arXiv:2611.02393v1 [cs.CV] 2 Nov 2026

IMUdrape: Generating Garment-Mounted Virtual IMU Data from Text through Physics-Based Cloth Simulation

ZIKANG LENG, School of Interactive Computing, Georgia Institute of Technology, USA
THOMAS PLÖTZ, School of Interactive Computing, Georgia Institute of Technology, USA

Wearable computing was envisioned as computation woven into what people literally wear, and inertial sensors are now moving into clothing itself, from sensorized sportswear to garments instrumented with hundreds of IMUs. Sewn into clothing, a sensor can be comfortable and unobtrusive enough for continuous use, but it no longer moves rigidly with the body. Loose fabric slides, twists and swings, so a garment-mounted sensor records the motion of the fabric as well as that of the body, and this extra motion differs from one garment to the next. A model for a garment-mounted sensor therefore needs training data recorded with the garment it is deployed on, yet such labeled data are scarce, as they are for sensor-based activity recognition in general. For body-attached sensors, virtual IMU data generated from videos or text have eased this scarcity. Existing generators, however, attach their virtual sensors rigidly to the body, so they cannot reproduce how clothing moves. We present IMUdrape, which closes this gap by generating garment-mounted virtual IMU data from text through physics-based clothes simulation. IMUdrape synthesizes human motion from activity descriptions with a text-to-motion model, reconstructs the target deployment's garment as a sewing pattern from photographs and dresses the body in it, places virtual IMUs at the deployment's real sensor positions on the garment, and simulates the garment with an intersection-free contact solver, so that every simulation yields paired garment and body signals for the same motion. On four real datasets with clothing-mounted sensors, IMUdrape improves on existing virtual-data approaches for both activity recognition and full-body pose estimation. With one real subject, its data give a relative improvement in recognition macro-F1 of 7% to 46% over real data alone and of 2% to 11% over body-mounted virtual data from the same motion, and they remain ahead once all real subjects are used. When the garment is simulated on recorded motion, the pose error from loose-garment IMUs without any real training data falls by 30% to 33% relative to generated motion, and garment data remain 12% more accurate than body data. IMUdrape thus reduces the amount of real data that garment-mounted sensing requires, and it lays the groundwork for researchers who want to study smart-clothing sensing but have no recorded data of their own.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → *Machine learning*; *Physical simulation.*

Additional Key Words and Phrases: Virtual IMU, Smart Clothing, Garment-Mounted Sensors, Cloth Simulation, Human Activity Recognition, Wearables

1 Introduction

Body-worn inertial measurement units (IMUs) are one of the central sensing modalities of mobile and ubiquitous computing, supporting human activity recognition (HAR), full-body pose estimation and health monitoring [14]. Increasingly, these sensors are built into clothing, from commercial smart shirts and motion-capture suits [28, 65, 79] to research garments that carry one to hundreds of IMUs [16, 19, 37, 54, 63, 92], because clothing is comfortable, unobtrusive and suited to long-term wear. Most inertial-sensing methods, however, from sparse-IMU pose estimators [30, 100, 101] to generators of virtual IMU data [44, 48, 103], assume that each sensor is fixed tightly to the body segment it measures, typically with straps [110]. A garment-mounted sensor instead moves

Authors' Contact Information: Zikang Leng, zleng7@gatech.edu, School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, USA; Thomas Plötz, thomas.ploetz@gatech.edu, School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, USA.



This work is licensed under a Creative Commons Attribution 4.0 International License.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

with the fabric, which slides, rotates, folds and swings, so its signal depends on the garment's cut, the fabric, the sensor's position and the activity [33, 57, 110].

This garment dependence sharpens the problem that limits all of sensor-based HAR, the lack of large, labeled datasets. Data recorded with one garment may not generalize to another, even at the same body position and for the same activity [33, 57], so a new garment may need new recordings. For body-worn sensors, virtual IMU data derived from videos [44, 78, 83, 106], motion capture [30, 93, 107] or text [23, 46, 48] have eased this scarcity. For garment-mounted sensors, existing approaches fall short. Virtual IMU generators place their sensors rigidly on the body, so their data describe a body-mounted sensor rather than the garment-mounted sensor that is deployed. Denoisers that map garment signals to body signals are learned from *paired* garment and body recordings [19, 57], which require laborious data collection for every garment. Learned garment models [71, 110] need no recordings, but they predict the garment's shape from the pose without simulating its dynamics and cannot be matched to an arbitrary target garment.

In this paper we present IMUdrape, a pipeline that generates *garment-mounted* virtual IMU data from text through physical simulation of the garment. IMUdrape generates human motion from a textual activity description, dresses an SMPL body, a standard parametric model of the human body [56], in a garment reconstructed to match the clothing of a target deployment [9, 41], places virtual IMUs at that deployment's real sensor positions, and simulates the garment with an intersection-free contact solver [3, 49]. The virtual IMU signals are computed from small rigid patches of the simulated fabric at each sensor position. Because both the motion and the garment are generated, IMUdrape can produce data for a wide variety of motions and garments. Every simulation also yields *paired* data, IMU signals on the body and on the garment for the same motion, which separates the effect of the garment from that of the motion.

We evaluate IMUdrape on four real datasets with clothing-mounted sensors, for activity recognition [19, 92, 108, 110] and pose estimation [19, 108, 110], with three recognition models and two pose networks. We compare it with real data alone, with body-mounted virtual IMU data from IMUGPT [46, 48] and with garment data from TailorNet [71], a learned garment model without dynamics. IMUdrape itself provides three sources, the body signal (IMUdrape Body), the garment signal (IMUdrape Garment) and clips drawn from either (IMUdrape Body + Garment).

IMUdrape data are effective in both tasks. For activity recognition with one real subject, IMUdrape Body + Garment improves macro-F1 by a relative 7.2% to 46.3% over real data alone across the four datasets, and by 2.1% to 44.9% over the other virtual sources, IMUGPT, TailorNet and IMUdrape Body. Even with all real subjects, the best IMUdrape source remains 1.0% to 5.5% above real data alone. For pose estimation without real data, IMUdrape data give the lowest joint angle error on all three datasets, 7.8% to 11.5% below IMUGPT and 3.2% to 10.0% below TailorNet. When the garment is simulated on GarMoCap's recorded instead of generated motion, this error falls by a further 29.7% to 33.3%, and IMUdrape Garment stays 12.1% below IMUdrape Body. Across both tasks, garment physics matters most when real data are scarce and when the simulated motion is realistic.

Our contributions are as follows:

- We present IMUdrape, a pipeline that generates garment-mounted virtual IMU data from text through physics-based cloth simulation, together with garments and sensor maps matched to the clothing and sensor layouts of four public datasets.
- We conduct a large-scale evaluation of garment-mounted virtual IMU data for activity recognition and pose estimation, against body-mounted and learned-garment virtual data, and characterize when garment physics matters: with scarce real data and with realistic motion.
- We provide empirical insights for practitioners on how much and which real and virtual data to collect, where on a garment and with how many sensors physics matters most, and how realistic motion and a matched garment add to the benefit.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 1. Overview of approaches that generate virtual inertial data, and how they model the attachment of the sensor. Sensor: where the virtual sensor sits; Body: rigidly attached to a body segment; Garment: on modeled clothing. Dynamics: the garment's motion relative to the body is simulated. Matched: the garment can be matched to the clothing of a target deployment.

System	Motion source	Sensor	Garment model	Dynamics	Matched	Tasks
IMUTube [44, 45]	Video	Body	–	–	–	HAR
Rey et al. [78]	Video	Body	–	–	–	HAR
IMUGPT [46, 48]	Text	Body	–	–	–	HAR
UniMTS [107]	Motion capture	Body	–	–	–	HAR
MOSAIC [96]	Text	Body	–	–	–	HAR
LIP synthetic data [110]	Motion capture	Garment	Learned (TailorNet)	–	–	Pose
GaIP [31]	Motion capture	Garment	Learned (TailorNet) + diffusion	–	–	Pose
Haeusler et al. [22]	Text	Body, garment	Learned (TailorNet)	–	–	HAR
IMUdrape (ours)	Text or recorded motion	Body, garment	Sewing pattern + physics	yes	yes	HAR, pose

2 Related Work

Sensor-based human activity recognition (HAR) with body-worn inertial sensors is one of the central pillars of ubiquitous computing [11, 14, 26]. Despite advances in deep [1, 24, 66, 69, 109] and self-supervised learning [25, 80, 89, 99] and in domain adaptation [12], its most persistent obstacle remains the lack of labeled training data, because recording and annotating body-worn sensor data is expensive, time-consuming and raises privacy concerns [72, 73]. Generating virtual training data through cross-modality transfer [39] is a promising response, and the one we follow. Table 1 positions IMUdrape relative to the most closely related virtual-data approaches.

2.1 Cross-Modality Transfer for Human Activity Recognition

Cross-modality transfer approaches address data scarcity by converting data from a modality that is plentiful into training data for a sensing modality that is not, most often inertial sensors. IMUTube [44, 45] extracts three-dimensional human motion from online videos with pose estimation and computes virtual accelerometry at body locations, and Rey et al. [78] derive training data for wearable activity recognition from monocular videos in a similar way. Follow-up work improved the quality and calibration of video-derived data, showed that large virtual corpora enable more complex recognition models [43, 45], and found that their benefit shrinks for fine-grained activities [47]. Accelerometry has also been synthesized from video with generative models [106], video-derived data have supported eating detection, exercise and sign-language recognition [32, 83, 98], and the same principle has produced radar data [2, 7] and pre-training data from motion-capture archives [60, 107] and image corpora [102].

With the emergence of large language models and text-driven motion synthesis [21, 35, 91, 104, 105], language has become a source modality as well. IMUGPT [46, 48] uses a language model to write diverse descriptions of how an activity can be performed, converts them into motion sequences with a text-to-motion model [104], and derives virtual IMU data with a biomechanical sensor simulator [103]. Text2IMU [23] animates body-surface models of different morphologies with text-generated motion, and Wonderwall [61] pre-trains a foundation model on physics-based simulated motion before refining it on real data. Language has also been fused with sensor data for zero-shot recognition [87], and MOSAIC [96] trains a sensor–language model on a synthetic IMU–text corpus generated in the same way as IMUGPT's.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Across these approaches, the motion source varies, but the virtual sensor is rigidly attached to the body, at a joint of the skeleton [103] or on the body surface [93], with WIMUSim [68] adding variability of body shape, sensor placement and hardware (Table 1). This assumption is reasonable for strapped sensors but not for sensors on clothing, whose signal depends on how the garment moves relative to the body. Like IMUGPT, IMUdrape generates motion from text, but it senses that motion through a physically simulated garment matched to the target deployment.

2.2 Garment-Mounted and Loose-Clothing Inertial Sensing

Smart textiles integrate sensing into everyday clothing [15, 86], and garments with embedded inertial sensors have been used to recognize activities and postures and to capture whole-body motion [16, 19, 37, 54, 63, 92]. Where a sensor sits, and how far it shifts, matters even for strapped sensors. A long line of work has studied how the placement and displacement of on-body sensors affect activity recognition [4, 42], typically with benchmark datasets recorded with strapped sensors [13, 77, 88]. Clothing adds a further displacement, one that changes with the motion itself. Paired recordings of garment-attached and strapped sensors show that clothing reduces the correlation between garment and body signals and adds high-frequency content, with effects that depend on the garment and the body location [33, 34, 57, 85]. This additional motion does not necessarily harm recognition. Sensors mounted on fabric distinguish similar movements better than rigidly attached ones and can raise classification accuracy [62], which a probabilistic model explains by the larger statistical distance between movements that fabric motion creates [84]. Earlier work also estimated the posture-recognition performance of sensing garments by modeling garment wrinkles geometrically [27], but such models address static postures and do not generate training data for dynamic activities.

The effect of clothing is most visible in sparse inertial pose estimation, which reconstructs full-body motion from a handful of IMUs [30, 36, 64, 94, 100, 101]. These methods assume that the IMUs are strapped tightly to the body, and they are trained largely on virtual IMU data synthesized from motion capture [60]. Loose Inertial Poser [110] and related work [70, 108] show that their pose error rises markedly when the same IMUs are instead attached to loose garments, because the sensors then rotate and shift relative to the limbs they are meant to track. To close this gap, LIP trains partly on loose-garment data synthesized with TailorNet [71]. TailorNet is a neural network that predicts how a garment of one of a few types drapes for the current body pose and shape. Trained on static poses, it models neither the garment's dynamics nor its contact with the body. GaIP [31] generates loose and tight IMU pairs on motion-capture data, using TailorNet garments for the sensor orientation and a diffusion model for the loose signal. Closest to ours, Haeusler et al. [22] dress SMPL bodies of eight sizes in tight and loose TailorNet garments, a long-sleeve shirt and long pants, animate them with text-generated motion and place virtual IMUs on the body and on the garment. They show that clothing changes the synthesized signals, most for loose garments and fast activities, and that a recognition model trained on garment data transfers better to the RealWorld dataset [88] than one trained on body data. Because TailorNet predicts each frame from the pose, their garments have no inertia or lag, and the authors name physical modeling of gravity and inertia as future work. They also do not match the garment to the clothing of a target deployment, and they address activity recognition only. A different approach removes the garment's effect instead of synthesizing it. GarMoCap [19] learns a garment-IMU denoiser (GID) that maps garment signals to body signals, from paired real recordings. These approaches require real paired data or rely on learned garment shapes without dynamics, and none matches the garment of an arbitrary target deployment. IMUdrape simulates the garment's dynamics from a sewing pattern, so it needs no real paired data and can represent virtually any everyday garment.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

2.3 Garment Modeling and Cloth Simulation

Simulating a specific garment first requires a model of its shape, and sewing patterns provide one. Like a tailor's pattern, they describe a garment as flat fabric panels and the seams that join them. Parametric systems such as GarmentCode [41] describe garments by programmable patterns, and learned models reconstruct such patterns from three-dimensional scans [40] or, with vision–language models such as ChatGarment [9], from photographs, which makes it possible to reconstruct a specific garment; Dress-1-to-3 [51] further refines such patterns with a differentiable simulator into simulation-ready garments from a single image.

Learned garment models predict the shape of clothing directly, from images [8], as an offset from the body surface [58] or for a given body pose [71, 81]. They capture the shape a garment would settle into, but not how it lags, swings and slides. Neural simulators learn these dynamics as well [6, 20, 82], but all of these models are limited to the garments and motions they were trained on, which is why we turn to physics-based simulation.

Physics-based cloth simulation instead computes the motion of the fabric from its material properties, gravity and contact, and has a long history in computer graphics [5, 10, 59, 67]. Incremental Potential Contact (IPC) [49] treats contact as an energy that grows sharply as two surfaces approach, which guarantees that they never pass through each other. Its extension to thin surfaces [50] brings this guarantee to fabric, and GPU solvers of this family [3] make it fast enough for large-scale use. Closer to our setting, simulated loose garments have been used to study how clothing affects optical motion capture [76]. IMUdrape combines sewing-pattern reconstruction with physics-based simulation. It reconstructs each target garment as a GarmentCode pattern [41] with ChatGarment [9] and simulates it with the ZOZO contact solver [3], a GPU solver of the IPC family [49, 50], which keeps the garment intersection-free even in fast motion and in close contact with the body.

3 IMUdrape: Generating Garment-Mounted Virtual IMU Data

IMUdrape turns a textual description of an activity into paired body- and garment-mounted virtual IMU data. Figure 1 gives an overview of the pipeline. Like IMUGPT [46, 48], it starts from language. A large language model writes varied descriptions of each target activity (for example, *"a person jogs at an easy pace, arms swinging loosely"*), and a text-to-motion model turns each description into a 3D motion of the body. We use HY-Motion 1.0 [90], a flow-matching text-to-motion model whose output we map to SMPL [56], a parametric model of the human body surface. Everything after the body motion is new in IMUdrape and is what body-mounted virtual IMU generators lack. *Garment generation* (Section 3.1) builds a simulation-ready garment from a photograph or a description of the target garment. The garment consists of a sewing pattern, whose flat panels are stitched together around the body into a 3D garment, and the mechanical properties of its fabric. *Garment simulation* (Section 3.2) moves the garment physically with the generated body, so that it slides, twists and swings relative to the body as a real garment does. *Garment IMU* (Section 3.3) reads virtual sensors off the simulated garment at any point of the garment, including the sensor positions of a real smart garment, and pairs each with a sensor on the body beneath it. The pipeline runs without manual intervention. Further details of each step and all parameter values are given in Appendix A.

3.1 Garment Generation

Unlike in body-mounted virtual IMU data, the garment is a free variable of IMUdrape. The same motion can be simulated with any garment, so the virtual wearer can be dressed in a garment that matches the deployment. To simulate such a garment, we use a cloth simulator. It needs three inputs: (i) the garment's sewing pattern, which fixes its cut and fit, (ii) the 3D garment on the body, stitched together from the pattern, and (iii) the mechanical properties of its fabric. The following subsections describe how IMUdrape produces each of them.

3.1.1 Sewing-Pattern Estimation. We estimate the sewing pattern with ChatGarment [9], a vision–language model that reads a photograph of a garment, a text description, or both, and outputs the design of the garment.

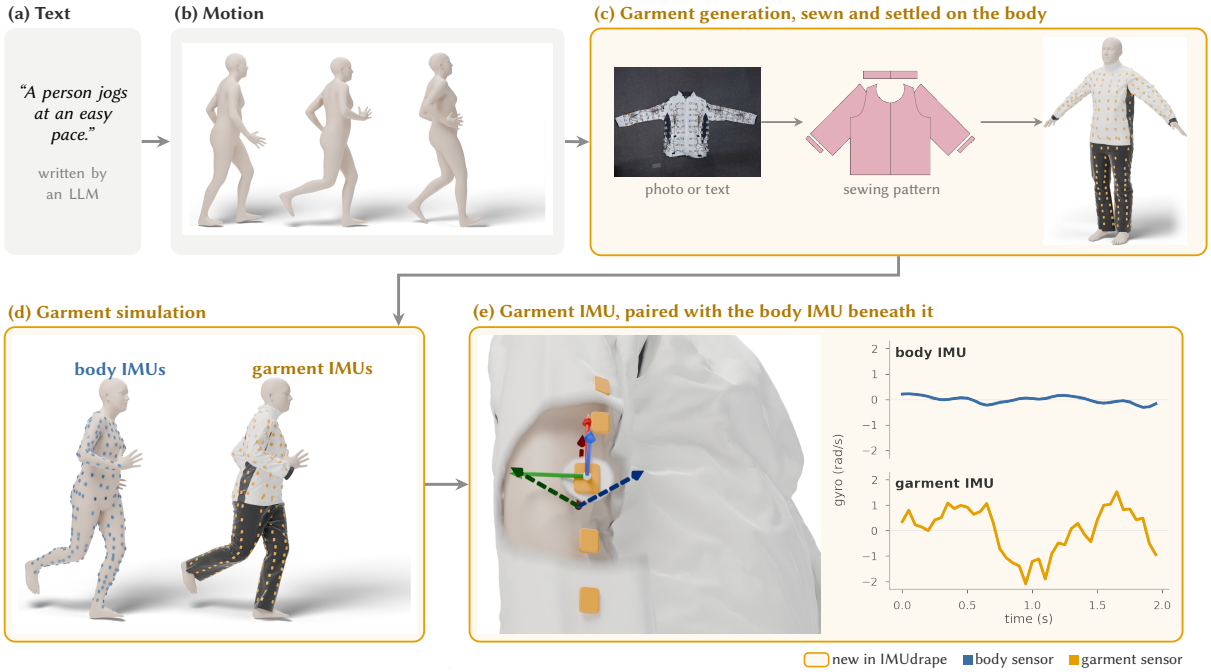


Fig. 1. IMUdrape generates *garment-mounted* virtual IMU data from text. A text-to-motion model turns an activity description into body motion (a, b), as in body-mounted generators. IMUdrape then generates a garment from a photo or a description of the target garment (c), simulates it on the moving body (d), and reads virtual IMUs off the simulated garment at any point of the garment, each paired with an IMU on the body beneath it (e). Orange panels mark the parts that are new in IMUdrape, and blue and orange squares mark body and garment sensors. The garment sensor slides and twists relative to the body. Within 1.6 s, its frame (solid axes) turns from 11° to 42° away from the frame of the body sensor beneath it (dashed axes), so the paired signals of the same motion differ strongly.

The design has two parts. Its discrete choices, such as the garment type, the collar, the cuffs and whether the front is open, are written as text. Its 76 continuous values $\mathbf{n}$, such as lengths, widths and angles, are regressed from the model's internal state by a small network, because language models are poor at producing precise numbers (Appendix A.3). For a photograph, the model first describes the garment in words, part by part, and then outputs the design given both the photograph and this description.

3.1.2 Sewing-Pattern Program. The design θ (Figure 2a–c), which consists of the categorical choices and the continuous values $\mathbf{n}$, is the input of a GarmentCode program [41]. GarmentCode is a programming language designed specifically for sewing patterns. It builds a garment from components, such as a torso, two sleeves and a collar, each with its own parameters, and builds components from panels and panels from edges. The program $\mathcal{G}$ maps the design and a vector $\mathbf{b}$ of body measurements (girths, lengths and heights) to a sewing pattern, a set of panels and stitches,

$$\mathcal{P} = \mathcal{G}(\theta, \mathbf{b}) = \big(\{(\Omega_k, R_k, \mathbf{t}_k)\}_k, \{s_l\}_l\big). \tag{1}$$

Each panel $\Omega_k \subset \mathbb{R}^2$ is a closed outline of straight edges and Bézier curves in the plane of the fabric. The rotation R_k and translation $\mathbf{t}_k$ place the panel in 3D around the body, so that a point $(u, v) \in \Omega_k$ starts at

$$\mathbf{x} = R_k\,(u, v, 0)^\top + \mathbf{t}_k. \tag{2}$$

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

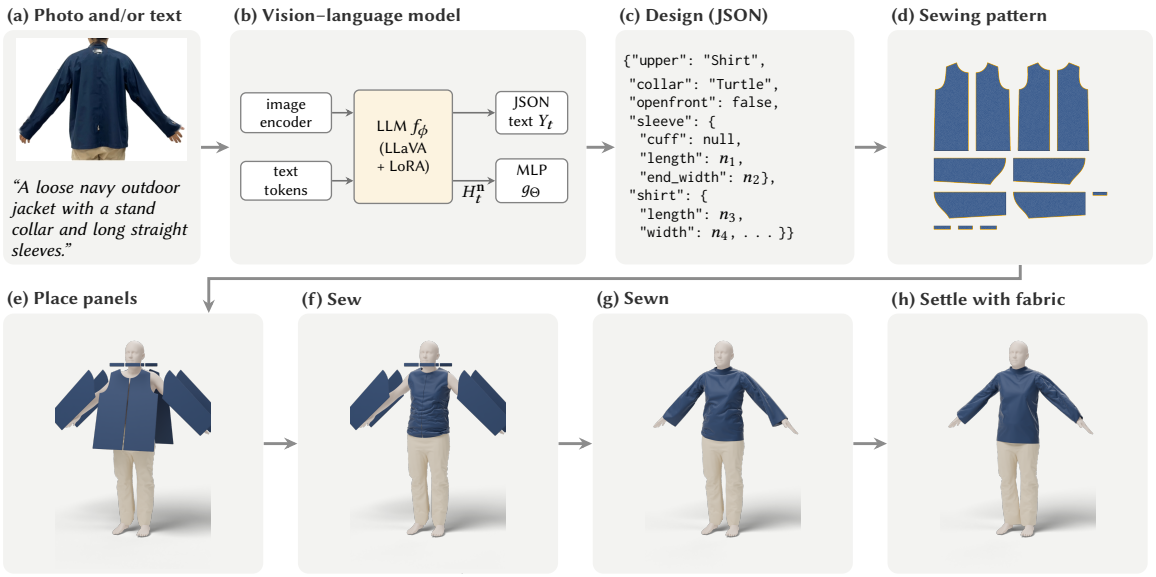


Fig. 2. Garment generation. (a–c) A vision–language model [9] turns a photograph and/or a description of the garment into a design (photograph adapted from LIP [110], annotations removed). The design consists of categorical choices (garment type, collar, cuffs) as JSON text and 76 continuous values (lengths, widths, angles) regressed from the hidden state of an end token. (d) A GarmentCode program [41] turns the design and the body measurements into 2D sewing panels. Orange edges are sewn to other panels. (e–h) The panels are placed around the body, sewn shut in the cloth simulator without inertia or gravity, and settled under gravity with the fabric's properties. The settled garment is the start state of every simulation. The garment shown is the LIP jacket, sewn over trousers.

Each stitch s_l joins an edge of one panel to an edge of another (or the same) panel. Because the measurements $\mathbf{b}$ are computed from the SMPL body mesh, the same design is automatically sized to the simulated body (Figure 2d).

3.1.3 Virtual Sewing. Each panel is triangulated as a flat mesh. This flat 2D geometry is the *rest shape* of the fabric, the shape in which it is unstretched, and the simulator measures stretch and bending relative to it. Each stitch is turned into many small links that join every vertex on one edge to the matching point on the other edge. We close these links in the same cloth simulator that later moves the garment (Section 3.2), with the links acting as zero-length springs (Appendix A.3). The body stays static in the dressing pose, and inertia and gravity are switched off, so that the panels are pulled shut slowly and evenly. Sewing succeeds only if the panels start outside the body and on both sides of the limb they wrap, and IMUdrape places them this way automatically. For the same reason, inner garments are sewn before outer ones, so that each garment starts outside those beneath it. The jacket, for example, is sewn over the already-sewn trousers, which act as an obstacle.

The fabric parameters (density, bending stiffness and damping) are set from the fabric type of the garment, using measured properties of typical fabrics of that type (Appendix A). With these parameters, the sewn garment then settles under gravity with friction against the body. The settled garment is the start state of every simulation with that garment (Figure 2h). The garments we generated for the datasets of this paper are shown in Appendix A.1.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

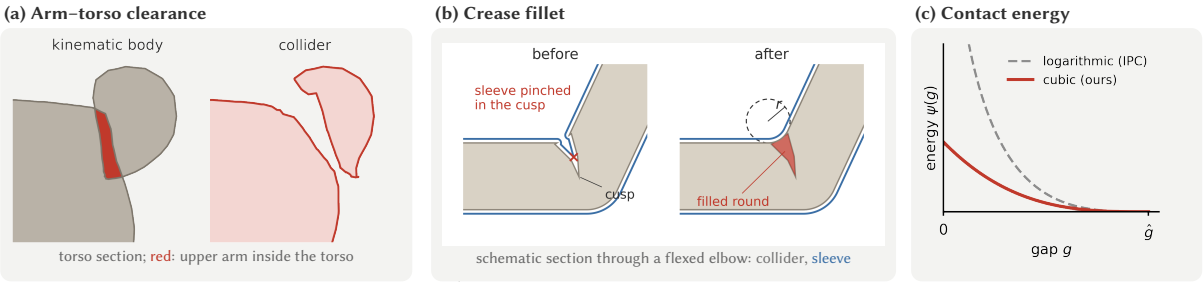


Fig. 3. Garment simulation. (a) In a generated motion (drumming), the upper arm of the kinematic body, which carries the body IMU, lies inside the torso (red, left). In the collider, the surface that the garment touches, arm and torso are pushed apart (right). (b) Schematic: skinning folds the body surface at a flexed elbow into a narrow cusp, and the sleeve is pinched in it, where the solver cannot resolve the contact (left). A ball of radius r rolled over the limbs cannot enter the crease; the region it cannot reach (red) is added to the collider, and the sleeve runs over the smooth round (right). (c) Contact is enforced by a barrier that is zero beyond the gap $\hat{g}$. The cubic barrier [3] stays finite as the gap closes, unlike the logarithmic barrier of IPC [49].

3.2 Garment Simulation

The garment is simulated on the moving body with a contact solver that never lets garment and body interpenetrate. Such a solver needs a physically clean body, and generated motion does not always provide one. We therefore first describe how IMUdrape prepares the body (Figure 3) and then the solver.

3.2.1 Kinematic Body and Collider. Two defects of generated motion have to be handled before the garment can be simulated. First, body parts pass through each other. With the default SMPL body shape, the upper arm of a generated walk sits inside the torso in most frames, because the motion model was trained on a slimmer body than SMPL (Appendix A.4). Fabric caught between two overlapping body parts has no valid position. Second, linear blend skinning, the standard way of deforming the body mesh with the skeleton, folds the body surface at a flexed elbow, hip or crotch into a sharp, hook-shaped cusp. Garment vertices get trapped in this cusp.

IMUdrape therefore uses two copies of the body. The *kinematic body* follows the generated motion exactly and carries the body sensors. The *collider* is the surface that the garment touches. It starts as a copy of the kinematic body, and wherever the motion is not physically clean, it is deformed in that frame, much as soft tissue would be (Figure 3a,b). Because the body sensors stay on the kinematic body, these deformations do not change the body IMU.

The collider removes the first defect by pushing arm and torso apart. Wherever an arm vertex is closer to the torso than a target clearance, the missing clearance is split between arm and torso, which move apart along the line between them (Appendix A.3). These displacements are smoothed over the surrounding surface, so that the collider gets a gentle dent that moves with the limb, and the procedure is repeated until arm and torso are clear in every frame.

The collider removes the second defect by filling the joint creases. Imagine rolling a ball of radius r over the outside of the two limbs that meet at a flexed joint. The ball cannot enter a crease narrower than itself. The surface it touches is therefore the limbs, with the crease filled by a smooth round of radius r, and this filled shape becomes the collider at the joint. In practice, every collider vertex at the bottom of the crease, where the ball cannot reach, is moved outward onto the round. In image processing, this operation is called a morphological closing. As the joint flexes, the round grows smoothly and pushes the fabric out of the crease instead of pinching

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

it. Finally, because the garment is sewn in a fixed dressing pose, each motion is preceded by a short transition from that pose into the motion's first frame.

3.2.2 Contact-Resolving Garment Dynamics. We simulate the garment with the ZOZO contact solver [3], a GPU solver from the family of Incremental Potential Contact (IPC) [49, 50]. These solvers advance the simulation by one time step by minimizing an energy. For vertex positions $\mathbf{x}$ and time step Δt, each step finds the minimizer of

$$\psi_{\text{total}}(\mathbf{x}) = \psi_{\text{dyn}}(\Delta t, \mathbf{x}) + \sum_i \psi_{\text{weak}}\big(g_i, \hat{g}, \bar{\kappa}_i\big). \tag{3}$$

The first term moves the fabric as momentum and its own elasticity dictate,

$$\psi_{\text{dyn}}(\Delta t, \mathbf{x}) = \frac{1}{2\Delta t^2}\,\|\mathbf{x} - \tilde{\mathbf{x}}\|_M^2 + \Psi_{\text{el}}(\mathbf{x}). \tag{4}$$

Here $\tilde{\mathbf{x}}$ is where each vertex would end up if it kept its current velocity under gravity alone [49], and the first part penalizes leaving that path, weighted by the vertex masses M. The elastic energy Ψ_{el} penalizes stretching, shearing [5] and bending the fabric. The minimum is where the fabric keeps as much of its momentum as its elasticity allows.

The second term keeps garment and body apart. Every contact i, between garment and body or between two parts of the garment, adds a barrier on its gap g_i that grows as the gap closes. IPC uses a logarithmic barrier, which becomes infinitely stiff as the gap closes and slows the solver down. The ZOZO solver uses a cubic barrier instead [3],

$$\psi_{\text{weak}}(g_i, \hat{g}, \bar{\kappa}_i) = \begin{cases} -\dfrac{2\bar{\kappa}_i}{3\hat{g}}\,(g_i - \hat{g})^3, & g_i \le \hat{g}, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

which is zero beyond the activation gap $\hat{g}$ and rises smoothly below it (Figure 3c). The solver sets the stiffness $\bar{\kappa}_i$ of each barrier automatically from the mass and the stiffness of the fabric at the contact, so that the barrier is strong enough to stop the fabric's momentum and to resist the fabric being pressed onto the body [3] (Appendix A). No manual tuning is needed. Each step is solved by Newton's method. Before accepting a step, the solver checks along the whole path of every vertex that no part of the garment passes through the body or through the garment itself, so every accepted step is free of intersections. Beyond contact, three settings shape how the garment moves on the body. Friction between garment and body follows the lagged dissipative model of IPC [49], stiffness-proportional damping suppresses unphysical high-frequency vibration of the fabric, and a soft pin holds the trousers at the waist, as a belt would.

3.3 Garment IMU

A real garment IMU sits on a small rigid circuit board sewn onto the fabric. We therefore model each virtual garment sensor as a small rigid patch of fabric (Figure 4). To reproduce the sensor layout of a real smart garment, each virtual sensor is placed at the point of the garment closest to the body vertex under the corresponding real sensor. This point is chosen in the settled start state, as if the sensor had been sewn over that spot while the wearer stood.

The fabric under a sensor deforms in every frame, so its vertices do not share one rotation and position. Like a rigid board sewn onto the fabric, the virtual sensor follows the rigid motion that best matches the fabric beneath it. Let $\mathcal{P}$ be the garment vertices within radius r of the sensor point, $\mathbf{x}_j^0$ their positions in the start state, $\mathbf{x}_j(t)$ their positions at time t and a_j their areas. The sensor's rotation $R(t)$ and position $\mathbf{p}(t)$ minimize the area-weighted distance between the moved patch and the fabric,

$$\big(R(t), \mathbf{p}(t)\big) = \arg\min_{R \in SO(3),\, \mathbf{p}} \sum_{j \in \mathcal{P}} a_j \left\| R\,(\mathbf{x}_j^0 - \bar{\mathbf{x}}^0) + \mathbf{p} - \mathbf{x}_j(t) \right\|^2, \tag{6}$$

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

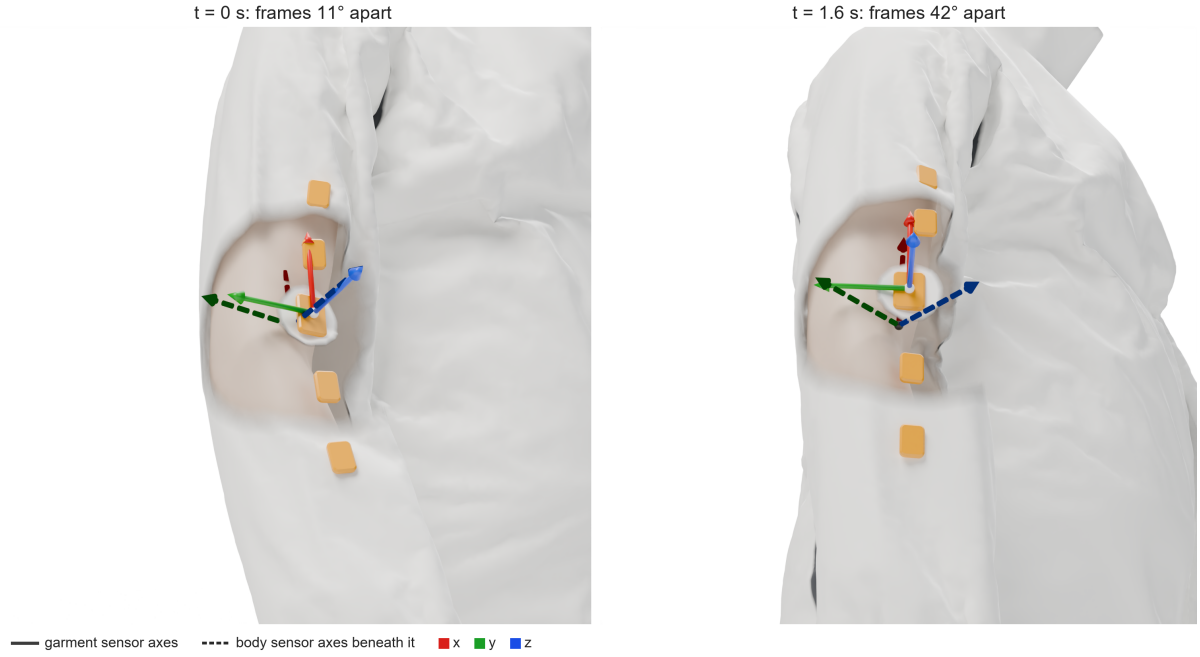


Fig. 4. A garment sensor on the upper sleeve (solid axes) and the body sensor beneath it (dashed axes), seen through a cut-away in a generated folding motion. The garment sensor's z axis is the outward normal of the fabric. At the start of the motion, the two frames are 11° apart (left). As the sleeve slides and twists, the garment frame rotates 42° away from the body frame within 1.6 s (right).

with $\bar{\mathbf{x}}^0$ the patch centroid in the start state. This is the Kabsch problem [38], which has a closed-form solution from the singular value decomposition of the weighted cross-covariance of the patch. The radius r sets how much the patch smooths the fabric motion. A single vertex would overestimate how much the fabric rotates locally, and a larger patch over-smooths it. We use $r = 2.5$ cm, about the size of a small IMU board. The sensor axes are fixed to the fabric (Figure 4). The z axis is the outward normal of the fabric, and the x axis points along the vertical direction of the flat panel, so that a sensor sewn upright on the garment stays aligned with the fabric as it moves. With S^0 being the sensor axes in the start state, $S(t) = R(t)\,S^0$ is the sensor frame at time t. With the sensor's position and frame known over time, the accelerometer measures the sensor's acceleration minus gravity, $\mathbf{a}(t)$, and the gyroscope its angular velocity, $\boldsymbol{\omega}(t)$,

$$\mathbf{a}(t) = S(t)^\top\big(\ddot{\mathbf{p}}(t) - \mathbf{g}\big), \qquad [\boldsymbol{\omega}(t)]_\times = S(t)^\top \dot{S}(t), \tag{7}$$

with derivatives computed from the 120 Hz simulation output and $\mathbf{g}$ the gravity vector. The paired *body* sensor is a rigid patch of the same radius on the kinematic body at the same body vertex, and its accelerometer and gyroscope signals are computed in the same way. Every simulation therefore yields garment and body IMU signals for exactly the same motion (Appendix A).

Pose estimation also needs each sensor's orientation relative to the bone beneath it, such as the forearm for a sensor on the sleeve cuff. Pose datasets obtain it by calibration. The wearer holds a known pose at the start of a recording, and the rotation between each sensor and its bone in that pose is stored and treated as fixed. For a

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

garment sensor, this rotation depends on how the garment hangs during the pose, so IMUdrape reproduces the calibration with a simulated T-pose hold before each motion (Appendix A.3).

Finally, the signals so far are those of an ideal sensor, and real IMUs have noise and other errors. IMUdrape adds the standard error model of a MEMS IMU [17, 97],

$$\tilde{\mathbf{a}}(t) = (I + \mathrm{diag}(\mathbf{s}))\, M_a\, \mathbf{a}(t) + \mathbf{b} + \boldsymbol{\eta}(t), \tag{8}$$

with scale errors $\mathbf{s}$, axis misalignment M_a, bias $\mathbf{b}$ and white noise $\boldsymbol{\eta}$, and likewise for the gyroscope. The error magnitudes are fitted to still windows of real recordings. Because a real sensor is never sewn on exactly as designed, IMUdrape also rotates each garment sensor by a small random mounting rotation.

4 On the Effectiveness of Garment-Mounted Virtual IMU Data

A sensor sewn into a garment moves with the fabric, so a model for it needs training data recorded with that garment, and such data are scarce. In this section, we test whether models trained with IMUdrape's virtual data perform well on the IMU data of real garments. We consider a deployment in which a garment with built-in IMUs should recognize its wearer's activities or estimate the wearer's full-body pose, for example to follow rehabilitation exercises at home or an athlete's technique without a motion-capture suit (Section 4.1), but labeled recordings exist for only a few people, or none. IMUdrape generates virtual IMU data for the activities of interest, on a simulated copy of the garment and at the positions of its sensors. We train a model on the virtual data together with whatever real data are available, and test it on real recordings of a person who was not part of the training data and who wears the real sensor garment. We compare it with a model trained on real data alone and with models trained on virtual data from existing generators. Section 4.1 describes the datasets, the virtual data and the protocol that both tasks share, and Sections 4.2 and 4.3 report the results of each task.

4.1 Experimental Setup

4.1.1 Datasets. We evaluate on four public datasets recorded with inertial sensors on everyday garments. The MIG dataset [92] was recorded with a jacket and trousers carrying 396 sewn-in IMUs. The LIP [110], FIP [108] and GarMoCap [19] datasets were recorded with four to six IMUs on loose jackets, and in GarMoCap also on loose trousers, together with the wearer's pose. The activities range from locomotion (walking, running, stairs) and sports (boxing, ping-pong, basketball, jumping rope) to everyday tasks (eating, typing, ironing, vacuuming). MIG has 14 activity classes, LIP 5, and FIP and GarMoCap 10 each. Table 11 in Appendix B lists the subjects, sensors, activities and recorded minutes of every dataset, and the amount of virtual data we generate for each.

All four datasets label the activities, so we use all of them for activity recognition, the established benchmark for virtual IMU data [44, 48]. In addition to activity recognition, we consider full-body pose estimation on the LIP, FIP and GarMoCap datasets, which also record the wearer's pose. Full-body pose is the most detailed description of how a person moves. It is the basis of motion capture for sports and rehabilitation, of avatars in virtual and augmented reality, and of finer-grained activity and health analyses than a fixed set of activity labels allows [30, 101]. Sensors built into everyday clothing would make such motion capture available outside the lab, without straps or a suit [19, 110]. Pose labels, however, require an optical or inertial motion-capture system during recording, so labeled garment data for pose estimation are even scarcer than for activity recognition. This makes pose estimation a demanding test of whether virtual garment data can stand in for real recordings.

4.1.2 Virtual Data and Protocol. For every dataset, we generate 20 text descriptions of each of its activities, generate the motion and simulate it on the garment matched to the dataset's clothing (Table 9 in Appendix A.1). This gives 89 minutes of virtual data for MIG, 12 for LIP, 24 for FIP and 21 for GarMoCap, 146 minutes in total (Table 11 in Appendix B). The virtual data are thus smaller than the real recordings of each dataset, between about a quarter and a tenth of their size.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Existing virtual IMU data place the sensors on the body, and the only established way to obtain garment motion without physics simulation is a learned garment model such as TailorNet [71]. We therefore compare IMUdrape with one baseline for each approach. IMUdrape itself provides three sources from the same simulations:

- **IMUGPT** [46, 48] generates motion from the same descriptions with its motion synthesis model T2M-GPT [104] and computes body IMU at the skeleton joints with IMUSim [103]. As calibration, IMUGPT fits the signal distribution of each activity class to the real training data; without real training data, its data are used uncalibrated.
- **TailorNet** [71] is a network that predicts the garment mesh from the body's pose and shape. It was trained on physics simulations of garments on single poses and predicts each frame independently, so the garment follows the body without inertia or delay. LIP's synthetic loose-garment data are built on it [110]. We run it on our motion and read the sensors from the predicted garment as in IMUdrape.
- **IMUdrape Body** uses the signal of a sensor on the body at the same position, which every IMUdrape simulation also gives. It separates the effect of the garment from that of the motion. A body-worn sensor is strapped in place, so it receives the sensor errors but not the mounting rotation of a garment sensor.
- **IMUdrape Garment** simulates the garment on our motion and reads the sensors from the simulated fabric.
- **IMUdrape Body + Garment** pools the body and garment data of the same clips and subsamples the pool to the size of the other sources, so that each motion appears once, as either its body or its garment signal.

Every source has the same number of clips per activity, where a clip is one generated motion sequence with its virtual IMU signals, so all sources have the same amount of virtual data. Every dataset uses leave-one-subject-out cross-validation over all of its subjects. We train with real data from 0 (virtual data only), 1, 2 or all training subjects. No real subjects corresponds to zero-shot and one or two subjects to few-shot learning, the low-data regimes in which a new garment is deployed. All subjects show whether virtual data still help once real data are plentiful. For each model, we repeat the cross-validation with three random seeds and report the average and the standard deviation across the three runs.

4.2 Activity Recognition

4.2.1 Experimental Setting. We classify 2 s windows of garment IMU signal into the activities of each dataset. The LIP, FIP and GarMoCap datasets provide accelerometer and orientation data without a gyroscope, so on these datasets we use the accelerometer signals for recognition. On the MIG dataset, we use six of the 396 sensors, on the right side where the fabric hangs free: the back of the forearm sleeve, the side panel below the armpit, the front hip below the belt, the buttock and the front and back of the ankle. We use three recognition models, DeepConvLSTM [69], DeepConvLSTM with attention [66] and TinyHAR [109]. Appendix C describes the models, and Appendix A.2 the sensor mapping of each dataset. The model is trained on the real and virtual windows together, as in earlier work on virtual IMU data for activity recognition [44, 48].

Sensor signals differ between people and with how each person wears the garment, so we use test-time adaptation [53, 95] by per-subject standardization. Every subject's channels are standardized with its own mean and standard deviation, and the virtual data with theirs. For the test subject, these statistics come from the unlabeled earlier half of each recording, and the later half is used for evaluation. This mirrors a real deployment, in which a system can collect unlabeled data from a new user before it has to recognize that user's activities, as in unsupervised adaptation of activity recognition to new users [12]. We report the macro-F1 score.

4.2.2 Results. Table 2 shows the results.

An IMUdrape source gives the best result in every setting, and the gain is largest when real data are scarce. With one real subject, IMUdrape Body + Garment improves macro-F1 over real data alone by a relative 19.6%, 46.3%, 7.2% and 24.9% on the LIP, FIP, GarMoCap and MIG datasets. It also improves on IMUdrape Body data from the

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 2. Activity recognition performance (macro-F1, %), averaged over the test subjects and 3 recognition models (per-model results in Appendix D). The best result in each block and column is highlighted.

Training data	LIP	FIP	GarMoCap	MIG
Virtual data only				
IMUGPT	51.8 ± 2.3	16.2 ± 1.0	24.3 ± 1.8	7.2 ± 0.2
TailorNet	65.2 ± 2.1	20.8 ± 0.8	28.7 ± 1.2	27.5 ± 1.2
IMUdrape Body	65.7 ± 3.3	18.7 ± 1.7	29.1 ± 0.4	37.7 ± 0.3
IMUdrape Garment	58.8 ± 3.1	21.4 ± 0.9	26.6 ± 1.6	43.5 ± 1.0
IMUdrape Body + Garment	71.9 ± 1.4	23.1 ± 0.6	27.6 ± 1.3	45.6 ± 1.2
1 real subject				
Real data only	66.9 ± 4.5	17.7 ± 4.3	45.8 ± 2.8	39.8 ± 2.7
IMUGPT	66.4 ± 3.4	18.0 ± 3.4	41.2 ± 1.3	34.3 ± 3.9
TailorNet	76.5 ± 1.5	23.9 ± 1.6	47.5 ± 1.8	43.7 ± 2.7
IMUdrape Body	78.2 ± 4.0	23.3 ± 2.4	47.2 ± 0.9	48.7 ± 2.1
IMUdrape Garment	77.8 ± 0.9	25.6 ± 2.1	47.9 ± 1.6	49.8 ± 3.3
IMUdrape Body + Garment	80.0 ± 1.3	25.9 ± 0.9	49.1 ± 1.7	49.7 ± 6.1
2 real subjects				
Real data only	80.1 ± 3.0	29.0 ± 4.0	56.6 ± 2.6	49.8 ± 2.2
IMUGPT	78.5 ± 2.8	26.9 ± 3.5	54.5 ± 1.0	46.9 ± 2.5
TailorNet	82.2 ± 1.0	32.6 ± 0.6	56.8 ± 1.2	50.9 ± 2.1
IMUdrape Body	84.7 ± 0.7	33.1 ± 1.8	56.8 ± 1.2	52.6 ± 1.9
IMUdrape Garment	85.0 ± 1.5	35.2 ± 2.4	58.1 ± 1.3	57.3 ± 0.6
IMUdrape Body + Garment	84.8 ± 1.5	35.5 ± 0.8	58.6 ± 1.9	56.4 ± 0.9
All real subjects				
Real data only	87.5 ± 0.3	55.7 ± 0.6	74.8 ± 0.9	67.3 ± 0.6
IMUGPT	87.3 ± 1.0	54.1 ± 1.2	74.5 ± 0.8	65.4 ± 1.7
TailorNet	88.2 ± 0.8	55.9 ± 0.4	74.9 ± 0.6	68.1 ± 0.8
IMUdrape Body	88.3 ± 0.9	54.0 ± 0.8	74.0 ± 0.8	67.7 ± 1.1
IMUdrape Garment	88.4 ± 1.1	56.6 ± 0.6	76.8 ± 1.8	71.0 ± 1.4
IMUdrape Body + Garment	88.2 ± 1.5	57.2 ± 0.5	75.7 ± 1.4	70.6 ± 1.8

same clips, by a relative 2.1% to 11.2%, so the simulated garment adds to what the motion alone provides. Without real data, the gain over IMUdrape Body is largest on the loose sensors of the MIG dataset, where IMUdrape Garment reaches a macro-F1 of 43.5% against 37.7% for IMUdrape Body, a relative improvement of 15.4% and more than real data from one subject reach (39.8%). With all real subjects, the best IMUdrape source remains a relative 1.0% to 5.5% above real data alone.

IMUdrape also outperforms existing techniques for generating virtual IMU data. IMUGPT generates motion from text, as IMUdrape does, but senses it with rigid sensors on the skeleton, so its data describe a body-mounted sensor. TailorNet adds a garment, but one whose shape follows the pose without any dynamics of its own. With one real subject, IMUdrape Body + Garment improves on IMUGPT by a relative 19.2% to 44.9% and on TailorNet by a relative 3.4% to 13.7%. The gain over TailorNet shows that the benefit comes from simulating how the garment moves, beyond a garment shape that follows the pose. Without real data, the gap to IMUGPT is largest,

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 3. Pose estimation performance: mean angular error over 15 major joints (degrees) and mean joint position error (cm), lower is better, averaged over 2 pose networks. The best result in each block and column is highlighted.

	Angular error (°)			Position error (cm)		
Training data	LIP	FIP	GarMoCap	LIP	FIP	GarMoCap
Virtual data only						
IMUGPT	31.4 ± 0.48	26.7 ± 0.31	22.9 ± 0.17	11.3 ± 0.23	8.6 ± 0.08	10.8 ± 0.08
TailorNet	29.3 ± 0.27	25.5 ± 0.62	23.4 ± 0.31	10.5 ± 0.14	8.6 ± 0.15	10.5 ± 0.19
IMUdrape Body	28.6 ± 0.28	25.8 ± 0.21	23.5 ± 0.27	10.3 ± 0.08	8.8 ± 0.11	10.3 ± 0.18
IMUdrape Garment	28.6 ± 0.47	23.7 ± 0.39	21.1 ± 0.18	9.7 ± 0.11	8.0 ± 0.24	9.5 ± 0.15
IMUdrape Body + Garment	28.3 ± 0.31	24.4 ± 0.45	21.1 ± 0.08	9.9 ± 0.06	8.2 ± 0.13	9.5 ± 0.10
1 real subject						
Real data only	22.9 ± 1.28	15.1 ± 1.04	13.9 ± 0.59	7.7 ± 0.37	4.8 ± 0.43	6.8 ± 0.34
IMUGPT	22.7 ± 1.22	15.3 ± 1.00	13.7 ± 0.46	7.7 ± 0.37	4.9 ± 0.44	6.7 ± 0.25
TailorNet	22.7 ± 1.17	15.1 ± 0.94	13.7 ± 0.50	7.7 ± 0.38	4.9 ± 0.40	6.7 ± 0.35
IMUdrape Body	22.7 ± 1.13	15.4 ± 1.29	13.7 ± 0.47	7.7 ± 0.38	5.0 ± 0.57	6.7 ± 0.33
IMUdrape Garment	22.9 ± 1.26	14.7 ± 0.68	13.4 ± 0.54	7.8 ± 0.37	4.7 ± 0.30	6.5 ± 0.31
IMUdrape Body + Garment	22.8 ± 1.33	14.9 ± 0.80	13.2 ± 0.48	7.7 ± 0.45	4.8 ± 0.37	6.4 ± 0.30
2 real subjects						
Real data only	20.7 ± 0.57	13.9 ± 0.27	12.4 ± 0.19	7.0 ± 0.10	4.3 ± 0.12	6.0 ± 0.06
IMUGPT	20.5 ± 0.37	13.9 ± 0.30	12.3 ± 0.18	7.0 ± 0.03	4.3 ± 0.15	5.9 ± 0.05
TailorNet	20.5 ± 0.44	13.9 ± 0.38	12.2 ± 0.07	7.0 ± 0.11	4.3 ± 0.13	5.9 ± 0.11
IMUdrape Body	20.6 ± 0.38	13.8 ± 0.30	12.2 ± 0.17	7.1 ± 0.04	4.3 ± 0.14	5.9 ± 0.07
IMUdrape Garment	20.5 ± 0.37	13.7 ± 0.28	12.3 ± 0.27	7.0 ± 0.09	4.3 ± 0.10	5.9 ± 0.11
IMUdrape Body + Garment	20.6 ± 0.38	13.7 ± 0.35	12.0 ± 0.14	7.0 ± 0.05	4.2 ± 0.15	5.8 ± 0.11
All real subjects						
Real data only	19.1 ± 0.09	12.5 ± 0.05	10.5 ± 0.08	6.5 ± 0.04	3.8 ± 0.01	5.1 ± 0.02
IMUGPT	19.1 ± 0.11	12.4 ± 0.05	10.7 ± 0.09	6.6 ± 0.03	3.8 ± 0.02	5.2 ± 0.03
TailorNet	19.2 ± 0.01	12.3 ± 0.06	10.7 ± 0.02	6.5 ± 0.02	3.8 ± 0.02	5.2 ± 0.02
IMUdrape Body	19.1 ± 0.01	12.3 ± 0.07	10.6 ± 0.01	6.5 ± 0.01	3.8 ± 0.02	5.1 ± 0.01
IMUdrape Garment	19.1 ± 0.04	12.2 ± 0.05	10.6 ± 0.07	6.5 ± 0.04	3.8 ± 0.02	5.1 ± 0.03
IMUdrape Body + Garment	19.1 ± 0.01	12.2 ± 0.05	10.4 ± 0.03	6.6 ± 0.03	3.7 ± 0.03	5.0 ± 0.01

with a macro-F1 of 45.6% on the MIG dataset against 7.2% for IMUGPT. Among the IMUdrape sources, IMUdrape Garment and IMUdrape Body + Garment perform similarly. IMUdrape Body + Garment leads without real data on the LIP, FIP and MIG datasets, and IMUdrape Garment leads on the LIP and MIG datasets with two real subjects. Tables 12 and 13 in Appendix D give the results of each model.

4.3 Pose Estimation

4.3.1 Experimental Setting. We estimate full-body pose from the garment IMUs of the LIP, FIP and GarMoCap datasets. We use two networks, an LSTM poser and TransPose [101]. The networks are pre-trained on virtual data and then fine-tuned on the real training data, as in sparse inertial pose estimation [30, 100, 101]. Without real data, the pre-trained network is tested directly. As in recognition, we standardize the accelerations of each subject with its own statistics, those of the test subject from the earlier half of each recording, and score the

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

later half. We report the mean angular error over 15 major joints in degrees and the mean joint position error in centimeters.

4.3.2 Results. Without real data, IMUdrape data give the lowest joint angle error and the lowest joint position error on all three datasets (Table 3). The best IMUdrape source reduces the angle error relative to IMUGPT by 7.8% to 11.5% and relative to TailorNet by 3.2% to 10.0%. Simulating the garment helps most on the FIP and GarMoCap datasets. Relative to IMUdrape Body, IMUdrape Garment reduces the angle error by 8.4% on FIP and by 10.2% on GarMoCap. On LIP, IMUdrape Garment and IMUdrape Body give the same error, and IMUdrape Body + Garment is 0.9% lower. Pre-training on IMUdrape data also helps with one real subject, reducing the error relative to real data alone by 0.9% on LIP, 2.2% on FIP and 5.1% on GarMoCap. With two or all real subjects, the real data dominate, and every source is within about 4% of real data alone.

Tables 14 and 15 in Appendix D give the angle and position errors of each network.

5 Discussion

The evaluation shows that simulated garment data improve activity recognition and pose estimation on real garments, beyond real data alone and beyond virtual data from body-mounted sensors (IMUGPT, IMUdrape Body) or from a learned garment model without dynamics (TailorNet). The benefit is largest when real data are scarce. In this section we first analyze when and why garment data help (Section 5.1). We then ask how the benefit changes when the garment is simulated on recorded instead of generated motion (Section 5.2) and when it is simulated on a garment other than the deployed one (Section 5.3). Finally, we summarize what these findings mean for practitioners (Section 5.4) and discuss limitations (Section 5.5).

5.1 When and Why Do Garment Data Help?

5.1.1 How Much Virtual Data Is Needed? IMUdrape can generate virtual data in any amount, at the cost of simulation time. A practical question is therefore how much to generate, and whether generating more would still pay off. We study this on activity recognition on the MIG dataset, with the setting of Section 4.2. We train on 25%, 50% or all of the virtual clips, that is, about 22, 45 or 89 minutes of virtual data, with 0, 1, 2 or all real training subjects. The smaller sets are drawn per activity, so each keeps the activity mix of the full set. The setting without real data is the one IMUdrape ultimately aims at, a new garment for which nothing has been recorded. The settings with real data show whether generating more virtual data still pays off once some recordings exist.

We find that more virtual data help most without real data (Table 4). There, going from 25% to 100% of the clips improves macro-F1 by a relative 22.0% for IMUdrape Body, 19.8% for IMUdrape Garment and 17.8% for IMUdrape Body + Garment data, and every source improves at each step. With one or two real subjects, more virtual data still help the garment sources, by up to 4.4% relative, and with all real subjects the three amounts perform alike. Without real data, the gains have not leveled off at the largest set, so generating more virtual data could improve recognition further. We therefore expect generation at a larger scale, with more descriptions per activity and more varied motions, to raise performance further, and we see it as a promising direction for future work.

5.1.2 Tight or Loose Sensors? A sensor on a loose part of a garment moves more relative to the body than one where the fabric is pressed against the body. One might therefore expect garment simulation to matter only for loose sensors, and body data to suffice for tight ones. We test this on the MIG dataset, whose jacket and trousers carry sensors at both kinds of positions. The main MIG results use six loose-fitting sensors, where the fabric hangs free (Section 4.2). Here we compare them with a tight set of six sensors in comparable regions, where the garment is pressed against the body: the forearm, upper arm, chest, waist, thigh and shank. Both sets are evaluated with the setting of Section 4.2, with every virtual data source and 0, 1, 2 or all real training subjects. Table 5 shows the results.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 4. Activity recognition on MIG (macro-F1, %) with 25, 50 or 100% of the virtual clips, for each amount of real training data. The best result in each column is highlighted.

		Real training subjects			
Virtual data	Amount	None	1 subject	2 subjects	All subjects
Real data only	–	–	39.8 ± 2.7	49.8 ± 2.2	67.3 ± 0.6
IMUdrape Body	25%	30.9 ± 2.2	47.1 ± 1.0	53.6 ± 0.8	68.8 ± 0.7
	50%	33.0 ± 0.5	45.9 ± 1.7	52.4 ± 2.1	69.0 ± 0.2
	100%	37.7 ± 0.3	48.7 ± 2.1	52.6 ± 1.9	67.7 ± 1.1
IMUdrape Garment	25%	36.3 ± 2.4	47.7 ± 3.7	55.7 ± 0.8	71.0 ± 0.6
	50%	42.2 ± 0.9	47.5 ± 2.6	54.7 ± 1.1	70.9 ± 1.5
	100%	43.5 ± 1.0	49.8 ± 3.3	57.3 ± 0.6	71.0 ± 1.4
IMUdrape Body + Garment	25%	38.7 ± 1.1	47.7 ± 3.3	56.1 ± 3.5	70.1 ± 0.4
	50%	42.9 ± 2.0	48.6 ± 4.0	54.9 ± 1.6	69.5 ± 0.5
	100%	45.6 ± 1.2	49.7 ± 6.1	56.4 ± 0.9	70.6 ± 1.8

Table 5. Activity recognition on MIG with the tight and the loose sensor set (macro-F1, %), for each amount of real training data. The best result in each column is highlighted.

	Tight sensor set				Loose sensor set			
Training data	None	1 subject	2 subjects	All subjects	None	1 subject	2 subjects	All subjects
Real data only	–	42.4 ± 3.0	53.5 ± 2.9	66.8 ± 0.2	–	39.8 ± 2.7	49.8 ± 2.2	67.3 ± 0.6
IMUGPT	6.7 ± 0.5	43.5 ± 1.2	54.6 ± 2.7	67.1 ± 0.6	7.2 ± 0.2	34.3 ± 3.9	46.9 ± 2.5	65.4 ± 1.7
TailorNet	26.7 ± 2.9	48.2 ± 1.4	54.9 ± 0.4	66.7 ± 1.0	27.5 ± 1.2	43.7 ± 2.7	50.9 ± 2.1	68.1 ± 0.8
IMUdrape Body	32.6 ± 1.9	49.0 ± 0.4	55.8 ± 2.6	67.0 ± 2.0	37.7 ± 0.3	48.7 ± 2.1	52.6 ± 1.9	67.7 ± 1.1
IMUdrape Garment	52.0 ± 1.0	53.9 ± 2.7	60.0 ± 0.6	69.9 ± 0.6	43.5 ± 1.0	49.8 ± 3.3	57.3 ± 0.6	71.0 ± 1.4
IMUdrape Body + Garment	52.8 ± 0.7	54.6 ± 1.7	60.8 ± 2.9	70.6 ± 0.9	45.6 ± 1.2	49.7 ± 6.1	56.4 ± 0.9	70.6 ± 1.8

On both sets, IMUdrape Garment data outperform IMUdrape Body data at every amount of real data. The gain is larger on the tight set when real data are scarce. Without real data, IMUdrape Garment improves on IMUdrape Body by a relative 59.5% on the tight set against 15.4% on the loose set, and with one real subject by 10.0% against 2.3%. With two or all real subjects, the loose set gains slightly more, by 8.9% and 4.9% on the loose set against 7.5% and 4.3% on the tight set. On the tight set, IMUdrape Body + Garment data are the best source at every amount of real data, and IMUdrape Garment data are within 0.8 points of it. Simulating the garment therefore pays off for sensors on tight-fitting as well as loose-fitting parts of a garment, and Section 5.1.3 examines where on the garment it matters most.

5.1.3 Where on a Garment Does Physics Matter? A densely instrumented simulated garment lets us ask where on a garment simulating the garment makes a difference. Sensor placement has been studied on the real MIG data with deep models [18] and in simulation for body-mounted sensors on motion-capture data [75], but not for the difference between garment and body sensors. We trained single-sensor recognition models at each of the 396 MIG sensor positions, once on simulated garment data and once on body data from the same position, without real training data, and tested them on the real garment data of the held-out subjects. We used the three recognition models of Section 4.2. Each model is evaluated with the protocol of Section 4.2 over all 12 leave-one-subject-out folds.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 6. Activity recognition on MIG without real training data, with 1 to 6 sensors drawn from the tight and the loose sensor set (macro-F1, %, mean over the three recognition models and over the sensor subsets of each size). Improvement over Body is the relative improvement of IMUdrape Garment over IMUdrape Body.

	Tight set			Loose set		
Sensors	Body	Garment	Improvement over Body	Body	Garment	Improvement over Body
1	26.2 ± 0.7	31.6 ± 0.2	+20.7%	25.2 ± 0.2	26.5 ± 0.2	+5.2%
2	30.3 ± 0.4	42.1 ± 0.3	+38.9%	28.5 ± 0.7	32.5 ± 0.2	+14.0%
3	30.5 ± 0.7	42.0 ± 0.4	+37.5%	32.0 ± 0.8	37.7 ± 0.7	+17.7%
4	31.6 ± 0.9	45.7 ± 0.5	+44.6%	33.0 ± 0.4	39.1 ± 0.5	+18.7%
5	31.2 ± 0.8	48.2 ± 0.8	+54.9%	34.5 ± 1.3	40.3 ± 0.9	+16.9%
6	32.7 ± 0.7	51.0 ± 1.6	+56.2%	37.5 ± 1.1	44.1 ± 1.6	+18.0%

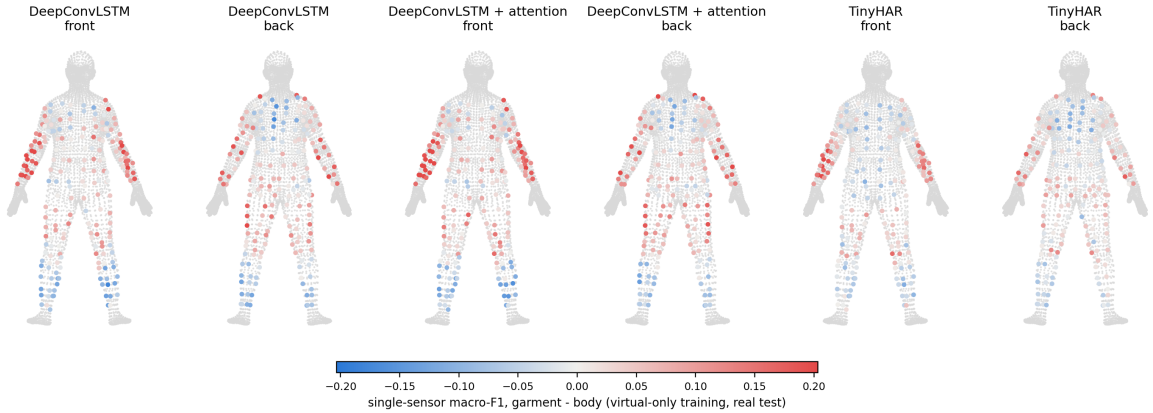


Fig. 5. Per-position difference in single-sensor macro-F1 between training on simulated garment IMU data and on body IMU data from the same position (virtual data only, tested on the real MIG data). Red: garment data better; blue: body data better. The front view faces the wearer.

Figure 5 shows that garment physics helps at most positions. Garment data beat body data at 255, 264 and 212 of the 396 positions in the three models. Averaged over all positions, they improve macro-F1 by a relative 11.1%, 14.4% and 4.8%. The gain holds in every model on the sleeves and the thighs. It is largest on the forearm sleeves, by a relative 59.5% to 105.7%, where garment data win at 47 or 48 of the 48 positions. There the sleeve turns around the arm and slides along it, and body data cannot reproduce this motion. On the jacket trunk, garment and body data are close (−8.3% to +3.2%). Body data remain better on the lower trouser legs in every model (−4.6% to −18.8%). We also combined the sensors of the tight and the loose set of Section 5.1.2, using one to six sensors of each set at once with the same protocol and no real training data. For two to five sensors, we averaged over several subsets of the set (Table 6). The advantage of garment data over body data grows as sensors are combined. On the tight set, it rises from a relative 20.7% with one sensor to 56.2% with all six. Here garment data improve from 31.6% to 51.0% macro-F1 as sensors are added, whereas body data gain little (26.2% to 32.7%). On the loose set, the advantage grows from 5.2% with one sensor to between 14.0% and 18.7% with two to six sensors. Simulating the garment therefore matters even more for garments that carry several sensors than for a single sensor.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 7. GarMoCap pose estimation and activity recognition when the garment is simulated on the dataset's *recorded* training motion or on the same amount of generated motion, both simulated on the matched garment. The best result in each column is highlighted.

		Pose: joint angle error (°)				Recognition: macro-F1 (%)			
	Training data	None	1 subject	2 subjects	All	None	1 subject	2 subjects	All
	Real data only	–	13.9 ± 0.59	12.4 ± 0.19	10.5 ± 0.08	–	45.8 ± 2.8	56.6 ± 2.6	74.8 ± 0.9
Recorded	IMUdrape Body	16.8 ± 0.04	13.3 ± 0.41	12.1 ± 0.12	10.5 ± 0.06	29.1 ± 1.6	52.7 ± 0.5	61.4 ± 0.8	76.3 ± 0.9
	IMUdrape Garment	14.8 ± 0.12	13.0 ± 0.65	12.0 ± 0.30	10.5 ± 0.05	32.1 ± 2.2	55.3 ± 2.7	63.7 ± 1.1	77.4 ± 1.1
	IMUdrape Body + Garment	14.3 ± 0.11	12.9 ± 0.60	11.9 ± 0.29	10.4 ± 0.05	32.9 ± 0.4	56.5 ± 1.6	63.8 ± 1.4	76.0 ± 0.8
Generated	IMUdrape Body	23.9 ± 0.09	13.9 ± 0.31	12.4 ± 0.11	10.7 ± 0.10	25.0 ± 0.6	47.0 ± 0.8	57.9 ± 1.3	76.0 ± 1.1
	IMUdrape Garment	21.2 ± 0.10	13.5 ± 0.57	12.3 ± 0.27	10.5 ± 0.03	25.9 ± 2.4	48.9 ± 0.6	58.7 ± 1.6	76.4 ± 1.4
	IMUdrape Body + Garment	21.5 ± 0.23	13.4 ± 0.59	12.1 ± 0.17	10.5 ± 0.07	27.4 ± 1.1	49.2 ± 0.3	59.9 ± 0.4	76.8 ± 0.1

5.2 Recorded Instead of Generated Motion

Generated motion differs from real recordings, so part of the remaining error may come from the motion rather than from the garment. To separate the two, we simulate the garment on recorded motion instead of generated motion. We use GarMoCap [19], the one dataset that records the wearer's full-body motion together with loose-garment IMUs on both a jacket and trousers. We simulate the matched GarMoCap garment on the recorded motion of every subject. In every leave-one-subject-out fold, the virtual data come only from the recorded motion of that fold's training subjects, so the test subject's motion is never simulated for training. For a fair comparison, we draw generated motion at random from the generated data of Section 4, simulated on the same garment, so that the virtual data from recorded and from generated motion are of the same size. Training and evaluation follow the setting of Sections 4.2 and 4.3. Table 7 shows the results.

Recorded motion improves every virtual source when real data are scarce, most of all without real data. Without real data, it lowers the joint angle error relative to generated motion by 29.7% for IMUdrape Body, 30.5% for IMUdrape Garment and 33.3% for IMUdrape Body + Garment data, and it raises recognition macro-F1 by 16.4%, 23.9% and 20.1%. Both pose networks show this drop, by 28.4% to 32.9% for the LSTM poser and by 31.0% to 33.7% for TransPose. With one real subject, IMUdrape Body + Garment data from recorded motion raise macro-F1 from 45.8% with real data alone to 56.5%, against 49.2% with generated motion, and lower the joint angle error from 13.9° to 12.9°, against 13.4°. IMUdrape Body + Garment data from recorded motion give the lowest pose error at every amount of real data. With all real subjects, recorded and generated motion perform alike. Garment data outperform body data with either motion. Without real data, IMUdrape Garment lowers the joint angle error relative to IMUdrape Body by 12.1% with recorded and by 11.1% with generated motion. The gains from realistic motion and from garment simulation therefore add up, and more realistic and more varied generated motion is a direct way to improve IMUdrape further (Section 5.5).

5.3 Matched or Generic Garment

So far, each dataset was simulated with its matched garment, recreated from photographs of the garment the participants wore. Generating the garment of a deployment from a photograph or a description is a capability that learned garment models such as TailorNet lack, as they are limited to the few garment templates they were trained on. Here we ask how much this matching is worth, by comparing the matched garment with a different garment of the same type. Every garment in IMUdrape carries the sensor positions of all datasets, so the same sensors can be read on a different garment. We simulate the recorded GarMoCap clips of Section 5.2 on the matched GarMoCap garment and on the MIG jacket and trousers. We then train and evaluate the two pose

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 8. GarMoCap pose estimation and activity recognition with IMUdrape Garment data from the matched GarMoCap garment and from the generic MIG garment, both simulated on the same recorded motion clips. The best result in each column is highlighted.

		Pose: joint angle error (°)				Recognition: macro-F1 (%)			
	Training data	None	1 subject	2 subjects	All	None	1 subject	2 subjects	All
	Real data only	–	13.9 ± 0.59	12.4 ± 0.19	10.5 ± 0.08	–	45.8 ± 2.8	56.6 ± 2.6	74.8 ± 0.9
Matched	IMUdrape Garment	14.8 ± 0.05	13.1 ± 0.56	12.0 ± 0.30	10.5 ± 0.04	32.3 ± 1.4	54.9 ± 2.2	64.0 ± 0.4	77.7 ± 0.9
Generic	IMUdrape Garment	15.1 ± 0.17	13.3 ± 0.50	12.2 ± 0.19	10.6 ± 0.01	28.1 ± 2.8	53.9 ± 1.3	63.9 ± 1.3	76.7 ± 0.4

networks and the three recognition models on IMUdrape Garment data from each garment, as in Section 5.2. Table 8 shows the results.

For pose estimation, the matched garment gives a lower joint angle error than the MIG garment at every amount of real data. The difference is largest without real data, 1.7%, and it appears with both pose networks. Without real data, both garments lower the error relative to body data from the same recorded motion by 10% to 12% (Table 7). For recognition, the matched garment helps most without real data, where it improves macro-F1 by 14.9%, from 28.1% to 32.3%. With real data, it stays ahead of the MIG garment by up to 1.0 point. The matched garment gives the best result in every setting, tied with real data alone for pose estimation with all real subjects. Matching the garment to the deployment therefore pays off most when no real data exist, which is exactly when a new smart garment is introduced. At the same time, a different garment of the right type still keeps the benefit of garment simulation over body data, so IMUdrape remains useful when the exact garment is not known.

5.4 Guidelines for Practitioners

Consider two teams who want to use inertial sensors built into clothing. The first is developing a new smart garment, for example a workout shirt or a jacket with sensors in the sleeves, and needs an activity recognition model before the product reaches its users. The second wants to monitor patients in everyday clothes, for example the daily activity of older adults or the exercises of rehabilitation patients, without asking them to strap on sensors. Both face the same problem. A model for a garment-mounted sensor needs training data recorded with that garment, at those sensor positions, and recording such data means recruiting participants, dressing them in the garment and labeling what they do. For a new product this has to be repeated for every cut and sensor layout, and for patients every recording session is a burden. Existing virtual IMU data do not solve this problem, because they describe sensors strapped to the body.

IMUdrape offers these teams a different starting point. From a photograph or a description of their garment, it generates a matching garment, places the virtual sensors where the real ones will sit and produces training data for the activities they describe in text. Our results show how to make the most of these data.

- **Start before any recording.** Without real data, IMUdrape data already outperform the body-mounted virtual data of IMUGPT and the learned garment model TailorNet on every dataset, for recognition and for pose estimation (Sections 4.2 and 4.3). A team can therefore train a first model before recruiting a single participant.
- **Record a few people.** The gains are largest with one or two real subjects, the situation of a pilot study or a small patient cohort. For recognition, IMUdrape Garment and IMUdrape Body + Garment data work best, and pre-training on IMUdrape data before fine-tuning on the real recordings helps pose estimation (Sections 4.2 and 4.3). For recognition, virtual data still add a small gain even with all real subjects.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

- **Simulate the garment that will be worn.** Generating the deployed garment instead of a different one of the same type helps most when no real data exist (Section 5.3), so the photograph of the product or of the patients' usual clothing is worth providing.
- **Use several sensors.** The advantage of garment data over body data grows when sensors are combined (Section 5.1.3), so simulation is most valuable for garments that carry several sensors.
- **Sensors on tight and on loose fabric both benefit** (Section 5.1.2), so a designer does not need to restrict sensors to tight-fitting parts to profit from simulation. The forearm sleeves gain most from physics (Section 5.1.3).
- **Generate more, and use real motion when it exists.** More virtual data keep improving recognition, most of all before real data are available (Section 5.1.1). If the team can record the wearers' motion, for example with a camera-based motion capture of the target exercises, simulating the garment on this motion adds to the benefit of the garment simulation (Section 5.2).

In both scenarios, IMUdrape thus turns a garment design and a list of activities into a working model with little or no real data, and makes each real recording that is collected count for more.

5.5 Limitations and Future Work

Scale and diversity of the virtual data. Our virtual datasets are still small. Each dataset receives 20 generated descriptions per activity, all simulated on a single body shape, the average SMPL body ($\beta = 0$). Three of our findings suggest that larger and more varied virtual data would increase the benefit of IMUdrape. First, without real data, more virtual data still improve recognition at the largest amount we generated (Section 5.1.1). Second, recorded motion lowers the pose error of every virtual source by more than a quarter without real data (Section 5.2). Generated motion still differs from real recordings and does not yet span the variety of ways in which the recorded participants move. Third, virtual data help most when real data come from only one or two people (Section 4.2), where they add variation between people that the real data lack. Real wearers also differ in build, and the same garment fits them differently. A jacket that is loose on a slim person can be tight on a broader one, which changes how far the fabric moves relative to the body. Generating more descriptions per activity, adopting better motion generators as they appear, and simulating a range of body shapes, each dressed in the garment at its own size or at a fixed size, are therefore natural next steps.

Designing smart garments in simulation. Because a garment in IMUdrape is a sewing pattern, its cut and its sensor layout can be changed and simulated before the garment is produced, so candidate designs for a planned smart garment can be compared before any participant is recorded. Garment design in simulation is already established in computer graphics, where sewing patterns are generated programmatically [41] and optimized with differentiable cloth simulation [52], and in wearable computing, simulation has begun to guide where IMUs are placed on the body [75] and where textile energy harvesters are placed on garments [74]. Optimizing the cut of a garment together with its sensor layout for a downstream task, such as recognizing a given set of activities, has not been explored, and IMUdrape provides the data needed to do so.

Further sensors and modalities. A smartphone carried in a trouser pocket is itself a garment-mounted sensor. Many smartphone HAR datasets were recorded this way, and simulating a phone inside a pocket, as a rigid body moving within the fabric, would extend IMUdrape to them. The simulation also describes the whole garment, not only the points where IMUs sit. It gives the stretch and bending of the fabric everywhere, which is what textile strain and pressure sensors measure, so the same simulations could provide virtual data for these sensors as well. Finally, the generated clips carry the text descriptions they were generated from, so they form IMU–text pairs. Sensor–language models are currently trained on body-mounted virtual data [96], and pre-training them on garment-mounted data could carry their zero-shot recognition over to smart clothing.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

6 Conclusion

We presented IMUdrape, a pipeline that generates garment-mounted virtual IMU data from text by physically simulating garments that are matched to the clothing and sensor layout of a target deployment, and that yields paired garment and body signals for every simulated motion. In a large-scale evaluation on four real datasets and two tasks, IMUdrape Body + Garment data improved activity recognition over body-mounted virtual data when real data came from one subject, and IMUdrape Garment data lowered pose error from real loose-garment IMUs by 12.1% relative to body-mounted virtual data when the garment was simulated on recorded motion. Garment physics matters most when real data are scarce, realistic motion increases its benefit further, and for activity recognition, IMUdrape Garment and IMUdrape Body + Garment data remain the strongest virtual sources even when real data are plentiful. IMUdrape thus turns virtual IMU data from a proxy for body-mounted sensors into a model of the smart garment itself. It reduces the amount of real data that garment-mounted sensing requires, and it lays the groundwork for researchers who want to study smart-clothing sensing but have no recorded data of their own.

Acknowledgments

This work was partially supported by the NSF Research Fellowship under Grant No. DGE-2039655. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Alireza Abedin, Mahsa Ehsanpour, Qinfeng Shi, Hamid Rezatofighi, and Damith C. Ranasinghe. 2021. Attend and Discriminate: Beyond the State-of-the-Art for Human Activity Recognition Using Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–22. doi:10.1145/3448083

[2] Karan Ahuja, Yue Jiang, Mayank Goel, and Chris Harrison. 2021. Vid2Doppler: Synthesizing Doppler Radar Data from Videos for Training Privacy-Preserving Activity Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, 1–10. doi:10.1145/3411764.3445138

[3] Ryoichi Ando. 2024. A Cubic Barrier with Elasticity-Inclusive Dynamic Stiffness. *ACM Transactions on Graphics* 43, 6, Article 224 (Dec. 2024), 13 pages. doi:10.1145/3687908

[4] Oresti Banos, Mate Toth, Miguel Damas, Hector Pomares, and Ignacio Rojas. 2014. Dealing with the Effects of Sensor Displacement in Wearable Activity Recognition. *Sensors* 14, 6 (June 2014), 9995–10023. doi:10.3390/s140609995

[5] David Baraff and Andrew Witkin. 1998. Large steps in cloth simulation. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98 (SIGGRAPH '98)*. ACM Press, 43–54. doi:10.1145/280814.280821

[6] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. 2022. Neural Cloth Simulation. *ACM Transactions on Graphics* 41, 6 (Nov. 2022), 1–14. doi:10.1145/3550454.3555491

[7] Sejal Bhalla, Mayank Goel, and Rushil Khurana. 2021. IMU2Doppler: Cross-Modal Domain Adaptation for Doppler-based Activity Recognition Using IMU Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (Dec. 2021), 1–20. doi:10.1145/3494994

[8] Bharat Bhatnagar, Garvita Tiwari, Christian Theobalt, and Gerard Pons-Moll. 2019. Multi-Garment Net: Learning to Dress 3D People From Images. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 5419–5429. doi:10.1109/iccv.2019.00552

[9] Siyuan Bian, Chenghao Xu, Yuliang Xiu, Artur Grigorev, Zhen Liu, Cewu Lu, Michael J. Black, and Yao Feng. 2025. ChatGarment: Garment Estimation, Generation and Editing via Large Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2924–2934. doi:10.1109/CVPR52734.2025.00278

[10] Robert Bridson, Ronald Fedkiw, and John Anderson. 2002. Robust treatment of collisions, contact and friction for cloth animation. *ACM Transactions on Graphics* 21, 3 (July 2002), 594–603. doi:10.1145/566654.566623

[11] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *Comput. Surveys* 46, 3 (Jan. 2014), 1–33. doi:10.1145/2499621

[12] Youngjae Chang, Akhil Mathur, Anton Isopoussu, Junehwa Song, and Fahim Kawsar. 2020. A Systematic Study of Unsupervised Domain Adaptation for Robust Human-Activity Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (March 2020), 1–30. doi:10.1145/3380985

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

[13] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R. Millán, and Daniel Roggen. 2013. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* 34, 15 (Nov. 2013), 2033–2042. doi:10.1016/j.patrec.2012.12.014

[14] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2021. Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities. *Comput. Surveys* 54, 4, Article 77 (2021), 40 pages. doi:10.1145/3447744

[15] Kunigunde Cherenack and Liesbeth van Pieterson. 2012. Smart textiles: Challenges and opportunities. *Journal of Applied Physics* 112, 9 (Nov. 2012). doi:10.1063/1.4742728

[16] Ishita Datta, Zikang Leng, Rachit Bhayana, and Thomas Plötz. 2026. StitchSense: Smart-Garment Sensing as a Source Domain for Watch-Based Activity Recognition. In *Proceedings of the 2026 ACM International Symposium on Wearable Computers*. ACM, 22–28. doi:10.1145/3830727.3834837

[17] Naser El-Sheimy, Haiying Hou, and Xiaoji Niu. 2008. Analysis and Modeling of Inertial Sensors Using Allan Variance. *IEEE Transactions on Instrumentation and Measurement* 57, 1 (2008), 140–149.

[18] Yu Enokibori, Takahiro Sato, and Kenji Mase. 2025. Where Are the Best Positions of IMUs for HAR? - Investigation with Four DNN Models of Different Characteristics. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1089–1095. doi:10.1145/3714394.3756334

[19] Jiawei Fang, Ruonan Zheng, Xiaoxia Gao, Shifan Jiang, Anjun Chen, Qi Ye, and Shihui Guo. 2026. Garment Inertial Denoiser (GID): Endowing Accurate Motion Capture via Loose IMU Denoiser. arXiv preprint arXiv:2601.01360. arXiv:2601.01360 [cs.CV]

[20] Artur Grigorev, Michael J. Black, and Otmar Hilliges. 2023. HOOD: Hierarchical Graphs for Generalized Modelling of Clothing Dynamics. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 16965–16974. doi:10.1109/cvpr52729.2023.01627

[21] Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. 2022. Generating Diverse and Natural 3D Human Motions from Text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 5142–5151. doi:10.1109/CVPR52688.2022.00509

[22] Lars Ole Haeusler, Lena Uhlenberg, and Oliver Amft. 2025. Dynamic Body-Garment Simulation to Characterise Wearable Activity Recognition Performance. *3DBODY.TECH Journal* 2, 40 (Dec. 2025). doi:10.15221/25.40

[23] Lars Ole Haeusler, Lena Uhlenberg, and Oliver Amft. 2025. Text2IMU: Advancing Human Activity Recognition by Text-Driven IMU Data Synthesis. In *2025 IEEE 21st International Conference on Body Sensor Networks (BSN)*. IEEE, 1–4. doi:10.1109/bsn66969.2025.11337798

[24] Nils Y. Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. arXiv:1604.08880 [cs.LG] https://arxiv.org/abs/1604.08880

[25] Harish Haresamudram, Irfan Essa, and Thomas Plötz. 2022. Assessing the State of Self-Supervised Human Activity Recognition Using Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (Sept. 2022), 1–47. doi:10.1145/3550299

[26] Harish Haresamudram, Chi Ian Tang, Sungho Suh, Paul Lukowicz, and Thomas Plötz. 2025. Past, Present, and Future of Sensor-based Human Activity Recognition Using Wearables: A Surveying Tutorial on a Still Challenging Task. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 2 (June 2025), 1–44. doi:10.1145/3729467

[27] Holger Harms, Oliver Amft, and Gerhard Tröster. 2010. Estimating Posture-Recognition Performance in Sensing Garments Using Geometric Wrinkle Modeling. *IEEE Transactions on Information Technology in Biomedicine* 14, 6 (2010), 1436–1445. doi:10.1109/TITB.2010.2076822

[28] Hexoskin. 2026. Placement of Hexoskin Sensors: Smart Shirts and ProShirts. Retrieved October 25, 2026 from https://support.hexoskin.com/placement-of-sensors-hexoskin-smart-shirts-proshirts

[29] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2106.09685

[30] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J. Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 1–15. doi:10.1145/3272127.3275108

[31] Andela Ilic, Jiaxi Jiang, Paul Streli, Xintong Liu, and Christian Holz. 2025. Human Motion Capture from Loose and Sparse Inertial Sensors with Garment-aware Diffusion Models. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*. International Joint Conferences on Artificial Intelligence Organization, 1206–1214. doi:10.24963/ijcai.2025/135

[32] Yash Jain, Hyeokhyen Kwon, and Thomas Plötz. 2022. On the Effectiveness of Virtual IMU Data for Eating Detection with Wrist Sensors. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22)*. ACM, 50–52. doi:10.1145/3544793.3560337

[33] Udeni Jayasinghe, William S. Harwin, and Faustina Hwang. 2020. Comparing Clothing-Mounted Sensors with Wearable Sensors for Movement Analysis and Activity Classification. *Sensors* 20, 1, Article 82 (2020). doi:10.3390/s20010082

[34] Udeni Jayasinghe, Faustina Hwang, and William S. Harwin. 2023. Inertial Measurement Data from Loose Clothing Worn on the Lower Body during Everyday Activities. *Scientific Data* 10, 1, Article 709 (2023). doi:10.1038/s41597-023-02567-4

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

[35] Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, and Tao Chen. 2023. MotionGPT: Human Motion as a Foreign Language. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 20067–20079. doi:10.52202/075280-0880

[36] Yifeng Jiang, Yuting Ye, Deepak Gopinath, Jungdam Won, Alexander W. Winkler, and C. Karen Liu. 2022. Transformer Inertial Poser: Real-time Human Motion Reconstruction from Sparse IMUs with Simultaneous Terrain Generation. In *SIGGRAPH Asia 2022 Conference Papers (SA '22)*. ACM, 1–9. doi:10.1145/3550469.3555428

[37] Peng Jin, Runtian Jiang, Rong Zheng, Qing Chen, and Jintu Fan. 2025. Smart Bodysuit Integrating Digital Twin Technology for Real-Time Human Motion Monitoring and Visualization. *IEEE Sensors Journal* 25, 20 (Oct. 2025), 38693–38706. doi:10.1109/jsen.2025.3601187

[38] W. Kabsch. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* 32, 5 (Sept. 1976), 922–923. doi:10.1107/s0567739476001873

[39] Abhi Kamboj and Minh Do. 2024. A Survey of IMU Based Cross-Modal Transfer Learning in Human Activity Recognition. arXiv:2403.15444 [eess.SP] https://arxiv.org/abs/2403.15444

[40] Maria Korosteleva and Sung-Hee Lee. 2022. NeuralTailor: reconstructing sewing pattern structures from 3D point clouds of garments. *ACM Transactions on Graphics* 41, 4 (July 2022), 1–16. doi:10.1145/3528223.3530179

[41] Maria Korosteleva and Olga Sorkine-Hornung. 2023. GarmentCode: Programming Parametric Sewing Patterns. *ACM Transactions on Graphics* 42, 6, Article 199 (Dec. 2023), 15 pages. doi:10.1145/3618351

[42] Kai Kunze and Paul Lukowicz. 2008. Dealing with sensor displacement in motion-based onbody activity recognition systems. In *Proceedings of the 10th international conference on Ubiquitous computing (UbiComp08)*. ACM, 20–29. doi:10.1145/1409635.1409639

[43] Hyeokhyen Kwon, Gregory D. Abowd, and Thomas Plötz. 2021. Complex Deep Neural Networks from Large Scale Virtual IMU Data for Effective Human Activity Recognition Using Wearables. *Sensors* 21, 24 (Dec. 2021), 8337. doi:10.3390/s21248337

[44] Hyeokhyen Kwon, Catherine Tong, Harish Haresamudram, Yan Gao, Gregory D. Abowd, Nicholas D. Lane, and Thomas Plötz. 2020. IMUTube: Automatic Extraction of Virtual on-body Accelerometry from Video for Human Activity Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3, Article 87 (Sept. 2020), 29 pages. doi:10.1145/3411841

[45] Hyeokhyen Kwon, Bingyao Wang, Gregory D. Abowd, and Thomas Plötz. 2021. Approaching the Real-World: Supporting Activity Recognition Training with Virtual IMU Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3, Article 111 (Sept. 2021), 32 pages. doi:10.1145/3478096

[46] Zikang Leng, Amitrajit Bhattacharjee, Hrudhai Rajasekhar, Lizhe Zhang, Elizabeth Bruda, Hyeokhyen Kwon, and Thomas Plötz. 2024. IMUGPT 2.0: Language-Based Cross Modality Transfer for Sensor-Based Human Activity Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3, Article 112 (Sept. 2024), 32 pages. doi:10.1145/3678545

[47] Zikang Leng, Yash Jain, Hyeokhyen Kwon, and Thomas Plötz. 2023. On the Utility of Virtual On-body Acceleration Data for Fine-grained Human Activity Recognition. In *Proceedings of the 2023 International Symposium on Wearable Computers (UbiComp/ISWC '23)*. ACM, 55–59. doi:10.1145/3594738.3611364

[48] Zikang Leng, Hyeokhyen Kwon, and Thomas Plötz. 2023. Generating Virtual On-body Accelerometer Data from Virtual Textual Descriptions for Human Activity Recognition. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers* (Cancun, Quintana Roo, Mexico) *(ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 39–43. doi:10.1145/3594738.3611361

[49] Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M. Kaufman. 2020. Incremental Potential Contact: Intersection- and Inversion-free, Large-Deformation Dynamics. *ACM Transactions on Graphics* 39, 4, Article 49 (2020). doi:10.1145/3386569.3392425

[50] Minchen Li, Danny M. Kaufman, and Chenfanfu Jiang. 2021. Codimensional incremental potential contact. *ACM Transactions on Graphics* 40, 4 (July 2021), 1–24. doi:10.1145/3450626.3459767

[51] Xuan Li, Chang Yu, Wenxin Du, Ying Jiang, Tianyi Xie, Yunuo Chen, Yin Yang, and Chenfanfu Jiang. 2025. Dress-1-to-3: Single Image to Simulation-Ready 3D Outfit with Diffusion Prior and Differentiable Physics. *ACM Transactions on Graphics* 44, 4 (July 2025), 1–16. doi:10.1145/3731177

[52] Yifei Li, Hsiao-yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, and Tuur Stuyck. 2024. DiffAvatar: Simulation-Ready Garment Optimization with Differentiable Simulation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4368–4378. doi:10.1109/cvpr52733.2024.00418

[53] Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. 2018. Adaptive Batch Normalization for practical domain adaptation. *Pattern Recognition* 80 (Aug. 2018), 109–117. doi:10.1016/j.patcog.2018.03.005

[54] Qi Lin, Shuhua Peng, Yuezhong Wu, Jun Liu, Hong Jia, Wen Hu, Mahbub Hassan, Aruna Seneviratne, and Chun H. Wang. 2023. Subject-adaptive Loose-fitting Smart Garment Platform for Human Activity Recognition. *ACM Transactions on Sensor Networks* 19, 4 (May 2023), 1–23. doi:10.1145/3584986

[55] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual Instruction Tuning. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 36. https://arxiv.org/abs/2304.08485

[56] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Transactions on Graphics* 34, 6, Article 248 (2015), 16 pages. doi:10.1145/2816795.2818013

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

[57] Michael Lorenz, Gabriele Bleser, Takayuki Akiyama, Takehiro Niikura, Didier Stricker, and Bertram Taetz. 2022. Towards Artefact Aware Human Motion Capture using Inertial Sensors Integrated into Loose Clothing. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 1682–1688. doi:10.1109/ICRA46639.2022.9811933

[58] Qianli Ma, Jinlong Yang, Anurag Ranjan, Sergi Pujades, Gerard Pons-Moll, Siyu Tang, and Michael J. Black. 2020. Learning to Dress 3D People in Generative Clothing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6468–6477. doi:10.1109/cvpr42600.2020.00650

[59] Miles Macklin, Matthias Müller, and Nuttapong Chentanez. 2016. XPBD: position-based simulation of compliant constrained dynamics. In *Proceedings of the 9th International Conference on Motion in Games (MiG '16)*. ACM, 49–54. doi:10.1145/2994258.2994272

[60] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael Black. 2019. AMASS: Archive of Motion Capture As Surface Shapes. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 5441–5450. doi:10.1109/iccv.2019.00554

[61] Shenghuan Miao and Ling Chen. 2026. Wonderwall: A Virtual-to-Real Foundation Model for IMU-based HAR. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 10, 1 (March 2026), 1–31. doi:10.1145/3789688

[62] Brendan Michael and Matthew Howard. 2017. Activity Recognition with Wearable Sensors on Loose Clothing. *PLoS ONE* 12, 10 (2017), e0184642. doi:10.1371/journal.pone.0184642

[63] Mohammad Iman Mokhlespour Esfahani and Maury A. Nussbaum. 2019. Classifying Diverse Physical Activities Using "Smart Garments". *Sensors* 19, 14 (July 2019), 3133. doi:10.3390/s19143133

[64] Vimal Mollyn, Riku Arakawa, Mayank Goel, Chris Harrison, and Karan Ahuja. 2023. IMUPoser: Full-Body Pose Estimation using IMUs in Phones, Watches, and Earbuds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, 1–12. doi:10.1145/3544548.3581392

[65] Movella. 2026. Xsens Link: High-Performance Motion Capture Suit for Professionals. Retrieved October 25, 2026 from https://www.xsens.com/motion-capture/xsens-link-specifications

[66] Vishvak S. Murahari and Thomas Plötz. 2018. On attention models for human activity recognition. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (UbiComp '18)*. ACM, 100–103. doi:10.1145/3267242.3267287

[67] Rahul Narain, Armin Samii, and James F. O'Brien. 2012. Adaptive anisotropic remeshing for cloth simulation. *ACM Transactions on Graphics* 31, 6 (Nov. 2012), 1–10. doi:10.1145/2366145.2366171

[68] Nobuyuki Oishi, Phil Birch, Daniel Roggen, and Paula Lago. 2025. WIMUSim: Simulating Realistic Variabilities in Wearable IMUs for Human Activity Recognition. *Frontiers in Computer Science* 7, Article 1514933 (2025). doi:10.3389/fcomp.2025.1514933

[69] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1, Article 115 (2016). doi:10.3390/s16010115

[70] Manuel Palermo, Sara M. Cerqueira, João André, António Pereira, and Cristina P. Santos. 2022. From Raw Measurements to Human Pose – a Dataset with Low-Cost and High-End Inertial-Magnetic Sensor Data. *Scientific Data* 9, 1, Article 591 (2022). doi:10.1038/s41597-022-01690-y

[71] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. 2020. TailorNet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7363–7373. doi:10.1109/CVPR42600.2020.00739

[72] Thomas Plötz. 2021. Applying Machine Learning for Sensor Data Analysis in Interactive Systems: Common Pitfalls of Pragmatic Use and Ways to Avoid Them. *Comput. Surveys* 54, 6 (July 2021), 1–25. doi:10.1145/3459666

[73] Thomas Plötz. 2023. If only we had more data!: Sensor-Based Human Activity Recognition in Challenging Scenarios. In *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 565–570. doi:10.1109/percomworkshops56833.2023.10150267

[74] Lala Shakti Swarup Ray, Ramyah Gowrishankar, Mengxi Liu, Deepika Gurung, Paul Lukowicz, and Bo Zhou. 2026. SPOTTEG: Simulation-guided Placement Optimization for Textile Triboelectric Energy Generators in Dynamic Wearable Scenarios. In *Proceedings of the 2026 ACM International Symposium on Wearable Computers*. ACM, 142–148. doi:10.1145/3830727.3834836

[75] Lala Shakti Swarup Ray, Bo Zhou, and Paul Lukowicz. 2025. W2W: A Simulated Exploration of IMU Placement Across the Human Body for Designing Smarter Wearable. In *Proceedings of the 2025 ACM International Symposium on Wearable Computers (UbiComp '25)*. ACM, 170–176. doi:10.1145/3715071.3750417

[76] Lala Shakti Swarup Ray, Bo Zhou, Sungho Suh, and Paul Lukowicz. 2023. Selecting the Motion Ground Truth for Loose-fitting Wearables: Benchmarking Optical MoCap Methods. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers* (Cancun, Quintana Roo, Mexico) *(ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 27–32. doi:10.1145/3594738.3611359

[77] Attila Reiss and Didier Stricker. 2012. Introducing a New Benchmarked Dataset for Activity Monitoring. In *2012 16th International Symposium on Wearable Computers*. IEEE, 108–109. doi:10.1109/iswc.2012.13

[78] Vitor Fortes Rey, Peter Hevesi, Onorina Kovalenko, and Paul Lukowicz. 2019. Let there be IMU data: generating training data for wearable, motion sensor based activity recognition from monocular RGB videos. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp '19)*. ACM, 699–708. doi:10.1145/3341162.3345590

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

[79] Rokoko. 2026. Smartsuit Pro II: Professional Body Motion Capture in One Mobile Mocap Suit. Retrieved October 25, 2026 from https://www.rokoko.com/products/smartsuit-pro

[80] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. 2019. Multi-task Self-Supervised Learning for Human Activity Detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (June 2019), 1–30. doi:10.1145/3328932

[81] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. 2019. Learning-Based Animation of Clothing for Virtual Try-On. *Computer Graphics Forum* 38, 2 (May 2019), 355–366. doi:10.1111/cgf.13643

[82] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. 2022. SNUG: Self-Supervised Neural Dynamic Garments. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 8130–8140. doi:10.1109/cvpr52688.2022.00797

[83] Panneer Selvam Santhalingam, Parth Pathak, Huzefa Rangwala, and Jana Kosecka. 2023. Synthetic Smartwatch IMU Data Generation from In-the-wild ASL Videos. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (June 2023), 1–34. doi:10.1145/3596261

[84] Tianchen Shen, Irene Di Giulio, and Matthew Howard. 2023. A Probabilistic Model of Human Activity Recognition with Loose Clothing. *Sensors* 23, 10 (May 2023), 4669. doi:10.3390/s23104669

[85] Tianchen Shen, Sacha Morris, Irene Di Giulio, and Matthew Howard. 2026. Human Motion Recognition and Prediction Using Loose Cloth. *Nature Communications* 17, 1, Article 807 (2026). doi:10.1038/s41467-025-67509-7

[86] Matteo Stoppa and Alessandro Chiolerio. 2014. Wearable Electronics and Smart Textiles: A Critical Review. *Sensors* 14, 7 (July 2014), 11957–11992. doi:10.3390/s140711957

[87] Jie Su, Fengtong Ge, Zhenyu Wen, Taotao Li, Yang Bai, Yejian Zhou, and Xiaoqin Zhang. 2025. IMUZero: Zero-Shot Human Activity Recognition by Language-Based Cross Modality Fusion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (Dec. 2025), 1–28. doi:10.1145/3770669

[88] Timo Sztyler and Heiner Stuckenschmidt. 2016. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–9. doi:10.1109/percom.2016.7456521

[89] Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, Soren Brage, Nick Wareham, and Cecilia Mascolo. 2021. SelfHAR: Improving Human Activity Recognition through Self-training with Unlabeled Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–30. doi:10.1145/3448112

[90] Tencent Hunyuan 3D Digital Human Team. 2025. HY-Motion 1.0: Scaling Flow Matching Models for Text-To-Motion Generation. arXiv preprint arXiv:2512.23464. arXiv:2512.23464 [cs.CV] https://github.com/Tencent-Hunyuan/HY-Motion-1.0

[91] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H. Bermano. 2022. Human Motion Diffusion Model. arXiv:2209.14916 [cs.CV] https://arxiv.org/abs/2209.14916

[92] Akihisa Tsukamoto, Naoto Yoshida, Tomoko Yonezawa, Kenji Mase, and Yu Enokibori. 2023. Where Are the Best Positions of IMU Sensors for HAR? - Approach by a Garment Device with Fine-Grained Grid IMUs -. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico) *(UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 445–450. doi:10.1145/3594739.3610736

[93] Lena Uhlenberg and Oliver Amft. 2022. Comparison of Surface Models and Skeletal Models for Inertial Sensor Data Synthesis. In *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–5. doi:10.1109/bsn56160.2022.9928504

[94] T. von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll. 2017. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Computer Graphics Forum* 36, 2 (May 2017), 349–360. doi:10.1111/cgf.13131

[95] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. 2021. Tent: Fully Test-time Adaptation by Entropy Minimization. arXiv:2006.10726 [cs.LG] https://arxiv.org/abs/2006.10726

[96] Qingxin Wei, Kai Hu, Jiaming Huang, Cheng Guo, Yi Gao, and Wei Dong. 2026. Open-ended Human Activity Understanding via LLM-assisted Motion Decomposition and Semantic Fusion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 10, 3, Article 168 (Sept. 2026), 25 pages. doi:10.1145/3831643

[97] Oliver J. Woodman. 2007. *An Introduction to Inertial Navigation.* Technical Report UCAM-CL-TR-696. University of Cambridge, Computer Laboratory.

[98] Chengshuo Xia and Yuta Sugiura. 2022. Virtual IMU Data Augmentation by Spring-Joint Model for Motion Exercises Recognition without Using Real Data. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers (UbiComp/ISWC '22)*. ACM, 79–83. doi:10.1145/3544794.3558460

[99] Huatao Xu, Pengfei Zhou, Rui Tan, Mo Li, and Guobin Shen. 2021. LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21)*. ACM, 220–233. doi:10.1145/3485730.3485937

[100] Xinyu Yi, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, and Feng Xu. 2022. Physical Inertial Poser (PIP): Physics-aware Real-time Human Motion Tracking from Sparse Inertial Sensors. In *2022 IEEE/CVF Conference on Computer*

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Vision and Pattern Recognition (CVPR). IEEE, 13157–13168. doi:10.1109/cvpr52688.2022.01282

[101] Xinyu Yi, Yuxiao Zhou, and Feng Xu. 2021. TransPose: real-time 3D human translation and pose estimation with six inertial sensors. *ACM Transactions on Graphics* 40, 4 (July 2021), 1–13. doi:10.1145/3450626.3459786

[102] Hyungjun Yoon, Hyeongheon Cha, Hoang C. Nguyen, Taesik Gong, and Sung-Ju Lee. 2025. From Vision to Motion: Translating Large-Scale Knowledge for Data-Scarce IMU Applications. *IEEE Transactions on Mobile Computing* 24, 9 (Sept. 2025), 8656–8667. doi:10.1109/tmc.2025.3556998

[103] Alexander D. Young, Martin J. Ling, and Damal K. Arvind. 2011. IMUSim: A Simulation Environment for Inertial Sensing Algorithm Design and Evaluation. In *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 199–210.

[104] Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Yong Zhang, Hongwei Zhao, Hongtao Lu, Xi Shen, and Ying Shan. 2023. Generating Human Motion from Textual Descriptions with Discrete Representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 14730–14740. doi:10.1109/CVPR52729.2023.01415

[105] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. 2024. MotionDiffuse: Text-Driven Human Motion Generation With Diffusion Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 6 (June 2024), 4115–4128. doi:10.1109/tpami.2024.3355414

[106] Shibo Zhang and Nabil Alshurafa. 2020. Deep generative cross-modal on-body accelerometer data synthesis from videos. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20)*. ACM, 223–227. doi:10.1145/3410530.3414329

[107] Xiyuan Zhang, Diyan Teng, Ranak Chowdhury, Shuheng Li, Dezhi Hong, Rajesh Gupta, and Jingbo Shang. 2024. UniMTS: Unified Pre-training for Motion Time Series. In *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 107469–107493. doi:10.52202/079017-3414

[108] Ruonan Zheng, Jiawei Fang, Yuan Yao, Xiaoxia Gao, Chengxu Zuo, Shihui Guo, and Yiyue Luo. 2025. FIP: Endowing Robust Motion Capture on Daily Garment by Fusing Flex and Inertial Sensors. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 927, 21 pages. doi:10.1145/3706598.3714140

[109] Yexu Zhou, Haibin Zhao, Yiran Huang, Till Riedel, Michael Hefenbrock, and Michael Beigl. 2022. TinyHAR: A Lightweight Deep Learning Model Designed for Human Activity Recognition. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers (UbiComp/ISWC '22)*. ACM, 89–93. doi:10.1145/3544794.3558467

[110] Chengxu Zuo, Yiming Wang, Lishuang Zhan, Shihui Guo, Xinyu Yi, Feng Xu, and Yipeng Qin. 2024. Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2209–2219. doi:10.1109/CVPR52733.2024.00215

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

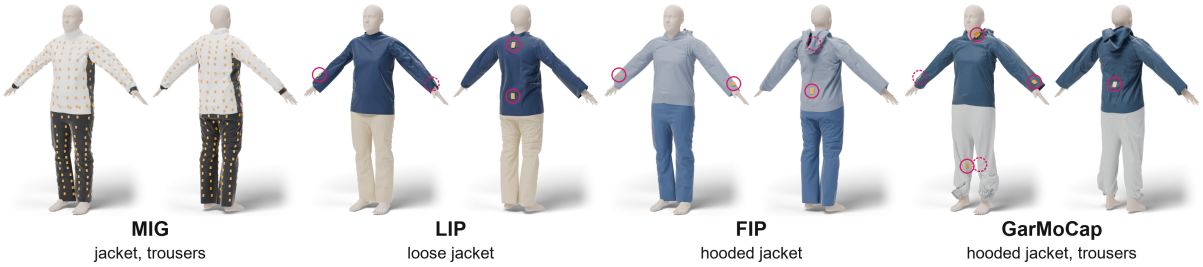


Fig. 6. The garments matched to the four datasets, each sewn onto the body and settled in the A-pose, in a front and a back view. Orange boards mark the sensor positions of the matched dataset (all 396 for MIG). For LIP, FIP and GarMoCap, each position is ringed once: solid in the view where its board is visible, and dashed where the board is hidden in both views, such as on the underside of a sleeve or under the hood, in the view it faces most. MIG: sensor jacket and trousers. LIP: loose outdoor jacket. FIP: loose hooded jacket. GarMoCap: hooded jacket and cuffed trousers.

Table 9. Garments matched to the four datasets. Each garment is reconstructed from the dataset's garment and carries the dataset's sensor positions.

Matched dataset	Garment	Real sensors
MIG [92]	jacket + trousers	396 sewn IMUs
LIP [110]	loose jacket + trousers	4 (cuffs, back, low back)
FIP [108]	loose hooded jacket + trousers	4 (forearms, back, waist)
GarMoCap [19]	loose hooded jacket + trousers	6 (cuffs, back, hem, ankles)

A Simulation Details

This appendix gives further details and the key settings behind Section 3.

A.1 Garments

For every dataset, we generated a garment that matches what its participants wore (Figure 6, Table 9), from photographs of the dataset's garment or, where no usable photograph exists, from the published description. Patterns are sized to the SMPL body with $\beta = 0$, and tight sleeves keep at least 3 cm of ease down to the wrist. Fabric parameters are set per garment from the fabric type (Table 10).

A.2 Sensor Mapping

Each real sensor is defined by a vertex of the SMPL body mesh and the garment it is sewn onto, and its virtual counterpart is the point of that garment closest to the body vertex in the start state. For MIG, we place all 396 sensors from the dataset's sensor-position photographs. The sensor positions of all four datasets are marked on their garments in Figure 6, in a front and a back view, since many of them sit on the back and on the outer forearm.

A.3 Pipeline Details

Sewing-pattern estimation. ChatGarment [9] is the LLaVA vision–language model [55], fine-tuned with LoRA [29]. Given the image X_v and the text prompt X_q, the language model f_ϕ writes a response $Y_t = f_\phi(X_v, X_q)$, or $Y_t = f_\phi(X_q)$ when there is no image. A small network g_Θ reads the continuous values off the model's internal state H_t at the end of the response, $\mathbf{n} = g_\Theta(H_t) \in [0,1]^{76}$, with every value scaled to $[0,1]$. The model was trained

Table 10. Simulation parameters. Top: fabric of the shell of each matched garment (areal density, bending stiffness B) and of the trousers. Bottom: solver and sensor settings shared by all garments.

Garment	Shell fabric	Density (g/m^2)	B (10^{-6} N m)
MIG jacket	Polyester windbreaker taffeta	90	4.7
LIP jacket	Light polyester woven	112	4.6
FIP jacket	Nylon shell with polyester lining	200	11
GarMoCap jacket	Polyester windbreaker taffeta	90	4.7
Trousers (all)	Polyester–cotton twill	230	24
Time step 1/240 s, output 120 Hz; friction coefficient 0.5; stiffness-proportional damping 0.02 s			
Settling 0.2 s; 1.5 s transition from the dressing pose; trousers pinned at the waist like a belt			
Arm–torso clearance of the collider at least 8 mm; elbow and hip crease fillets of 12–35 mm radius			
Rigid sensor patch of radius 2.5 cm on garment and body			

on 40,000 synthetic garments with known designs, with a text loss on the discrete choices and an L_1 loss on the continuous values.

Virtual sewing. Each vertex i on one stitched edge is linked to the matching point on the other edge, which usually falls between two vertices j and k of that edge, at $(1-w)\,\mathbf{x}_j + w\,\mathbf{x}_k$ with $w \in [0,1]$. Each solver step of the sewing stage finds

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \; \Psi_{\mathrm{el}}(\mathbf{x}) + \sum_i \psi_{\mathrm{weak}}(g_i, \hat{g}, \bar{\kappa}_i) + \sum_{(i,j,k,w)} \frac{\kappa_s}{2} \left\| \mathbf{x}_i - (1-w)\,\mathbf{x}_j - w\,\mathbf{x}_k \right\|^2, \tag{9}$$

where Ψ_{el} is the elastic energy of the fabric, the barriers ψ_{weak} keep the garment outside the body (Equation 5) and the last term pulls every stitch shut with stiffness κ_s.

Arm–torso clearance. Let $d_v(t)$ be the distance from an arm vertex v to the torso in frame t, $\mathbf{n}_v$ the unit vector from v toward the closest torso point and d^* the target clearance. Where the arm is closer to the torso than d^*, the arm vertex moves away from the torso by a fraction α of the missing clearance, and the closest torso vertex u moves away from the arm by the remaining fraction,

$$\Delta\mathbf{x}_v(t) = -\,\alpha \left[d^* - d_v(t) \right]_+ \mathbf{n}_v, \qquad \Delta\mathbf{x}_u(t) = -\,(1-\alpha) \left[d^* - d_u(t) \right]_+ \mathbf{n}_u, \tag{10}$$

where $\mathbf{n}_u$ points from u toward the arm, $[\cdot]_+ = \max(\cdot, 0)$ and $\alpha = 0.5$ shares the correction equally.

T-pose calibration. The virtual wearer holds a T-pose before the motion, and the garment settles on the raised arms. The stored rotation O_T is the average of $B(t)^\top S(t)$ at the end of that hold, where $S(t)$ is the orientation of the garment sensor and $B(t)$ that of the body sensor beneath it, and the reported orientation is $S(t)\,O_T^\top$. Because raising the arms changes how a sleeve hangs, this simulated hold matters most for sensors on the sleeves.

A.4 Simulation Parameters

Table 10 lists the fabric of each garment and the settings shared by all simulations. The collider settings follow from the body. With the default SMPL body shape, the upper arm of a generated walk sits about 22 mm inside the torso, because SMPL's torso is fuller than that of the body the motion model was trained on. The collider therefore pushes arm and torso apart to at least 8 mm clearance, without moving any sensor position. The creases that linear blend skinning forms at flexed elbows, at the hip and at the crotch are rounded with a radius of 12–22 mm at the elbow and 15–35 mm at the hip, which changes the garment IMU only within the run-to-run noise of the solver. Each motion is preceded by a 1.5 s transition from the dressing pose into its first frame. The solver

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

uses a time step of 1/240 s with output at 120 Hz, a friction coefficient of 0.5, stiffness-proportional damping of 0.02 s and a pin that holds the trousers at the waist like a belt. It sets the stiffness of the barrier of each contact i (Equation 5) as

$$\bar{\kappa}_i = \frac{m_i}{g_i^2} + \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \cdot \left(H \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \right), \tag{11}$$

where m_i is the mass of the contacting vertices, $\mathbf{w}_i$ the contact direction distributed to them and H the Hessian of the elastic energy [3]. The first term stops the fabric's momentum within the gap, and the second resists the elastic force that presses the fabric onto the body.

B Datasets

Table 11 summarizes the four datasets of Section 4.1, their activities and the amount of real and virtual data. Figure 7 shows examples of the virtual data, the simulated garments in motion.

Table 11. Datasets used in the evaluation and the virtual data generated for each. Real: recorded data with garment-mounted IMUs. Virtual: simulated motion on the garment matched to the dataset's clothing. Real and virtual amounts in minutes.

Dataset	Subjects	IMUs	Modalities	Activities	Real	Virtual
MIG [92]	12	396 (6 used)	Acc., gyro.	Eating a cookie, eating pasta, folding, going downstairs, going upstairs, ironing, jogging, sitting, standing, toothbrushing, typing, vacuuming, walking, writing	432	89
LIP [110]	5	4	Acc., orientation	Running, walking, jumping, boxing, ping-pong	118	12
FIP [108]	10	4	Acc., orientation	Walking, running, boxing, jumping rope, chest expansion, lifting weights, dribbling, drumming, crossing hands, ping-pong	103	24
GarMoCap [19]	9	6	Acc., orientation	Basketball, boxing, casual movement, jumping, jumping rope, ping-pong, rotating, running, tennis, walking	93	21

C Evaluation Details

We use three activity recognition models, DeepConvLSTM [69], DeepConvLSTM with attention [66] and TinyHAR [109], and two pose estimation networks, an LSTM poser and TransPose [101]. This section describes their inputs and architectures and how they are trained.

Activity recognition. Real and virtual signals are resampled to 20 Hz and cut into 2 s windows (40 samples) with 50% overlap. Every sensor contributes its accelerometer (LIP, FIP and GarMoCap) or its accelerometer and gyroscope (MIG) in its own frame. The datasets' activity classes are used as published, and every generated clip is labeled with the class of its description. DeepConvLSTM has four convolutional layers of 64 filters with kernel size 5, followed by a two-layer LSTM with 128 units and dropout of 0.5. DeepConvLSTM with attention uses the same trunk and adds a temporal self-attention head over the LSTM outputs, whose result is added to the output of the last time step. TinyHAR processes each channel with four convolutional layers of 20 filters (kernel size 5, strides 1, 2, 1 and 2, batch normalization), lets the channels interact through a transformer layer, fuses them with a fully connected layer and aggregates over time with an LSTM and temporal attention. All three are trained with Adam at a learning rate of 10^{-3} and a batch size of 128, for 30 epochs on the real and virtual windows together.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information



Fig. 7. IMUdrape garments in motion. Each row shows frames of one generated activity with the physics-simulated jacket and trousers.

Pose estimation. LIP and FIP sequences are used at 30 Hz and GarMoCap sequences at 60 Hz. The input of every frame is the free acceleration and the orientation of each garment sensor relative to the root sensor, and the target is the SMPL joint rotations and joint positions. Virtual orientations are calibrated as the datasets calibrate their real ones, with a simulated T-pose calibration on all three datasets (Section 3.3). The LSTM poser has the shape of the network of Loose Inertial Poser [110]. A linear layer with 256 units, tanh activation and dropout of 0.2 feeds a two-layer LSTM with 256 units, which regresses the joint rotations and positions of every frame. TransPose is used with its pose branch of three bidirectional LSTM stages (hidden sizes 256, 64 and 128). They estimate the leaf-joint positions, then all joint positions, then the joint rotations, with noise added to the intermediate

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 12. Activity recognition performance of each model on the LIP and FIP datasets (macro-F1, %): DeepConvLSTM (DCL), DeepConvLSTM with attention (DCL-Att) and TinyHAR, averaged over the test subjects. The best result in each block and column is highlighted.

	LIP			FIP		
Training data	DCL	TinyHAR	DCL-Att	DCL	TinyHAR	DCL-Att
Virtual data only						
IMUGPT	48.3 ± 1.8	59.5 ± 2.5	47.7 ± 6.4	15.2 ± 1.1	17.5 ± 0.9	15.8 ± 1.4
TailorNet	66.9 ± 1.8	66.1 ± 2.7	62.5 ± 5.5	21.1 ± 1.9	20.5 ± 1.0	20.8 ± 0.5
IMUdrape Body	65.8 ± 0.9	66.0 ± 3.6	65.3 ± 9.7	16.8 ± 2.6	20.5 ± 1.1	18.8 ± 1.6
IMUdrape Garment	54.3 ± 3.9	64.4 ± 4.6	57.6 ± 4.3	20.6 ± 2.0	22.9 ± 1.1	20.7 ± 2.2
IMUdrape Body + Garment	71.4 ± 2.8	71.5 ± 2.2	73.0 ± 5.5	23.4 ± 0.8	24.0 ± 0.7	22.0 ± 1.1
1 real subject						
Real data only	61.3 ± 3.0	75.5 ± 5.8	63.9 ± 4.8	15.8 ± 2.3	21.2 ± 8.5	16.0 ± 2.7
IMUGPT	58.0 ± 2.4	74.3 ± 6.5	66.8 ± 6.0	15.7 ± 2.5	20.7 ± 6.2	17.8 ± 1.6
TailorNet	75.9 ± 4.9	76.2 ± 1.9	77.3 ± 1.5	22.8 ± 1.9	27.1 ± 2.9	21.9 ± 0.2
IMUdrape Body	79.4 ± 8.0	78.7 ± 1.3	76.5 ± 6.5	22.8 ± 0.6	26.7 ± 5.4	20.4 ± 1.8
IMUdrape Garment	76.3 ± 2.1	81.6 ± 4.5	75.6 ± 5.0	24.2 ± 1.4	30.9 ± 2.6	21.7 ± 2.1
IMUdrape Body + Garment	80.0 ± 3.6	80.1 ± 3.3	79.8 ± 4.5	25.3 ± 1.0	29.9 ± 3.0	22.5 ± 1.0
2 real subjects						
Real data only	76.9 ± 3.6	83.1 ± 2.6	80.2 ± 4.3	27.9 ± 4.1	32.7 ± 3.6	26.3 ± 5.0
IMUGPT	75.4 ± 2.8	84.3 ± 2.6	75.8 ± 3.9	25.7 ± 3.2	30.6 ± 6.3	24.6 ± 1.1
TailorNet	82.9 ± 1.4	83.3 ± 3.9	80.4 ± 2.6	32.1 ± 2.7	34.2 ± 4.3	31.4 ± 4.9
IMUdrape Body	86.0 ± 0.7	83.9 ± 2.8	84.3 ± 1.9	33.4 ± 3.3	34.6 ± 2.8	31.3 ± 0.7
IMUdrape Garment	84.6 ± 2.2	87.2 ± 2.1	83.1 ± 5.4	33.2 ± 2.6	38.5 ± 2.7	33.9 ± 2.1
IMUdrape Body + Garment	84.2 ± 1.8	86.0 ± 0.8	84.2 ± 2.9	34.9 ± 2.3	37.0 ± 2.3	34.6 ± 1.9
All real subjects						
Real data only	85.7 ± 1.2	89.4 ± 1.3	87.2 ± 0.3	56.8 ± 0.7	54.5 ± 0.7	55.9 ± 1.9
IMUGPT	85.5 ± 1.5	90.7 ± 1.5	85.8 ± 4.4	55.2 ± 1.1	53.7 ± 1.6	53.4 ± 1.4
TailorNet	86.6 ± 1.8	90.5 ± 1.1	87.6 ± 1.4	57.8 ± 0.6	53.4 ± 0.6	56.6 ± 1.2
IMUdrape Body	87.6 ± 0.7	90.6 ± 0.6	86.7 ± 2.7	56.5 ± 2.1	51.5 ± 1.5	54.0 ± 2.0
IMUdrape Garment	87.6 ± 1.6	89.8 ± 1.0	88.0 ± 0.7	56.3 ± 2.7	55.7 ± 0.7	57.7 ± 0.5
IMUdrape Body + Garment	87.1 ± 0.3	89.6 ± 1.5	87.9 ± 2.8	59.0 ± 2.1	55.6 ± 1.8	56.9 ± 0.6

positions during training as in the original work. It runs online with a window of 20 past and 5 future frames. Both networks are trained on random 128-frame crops with a batch size of 64 and Adam at a learning rate of 10^{-3}, reduced by a factor of 0.3 halfway. The loss is the error of the rotations plus ten times the error of the positions. On LIP and FIP for both networks, and on GarMoCap for TransPose, pre-training runs for 2,000 or 5,000 iterations and fine-tuning on real data for 2,000 or 4,000 iterations at 0.1 or 0.3 times the learning rate. These three settings are tuned as hyperparameters per network and amount of real data; without real data, 2,000 iterations of pre-training are used throughout. On GarMoCap, the LSTM poser is instead trained on 256-frame crops for 6,000 iterations and fine-tuned for 3,000 iterations at a learning rate of 3×10^{-4}.

D Per-Model Results

Tables 12 and 13 (recognition) and Tables 14 and 15 (pose: joint angle error and MPJPE) report the results of each recognition model and each pose network behind Tables 2 and 3.

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 13. Activity recognition performance of each model on the GarMoCap and MIG datasets (macro-F1, %): DeepConvLSTM (DCL), DeepConvLSTM with attention (DCL-Att) and TinyHAR, averaged over the test subjects. The best result in each block and column is highlighted.

	GarMoCap			MIG		
Training data	DCL	TinyHAR	DCL-Att	DCL	TinyHAR	DCL-Att
Virtual data only						
IMUGPT	20.9 ± 1.2	30.7 ± 3.7	21.3 ± 0.8	5.0 ± 0.4	11.0 ± 0.6	5.7 ± 1.5
TailorNet	25.1 ± 1.4	33.8 ± 1.9	27.2 ± 1.9	20.2 ± 2.7	39.9 ± 3.7	22.4 ± 4.9
IMUdrape Body	28.0 ± 1.4	33.0 ± 3.6	26.3 ± 3.1	35.0 ± 0.9	44.9 ± 0.4	33.3 ± 1.5
IMUdrape Garment	24.9 ± 2.9	33.5 ± 0.5	21.5 ± 1.6	40.9 ± 0.8	47.1 ± 3.1	42.6 ± 0.9
IMUdrape Body + Garment	25.8 ± 4.8	32.0 ± 2.3	24.8 ± 1.6	41.0 ± 1.3	52.4 ± 0.8	43.5 ± 1.7
1 real subject						
Real data only	40.3 ± 2.6	54.3 ± 4.3	42.8 ± 2.4	34.1 ± 0.7	48.0 ± 3.8	37.3 ± 4.6
IMUGPT	37.6 ± 1.8	49.2 ± 3.4	36.8 ± 1.9	27.5 ± 3.1	47.0 ± 4.1	28.3 ± 4.4
TailorNet	45.8 ± 1.1	52.4 ± 1.7	44.4 ± 3.6	38.8 ± 2.3	51.9 ± 3.5	40.4 ± 3.2
IMUdrape Body	43.5 ± 1.2	53.6 ± 1.5	44.7 ± 0.6	45.4 ± 0.8	55.3 ± 3.5	45.4 ± 2.3
IMUdrape Garment	46.1 ± 1.2	52.6 ± 0.3	45.2 ± 3.9	45.9 ± 4.5	55.4 ± 4.3	48.2 ± 1.8
IMUdrape Body + Garment	48.5 ± 2.9	53.2 ± 1.3	45.6 ± 3.6	45.2 ± 6.5	56.1 ± 5.2	47.8 ± 6.6
2 real subjects						
Real data only	53.8 ± 4.7	61.6 ± 2.0	54.4 ± 2.5	44.9 ± 4.6	59.6 ± 1.5	45.1 ± 1.0
IMUGPT	51.7 ± 1.2	60.4 ± 1.9	51.3 ± 2.7	40.4 ± 0.8	57.9 ± 2.8	42.3 ± 4.2
TailorNet	53.3 ± 0.7	61.2 ± 3.9	56.0 ± 3.6	48.9 ± 1.5	58.4 ± 1.9	45.4 ± 5.2
IMUdrape Body	53.9 ± 3.4	60.6 ± 1.9	55.9 ± 2.4	47.9 ± 1.3	60.5 ± 2.4	49.4 ± 2.5
IMUdrape Garment	55.4 ± 1.9	64.1 ± 1.8	54.8 ± 4.5	53.7 ± 0.9	63.3 ± 1.6	54.8 ± 2.1
IMUdrape Body + Garment	53.3 ± 4.6	64.3 ± 1.5	58.3 ± 0.6	52.5 ± 2.6	62.6 ± 1.5	54.2 ± 1.1
All real subjects						
Real data only	75.3 ± 2.1	75.3 ± 0.7	73.7 ± 1.1	65.8 ± 1.8	69.2 ± 0.4	66.9 ± 1.0
IMUGPT	73.2 ± 1.9	76.4 ± 0.8	73.8 ± 0.5	64.2 ± 1.0	69.7 ± 1.6	62.4 ± 2.5
TailorNet	74.1 ± 2.2	76.0 ± 1.8	74.6 ± 0.7	65.7 ± 1.1	71.3 ± 0.6	67.2 ± 2.5
IMUdrape Body	73.6 ± 1.0	75.2 ± 1.1	73.3 ± 2.3	66.6 ± 3.5	71.2 ± 0.7	65.5 ± 0.4
IMUdrape Garment	76.3 ± 3.2	77.8 ± 0.3	76.2 ± 2.5	69.3 ± 3.4	73.1 ± 1.4	70.5 ± 0.9
IMUdrape Body + Garment	76.1 ± 2.5	75.9 ± 1.8	75.2 ± 1.3	67.8 ± 2.0	73.6 ± 1.4	70.3 ± 2.6

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 14. Pose estimation performance of each network: mean angular error over 15 major joints (degrees), lower is better. The best result in each block and column is highlighted.

	LIP		FIP		GarMoCap	
Training data	LSTM	TransPose	LSTM	TransPose	LSTM	TransPose
Virtual data only						
IMUGPT	33.5 ± 1.24	29.2 ± 1.07	28.3 ± 0.38	25.2 ± 0.49	24.2 ± 0.44	21.5 ± 0.18
TailorNet	29.8 ± 0.38	28.7 ± 0.16	25.4 ± 0.50	25.6 ± 0.86	24.1 ± 0.77	22.7 ± 0.24
IMUdrape Body	28.7 ± 0.57	28.5 ± 0.33	26.2 ± 0.53	25.5 ± 0.16	23.7 ± 0.62	23.2 ± 0.30
IMUdrape Garment	28.8 ± 0.39	28.4 ± 0.59	23.1 ± 0.68	24.2 ± 0.25	21.9 ± 0.29	20.3 ± 0.22
IMUdrape Body + Garment	28.3 ± 0.32	28.3 ± 0.31	24.0 ± 0.48	24.7 ± 0.45	21.6 ± 0.09	20.6 ± 0.11
1 real subject						
Real data only	23.3 ± 1.17	22.5 ± 1.40	15.2 ± 1.11	14.9 ± 0.97	14.4 ± 0.61	13.5 ± 0.60
IMUGPT	23.1 ± 1.20	22.3 ± 1.23	15.5 ± 1.07	15.0 ± 0.94	14.4 ± 0.40	13.1 ± 0.53
TailorNet	22.9 ± 1.25	22.4 ± 1.09	15.4 ± 1.18	14.8 ± 0.71	14.2 ± 0.48	13.2 ± 0.55
IMUdrape Body	23.0 ± 1.18	22.3 ± 1.09	15.6 ± 1.47	15.1 ± 1.12	14.3 ± 0.49	13.2 ± 0.46
IMUdrape Garment	23.3 ± 1.33	22.4 ± 1.19	14.8 ± 0.64	14.6 ± 0.73	13.8 ± 0.58	12.9 ± 0.50
IMUdrape Body + Garment	23.1 ± 1.36	22.5 ± 1.29	15.1 ± 0.89	14.8 ± 0.71	13.6 ± 0.56	12.9 ± 0.40
2 real subjects						
Real data only	21.0 ± 0.48	20.3 ± 0.66	14.0 ± 0.38	13.7 ± 0.15	12.7 ± 0.11	12.1 ± 0.27
IMUGPT	20.9 ± 0.35	20.0 ± 0.41	14.1 ± 0.40	13.7 ± 0.22	12.6 ± 0.13	11.9 ± 0.24
TailorNet	20.9 ± 0.44	20.1 ± 0.44	14.0 ± 0.43	13.7 ± 0.34	12.7 ± 0.23	11.8 ± 0.15
IMUdrape Body	21.0 ± 0.45	20.3 ± 0.33	13.9 ± 0.33	13.7 ± 0.28	12.6 ± 0.15	11.8 ± 0.19
IMUdrape Garment	20.8 ± 0.46	20.1 ± 0.30	13.9 ± 0.26	13.5 ± 0.30	12.7 ± 0.26	11.8 ± 0.30
IMUdrape Body + Garment	20.9 ± 0.37	20.2 ± 0.39	13.9 ± 0.35	13.5 ± 0.36	12.4 ± 0.09	11.6 ± 0.23
All real subjects						
Real data only	19.4 ± 0.11	18.8 ± 0.07	12.6 ± 0.09	12.3 ± 0.02	11.0 ± 0.13	10.1 ± 0.04
IMUGPT	19.3 ± 0.03	18.8 ± 0.21	12.6 ± 0.12	12.2 ± 0.04	11.1 ± 0.20	10.3 ± 0.08
TailorNet	19.5 ± 0.07	18.9 ± 0.09	12.5 ± 0.11	12.2 ± 0.03	11.2 ± 0.04	10.3 ± 0.06
IMUdrape Body	19.3 ± 0.09	18.8 ± 0.10	12.6 ± 0.08	12.1 ± 0.06	10.9 ± 0.06	10.3 ± 0.03
IMUdrape Garment	19.5 ± 0.16	18.7 ± 0.08	12.4 ± 0.06	12.0 ± 0.07	11.0 ± 0.08	10.2 ± 0.07
IMUdrape Body + Garment	19.4 ± 0.04	18.8 ± 0.02	12.4 ± 0.12	12.1 ± 0.06	10.8 ± 0.09	10.1 ± 0.14

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information

Table 15. Pose estimation performance of each network: mean joint position error (MPJPE, cm), lower is better. The best result in each block and column is highlighted.

	LIP		FIP		GarMoCap	
Training data	LSTM	TransPose	LSTM	TransPose	LSTM	TransPose
Virtual data only						
IMUGPT	12.2 ± 0.61	10.4 ± 0.48	9.1 ± 0.08	8.1 ± 0.23	11.6 ± 0.17	10.0 ± 0.04
TailorNet	10.7 ± 0.18	10.3 ± 0.12	8.6 ± 0.11	8.6 ± 0.25	10.7 ± 0.40	10.3 ± 0.09
IMUdrape Body	10.4 ± 0.29	10.3 ± 0.16	8.9 ± 0.21	8.6 ± 0.19	10.4 ± 0.15	10.2 ± 0.33
IMUdrape Garment	9.7 ± 0.11	9.6 ± 0.10	7.9 ± 0.38	8.1 ± 0.13	9.9 ± 0.09	9.2 ± 0.26
IMUdrape Body + Garment	9.9 ± 0.04	9.9 ± 0.09	8.2 ± 0.09	8.3 ± 0.20	9.7 ± 0.04	9.2 ± 0.17
1 real subject						
Real data only	7.9 ± 0.35	7.6 ± 0.39	4.9 ± 0.47	4.7 ± 0.39	7.1 ± 0.33	6.6 ± 0.34
IMUGPT	7.9 ± 0.38	7.5 ± 0.37	5.1 ± 0.47	4.8 ± 0.41	7.0 ± 0.26	6.4 ± 0.24
TailorNet	7.8 ± 0.37	7.5 ± 0.38	5.0 ± 0.51	4.7 ± 0.30	6.9 ± 0.34	6.4 ± 0.38
IMUdrape Body	7.9 ± 0.39	7.6 ± 0.37	5.1 ± 0.63	4.9 ± 0.52	7.0 ± 0.35	6.4 ± 0.34
IMUdrape Garment	8.0 ± 0.43	7.5 ± 0.32	4.7 ± 0.27	4.6 ± 0.32	6.8 ± 0.31	6.3 ± 0.31
IMUdrape Body + Garment	7.9 ± 0.48	7.6 ± 0.44	4.8 ± 0.42	4.7 ± 0.32	6.7 ± 0.34	6.2 ± 0.26
2 real subjects						
Real data only	7.2 ± 0.07	6.8 ± 0.16	4.3 ± 0.15	4.2 ± 0.09	6.2 ± 0.05	5.8 ± 0.06
IMUGPT	7.2 ± 0.02	6.9 ± 0.06	4.4 ± 0.23	4.3 ± 0.09	6.1 ± 0.05	5.6 ± 0.04
TailorNet	7.2 ± 0.12	6.8 ± 0.10	4.4 ± 0.13	4.3 ± 0.13	6.2 ± 0.16	5.6 ± 0.08
IMUdrape Body	7.2 ± 0.06	6.9 ± 0.03	4.3 ± 0.19	4.3 ± 0.09	6.2 ± 0.06	5.6 ± 0.08
IMUdrape Garment	7.2 ± 0.12	6.8 ± 0.07	4.3 ± 0.09	4.2 ± 0.11	6.1 ± 0.18	5.6 ± 0.05
IMUdrape Body + Garment	7.2 ± 0.02	6.9 ± 0.08	4.3 ± 0.16	4.2 ± 0.14	6.0 ± 0.13	5.5 ± 0.09
All real subjects						
Real data only	6.6 ± 0.03	6.4 ± 0.05	3.8 ± 0.03	3.7 ± 0.01	5.3 ± 0.03	4.9 ± 0.01
IMUGPT	6.7 ± 0.05	6.4 ± 0.07	3.9 ± 0.04	3.7 ± 0.01	5.3 ± 0.08	5.0 ± 0.01
TailorNet	6.7 ± 0.02	6.4 ± 0.04	3.9 ± 0.04	3.7 ± 0.02	5.4 ± 0.04	5.0 ± 0.01
IMUdrape Body	6.6 ± 0.03	6.4 ± 0.01	3.9 ± 0.01	3.7 ± 0.02	5.3 ± 0.04	5.0 ± 0.04
IMUdrape Garment	6.7 ± 0.08	6.4 ± 0.06	3.8 ± 0.02	3.7 ± 0.03	5.3 ± 0.06	4.9 ± 0.01
IMUdrape Body + Garment	6.7 ± 0.03	6.4 ± 0.04	3.8 ± 0.03	3.7 ± 0.03	5.2 ± 0.05	4.9 ± 0.02

This manuscript is under review. Please write to zleng7@gatech.edu for up-to-date information