InterPhys: Physics-aware Human Motion Synthesis in a Dynamic Scene

Supplementary Material

1. Supplementary Material Overview

The supplementary material provides details about optimization and more visualization results. Specifically, Sec. 2 describes the construction of *Phys-SMPL* and *Phys-Object*, where we extend geometric meshes with physical attributes such as mass, inertia tensors, and volumes. Sec. 3 derives the explicit forms of the generalized mass matrix $\mathbf{M}$, gravitational force $\mathbf{G}$, and bias force term $\mathbf{C}$ for both articulated human bodies and rigid objects. Sec. 4 details our two-stage dynamic optimization procedure for estimating physics coefficients and joint torques consistent with motion sequences. Sec. 5 explains the selection of candidate contact points on the body and hands, which form the basis for our continuous contact model. Finally, we provide additional synthesized motion results on OMOMO and Trumans datasets.

2. Phys-SMPL and Phys-Object Construction

2.1. Phys-SMPL: Physics-aware Human Body

The SMPL model provides the human geometry as a triangulated mesh but does not encode physical properties. To equip SMPL with physics information, we compute body part volumes, masses, and inertia tensors following [6]. We follow Physpt [6] to define the Physics-aware Human Body.

Body part volume. Each body part mesh is closed along its boundary to form a watertight surface. Each triangle, together with the body part centroid, defines a tetrahedron. The total volume of the i-th body part is

$$V_i = \sum_{j=1}^{n_j} \left| \det(\mathbf{P}_{i,j,1}, \mathbf{P}_{i,j,2}, \mathbf{P}_{i,j,3}) \right|, \tag{1}$$

where $\mathbf{P}_{i,j,1}, \mathbf{P}_{i,j,2}, \mathbf{P}_{i,j,3}$ are the vertices of the j-th triangle.

Body part mass. For the mean SMPL shape, the total mass is set to 70 kg. Mass is distributed across body parts according to anatomical weight distribution tables. For subjects with different shape parameters $\boldsymbol{\beta}$, body part mass is scaled proportionally to the part volume while incorporating density scaling factors to distinguish bone, muscle, and fat.

Body part inertia tensor. Each body part is modeled as a rigid solid with uniform density. The inertia tensor of part i relative to its center of mass is

$$\mathbf{I}_i = \frac{m_i}{V_i} \iiint_{(x,y,z)\in S_i} \mathbf{f}(x,y,z)\, dx\, dy\, dz, \tag{2}$$

$$\mathbf{f}(x,y,z) = \begin{bmatrix} y^2+z^2 & -xy & -xz \\ -xy & x^2+z^2 & -yz \\ -xz & -yz & x^2+y^2 \end{bmatrix}. \tag{3}$$

The integral is evaluated as a sum over tetrahedra of the part:

$$\frac{m_i}{V_i} \sum_{j=1}^{n_j} \iiint_{(x,y,z)\in \mathcal{S}_{i,j}} f(x,y,z)\, dx\, dy\, dz. \tag{4}$$

Each tetrahedron integral is computed in closed form based on its vertices.

2.2. Phys-Object: Physics-aware Rigid Object

For each object mesh $\mathcal{M}_o$, the same procedure is applied.

Object volume and centroid. The object mesh is closed and decomposed into tetrahedra. The total volume and centroid are

$$V_o = \sum_{j=1}^{n_o} v_{o,j}, \qquad \mathbf{c}_o = \frac{1}{V_o} \sum_{j=1}^{n_o} v_{o,j}\, \mathbf{c}_{o,j}, \tag{5}$$

where $v_{o,j}$ and $\mathbf{c}_{o,j}$ are the volume and centroid of the j-th tetrahedron.

Object mass. With uniform density ρ_o, the object mass is $m_o = \rho_o V_o$.

Object inertia tensor. The inertia tensor relative to the object centroid is

$$\mathbf{I}_o = \frac{m_o}{V_o} \sum_{j=1}^{n_o} \iiint_{(x,y,z)\in \mathcal{S}_{o,j}} f(x,y,z)\, dx\, dy\, dz, \tag{6}$$

where $f(x,y,z)$ follows the same form as in the human body part case.

3. Derivation of M, G, and C Terms

We detail the explicit forms of the generalized mass matrix $\mathbf{M}$, the gravitational force $\mathbf{G}$, and the bias force term $\mathbf{C}$ for both the articulated human body (Phys-SMPL) and the manipulated object (Phys-Object). For Phys-SMPL, we follow the formulation introduced in PhysPT [6].

3.1. Human Body (Phys-SMPL)

Let $\mathbf{q}$ denote the generalized pose of the SMPL model. Each body part $n \in \{1, \ldots, 24\}$ is associated with mass m_n, inertia tensor $\mathbf{I}_n$, world transformation $\mathbf{R}_{0,n}$, translational Jacobian $\mathbf{J}_{S,n}$, and rotational Jacobian $\mathbf{J}_{R,n}$.

Generalized Mass Matrix.

$$\mathbf{M}_h(\mathbf{q}) \;=\; \sum_{n=1}^{24} \mathbf{J}_{S,n}^{\top} m_n \mathbf{J}_{S,n} \;+\; \mathbf{J}_{R,n}^{\top} \mathbf{R}_{0,n} \mathbf{I}_n \mathbf{R}_{0,n}^{\top} \mathbf{J}_{R,n}. \tag{7}$$

where $\mathbf{J}_{S,n} \in \mathbb{R}^{3\times 75}$ maps generalized velocities to the linear velocity of part n, $m_n \in \mathbb{R}$ is the body part mass, $\mathbf{J}_{R,n} \in \mathbb{R}^{3\times 75}$ maps generalized velocities to the angular velocity of part n, $\mathbf{R}_{0,n} \in SO(3)$ is the world orientation of part n, and $\mathbf{I}_n \in \mathbb{R}^{3\times 3}$ is the inertia tensor of part n in its local frame.

Gravitational Term.

$$\mathbf{G}_h(\mathbf{q}) \;=\; -\sum_{n=1}^{24} \mathbf{J}_{S,n}^{\top} m_n \mathbf{g}, \tag{8}$$

where $\mathbf{g} = [0, 0, -9.81]^{\top}$ is the gravitational acceleration vector in the world frame.

Bias Force Term.

$$\begin{aligned} \mathbf{C}_h(\mathbf{q}, \dot{\mathbf{q}}) \;&=\; \sum_{n=1}^{24} \mathbf{J}_{S,n}^{\top} m_n \dot{\mathbf{J}}_{S,n} \dot{\mathbf{q}} \\ &+\; \mathbf{J}_{R,n}^{\top} \Big(\mathbf{R}_{0,n} \mathbf{I}_n \mathbf{R}_{0,n}^{\top} \dot{\mathbf{J}}_{R,n} \dot{\mathbf{q}} \\ &+\; \mathbf{J}_{R,n} \dot{\mathbf{q}} \times \mathbf{R}_{0,n} \mathbf{I}_n \mathbf{R}_{0,n}^{\top} \mathbf{J}_{R,n} \dot{\mathbf{q}} \Big), \end{aligned} \tag{9}$$

where $\dot{\mathbf{J}}_{S,n}$ and $\dot{\mathbf{J}}_{R,n}$ are the time derivatives of the translational and rotational Jacobians, respectively, and $\times$ denotes the cross product operator.

3.2. Rigid Object (Phys-Object)

Let $\mathbf{q}_o = [\mathbf{r}_o; \mathbf{R}_o]$ denote the object state with COM position $\mathbf{r}_o \in \mathbb{R}^3$ and orientation $\mathbf{R}_o \in SO(3)$. The object has mass m_o and inertia tensor at the COM $\mathbf{I}_{C,o}$.

Generalized Mass Matrix.

$$\mathbf{M}_o(\mathbf{q}_o) = \begin{bmatrix} m_o \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_o \mathbf{I}_{C,o} \mathbf{R}_o^{\top} \end{bmatrix}, \tag{10}$$

where $\mathbf{I}_3$ is the 3×3 identity matrix.

Gravitational Term.

$$\mathbf{G}_o(\mathbf{q}_o) = \begin{bmatrix} m_o \mathbf{g} \\ \mathbf{0} \end{bmatrix}, \tag{11}$$

where $m_o \mathbf{g}$ is the linear gravitational force acting on the COM, and no direct gravitational torque is applied.

Bias Force Term.

$$\mathbf{C}_o(\mathbf{q}_o, \dot{\mathbf{q}}_o) = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\omega}_o \times \left(\mathbf{R}_o \mathbf{I}_{C,o} \mathbf{R}_o^{\top} \boldsymbol{\omega}_o \right) \end{bmatrix}, \tag{12}$$

where $\boldsymbol{\omega}_o \in \mathbb{R}^3$ is the object angular velocity in the world frame.

4. Two-Stage Dynamic Optimization

4.1. Stage 1: Hand–Object Force Optimization

For each frame t, the object dynamics in Eq.3 in the main paper is used to recover the contact forces between the hands and the object. The optimization variables are the physics coefficients $\mathbf{b}_t \in \mathbb{R}^{C_o \times 4}$, where each entry corresponds to $\{\kappa, \delta, \rho, \mu\}$ at a hand contact point. The contact forces are represented as

$$\boldsymbol{\lambda}_o(\mathbf{b}_t) = \mathbf{A}_t^o \mathbf{b}_t, \tag{13}$$

where $\mathbf{A}_t^o \in \mathbb{R}^{3C_o \times 4C_o}$ encodes contact geometry terms (surface normals $\mathbf{n}(\mathbf{x})$, tangential directions, and relative velocities). The optimization is formulated as

$$\min_{\mathbf{b}_t} \quad \left\| \mathbf{M}_o(\bar{\mathbf{q}}_o^t) \ddot{\bar{\mathbf{q}}}_o^t + \mathbf{C}_o(\bar{\mathbf{q}}_o^t, \dot{\bar{\mathbf{q}}}_o^t) + \mathbf{G}_o(\bar{\mathbf{q}}_o^t) + \mathbf{J}_o^{\top} \mathbf{A}_t^o \mathbf{b}_t \right\|_2^2, \tag{14}$$

subject to element-wise box constraints

$$\bar{\mathbf{b}}_{\min} < \mathbf{b}_t < \bar{\mathbf{b}}_{\max}. \tag{15}$$

The resulting $\mathbf{b}_t$ gives the optimized physics coefficients, and $\boldsymbol{\lambda}_o(\mathbf{b}_t)$ are the hand–object forces consistent with the object trajectory.

4.2. Stage 2: Full-Body and Scene Optimization

The recovered hand–object forces are then introduced into the human dynamics in Eq.2 in the main paper. The optimization variables are the physics coefficients $\mathbf{a}_t \in \mathbb{R}^{C_s \times 4}$ for scene contact points and the internal joint torques $\boldsymbol{\tau}_t \in \mathbb{R}^{75}$. The human–scene contact forces are represented as

$$\boldsymbol{\lambda}_s(\mathbf{a}_t) = \mathbf{A}_t^s \mathbf{a}_t, \tag{16}$$

where $\mathbf{A}_t^s \in \mathbb{R}^{3C_s \times 4C_s}$ encodes scene geometry and velocity-dependent terms. The optimization is formulated as

$$\begin{aligned} \min_{\mathbf{a}_t, \boldsymbol{\tau}_t} \quad & \Big\| \mathbf{M}_h(\bar{\mathbf{q}}_t) \ddot{\bar{\mathbf{q}}}_t + \mathbf{C}_h(\bar{\mathbf{q}}_t, \dot{\bar{\mathbf{q}}}_t) + \mathbf{G}_h(\bar{\mathbf{q}}_t) \\ & - \mathbf{J}_{hs}^{\top} \mathbf{A}_t^s \mathbf{a}_t - \mathbf{J}_{ho}^{\top} \mathbf{A}_t^o \mathbf{b}_t - \boldsymbol{\tau}_t \Big\|_2^2, \end{aligned} \tag{17}$$

subject to

$$\bar{\mathbf{a}}_{\min} < \mathbf{a}_t < \bar{\mathbf{a}}_{\max}. \tag{18}$$

$$\bar{\boldsymbol{\tau}}_{t\min} < \boldsymbol{\tau}_t < \bar{\boldsymbol{\tau}}_{t\max}. \tag{19}$$

Given the ground-truth motion, we solve the above optimization problems using CVXPY following the formulation of [6].

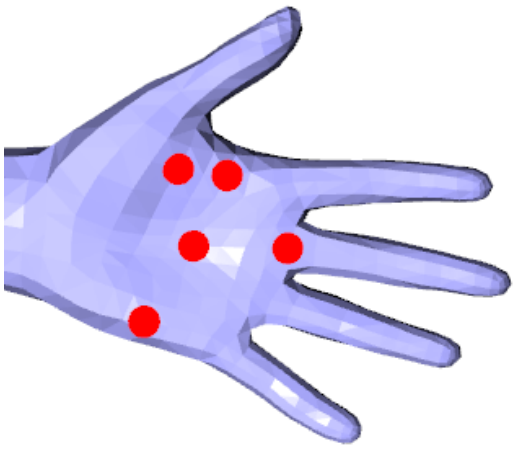

Figure 1. Selection of hand contact vertices. Five vertices are chosen per hand, including the palm center and fingertips.

5. Selection of Contact Points

5.1. Body Contact Points

For the body, we directly follow PhysPT [6], which predefines a set of candidate ground-contact vertices based on contact frequency analysis on AMASS sequences. We adopt exactly the same set of body contact points as in PhysPT without modification.

5.2. Hand Contact Points

For the hands, we manually select 5 vertices for each hand from the MANO mesh. The selected points correspond to the palm center and the fingertips, which are the most likely regions to contact with objects during interaction. These vertices are fixed across all sequences and serve as candidate hand–object contact points. The selected vertices are illustrated in Figure 1.

6. More Visualization

We provide more qualitative results to demonstrate the generality of our approach across different human–object interaction benchmarks. We provide additional qualitative comparisons on OMOMO in Figure 2 and Trumans in Figure 3.

References

[1] Nan Jiang, Zhiyuan Zhang, Hongjie Li, Xiaoxuan Ma, Zan Wang, Yixin Chen, Tengyu Liu, Yixin Zhu, and Siyuan Huang. Scaling up dynamic human-scene interaction modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1737–1747, 2024. 4

[2] Jiaman Li, Jiajun Wu, and C Karen Liu. Object motion guided human motion synthesis. *ACM Transactions on Graphics (TOG)*, 42(6):1–11, 2023. 4

[3] Jiaman Li, Alexander Clegg, Roozbeh Mottaghi, Jiajun Wu, Xavier Puig, and C. Karen Liu. Controllable human-object interaction synthesis. In *ECCV*, 2024. 4

[4] Sirui Xu, Zhengyuan Li, Yu-Xiong Wang, and Liang-Yan Gui. InterDiff: Generating 3d human-object interactions with physics-informed diffusion. In *ICCV*, 2023. 4

[5] Sirui Xu, Dongting Li, Yucheng Zhang, Xiyan Xu, Qi Long, Ziyin Wang, Yunzhi Lu, Shuchang Dong, Hezi Jiang, Akshat Gupta, et al. Interact: Advancing large-scale versatile 3d human-object interaction generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7048–7060, 2025. 4

[6] Yufei Zhang, Jeffrey O Kephart, Zijun Cui, and Qiang Ji. Physpt: Physics-aware pretrained transformer for estimating human dynamics from monocular videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2305–2317, 2024. 1, 2, 3

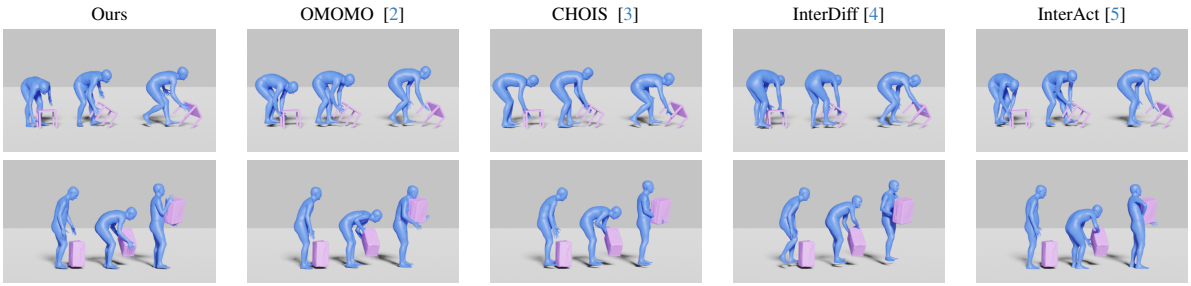

Figure 2. Qualitative comparison on OMOMO. From left to right: ground truth, our prediction, and predictions from OMOMO, CHOIS, InterDiff, and InterAct.

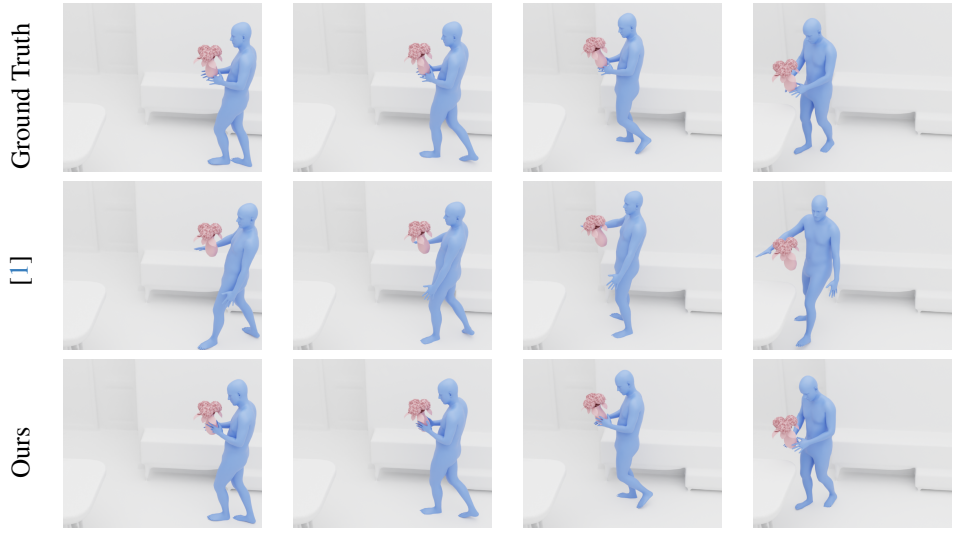

Figure 3. Qualitative comparison on Trumans [1] Dataset. Each row shows ground truth, the Turmans baseline, and our method.