldsymbol{q}}(0)), \tag{26}$$

with the initial state $[\boldsymbol{q}(0), \dot{\boldsymbol{q}}(0)]^T$. For the approximate solution, an ordinary differential equation (ODE) solver has to be used. These solvers are generally based on a Runge-Kutta method (Mayers and Sli 2003).

In our problem statement, the motion data exists as input information, e.g., recorded by a MoCap system and we want to estimate appropriate joint torques. Therefore, it is reasonable to define an objective function that includes the distance between modeled states and target states: $[\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)]^T_{\text{targ}}$,e.g., the quadratic L_2-norm of their difference:

$$f(\boldsymbol{\theta}) = \int_0^T \left\| \mathcal{D}(t, \boldsymbol{\theta}, \boldsymbol{q}(0), \dot{\boldsymbol{q}}(0)) - [\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)]^T_{\text{targ}} \right\|^2. \tag{27}$$

Since the optimization algorithm generally has to deal with a high-dimensional parameter space and the subspace that creates realistic motions is non-convex, careful regularization and constraints on the system are indispensable. A number of human performance measures can be used to support the optimization routine. These measures penalize motions with high energetic cost which are considered to be unnatural. Commonly used quantities are the **dynamic effort**:

$$f = \int_0^T ||\boldsymbol{\tau}||^2 \mathrm{d}t, \tag{28}$$

and the **jerk**:

$$f = \int_0^T ||\dot{\boldsymbol{\tau}}||^2 \mathrm{d}t. \tag{29}$$

Minimizing the jerk increases the smoothness of the simulated motion.

Furthermore, we can reduce the searched parameter space by incorporating prior knowledge as constraints g_i and h_i. Obviously, it is advisable to set boundaries on the model parameters, such as

$$\begin{aligned} 0 \leq \kappa_j \leq u_{\kappa_j}, \\ l_{q_j} \leq q_j^{(0)} \leq u_{q_j}, \\ 0 \leq d_j \leq u_{d_j}, \end{aligned} \tag{30}$$

with the respective lower and upper bounds l and u for the joint spring parameters. In addition to that, we might have knowledge about the motion of certain points of our model, e.g., the contact points during walking are not supposed to slide over the ground, but have to stay fixed. This information can be taken into account by the equality constraint

$$h_c = \int_0^T ||J_c \dot{\boldsymbol{q}}||^2 \mathrm{d}t = 0, \tag{31}$$

where $\boldsymbol{J}_c$ is the section of the Jacobian that is associated with the coordinates of the contact points.

In some scenarios, GRF have been recorded with force plates and can be used as ground truth data $\boldsymbol{F}_{\text{targ}}$. Then we can constrain the modeled contact force with

$$h_f = \int_0^T ||\boldsymbol{F}_c - \boldsymbol{F}_{\text{trag}}||^2 \mathrm{d}t = 0. \tag{32}$$

In practice, it is often beneficial to formulate equality constraints as soft inequality constraints by introducing an upper threshold or to include them as regularization terms. This way the convergence of the optimization algorithm is more likely.

Learning Force Patterns in 2D

In this section, we will present a statistical approach for the analysis of human gait (Zell and Rosenhahn 2015). The general idea of this work is to learn model parameters on a set of MoCap walking sequences from 115 subjects and then use this knowledge to infer the underlying joint torques of a new gait pattern directly from the motion data. The proposed model considers the 2D projection of walking in the sagittal plane (cf. Fig. 1 on the left-hand side). It consists of ten segments and nine DOF for the joint angles. The corresponding EOM can be derived as described in section "Formulating the Equations of Motion" and have the form

$$\boldsymbol{\mathcal{M}}(\boldsymbol{q})\ddot{\boldsymbol{q}} = \boldsymbol{F}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{\theta}). \tag{33}$$

Based on this, movement can be simulated by solving the related initial value problem according to Eq. (26).

Combined Statistical Model

For the direct regression of joint torques from motion, a statistical model has to be learned that combines motion characteristics with a physical representation. Following Troje (2002a), the authors represent walking as a linear combination of principle component postures $\boldsymbol{p}_i$ with sinusoidal variation of coefficients:

$$\begin{aligned}\boldsymbol{p}(t) = \boldsymbol{p}_0 + \boldsymbol{p}_1 \sin(\omega t) + \boldsymbol{p}_2 \sin(\omega t + \phi_2) \\ + \boldsymbol{p}_3 \sin(2\omega t + \phi_3) + \boldsymbol{p}_4 \sin(\omega t + \phi_4).\end{aligned} \tag{34}$$

Here, $\boldsymbol{\omega}$ is the fundamental frequency of the gait and (ϕ_2,ϕ_3,ϕ_4) are phase shifts. In Troje (2002a), a posture consists of 15 three-dimensional joint positions, resulting in a 45-dimensional vector $\boldsymbol{p}$. For a full motion representation, all characterizing parameters are summarized in one vector:

$$\boldsymbol{u} = [\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4, \omega, \phi_2, \phi_3, \phi_4]^T. \tag{35}$$

To include information about physical properties, the gait patterns from the training set are approximated using a forward dynamics optimization approach (cf. section "Optimization-Based Methods"). The EOM are integrated, resulting in a simulated motion which is compared to the target motion at a set of key times $\{t_k\}_k$. The authors optimize the physical model parameters $\boldsymbol{\theta}$ together with the initial state $\left[\boldsymbol{q}_0, \dot{\boldsymbol{q}}_0\right]^T$ by minimizing the sum of squared reconstruction errors and constraining the simulated GRF $\boldsymbol{F}_c$ to lie in the vicinity of ground truth data $\boldsymbol{F}_{\text{targ}}$. This way, realistic force patterns are ensured. The optimization problem is formulated as follows:

$$\begin{aligned}(\boldsymbol{q}_0, \dot{\boldsymbol{q}}_0, \boldsymbol{\theta}) = \quad & \underset{\boldsymbol{q}_o, \dot{\boldsymbol{q}}_0, \boldsymbol{\theta}}{\arg\min} \left\{ \sum_k \left\| \mathcal{D}(t_k, \boldsymbol{q}_0, \dot{\boldsymbol{q}}_0, \boldsymbol{\theta}) - [\boldsymbol{q}(t_k), \dot{\boldsymbol{q}}(t_k)]^T_{\text{targ}} \right\|^2 \right\}, \\ & \text{s.t.} \quad \left\| \boldsymbol{F}_c\left(\boldsymbol{q}(t_k), \dot{\boldsymbol{q}}(t_k), \ddot{\boldsymbol{q}}(t_k)\right) - \overline{\boldsymbol{F}}_{\text{targ}} \right\| \leq \eta_k,\end{aligned}$$

with thresholds η_k. The simulated GRF vectors are normalized with respect to body weight and compared to a normalized mean ground truth force $\overline{\boldsymbol{F}}_{\text{targ}}$, obtained from force plate measurements. The effective GRF is calculated via

$$\boldsymbol{F}_c(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}}) = \sum_i \frac{m_i}{M} (\boldsymbol{a}_i(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}}) - \boldsymbol{g}), \tag{37}$$

where $\boldsymbol{g}$ is the gravitational acceleration and $\boldsymbol{a}_i$ is the Cartesian acceleration of body segment i with respective to mass m_i. The total body mass is denoted by M.

The optimization problem can be solved by sequential quadratic programming (SQP) (Powell 1978) and the resulting physics-based model parameters are

$$\boldsymbol{v} = [\boldsymbol{q}_0, \dot{\boldsymbol{q}}_0, \boldsymbol{\theta}, M]^T. \tag{38}$$

Based on the combined parametrization of $\boldsymbol{u}$ and $\boldsymbol{v}$, it is possible to infer joint torques from joint angle trajectories and vice versa. The corresponding parameter vectors are optimized for each walking sequence in the training set and then stacked in one matrix:

$$\boldsymbol{W} = \underbrace{\begin{bmatrix} \boldsymbol{u}_1 \dots \boldsymbol{u}_{115} \\ \boldsymbol{v}_1 \dots \boldsymbol{v}_{115} \end{bmatrix}}_{115 \text{ subjects}} \begin{matrix} \in \mathbb{R}^{229} (\text{motion Eq.}(35)) \\ \in \mathbb{R}^{52} (\text{physical model Eq.}(38)) \end{matrix} \tag{39}$$

The combined statistical model encompasses geometrical properties, dynamical behavior, and the physical description of a walking motion. All of these features contribute to the characteristics of a gait pattern, and their mutual dependence can be exploited to infer missing information based on an incomplete parameter set. Two different regression methods are applied: a *k*-nearest neighbor (*k*-NN) regression and an asymmetrical projection into the principal component space (aPCA), as introduced by Al-Naser and Söderström (2012).

Force Estimation with Occluded Input

The proposed direct regression methods are evaluated regarding the deviation of estimated 2D GRF and stance knee torques from ground truth data and inversely calculated torques, respectively. For this purpose, MoCap sequences of three different subjects with synchronized force plate measurements were recorded. Based on the motion parameters $\boldsymbol{u}$ of the considered sequence, the physical parameters $\boldsymbol{v}$ are inferred and used as input to the physical model simulation. This yields the estimated GRF via Eq. (37) and the estimated joint torques through the respective spring torque equation, e.g., Eqs. (20) or (21).

In order to evaluate the robustness of the direct regression methods with respect to incomplete input information, joint trajectories are iteratively removed from the considered gait sequence. Starting with missing left-hand trajectory, the authors successively remove the trajectories of the left elbow, ankle, and knee, resulting in a final regression based solely on the right-hand side of the human body. The number of missing joint trajectories is denoted by N. Results for one example subject are illustrated in Fig. 2. For a quantitative evaluation, the symmetric mean absolute percentage errors (SMAPE) ε_F and ε_τ for GRF $\boldsymbol{F}_c$ and stance knee torques τ_{K_1}, respectively, are listed in Table 1 together with the arising computation times. To compare the performance of the direct regression methods to an optimization procedure, a third method (OPT) following the forward dynamics optimization in Eq. (36) was implemented. In the case of missing joint trajectories, an additional term, based on the dynamic effort, is minimized to compensate for the incomplete information.

With zero missing input trajectories, the best approximation is achieved by the optimization-based method OPT, as expected. Under these circumstances, the physical parameters are optimized, to create a simulated motion as close to the target

motion as possible with no additional energy minimization that affects the results. As the input information is reduced, the direct regression methods (k-NN, aPCA) increasingly outperform OPT. In particular, the estimates based on the k-NN method show consistently low SMAPE values. In regard to the computational cost, the optimization requires computation times that are at least three orders of magnitude higher than those for the direct regression methods. All of these values were generated using unoptimized Matlab code.

In summary, the combination of motion and force parameters in one statistical representation allows a fast and robust regression of joint torques from MoCap data. In this work, a limiting factor of the approximation quality is the rather simple 2D physical model.

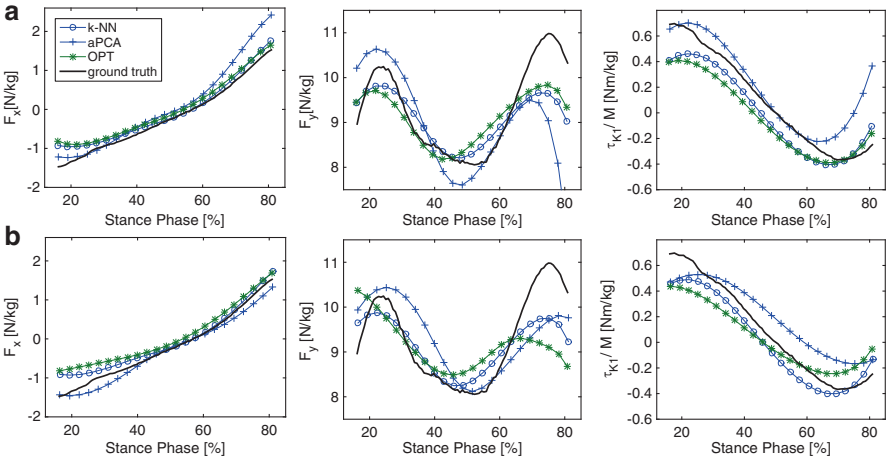

Fig. 2 Comparison between different regression and optimization methods concerning estimated GRF components F_y, F_x and knee torques τ_{K1}. Positive values correspond to flexor torques. The results are based on full joint trajectory information (in (**a**)) and on partial information with $N = 4$ (in (**b**)), respectively. In the case of GRF components, the *black* line illustrates ground truth data, and in the case of joint torques, it represents torques calculated via inverse dynamics. A quantitative evaluation can be found in Table 1

Table 1 SMAPE values ϵ_F and ϵ_τ for GRF and knee torque estimates with related computation times t_c

	k-NN			aPCA			OPT		
N	ε_F	ε_τ	t_c [s]	ε_F	ε_τ	t_c [s]	ε_F	ε_τ	t_c [s]
0	1.504	1.624	2.994	1.726	2.019	**0.881**	**1.483**	**1.420**	8103
1	**1.504**	**1.586**	2.117	1.780	2.052	**0.934**	1.811	1.851	4358
2	**1.496**	**1.582**	2.112	2.005	2.083	**0.913**	1.890	1.793	3096
3	**1.524**	**1.562**	2.583	1.606	2.169	**0.619**	1.843	2.069	3229
4	**1.528**	**1.565**	2.538	1.594	2.072	**0.669**	1.911	2.188	2092

N indicates the number of missing input joint trajectories

Video-Based Gait Analysis

As shown in the previous sections, methods based on motion capture records combined with additional force measurement achieve high-quality results. However, they require expensive and complex laboratory setups. With the increasing ease of capturing videos, for example, with mobile devices, the demand for purely video-based motion capture rises. In the combination of structure-from-motion methods with complex physical models, researchers found a tool to infer physical parameters directly from monocular input videos.

A Short Introduction to NRSfM

Structure-from-motion (SfM) describes techniques to infer a 3D structure of an object from its 2D projections into an image plane. Early work concerning the 3D reconstruction of rigid objects was done in the 90s by Tomasi and Kanade (1992). In their seminal work, they proposed to decompose a 2D measurement matrix $\boldsymbol{W}_{2D}$ into camera matrices $\boldsymbol{K}$ and a shape matrix $\boldsymbol{S}$:

$$\boldsymbol{W}_{2D} = \boldsymbol{K}\boldsymbol{S}. \tag{40}$$

With observations from multiple viewpoints, they were able to reconstruct the rigid 3D structure of the object. In 2000, Bregler et al. (2000) generalized Tomasi and Kanades work to the nonrigid case, where the observed object is allowed to deform over time. Their method rests upon the idea that a deformable shape $\boldsymbol{S}$ can be represented by a linear combination of $\boldsymbol{K}$-weighted base shapes $\boldsymbol{Q}_i$:

$$\boldsymbol{S} = \sum_{i=1}^{K} d_i \boldsymbol{Q}_i, \tag{41}$$

where d_i is the weighting factor for the i-th base shape. This leads to a generalization of Eq. (40):

$$\boldsymbol{W}_{2D} = \boldsymbol{K}\boldsymbol{D}\boldsymbol{Q}, \tag{42}$$

where $\boldsymbol{D}$ contains the weighting factors d_i for the base shapes in $\boldsymbol{Q}$. A common approach is to minimize a reprojection error e_r such as

$$e_r = \| \boldsymbol{W}_{2D} - \boldsymbol{K}\boldsymbol{D}\boldsymbol{Q} \| \tag{43}$$

for the different variable sets with respect to orthogonality constraints on the camera matrix. For further detail, the reader is referred to Bregler et al. (2000).

Since the base shapes in this mathematical representation are nonunique in the following years, numerous authors (e.g., Dai and Li 2012; Hamsici et al. 2011; Park et al. 2010; Torresani et al. 2003, 2008) proposed several constraints to convert this problem into a more feasible one.

NRSfM for 3D Reconstruction of Human Motion

Although the methods briefly introduced in section "A Short Introduction to NRSfM" were developed for arbitrary deformable objects, they also showed great success in the 3D reconstruction of human motion. However, they require a sufficient camera or object motion to achieve acceptable results. Therefore, they are not applicable for more realistic scenarios with limited camera motion such as in surveillance situations. This leads to the idea of including prior knowledge about the observed object into the problem formulation.

In 2002, N. Troje (2002a, b) showed that a single human pose during a gait motion can be described by the linear combination of specific weighted vectors. These vectors (named *eigenpostures*) are obtained by performing a PCA on multiple 3D gait sequences. It was shown that the first four eigenpostures were sufficient to cover more than 98% of the variance of a walking motion. N. Troje also discovered that for a walking motion, the weights for the eigenpostures describe a sine curve. This leads to the conclusion that the human gait can be described by just the four eigenpostures and the corresponding sine functions which are defined by their amplitude, frequency, and phase shift. In Troje (2002a, b), this discovery was used to identify and create gait patterns for different gait characteristics. Since Troje's formulation is in close proximity to Eqs. (41) and (42), the applicability for 3D reconstruction is obvious.

Multiple researchers used the idea of representing human poses in a PCA basis for the 3D reconstruction from single images, e.g., Akhter and Black (2015), Ramakrishna et al. (2012), and Wang et al. (2014). Since these learned bases also include many nonhuman poses, further anthropometric priors are applied, such as symmetry and predefined bone length relations. Wandt et al. (2015, 2016) were the first to combine temporal information of the skeletal structure with the learning of subspaces by employing a bone length constancy term. They also exploited the periodic behavior of the eigenpostures coefficients for periodic motions as proposed by Troje (2002a). Thus, their method appears to be well suited for gait reconstruction tasks in general, but not for a detailed analysis of idiosyncratic gait (e.g., pathological gait caused by neurological disorders) due to the limited number of used principle components.

Combining Physical Models with 3D Reconstruction

Here an algorithm is presented to jointly infer the 3D pose and physical parameters from monocular input data. The presented algorithm first performs the 3D pose reconstruction followed by the physical simulation to eliminate any ambiguities between camera and object motion as well as enforcing a physically plausible reconstruction. The dynamical description is based on a 3D physical mass-spring model (cf. Fig. 1 on the right-hand side) with spring and contact parameters $\boldsymbol{\theta}$ that result from the following optimization procedure:

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \left\{ \frac{w_0}{T} \sum_{t=1}^{T} \left\| [\boldsymbol{q},\dot{\boldsymbol{q}}]_{\mathrm{mod},t}{}^{T} - [\boldsymbol{q},\dot{\boldsymbol{q}}]_{\mathrm{targ},t}^{T} \right\|^2 + \frac{w_1}{T} \sum_{t=1}^{T} \left(\|\boldsymbol{\tau}_t\|^2 + \|\dot{\boldsymbol{\tau}}_t\|^2 \right) + \frac{w_2}{T} \sum_{t=1}^{T} \|J_c \dot{\boldsymbol{q}}_{\mathrm{mod},t}\|^2 \right\}, \tag{44}$$

with regularization weights (w_0,w_1,w_2). A description of the individual terms has been given in section "Optimization Problem Setup."

A joint parameter space is built consisting of the weighting matrix $\boldsymbol{D}$ from Eq. (43) and the physical parameters $\boldsymbol{\theta}$. As shown in Eq. (42), the 3D shape can be written as $\boldsymbol{S} = \boldsymbol{DQ}$. For known 3D shapes $\boldsymbol{S}$ from the training data, $\boldsymbol{D}$ can be directly calculated via

$$\boldsymbol{D} = \boldsymbol{SQ}^{+}, \tag{45}$$

where $\boldsymbol{Q}^{+}$ denotes the Moore-Penrose pseudoinverse of $\boldsymbol{Q}$. Finally, a vector $\boldsymbol{v_k}$ composed of the vectorized weighting matrix $\boldsymbol{D}$ and the physical parameters $\boldsymbol{\theta}$ is assigned to each sequence k in the training set:

$$\boldsymbol{v}_k = \begin{pmatrix} vec(\boldsymbol{D}) \\ \boldsymbol{\theta} \end{pmatrix}. \tag{46}$$

Here, $vec(\cdot)$ is the vectorization operator which stacks the columns of the matrix into a vector. We assume that a newly observed motion lies in the space spanned by the vectors $\boldsymbol{v}_k$ for each sequence k.

In the first step, the algorithm performs a 3D reconstruction of the observation matrix $\boldsymbol{W}_{2d}$ alternatingly minimizing Eq. (43) for the camera matrix $\boldsymbol{K}$ and coefficient matrix $\boldsymbol{D}$. The consequent 3D shapes are calculated by $\boldsymbol{S} = \boldsymbol{DQ}$. Since a linear model is used to represent nonlinear deformations, the estimated 3D motion is expected to differ from the real motion. This can be easily seen when analyzing the temporal behavior of the bone lengths, shown on the left in Fig. 3. After the 3D reconstruction, the bone lengths fluctuate heavily due to the above mentioned linear model. This issue is addressed by limiting the parameter space to physically valid motions; in other words, the weighting coefficients and physical parameters are inferred by means of a k-nearest neighbor (k-NN) regression in the space spanned by the vectors $\boldsymbol{v}_k$. As suggested by Zell and Rosenhahn (2015), a local k-NN regression can be used, which outperforms global approaches like PCA or asymmetric PCA for this particular problem.

The recovered physical parameters $\boldsymbol{\theta}$ are now employed to simulate a 3D motion by integrating the corresponding set of EOM. This step provides the resulting joint torques and contact forces and converts the rough 3D pose estimation to a physically feasible 3D reconstruction of the observed motion. Comparing the bone length variation before (cf. Fig. 3 left) and after physical simulation (cf. Fig. 3 right) indicates an improvement regarding plausibility.

Additionally, the use of the physical model allows for resolving the ambiguity between camera and object motion. Based on the knowledge gained from the physical simulation of the observed object, e.g., the forward movement during

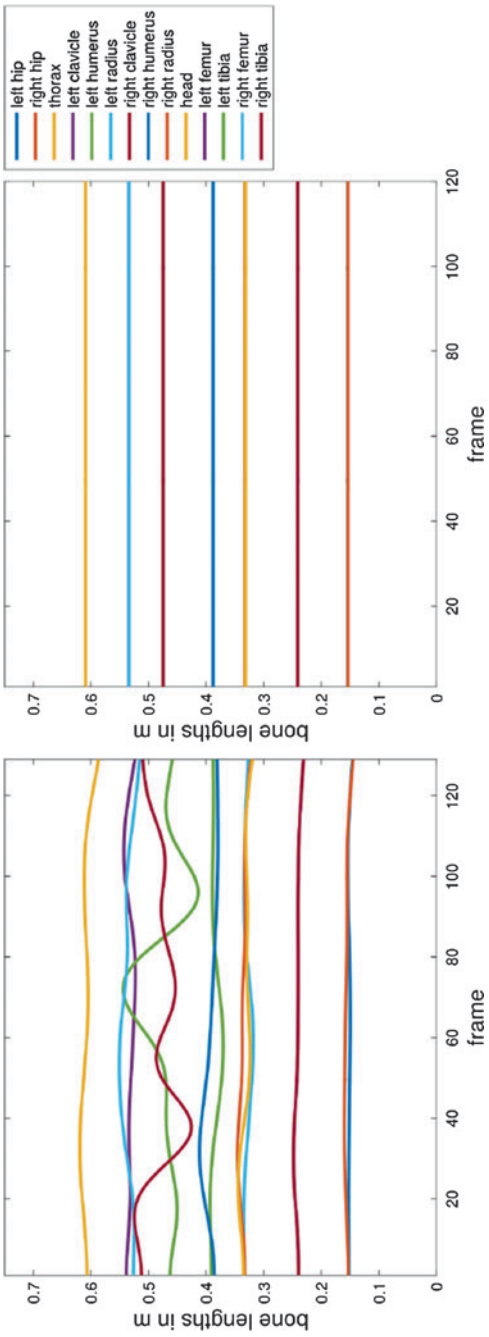

Fig. 3 Comparison of the temporal behavior of bone lengths after 3D reconstruction (*left*) and after applying the physical model (*right*). Obviously, the desired bone length constancy is assured

walking, a standard camera calibration technique with known 2D-3D point correspondences can be used to reconstruct the camera parameters.

Experiments on Force Estimation

The joint model is evaluated on a training set consisting of 45 MoCap walking sequences with synchronized force plate data. For each reconstruction, the considered sequence is excluded from the training set. First of all, the estimation of knee torques from video data is assessed. The torque profiles are generated as described in the previous section. Figure 4 shows the mean value of estimated knee torques for all reconstructed gait sequences together with the related standard deviation. The graph on the right-hand side of Fig. 4 displays the inversely calculated knee torque of an example sequence for comparison.

The torque was computed based on force plate data using a bottom-up procedure, i.e., calculating all acting lever arms, joint forces, and torques from the center of pressure (COP) on the ground to the knee joint. Because of this, only the torque during the stance phase is shown. The kinematic chain from the COP to the knee at the swing leg would be too long and consequently the accumulated error would become too high. It can be seen that the estimated torques are consistent for all reconstructed 3D motions and the absolute values are similar to inverse dynamics torques. The maximal extension torque is reached during the second half of the swing phase and not during double support.

This discrepancy is mainly due to the estimated contact model, that is very sensitive to the distance of contact points to the ground. Further error sources are model inaccuracies, e.g., concerning mass distribution, imprecision of the fitted skeleton, and of course the reconstruction error of $\boldsymbol{\theta}$.

To further analyze the joint model regression, vertical GRF are compared to ground truth data in Fig. 5. The absolute values of the extremal points are slightly high, but the overall curve progression resembles ground truth reaction forces. The experiment shows that the joint model provides a sound estimation of unobservable 3D torques from monocular videos without the need for tedious optimization.

To demonstrate the applicability of the algorithm to real-world scenarios, a sequence from the KTH football database (Kazemi et al. 2013) has been reconstructed. This dataset contains multiview sequences of a challenging noisy outdoor scene that shows a football player walking over a playfield. The 3D reconstruction from camera 1 with estimated torques is illustrated in Fig. 6. As expected, the reconstructions from the remaining two cameras yield very similar results with a maximal reconstruction error of 0.05 m.

Future Directions

In this chapter, we gave you an overview over motion, in particular gait analysis based on a physical simulation. We presented basic concepts of physical modeling, parameter optimization, and 3D reconstruction and showed how these methods can

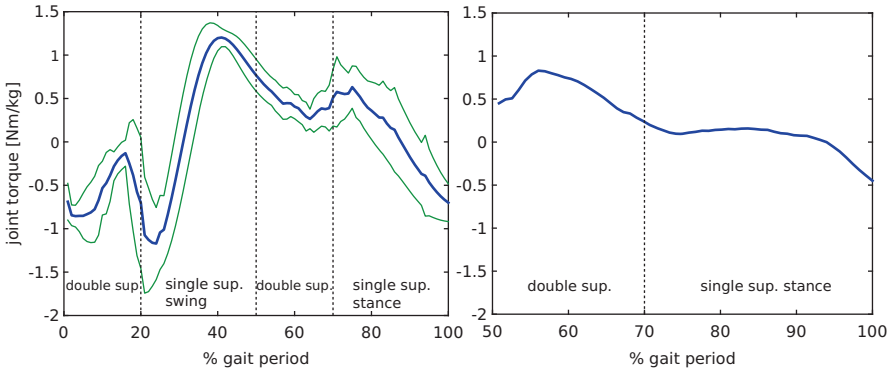

Fig. 4 Consistency of torques: Estimated knee torques for all reconstructed walking sequences (*left*). The torques are shown for a full gait cycle with mean value in *blue* and standard deviation in *green*. On the right-hand side, an example of the knee torque for the stance phase, calculated via inverse dynamics, is displayed

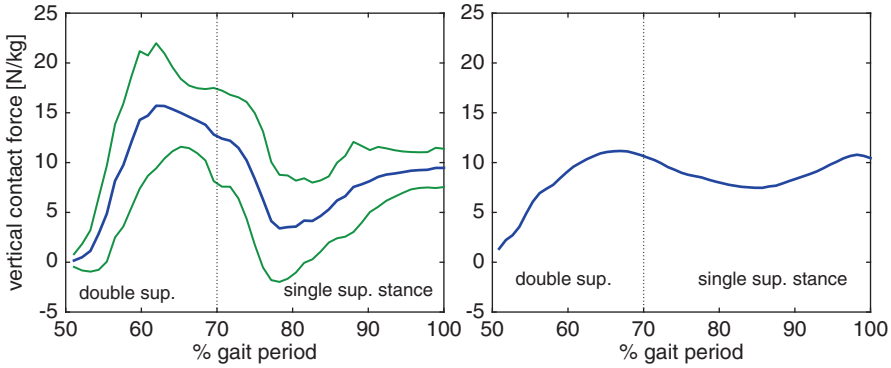

Fig. 5 Consistency of forces: Modeled vertical contact forces (*left*) with mean value in *blue* and standard deviation in *green* for the whole set of gait sequences. The *curve* on the *right*-hand side displays the vertical component of a measured GRF vector

be extended to data-driven learning approaches for force estimation. At this point, there are multiple promising directions for future research.

Considering the video-based gait analysis presented in section "Video-Based Gait Analysis," it is a straightforward conclusion to search for a method that does not combine the 3D reconstruction and the physical simulation in a pipeline as separate steps, but joins them in one model. A further goal in this context is the development of algorithms for handheld devices that would be able to jointly estimate the 3D motion and 3D forces from an integrated monocular camera view. Based on this, short video clips could be directly analyzed using the smartphone that recorded them. For this objective, we require a very robust algorithm concerning camera

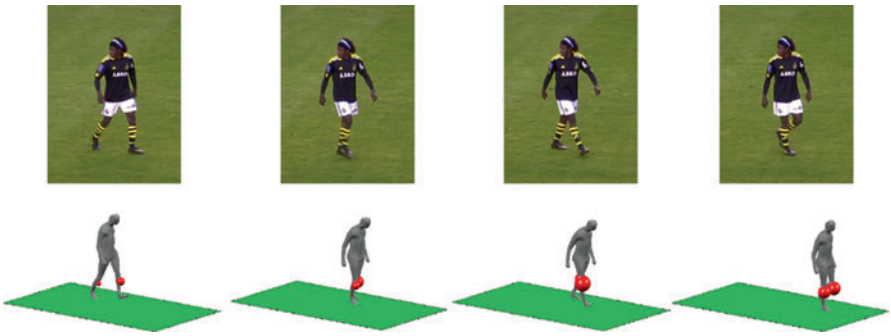

Fig. 6 3D reconstruction and estimated torques of the KTH football dataset. The reconstructions and torques (*red spheres*) appear to be plausible compared to the corresponding images and torques in Fig. 4

motion and an implementation that enables real-time processing, in order to ensure the applicability in everyday life.

Another interesting, sparsely treated research direction is the fusion of different sensor types to facilitate human motion analysis. For example, prior knowledge about the acceleration and the orientation of a subset of body segments could be supplied by including inertial measurement units (IMU). This way, the search space for underlying forces could be reduced significantly. In addition, when considering a muscular skeletal model, the fusion between video data and electromyograph (EMG) measurements might provide vital constraints on the high-dimensional system.

References

Akhter I, Black MJ (2015) Pose-conditioned joint angle limits for 3D human pose reconstruction. In: IEEE Conference on computer vision and pattern recognition (CVPR 2015). IEEE, pp 1446–1455

Al-Naser M, Söderström U (2012) Reconstruction of occluded facial images using asymmetrical principal component analysis. Integrated Comput Aided Eng 19(3):273–283

Bhat KS, Seitz SM, Popović J, Khosla PK (2002) Computer vision – ECCV 2002: 7th European conference on computer vision copenhagen, Denmark, 2002. In: Proceedings, Part I, chapter computing the physical parameters of rigid-body motion from video, Springer, Berlin/Heidelberg, pp 551–565, 28–31 May 2002

Blajer W, Dziewiecki K, Mazur Z (2007) Multibody modeling of human body for the inverse dynamics analysis of sagittal plane movements. Multibody Sys Dyn 18(2):217–232

Bregler C, Hertzmann A, Biermann H (2000) Recovering non-rigid 3D shape from image streams. In: IEEE Conference on computer vision and pattern recognition (CVPR). IEEE, pp 690–696

Brubaker MA, Fleet DJ (2008) The kneed walker for human pose tracking. In: IEEE conference on, computer vision and pattern recognition, 2008 (CVPR 2008). pp 1–8, June 2008

Brubaker MA, Sigal L, Fleet DJ (2009) Estimating contact dynamics. In: IEEE 12th international conference on Computer vision 2009. IEEE, pp 2389–2396

Dai Y, Li H (2012) A simple prior-free method for non-rigid structure-from-motion factorization. In: Conference on computer vision and pattern recognition (CVPR), CVPR'12, IEEE Computer Society, Washington DC, pp 2018–2025, 2012

Fang AC, Pollard NS (2003) Efficient synthesis of physically valid human motion. ACM Trans Graph 22(3):417–426

Fregly BJ, Reinbolt JA, Rooney KL, Mitchell KH, Chmielewski TL (2007) Design of patient-specific gait modifications for knee osteoarthritis rehabilitation. IEEE Trans Biomed Eng 54 (9):1687–1695

Hamsici O, Gotardo P, Martinez A (2011) Learning spatially-smooth mappings in non-rigid structure from motion. In: European conference on computer vision (ECCV). Springer, Berlin/Heidelberg

Johnson L, Ballard DH (2014) Efficient codes for inverse dynamics during walking. In: Proceedings of the twenty-eighth AAAI press conference on artificial intelligence, AAAI'14. AAAI Press, pp 343–349

Kazemi V, Burenius M, Azizpour H, Sullivan J (2013) Multi-view body part recognition with random forests. In: British machine vision conference (BMVC). BMVC Press, Bristol

Liu CK, Hertzmann A, Popović Z (2005) Learning physics-based motion style with nonlinear inverse optimization. ACM Trans Graph 24(3):1071–1081

Mayers D, Sli E (2003) An introduction to numerical analysis. Cambridge University Press, Cambridge

Park HS, Shiratori T, Matthews I, Sheikh Y (2010) 3D reconstruction of a moving point from a series of 2D projections. In: European conference on computer vision (ECCV). Springer, Berlin/Heidelberg

Powell MJD (1978) Numerical analysis. In: Proceedings of the Biennial Conference held at Dundee, chapter A fast algorithm for nonlinearly constrained optimization calculations, Springer, Berlin/Heidelberg, pp 144–157, June 28–July 1 1977

Powers CM (2010) The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. J Orthop Sports Phys Ther 40(2):42–51

Ramakrishna V, Kanade T, Sheikh YA (2012) Reconstructing 3D human pose from 2D image landmarks. In: European conference on computer vision (ECCV). Springer, Berlin/Heidelberg

Safonova A, Hodgins JK, Pollard NS (2004) Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. ACM Trans Graph 23(3):514–521

Schmalz T, Blumentritt S, Jarasch R (2002) Energy expenditure and biomechanical characteristics of lower limb amputee gait: the influence of prosthetic alignment and different prosthetic components. Gait Posture 16(3):255–263

Schwab AL, Delhaes GMJ (2009) Lecture notes multibody dynamics B, wb1413

Sok KW, Kim M, Lee J (2007) Simulating biped behaviors from human motion data. ACM Trans Graph 26(3):107:1–107:9

Spong M, Hutchinson S, Vidyasagar M (2005) Robot modeling and control. Wiley

Steinparz F (1985) Co-ordinate transformation and robot control with denavit-hartenberg matrices. J Microcomput Appl 8(4):303–316

Stelzer M, von Stryk O (2006) Efficient forward dynamics simulation and optimization of human body dynamics. ZAMM – J Appl Math Mech/Zeitschrift fr Angewandte Mathematik und Mechanik 86(10):828–840

Tomasi C, Kanade T (1992) Shape and motion from image streams under orthography: a factorization method. Int J Comput Vis 9:137–154

Torresani L, Hertzmann A, Bregler C (2003) Learning non-rigid 3D shape from 2D motion. In: Thrun S, Saul LK, Schölkopf B (eds) Neural information processing systems (NIPS). MIT Press, Cambridge, MA

Torresani L, Hertzmann A, Bregler C (2008) Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors. In: IEEE Transactions pattern analysis and machine intelligence, IEEE, 21 March 2008

Troje NF (2002a) Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. J Vis 2(5):371–387

Troje NF (2002b) The little difference: Fourier based synthesis of gender-specific biological motion. AKA Press, Berlin, pp 115–120

Vondrak M, Sigal L, Jenkins OC (2008) Physical simulation for probabilistic motion tracking. In: IEEE conference on computer vision and pattern recognition, 2008 (CVPR 2008), pp 1–8, June 2008. IEEE

Wandt B, Ackermann H, Rosenhahn B (2015) 3d human motion capture from monocular image sequences. In: IEEE conference on computer vision and pattern recognition workshops, June 2015. IEEE

Wandt B, Ackermann H, Rosenhahn B (2016) 3d reconstruction of human motion from monocular image sequences. IEEE Trans Pattern Anal Mach Intell 38(8):1505–1516

Wang C, Wang Y, Lin Z, Yuille A, Gao W (2014) Robust estimation of 3d human poses from a single image. In: IEEE Conference on computer vision and pattern recognition (CVPR). IEEE

Wei X, Min J, Chai J (2011) Physically valid statistical models for human motion generation. ACM Trans Graph 30(3):19:1–19:10

Wren CR, Pentland AP (1998) Dynamic models of human motion. In: Proceedings of the third IEEE internatonal conference on automatic face and gesture recognition, Nara, April 1998.

Xiang Y, Chung H-J, Kim JH, Bhatt R, Rahmatalla S, Yang J, Marler T, Arora JS, Abdel-Malek K (2010) Predictive dynamics: an optimization-based novel approach for human motion simulation. Struct Multidiscip Optim 41(3):465–479

Zell P, Rosenhahn B (2015) Pattern recognition: 37th German conference, GCPR 2015. In: Proceedings, chapter A physics-based statistical model for human gait analysis, Springer International Publishing, Aachen, Germany, October 7–10, 2015, pp 169–180.

Zordan VB, Majkowska A, Chiu B, Fast M (2005) Dynamic response for motion capture animation. ACM Trans Graph 24(3):697–701