Dissertation Proposal:
Dynamics-Inspired Hidden Parameter Learning for Simulation-Based Virtual Try-On

Junbang Liang
University of Maryland, College Park
College Park, MD 20785
liangjb@cs.umd.edu

March 30, 2020

1 Introduction

E-Commerce has been growing at a rapid pace in recent years. People are now more likely to shop online than to go to physical stores. Digital try-on systems, as one important way to improve the user experience and popularize online garment shopping, has drawn attention of many researchers [GCH+12]. However, the technology is still far from being practical and easy-to-use to replace physical try-on, mostly due to the gap in modeling and in demonstrating garment fitting between the digital and the real worlds.

There are several reasons why customers still prefer physical try-on. First, consumers are unsure if what they buy online will fit their bodies well. Although there exist general sizing systems for individuals, its lack of consistency and standardization across different brands and garment materials can often make it difficult to sizing the clothes, especially for those with non-standard body shapes and proportion. Accurate estimation of human body shapes is the key to make digital try-on work. Second, the fabric material is usually one of the key considerations when shopping for clothes. Different fabric materials affect how the garments look and fit on a body, how customers would wear it, and whether or not they would buy it. However, the correspondences between the actual material and its digital representation are not well understood, not to mention an accurate material estimation and digital cloning from the real-world examples.

Visual effects from the customers' view is as critical as other factors. There are two common presentations of garments: 2D image-based and 3D mesh with photo-realistic rendering. They have different advantages and drawbacks, but both need a large garment database for support. While creating a 3D garment model takes considerable labor, 2D images often suffer from the lack of variation and it is much more difficult to make customized changes. In either case, the try-on system would need a user-friendly design and manipulation backend to meet the customer's needs. Last, but not least, a fast and vivid animation of the garments in motion, along with the body movement, can considerably improve the user experience. Although it is not as critical as other factors, realistic visual rendering could effectively reduce the perceptual gap between the real-world and the virtual garments for online shopping.

Although previous methods have made some progresses on these under-constrained problems, learning-based approaches have shown tremendous potential in making notable impact. We propose to address the key open research issues above by adopting machine learning and optimization techniques. These include:

- Fast and realistic visual rendering of animated try-on results;
- Accurate reconstruction of human shapes and sizes through consumer devices;
- Faithful estimation of garment materials via learning and optimization; and
- User-friendly recovery of garment geometry for rapid prototyping.

1.1 Thesis Statement

Dynamic constraints can be effectively enforced in simulation acceleration, human body estimation, and garment reconstruction for virtual try-on systems, by coupling machine learning and optimization methods with cloth simulation.

2 Related Work

In this section, we discuss previous works on cloth simulation and acceleration, hidden parameter estimation, and differentiable physics.

2.1 Cloth Simulation

Simulation of cloth and deformable bodies has been extensively studied for a wide range of applications in different areas, from computer graphics, CAD/CAM, robotics and automation, to textile engineering. Due to their ability to take large time steps, implicit or semi-implicit methods [GHF+07, VMTF09, Zel05, BWK03] have been widely adopted after the seminal work by Baraff and Witkin [BW98]. However, most of these works focus on the serial simulation improvement and their runtime performances can be slow.

2.1.1 Parallel Cloth Simulation

Parallelization is a popular, practical way to achieve performance improvement. Several parallelization techniques for cloth simulation have been proposed. [WY16, FTP16] proposed GPU-based simulation methods for elastic bodies. [MRB+99, RRZ00, KB04, TB06, ZFV02] proposed different types of spatial parallelization but they all suffer from severe sub-linear scalability due to large communication overhead. [NKT15] improved the work from [AVGT12] using Asynchronous Contact Mechanics and reduced the communication by proposing a locality-aware task assignment, which first scaled more than 16 cores. [TWT+16] implemented a GPU-based simulation pipeline. Their method has achieved an impressive speedup of 58 times.

2.2 Human Body Reconstruction

Human body recovery has gained substantial interest due to its importance in a large variety of applications, such as virtual environments, computer animation, and garment modeling. However, the problem itself is naturally ambiguous, given limited input and occlusion. Recently a number of methods have been proposed to improve the 3D pose estimation with calibrated multi-view input, either using LSTM [TGM+17, NCV+19], auto-encoder [RSF18, TGHC18] or heat map refinement [PZDD17b, TTAR18]. They mainly focus on 3D joint positions without parameterization, thus not able to articulate and animate. Choy *et al*. [CXG+16] proposed an LSTM-based shape recovery network for general objects. Varol *et al*. [VCR+18] proposed a 2-step estimation on human pose and shape. However, both methods are largely limited by the resolution due to the voxel representation. Kanazawa *et al*. [KBJM18] used an iterative correction framework and regularized the model using a learned discriminator. Since they do not employ any supervision other than joint positions, the shape estimation can be inaccurate, especially, when the person is relatively over-weighted.

2.3 Use of Synthetic Dataset

Since it is often time- and labor-intensive to gather a dataset large enough for training a deep neural network, an increasing amount of attention is drawn to synthetic dataset generation. Recent studies [CWL+16, YLL17] have shown that using a synthetic dataset, if sufficiently close to the real-world data, is helpful in training neural networks for real tasks. Varol *et al*. [VRM+17] built up a dataset (SURREAL) which contains human motion sequences with clothing using the SMPL model and CMU MoCap data [CMU03]. Recent works [SPMF19, AMB+19] also generate synthetic data to assist training, but their datasets have only very limited variance on pose, shape, and textures to prevent from overfitting.

2.4 Differentiable Physics

With recent advances in deep learning, there has been increasing interest in differentiable physics simulation, which can be combined with other learning methods to provide physically consistent predictions. Belbute-Peres *et al*. [dABSA+18] and Degrave *et al*. [DHDw19] proposed rigid body simulators using a static formulation of the linear complementarity problem (LCP) [Cot09, Cli02]. Toussaint *et al*. [TAST18] developed a robot reasoning system that can achieve user-defined tasks and is based on differentiable primitives. Hu *et al*. [HLS+19] implemented a differentiable simulator for soft robots based on the Material Point Method (MPM). They store the object data at every simulation step so that the gradient can be computed out of the box. Schenck and Fox [SF18] embedded particle-based fluid dynamics into convolutional neural networks, with precomputed signed distance functions for collision handling. They solved or avoided collisions by assuming special object shapes, transferring to an Eulerian grid, or solving the corresponding collision constraint equation.

None of these methods can be applied to cloth simulation. First, cloth is a 2D surface in a 3D world; thus methods that use an Eulerian grid to compute material density, such as MPM [HLS+19], are not applicable. Second, the collision constraints in cloth simulation are more dynamic and complex given the high number of degrees of freedom; thus constructing a static dense LCP for the entire system [dABSA+18, DHDw19] or constructing the overall state transition graph [TAST18] is inefficient and usually impossible for cloth of common resolution, since contact can happen for every edge-edge or vertex-face pair. Lastly, the shape of cloth changes constantly so self-collision cannot be handled by precomputed signed distance functions [SF18].

2.5 Garment Geometry Modeling and Estimation

Garment model generation has attracted attention these days due to its importance in both real-world and virtual garment design application. Although professional tools, such as Marvelous Designer [2018], can help design high-quality garment models, it may take an excessive amount of time to use it. Several studies have addressed this issue by introducing an automatic generation pipeline to improve the efficiency. Assuming different priors, most previous studies lie in three categories: sketch-based, image-based, and depth-based.

Sketch-based methods. Generating garment models with sketches is one of the most popular ways. This approach takes one or more sketches as input and generates the garment model. Turquin *et al*. [TCH07] and Decaudin *et al*. [DJW+06] developed some of the early work in this area. They used grid and geometric methods to generate garment models with sketches. However, the garment models generated by these methods have limited visual quality. Later, Robson *et al*. [RMSC11] proposed a context-aware method to make the generated garment model more realistic based on a set of observations on key factors which could affect the shapes of garments. These models are, however, fixed to a given body shape. Jung *et al*. [JHR+15] proposed a method to model 3D developable surfaces with a multi-view sketch input. Recently, Huang *et al*. [HYZ16] proposed a realistic 3D garment generation algorithm based on front and back image sketches, but it cannot retarget the generated garments to other body shapes easily. Wang *et al*. [WCPM18] proposed an algorithm that can achieve retargeting conveniently, but is limited to very few topology, namely T-shirts or skirts. In addition, a common limitation using the sketch-based algorithm is that they require domain knowledge on garment sketching.

Image-based or depth-based methods. Other information such as images can also be used to generate a garment model. Bradley *et al*. [BPS+08] researched early on markerless image-based garment modeling using multi-view images. Later, Zhou *et al*. [ZCF+13] proposed a single-view image approach. In their work, a human shape was estimated from the image and the garment model was reconstructed with the garment outline. Jeong *et al*. [JHK15] created the garment model with a single photograph by detecting the landmark points of the garment. Yang [YPA+18] made a full use of garment and human body databases to generate the garment models from a single–view image. Daněřek *et al*. 's [DDÖ+17] method can estimate the 3D garment shape from a single image using deep neural networks. Tex2Shape [APMTM19] is an image-to-image translation model for detailed full-body geometry reconstruction. MGN [BTTPM19] predicts body shape and clothing, layered on top of the SMPL [LMR+15] model from a few (typically 1 - 8) frames of a video. Depth information can also be useful. Chen *et al*. [CZL+15] proposed a method to generate garment models given an RGBD sequence of a worn garment. However, these methods require photos or depth images from a real garment, which means they cannot generate a garment model from size parameters only.

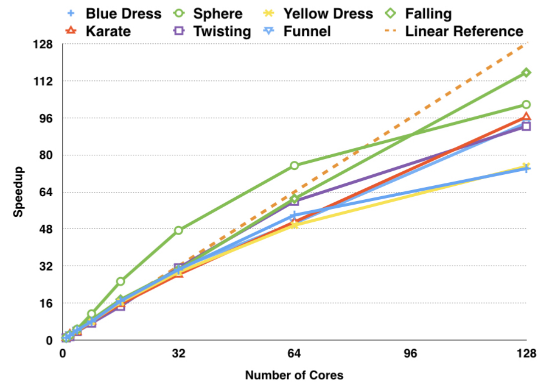


Figure 1: Performance scaling result with large low-res time step. A nearly linear scalability is achieved.

Table 1: Results on a higher-resolution mesh. We run our system on meshes of higher resolution. Values in the table are the corresponding speedup of 128-core compared to single-core running time.

Scenario	Blue Dress	Yellow Dress	Sphere	Falling	Karate	Twisting	Funnel
Original size speedup	74.1	75.0	102	116	96.4	92.3	93.8
4x large size speedup	99.6	109	178	119	103	101	108

3 Research Summary

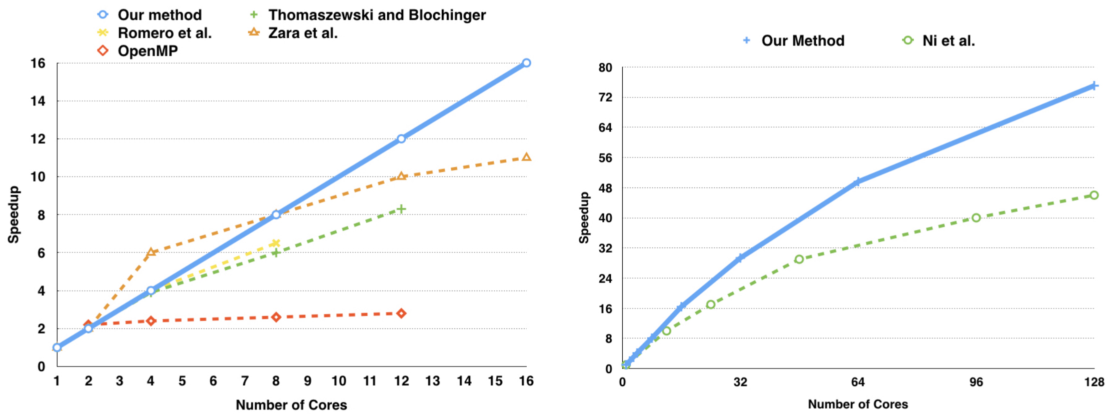
In this section, we discuss our work on cloth simulation and inverse problem. First, we propose an algorithm to parallelize cloth simulation, which divides the workload in *time* domain that minimizes the communication overhead (Sec. 3.1). Next, we propose a scalable neural network framework to reconstruct the 3D mesh of a human body from multi-view images and show that the learning benefits from the synthetic dataset generated from cloth simulation since it has good flexibility of variable control and can provide ground-truth for validation (Sec. 3.2). We then propose a differentiable cloth simulator that can be embedded as a layer in deep neural networks. We demonstrate the potential of differentiable cloth simulation in a number of application scenarios, such as physical parameter estimation and motion control of cloth (Sec. 3.3). Finally, we propose a pipeline for end-to-end learning of human appearance as a whole, including body parameters, garment geometry, and its material (Sec. 3.4).

3.1 Time-Domain Parallelization for Accelerating Cloth Simulation

Significant progress has been achieved in visual simulation of cloth over the past decades [GHF+07, ZY01, BFA02, BW98]. Numerous algorithms have been proposed that achieve high accuracy and robustness for various 3D graphics applications, though real-time simulation remains illusive for complex simulation scenarios. Several parallelization techniques for cloth simulation have been proposed. [MRB+99, RRZ00, KB04, TB06, ZFV02] proposed different types of spatial parallelization but they all suffer from severe sub-linear scalability due to large communication overhead.

We propose a novel method that divides the workload in *time* domain that minimizes the communication overhead, thereby achieving much better scalability and higher performance gain over previous methods. The key challenge in time-domain parallelization is to obtain or approximate the simulation states before the time-consuming simulation begins. We use a two-level mesh representation to address this time-dependency issue. Observing that a coarse-level mesh can be simulated at a much higher speed, our method runs a lower-resolution simulation using coarser meshes to approximate the state at each time step. After an appropriate remeshing process, the higher-resolution simulations using finer meshes can be run in parallel. To further refine the simulation results, we propose a practical technique to smooth the state transition from the low-resolution to high-resolution simulations. To recover the lost states, we make use of the coarse-level mesh and run several 'static' simulation steps before the high-resolution simulation starts.

We conduct experiments on various simulation scenarios to test the scalability of our method. As indicated in



(a) (b)

Figure 2: Small- and large-scale parallelization comparison. Our method (in blue solid line) achieves a linear speedup at small scale, while others are limited by the communication overhead due to spatial domain partitioning. Our method also achieves about 50% higher efficiency than [NKT15] at the large scale using dynamic workload balancing.

Fig. 1, our method achieves a good scalability with an increasing number of processors. The reason of the super-linear speedup in the 'Sphere' scene is that it contains rapidly changing contacts with obstacles. When the cloth is free from contact after the sphere passes through, the remeshing algorithm of ARCSim failed to simplify the mesh effectively, spending an unnecessarily large amount of time simulating simple flat cloth. However, due to the nature of our two-level structure, we maintain a reasonably small number of mesh elements while preserving the quality, and therefore outperform the serial approach significantly. We tested our method on a higher-resolution mesh and observed an even better speedup (Table 1) due to the same reason.

We also compare the performance of our method against other CPU parallelization techniques. Fig. 2(a) shows that in smaller-scale systems (less than 16 cores), our method can maintain a linear speedup with respect to the single-core system, scaling better compared to previous CPU-based methods using spatial-domain partitioning, e.g. 11x over 16 cores by [ZFV04]. For larger-scale systems (Fig. 2(b)), we achieved about 50% more efficiency than previous methods such as [NKT15]. In these methods, the processors need to send the information to each other, typically several times, when solving the linear system, resulting in large communication overhead and limited scalability. In contrast, our method only needs to share the states from low-resolution simulations once. Therefore, our method can achieve greater scalability and efficiency in comparison.

3.2 Shape-Aware Human Pose and Shape Reconstruction Using Multi-View Images

Human body reconstruction, consisting of pose and shape estimation, has been widely studied in a variety of areas, including digital surveillance, computer animation, special effects, and virtual/augmented environments. Yet, it remains a challenging partly because the problem itself is naturally ambiguous, given limited input and occlusion. Previous works reduce this ambiguity using different assumptions and input data. They consist of four main categories: pose from images [PZDD17a, TMNSF17, TRA17, ZHS+17], pose and shape from images under tight clothing [CKC10, DJÖ+16, HAR+10, JTST10], scanned meshes [PPHB17, HSR+09, WPB+14], and images with loose clothing [BB08, BKL+16, LRK+17]. One of the most important metrics used in these methods is the difference between the original and the estimated silhouette. As a result, these methods cannot be directly applied to images where the human wears loose garments, e.g. long coat, evening gown. For works that handle loose clothing, they either relaxed the loss on clothed regions, or incorporating a silhouette energy term on SMPLify [BKL+16], which can be easily degraded when the skin detector is not helpful, and has introduced errors inherently from the optimization.

To tackle this problem, we propose a learning-based *shape-aware* human body mesh reconstruction using SMPL parameters for both pose and shape estimation that is supervised directly on shape parameters. A scalable, end-to-end, multi-view multi-stage learning framework is developed to account for the ambiguity of the 3D human body (geometry) reconstruction problem from 2D images, achieving improved estimation results. Our proposed framework

(a) The input image. (b) Our result. (c) HMR.

Figure 3: Prediction results compared to HMR. Our model can better capture the shape of the human body. The recovered legs and chest are closer to the person in the image.

uses a recurrent structure, making it a universal model applicable to the input of any number of views. At the same time, it couples the shareable information across different views so that the human body pose and shape can be optimized using image features from all views. Since real-world datasets suffer from limited foreground/background textures and ground-truth pose and shape parameters, we make use of synthetic data as additional training samples so that the model can be trained to be more shape-aware. Compared to previous learning-based methods, our model is more shape-aware due to the extra supervision from our synthetic dataset.

We tested our model on datasets using multi-view images to demonstrate the strength of our framework. We use *Mean Per Joint Position Error* (MPJPE) of the 14 joints of the body. As shown in Table 2, under the same training condition, our model in single-view has similar, if not better, results in all experiments. Meanwhile, our model in multi-view achieves much higher accuracy. The reason why the accuracy drops after jointly-trained with synthetic data is that the training data and the test data are not on the same distribution.

Method	MPJPE w/ syn. training	MPJPE w/o syn. training
HMR [KBJM18]	60.14	58.1
Ours (single)	58.55	59.09
Ours (multi)	**45.13**	**44.4**

Table 2: Comparison results on Human3.6M using MPJPE. Smaller errors implies higher accuracy.

For shape estimation, other than MPJPE for joint accuracy, we use the Hausdorff distance between two meshes to capture the shape difference to the ground-truth. Our single-view results are not carefully tuned for the experiment. We directly used single-view input for our multi-view-trained model, so its accuracy may not be as good as the baseline. But as shown in Table 3, our model with multi-view input achieves the smallest error values, when compared to two other baselines.

Method	MPJPE/HD w/ syn. training	MPJPE/HD w/o syn. training
HMR	42/83	89/208
Ours (single)	44/65	102/283
Ours (multi)	**27/53**	**84**/273

Table 3: Comparison results on our synthetic dataset in MPJPE/Hausdorff Distance(HD). Better results have lower values.

After joint-training with synthetic data, all models perform better in shape estimation, while maintaining similar results using other metrics (Table 2), i.e. they do not overfit. The joint errors of the HMR [KBJM18] are fairly good, so they can still recognize the synthesized human in the image. However, a larger Hausdorff distance indicates that

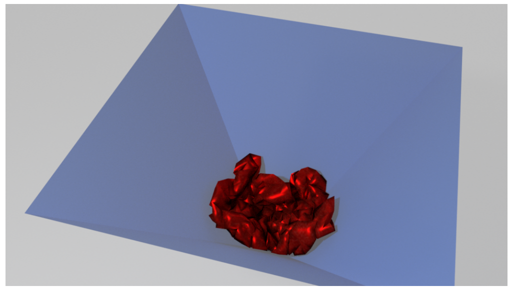


Figure 4: Example frame from the ablation study. A piece of cloth is crumpled inside a square pyramid, so as to generate a large number of collisions.

they lose precision on the shape recovery. Adding our synthetic datasets for training can effectively address this issue and thereby provide better shape estimation. We achieved a much smaller Hausdorff distance (with syn. training) even only using single view. This is because our refinement framework is effectively deeper, aiming at not only the pose but also the shape estimation, which is much more challenging than the pose-only estimation. With the same method, multi-view inputs can further improve the accuracy of shape recovery compared to results using only one single-view image.

We also tested our model on other online images, where no such measurement can be done. As shown in Fig. 3, HMR [KBJM18] can predict the body pose but fails on inferring the person's shape. On the contrary, our model not only refines the relative leg orientations but also largely respects and recovers the original shape of the body. More examples are shown in our supplemental document and video.

3.3 Differentiable Cloth Simulation for Inverse Problems

Differentiable physics simulation is a powerful family of techniques that applies gradient-based methods to learning and control of physical systems [dABSA+18, DHDw19, TAST18, HLS+19, SF18]. It can enable optimization for control, and can also be integrated into neural network frameworks for performing complex tasks. Our work focuses on cloth simulation, which relates to applications in robotics, computer vision, and computer graphics [CSM+11, MvdBF+12, BXBF13, YLL17, PPHB17, LCT18, GCS+19]. Our goal is to enable differentiable cloth simulation, which can provide a unified approach to a variety of inverse problems that involve cloth.

Differentiable cloth simulation is challenging due to a number of factors, which include the high dimensionality of cloth (as compared for example to rigid bodies [dABSA+18]) and the need to handle contacts and collision. For example, a simple 16×16 grid-based cloth mesh has 289 vertices, 867 variables, and 512 faces when triangulated. Typical resolutions for garments would be at least many thousands, if not millions, of vertices and faces. Previous work that tackled differentiable simulation with collisions set up a static linear solver to account for all constraints [dABSA+18]. In our simple example with cloth, the number of pairwise constraints would be at least $289\times512 = 140K$ for vertex-face collisions alone, which renders existing methods impractical even for this simple system. Even if a dynamic solver is applied upon collision, solving a dense linear system with such high dimensionality makes the gradient computation infeasible.

In this work, we propose a differentiable cloth simulation algorithm that overcomes the above difficulties. In general, we follow the computation flow of the common approach to cloth simulation: discretization using the finite element method [EKS03], integration using implicit Euler [BW98], and collision response on impact zones [NSO12, HVTG08].

Collision handling in our implementation is based on impact zone optimization [NSO12]. It finds all colliding instances using continuous collision detection and sets up the constraints for all collisions. In order to introduce minimum change to the original mesh state, we develop a Quadratic Programming (QP) problem to solve for the constraints. Since the signed distance function is linear in $\mathbf{x}$, the optimization takes a quadratic form.

We use implicit differentiation in the linear solve and the optimization in order to compute the gradient with respect to the input parameters. The discontinuity introduced by the collision response is negligible because the discontinuous

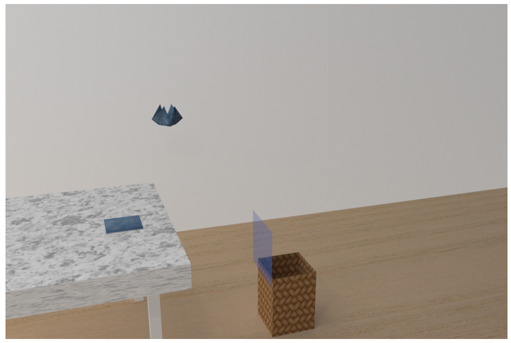


Figure 5: Example frame from the motion control experiment: dropping cloth into a basket.

states constitute a zero-measure set. During the backpropagation in the optimization, the gradient values can be directly computed after QR decomposition of the constraint matrix, which is much smaller than the original linear system and is often of low rank. The intuition behind the QR decomposition is as follows. When perturbing the original point in an optimization, the resulting displacement of the optimized output will be moving along the surface of the constraint, which will become perpendicular to the normal when the perturbation is small. This approach reduces the gradient computation to a linear system of a small upper triangular matrix (the R component of the decomposition), which enables fast simulation and backpropagation.

Mesh	Baseline		Ours		Speedup	
resolution	Matrix size	Runtime (s)	Matrix size	Runtime (s)	Matrix size	Runtime
16x16	599 ± 76	0.33 ± 0.13	$\mathbf{66 \pm 26}$	$\mathbf{0.013 \pm 0.0019}$	8.9	25
32x32	1326 ± 23	1.2 ± 0.10	$\mathbf{97 \pm 24}$	$\mathbf{0.011 \pm 0.0023}$	13	112
64x64	2024 ± 274	4.6 ± 0.33	$\mathbf{242 \pm 47}$	$\mathbf{0.072 \pm 0.011}$	8.3	64

Table 4: Statistics of the backward propagation with and without our method for various mesh resolutions. We report the average values in each cell with the corresponding standard deviations. By using our method, the runtime of gradient computation is reduced by up to *two orders of magnitude*.

We conduct an ablation study to verify this estimate in practice. In order to clearly measure the timing difference, we design a scenario with many collisions. We place a piece of cloth into an upside-down square pyramid, so that the cloth is forced to fold, come into frequent contact with the pyramid, and self collide, as shown in Fig. 4.

We measure the running time of backpropagation in each quadratic optimization and also the running time of the physics solver as a reference. With all other variables fixed, we compare to the baseline method where the gradients are computed by directly solving the linear system from the implicit differentiation. Timings are listed in Tab. 4. In this experiment, the backpropagation of the physics solve takes from 0.007s to 0.5s, which, together with the timings of the baseline, implies that the collision handling step is the critical bottleneck when there are many collisions in the

Method	Error (%)	Samples
Point mass	111	–
PPO [LLN+18]	432	10,000
Ours	**17**	**53**
Ours+FC	39	108

Table 5: Motion control results. The table reports the smallest distance to the target position, normalized by the size of the cloth, and the number of samples used during training. 'Ours+FC' means we added a simple fully-connected network for control instead of optimizing the signals directly.

scene. The results in Tab. 4 show that our proposed method can significantly decrease the matrix size required for computation and thus the actual running time, resolving the bottleneck in backpropagation.

We further demonstrate the capability of our differentiable simulator by optimizing control parameters. The task is to drop a piece of cloth into a basket, as shown in Fig. 5. Tab. 5 shows the performance of the different methods and their sample complexity. The error shown in the table is the distance defined above normalized by the size of the cloth. Our method achieves the best performance with a much lower number of simulation steps. The bottom of the basket in our setting has the same size as the cloth, so a normalized error of less than 50%, as our methods achieve, implies that the cloth is successfully dropped into the basket.

3.4 Vid2Avatar: Human Body Appearance Reconstruction from RGB Video (In Progress)

Recreating virtual human that is perceptually similar to the real human is an important and powerful technique that can be applied in many virtual applications, such as virtual try-on. Although capturing human poses and shapes has been extensively studied recently [LL19, KBJM18], approaches that estimate garment properties are very rare, partly due to the difficulty of modeling garments as a whole in the parameter space. Existing approaches either learn the garment model in each category [YFHWW16], or develop the garment based on the displacement from the body [APMTM19, BTTPM19], both of which suffer from lack of various garment topologies. Moreover, physical material is also a critical aspect when estimating the garment properties. Different materials can lead to very different garment motions and human perceptions, even when their geometry structures are the same to start with. Previous work [YLL17] learn cloth materials from a fixed and controlled scenario, which is not applicable to garment material captured from daily human motion in video. While these issues are closely intertwined with each other, no previous work has addressed them simultaneously.

Inspired from multi-tasking and iterative optimization techniques that are proven successful in other tasks, we propose to jointly learn human body, cloth geometry, and cloth materials together in an end-to-end fashion. The main challenges are:

1. A unified parametric representation for all types of clothing.
2. Information sharing between different tasks.

We first set up an auto-encoder for the cloth model. Since the model is designed not to assume fixed garment topology, we choose to use point cloud as the underlying representation. Previous point cloud auto-encoders such as AtlasNet [DGF+19] use MLPs to transform a 2D patch to a set of 3D points in the space. Their method performs well in point cloud datasets that include rigid objects such as airplanes or chairs, since the deformations presented in those objects are simple and regular. However, it cannot be directly applied to learn garment point clouds since garments have much larger variance in point cloud distribution due to its dynamic nature. For example, a simple dress can create different wrinkle structures under different external forces. As a result, one global auto-encoder cannot account for all detail structure, resulting in smooth and blurry point clouds.

Inspired from PointNet++ [QYSG17] where it used a two-level encoder for segmentation tasks, we propose a two-level auto-encoder for learning the latent space of the cloth. We first do a farthest point sampling to get the representative points, which is a subset of the entire point cloud. We start from the center of mass so that the sampled result is stable. Second, we group the other points by nearest neighbor to form a set of patches, each contains one representative point. We then feed the representative point cloud and the patches to different PointNets to obtain the global and local latent vectors. On the decoder side, we first use a global AtlasNet to recover the representative points. Then we use each of the points as a condition to decode the patches accordingly, using a local AtlasNet. Finally, we combine the patches and the representative points to get the full reconstructed point cloud.

The major differences between our method and an naive combination of PointNet++ and AtlasNet are: (a) we use a two-level decoder structure, where the local decoder is conditioned on the global output, (b) we start the farthest point sampling from the center of mass, which ensures stability of the sampled result, (c) our structure inherently separates different patches of the cloth, avoiding interwound or vanished results as seen in AtlasNet, and (d) we condition the global decoder on the human body parameters, which can help improve the accuracy of the reconstructed points.

In this task, we used Chamfer Distance between two point clouds as the loss:

$$d(\mathbf{P},\mathbf{Q}) = \frac{1}{|\mathbf{P}|}\sum_{\mathbf{p}\in\mathbf{P}}\min_{\mathbf{q}\in\mathbf{Q}}\|\mathbf{p}-\mathbf{q}\|^2 + \frac{1}{|\mathbf{Q}|}\sum_{\mathbf{q}\in\mathbf{Q}}\min_{\mathbf{p}\in\mathbf{P}}\|\mathbf{q}-\mathbf{p}\|^2 \tag{1}$$

After convergence on the auto-encoder, we train a model to estimate human body and cloth geometry given one single frame. We follow the network structure from the recent state-of-the-art [KBJM18] for the body estimation. For the cloth geometry estimation, we take as input the ResNet50 features of the image, sharing with the body estimation, and regress the latent parameters of the garment point cloud. We define the loss function using Chamfer distance for the supervised learning as well, but we fix the parameters in the decoder for consistent training. There is also one additive output for corrective values of the human body estimation from this branch, enabling bidirectional information sharing.

Lastly, we train the material parameters using temporal information given by the human body and the cloth geometry. Since material parameters are less likely to have impact on the body and cloth estimation, we do not introduce the same additive output here, as we do in the cloth estimation branch. In this task, we use TCN [BKK18] for temporal information gathering and processing. Following previous work [YLL17], we discretize the material parameter space into different categories according to their dynamics similarity, and treat the material estimation task as a classification problem.

In order to train such a model, we need a large number of samples that contain ground truth human body parameters, garment meshes, and the corresponding material parameters, which are nearly impossible to capture in real world. Therefore, we choose to create a large synthetic dataset that can control these variables and generate videos with the corresponding ground truth data. We sample human motion sequences and shape parameters in the CMU Mocap dataset [CMU03], collect different garment meshes online and register them onto the initial (T-Posed) body. By simulating the garment movement with the recorded body motion, we obtain a set of different garment geometry under different body motions, shapes, and garment materials. Finally, we render the sequence with random textures (both on human and garment surface) and background lighting to generate the synthetic videos.

The steps to be completed for this work are:

- Train and evaluate the garment estimation model.
- Train and evaluate the material estimation model.
- Test the model in synthetic dataset and real-world videos to evaluate the generalizability of this approach.

4 Expected Contributions

In summary, my research mainly focuses on improving efficiency, scalability, and capability of cloth simulation, and enabling accurate hidden parameter estimation by exploiting cloth simulation for supervised learning and gradient-based feedback control. My contributions can be categorized as follows:

- **Time-domain parallelized cloth simulation:** We introduce a novel temporal-domain parallelization method for practical cloth simulation such as rapid design prototyping. Taking the advantage of faster simulations on coarser meshes, we parallelize the cloth simulation in time with accelerated computation and minimal communication overhead. We also proposed an iterative detail recovery algorithm to minimize the visual artifacts due to the state transitioning from coarse to fine meshes. Our method outperforms existing CPU- and GPU-based parallelization techniques on a diverse set of benchmarks. It offers high efficiency and nearly linear scalability on large distributed systems, while maintaining high-fidelity visual simulation of the cloth. The scalability of our method is dependent on the ratio of low- to high-resolution simulation time, the length of the simulation, and persistence of contacts with obstacles. This work is already published and is avaiable at:

  http://gamma.cs.unc.edu/TParallelCloth/

- **Simulated data-assisted shape-aware human body estimation:** We proposed a novel multi-view multi-stage framework for pose and shape estimation. The framework is trained on datasets with at most 4 views but can be naturally extended to an arbitrary number of views. Moreover, we introduced a physically-based synthetic data generation pipeline to enrich the training data, which is very helpful for shape estimation and regularization of end effectors that traditional datasets do not capture. Experiments have shown that our trained model can provide equally good pose estimation as state-of-the-art using single-view images, while providing considerable improvement on pose estimation using multi-view inputs and a better shape estimation across all datasets. This work is already published and is avaiable at:

https://gamma.umd.edu/researchdirections/virtualtryon/humanmultiview

- **Differentiable cloth simulation for learning and control:** We presented a differentiable cloth simulator that can compute the analytical gradient of the simulation function with respect to the input parameters. We used dynamic collision handling and explicitly derived its gradient. Implicit differentiation is used in computing gradients of the linear solver and collision response. Experiments have demonstrated that our method accelerates backpropagation by up to two orders of magnitude. We have demonstrated the potential of differentiable cloth simulation in two application scenarios: material estimation and motion control. By making use of the gradients from the physically-aware simulation, our method can optimize the unknown parameters faster and more accurately than gradient-free baselines. Using differentiable simulation, we can learn the intrinsic properties of cloth from observation. This work is already published and is avaiable at:

  https://gamma.umd.edu/researchdirections/virtualtryon/differentiablecloth

- **Joint estimation of human body, cloth geometry and material from videos:** We propose a novel end-to-end learning framework that jointly estimates the human body, cloth geometry, and cloth material. By training the model in a multi-tasking manner, the accuracy of all tasks can be improved from each other. We are also the first to estimate the cloth geometry and its material in videos of common human actions, which is useful in a wide range of applications. Compared to previous optimization-based material estimation work, we have much faster inference speed. Compared to previous garment estimation work, we can support arbitrary topologies because of the generality of the garment representation. This work is still ongoing and is planned to submit to NeurIPS 2020.

5 Dissertation & Department Requirements Schedule

5.1 Current Progress

- Fall 2016
  - Worked on learning-based cloth material recovery methods.
- Spring 2017
  - Submitted "Learning-Based Cloth Material Recovery from Video" [YLL17] to International Conference on Computer Vision 2017, accepted.
  - Worked on time parallelization for cloth simulation.
- Summer 2017
  - Interned at Google (New York City, NY).
- Fall 2017
  - Continue to work on time parallelization for cloth simulation.
- Spring 2018
  - Submitted "Time Domain Parallelization for Cloth Simulation" [LL18] to SIGGRAPH 2018, rejected.
  - Submitted "Time Domain Parallelization for Cloth Simulation" [LL18] to Symposium on Computer Animation 2018, accepted.

- Worked on Shape-aware multi-view human body estimation.

- Summer 2018
  - Interned at Google (Sunnyvale, CA).
  - Presented "Time Domain Parallelization for Cloth Simulation" [LL18] at Symposium on Computer Animation 2018.

- Fall 2018
  - Submitted "Shape-Aware Human Pose and Shape Reconstruction Using Multi-View Images" [LL19] to Conference on Computer Vision and Pattern Recognition 2019, rejected.

- Spring 2019
  - Submitted "Shape-Aware Human Pose and Shape Reconstruction Using Multi-View Images" [LL19] to International Conference on Computer Vision 2019, accepted.
  - Worked on differentiable cloth simulation.
  - Submitted "Differentiable Cloth Simulation for Inverse Problems" to Conference on Neural Information Processing Systems 2019, accepted.

- Summer 2019
  - Interned at Facebook Reality Lab (Sausalito, CA).

- Fall 2019
  - Worked on generative model for garment geometry.
  - Worked on differentiable simulation for coupling rigid body and cloth.
  - Submitted "GAN based garment generation using sewing pattern image" to Conference on Computer Vision and Pattern Recognition 2020, rejected.

- Spring 2020
  - Submitted "Scalable Differentiable Physics for Learning and Control" to International Conference on Machine Learning 2020.
  - Worked on joint learning of human body, garment geometry and material from videos.
  - Submitted "GAN based garment generation using sewing pattern image" to European Conference on Computer Vision 2020.

5.2 Future Plans

- Spring 2020
  - Submit the proposal and pass the Preliminary Examination.
  - Submit the joint learning work to Conference on Neural Information Processing Systems 2020.

- Fall 2020
  - Work on latent space learning of cloth simulation.
  - Submit the latent learning work to Conference on Computer Vision and Pattern Recognition 2021.

- Spring 2021
  - Work on Dissertation and prepare for Dissertation Defense.
  - Defend Dissertation.

References

[AMB+19] Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, and Gerard Pons-Moll. Learning to reconstruct people in clothing from a single rgb camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1175–1186, 2019.

[APMTM19] Thiemo Alldieck, Gerard Pons-Moll, Christian Theobalt, and Marcus Magnor. Tex2shape: Detailed full human body geometry from a single image. pages 2293–2303, 2019.

[AVGT12] Samantha Ainsley, Etienne Vouga, Eitan Grinspun, and Rasmus Tamstorf. Speculative parallel asynchronous contact mechanics. *ACM Trans. Graph.*, 31(6):151:1–151:8, November 2012.

[BB08] Alexandru O Bălan and Michael J Black. The naked truth: Estimating body shape under clothing. In *European Conference on Computer Vision*, pages 15–29. Springer, 2008.

[BFA02] Robert Bridson, Ronald Fedkiw, and John Anderson. Robust treatment of collisions, contact and friction for cloth animation. *ACM Transactions on Graphics (ToG)*, 21(3):594–603, 2002.

[BKK18] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[BKL+16] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In *European Conference on Computer Vision*, pages 561–578. Springer, 2016.

[BPS+08] Derek Bradley, Tiberiu Popa, Alla Sheffer, Wolfgang Heidrich, and Tamy Boubekeur. Markerless garment capture. *ACM Trans. Graph.*, 27(3):99, 2008.

[BTTPM19] Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, and Gerard Pons-Moll. Multi-garment net: Learning to dress 3d people from images. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2019.

[BW98] David Baraff and Andrew Witkin. Large steps in cloth simulation. In *SIGGRAPH*, 1998.

[BWK03] David Baraff, Andrew Witkin, and Michael Kass. Untangling cloth. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 862–870. ACM, 2003.

[BXBF13] Katherine L. Bouman, Bei Xiao, Peter Battaglia, and William T. Freeman. Estimating the material properties of fabric from video. In *International Conference on Computer Vision (ICCV)*, 2013.

[CKC10] Yu Chen, Tae-Kyun Kim, and Roberto Cipolla. Inferring 3d shapes and deformations from single views. In *European Conference on Computer Vision*, pages 300–313. Springer, 2010.

[Cli02] Michael Bradley Cline. *Rigid Body Simulation with Contact and Constraints*. PhD thesis, University of British Columbia, 2002.

[CMU03] CMU. Carnegie-mellon mocap database. created with funding from nsf eia- 0196217, 2003.

[Cot09] Richard W Cottle. *Linear Complementarity Problem*. Springer, 2009.

[CSM+11] Marco Cusumano-Towner, Arjun Singh, Stephen Miller, James F. O'Brien, and Pieter Abbeel. Bringing clothing into desired configurations with limited perception. In *International Conference on Robotics and Automation (ICRA)*, 2011.

[CWL+16] Wenzheng Chen, Huan Wang, Yangyan Li, Hao Su, Zhenhua Wang, Changhe Tu, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. Synthesizing training images for boosting human 3d pose estimation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 479–488. IEEE, 2016.

[CXG+16] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *European conference on computer vision*, pages 628–644. Springer, 2016.

[CZL+15] Xiaowu Chen, Bin Zhou, Fei-Xiang Lu, Lin Wang, Lang Bi, and Ping Tan. Garment modeling with a depth camera. *ACM Trans. Graph.*, 34(6):203:1–203:12, 2015.

[dABSA+18] Filipe de Avila Belbute-Peres, Kevin A. Smith, Kelsey Allen, Josh Tenenbaum, and J. Zico Kolter. End-to-end differentiable physics for learning and control. In *Advances in Neural Information Processing Systems*, 2018.

[DDÖ+17] R. Danerek, Endri Dibra, A. Cengiz Öztireli, Remo Ziegler, and Markus H. Gross. Deepgarment : 3d garment shape estimation from a single image. *Comput. Graph. Forum*, 36(2):269–280, 2017.

[DGF+19] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir Kim, Bryan Russell, and Mathieu Aubry. Learning elementary structures for 3d shape generation and matching. In *Advances in Neural Information Processing Systems*, pages 7433–7443, 2019.

[DHDw19] Jonas Degrave, Michiel Hermans, Joni Dambre, and Francis wyffels. A differentiable physics engine for deep learning in robotics. *Frontiers in Neurorobotics*, 13, 2019.

[DJÖ+16] Endri Dibra, Himanshu Jain, Cengiz Öztireli, Remo Ziegler, and Markus Gross. Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 108–117. IEEE, 2016.

[DJW+06] Philippe Decaudin, Dan Julius, Jamie Wither, Laurence Boissieux, Alla Sheffer, and Marie-Paule Cani. Virtual garments: A fully geometric approach for clothing design. *Comput. Graph. Forum*, 25(3):625–634, 2006.

[EKS03] Olaf Etzmuß, Michael Keckeisen, and Wolfgang Straßer. A fast finite element solution for cloth modelling. In *Pacific Conference on Computer Graphics and Applications*, 2003.

[FTP16] Marco Fratarcangeli, Valentina Tibaldo, and Fabio Pellacini. Vivace: A practical gauss-seidel method for stable soft body dynamics. *ACM Transactions on Graphics (TOG)*, 35(6):214, 2016.

[GCH+12] Stevie Giovanni, Yeun Chul Choi, Jay Huang, Eng Tat Khoo, and KangKang Yin. Virtual try-on using kinect and hd camera. In *International Conference on Motion in Games*, pages 55–65. Springer, 2012.

[GCS+19] Erhan Gundogdu, Victor Constantin, Amrollah Seifoddini, Minh Dang, Mathieu Salzmann, and Pascal Fua. GarNet: A two-stream network for fast and accurate 3D cloth draping. In *International Conference on Computer Vision (ICCV)*, 2019.

[GHF+07] Rony Goldenthal, David Harmon, Raanan Fattal, Michel Bercovier, and Eitan Grinspun. Efficient simulation of inextensible cloth. *ACM Transactions on Graphics (TOG)*, 26(3):49, 2007.

[HAR+10] Nils Hasler, Hanno Ackermann, Bodo Rosenhahn, Thorsten Thormählen, and Hans-Peter Seidel. Multilinear pose and body shape estimation of dressed subjects from image sets. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1823–1830. IEEE, 2010.

[HLS+19] Yuanming Hu, Jiancheng Liu, Andrew Spielberg, Joshua B. Tenenbaum, William T. Freeman, Jiajun Wu, Daniela Rus, and Wojciech Matusik. ChainQueen: A real-time differentiable physical simulator for soft robotics. In *International Conference on Robotics and Automation (ICRA)*, 2019.

[HSR+09] Nils Hasler, Carsten Stoll, Bodo Rosenhahn, Thorsten Thormählen, and Hans-Peter Seidel. Estimating body shape of dressed humans. *Computers & Graphics*, 33(3):211–216, 2009.

[HVTG08] David Harmon, Etienne Vouga, Rasmus Tamstorf, and Eitan Grinspun. Robust treatment of simultaneous collisions. *ACM Trans. Graph.*, 27(3), 2008.

[HYZ16] Ping Huang, Junfeng Yao, and Hengheng Zhao. Automatic realistic 3d garment generation based on two images. *2016 International Conference on Virtual Reality and Visualization (ICVRV)*, 2016.

[JHK15] Moon-Hwan Jeong, Dong-Hoon Han, and Hyeong-Seok Ko. Garment capture from a photograph. *Journal of Visualization and Computer Animation*, 26(3-4):291–300, 2015.

[JHR+15] Amaury Jung, Stefanie Hahmann, Damien Rohmer, Antoine Bégault, Laurence Boissieux, and Marie-Paule Cani. Sketching folds: Developable surfaces from non-planar silhouettes. *ACM Trans. Graph.*, 34(5):155:1–155:12, 2015.

[JTST10] Arjun Jain, Thorsten Thormählen, Hans-Peter Seidel, and Christian Theobalt. Moviereshape: Tracking and reshaping of humans in videos. In *ACM Transactions on Graphics (TOG)*, volume 29, page 148. ACM, 2010.

[KB04] Michael Keckeisen and Wolfgang Blochinger. Parallel implicit integration for cloth animations on distributed memory architectures. In *Proceedings of the 5th Eurographics conference on Parallel Graphics and Visualization*, pages 119–126. Eurographics Association, 2004.

[KBJM18] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Regognition (CVPR)*, 2018.

[LCT18] Zorah Lähner, Daniel Cremers, and Tony Tung. Deepwrinkles: Accurate and realistic clothing modeling. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IV*, pages 698–715, 2018.

[LL18] Junbang Liang and Ming C Lin. Time-domain parallelization for accelerating cloth simulation. In *Computer Graphics Forum*, volume 37, pages 21–34. Wiley Online Library, 2018.

[LL19] Junbang Liang and Ming C Lin. Shape-aware human pose and shape reconstruction using multi-view images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4352–4362, 2019.

[LLN+18] Eric Liang, Richard Liaw, Robert Nishihara, Philipp Moritz, Roy Fox, Ken Goldberg, Joseph E. Gonzalez, Michael I. Jordan, and Ion Stoica. RLlib: Abstractions for distributed reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2018.

[LMR+15] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: a skinned multi-person linear model. *ACM Trans. Graph.*, 34(6):248:1–248:16, 2015.

[LRK+17] Christoph Lassner, Javier Romero, Martin Kiefel, Federica Bogo, Michael J Black, and Peter V Gehler. Unite the people: Closing the loop between 3d and 2d human representations. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 3, 2017.

[MRB+99] Bart Maerten, Dirk Roose, Achim Basermann, Jochen Fingberg, and Guy Lonsdale. Drama: A library for parallel dynamic load balancing of finite element applications. In *European Conference on Parallel Processing*, pages 313–316. Springer, 1999.

[MvdBF+12] Stephen Miller, Jur van den Berg, Mario Fritz, Trevor Darrell, Kenneth Y. Goldberg, and Pieter Abbeel. A geometric approach to robotic laundry folding. *I. J. Robotics Res.*, 31(2), 2012.

[NCV+19] Juan Carlos Núñez, Raúl Cabido, José F Vélez, Antonio S Montemayor, and Juan José Pantrigo. Multiview 3d human pose estimation using improved least-squares and lstm networks. *Neurocomputing*, 323:335–343, 2019.

[NKT15] Xiang Ni, Laxmikant V Kale, and Rasmus Tamstorf. Scalable asynchronous contact mechanics using charm++. In *Parallel and Distributed Processing Symposium (IPDPS), 2015 IEEE International*, pages 677–686. IEEE, 2015.

[NSO12] Rahul Narain, Armin Samii, and James F. O'Brien. Adaptive anisotropic remeshing for cloth simulation. *ACM Trans. Graph.*, 31(6), 2012.

[PPHB17] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J. Black. ClothCap: Seamless 4D clothing capture and retargeting. *ACM Trans. Graph.*, 36(4), 2017.

[PZDD17a] Georgios Pavlakos, Xiaowei Zhou, Konstantinos G Derpanis, and Kostas Daniilidis. Coarse-to-fine volumetric prediction for single-image 3d human pose. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 1263–1272. IEEE, 2017.

[PZDD17b] Georgios Pavlakos, Xiaowei Zhou, Konstantinos G Derpanis, and Kostas Daniilidis. Harvesting multiple views for marker-less 3d human pose annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6988–6997, 2017.

[QYSG17] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017.

[RMSC11] Cody Robson, Ron Maharik, Alla Sheffer, and Nathan Carr. Context-aware garment modeling from sketches. *Computers & Graphics*, 35(3):604–613, 2011.

[RRZ00] Sergio Romero, Luis F Romero, and Emilio L Zapata. Fast cloth simulation with parallel computers. In *European Conference on Parallel Processing*, pages 491–499. Springer, 2000.

[RSF18] Helge Rhodin, Mathieu Salzmann, and Pascal Fua. Unsupervised geometry-aware representation for 3d human pose estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 750–767, 2018.

[SF18] Connor Schenck and Dieter Fox. SPNets: Differentiable fluid dynamics for deep neural networks. In *Conference on Robot Learning (CoRL)*, 2018.

[SPMF19] Hosnieh Sattar, Gerard Pons-Moll, and Mario Fritz. Fashion is taking shape: Understanding clothing preference based on body shape from online sources. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 968–977. IEEE, 2019.

[TAST18] Marc Toussaint, Kelsey Allen, Kevin Smith, and Joshua Tenenbaum. Differentiable physics and stable modes for tool-use and manipulation planning. In *Robotics: Science and Systems (RSS)*, 2018.

[TB06] Bernhard Thomaszewski and Wolfgang Blochinger. Parallel simulation of cloth on distributed memory architectures. In *Proceedings of the 6th Eurographics conference on Parallel Graphics and Visualization*, pages 35–42. Eurographics Association, 2006.

[TCH07] Emmanuel Turquin, Marie-Paule Cani, and John F. Hughes. Sketching garments for virtual characters. In *34. International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2007, San Diego, California, USA, August 5-9, 2007, Courses*, page 28, 2007.

[TGHC18] Matthew Trumble, Andrew Gilbert, Adrian Hilton, and John Collomosse. Deep autoencoder for combined human pose estimation and body model upscaling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 784–800, 2018.

[TGM+17] Matthew Trumble, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John Collomosse. Total capture: 3d human pose estimation fusing video and inertial sensors. In *Proceedings of 28th British Machine Vision Conference*, pages 1–13, 2017.

[TMNSF17] Bugra Tekin, Pablo Marquez Neila, Mathieu Salzmann, and Pascal Fua. Learning to fuse 2d and 3d image cues for monocular body pose estimation. In *International Conference on Computer Vision (ICCV)*, number EPFL-CONF-230311, 2017.

[TRA17] Denis Tome, Christopher Russell, and Lourdes Agapito. Lifting from the deep: Convolutional 3d pose estimation from a single image. *CVPR 2017 Proceedings*, pages 2500–2509, 2017.

[TTAR18] Denis Tome, Matteo Toso, Lourdes Agapito, and Chris Russell. Rethinking pose in 3d: Multi-stage refinement and recovery for markerless motion capture. In *2018 International Conference on 3D Vision (3DV)*, pages 474–483. IEEE, 2018.

[TWT+16] Min Tang, Huamin Wang, Le Tang, Ruofeng Tong, and Dinesh Manocha. Cama: Contact-aware matrix assembly with unified collision handling for gpu-based cloth simulation. In *Computer Graphics Forum*, volume 35, pages 511–521. Wiley Online Library, 2016.

[VCR+18] Gul Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. Bodynet: Volumetric inference of 3d human body shapes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 20–36, 2018.

[VMTF09] Pascal Volino, Nadia Magnenat-Thalmann, and Francois Faure. A simple approach to nonlinear tensile stiffness for accurate cloth simulation. *ACM Transactions on Graphics*, 28(4):Article–No, 2009.

[VRM+17] Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017.

[WCPM18] Tuanfeng Y. Wang, Duygu Ceylan, Jovan Popovic, and Niloy J. Mitra. Learning a shared shape space for multimodal garment design. *CoRR*, abs/1806.11335, 2018.

[WPB+14] Stefanie Wuhrer, Leonid Pishchulin, Alan Brunton, Chang Shu, and Jochen Lang. Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding*, 127:31–42, 2014.

[WY16] Huamin Wang and Yin Yang. Descent methods for elastic body simulation on the gpu. *ACM Transactions on Graphics (TOG)*, 35(6):212, 2016.

[YFHWW16] Jinlong Yang, Jean-Sébastien Franco, Franck Hétroy-Wheeler, and Stefanie Wuhrer. Estimation of human body shape in motion with wide clothing. In *European Conference on Computer Vision*, pages 439–454. Springer, 2016.

[YLL17] Shan Yang, Junbang Liang, and Ming C Lin. Learning-based cloth material recovery from video. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4383–4393, 2017.

[YPA+18] Shan Yang, Zherong Pan, Tanya Amert, Ke Wang, Licheng Yu, Tamara Berg, and Ming C Lin. Physics-inspired garment recovery from a single-view image. *ACM Transactions on Graphics (TOG)*, 37(5):1–14, 2018.

[ZCF+13] Bin Zhou, Xiaowu Chen, Qiang Fu, Kan Guo, and Ping Tan. Garment modeling from a single image. *Comput. Graph. Forum*, 32(7):85–91, 2013.

[Zel05] Cyril Zeller. Cloth simulation on the gpu. In *ACM SIGGRAPH 2005 Sketches*, page 39. ACM, 2005.

[ZFV02] Florence Zara, François Faure, and Jean-Marc Vincent. Physical cloth simulation on a pc cluster. In *4h Eurographics Workshop on Parallel Graphics and Visualization*, 2002.

[ZFV04] Florence Zara, François Faure, and J-M Vincent. Parallel simulation of large dynamic system on a pc cluster: Application to cloth simulation. *International Journal of Computers and Applications*, 26(3):1–8, 2004.

[ZHS+17] Xingyi Zhou, Qixing Huang, Xiao Sun, Xiangyang Xue, and Yichen Wei. Weaklysupervised transfer for 3d human pose estimation in the wild. In *IEEE International Conference on Computer Vision*, volume 206, page 3, 2017.

[ZY01] Dongliang Zhang and Matthew MF Yuen. Cloth simulation using multilevel meshes. *Computers & Graphics*, 25(3):383–389, 2001.