Chapter 1: Introduction

Digital try-on system, as one important part of E-Commerce, has the potential to become one of the revolutionary technologies that change people's lives. However, its development is limited by some practical constraints, such as accurate sizing of the body and vivid try-on demonstrations.

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 that an accurate material estimation and digital cloning from the real-world examples is challenging.

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. I propose to address the key open research challenges above by adopting machine learning and optimization techniques. These include:

- Accurate reconstruction of human and garment through consumer devices;
- Faithful estimation of fabric materials via learning and optimization;
- User-friendly recovery of dressed garments;
- A real-time cloth simulation system for customized animation; and
- Fast and realistic visual rendering of animated try-on results.

1.1 Learning-Based Human Body and Garment Estimation

Human appearance reconstruction is one of the key techniques for building a vivid and interactive virtual world. It can be applied to create a virtual avatar for various applications, such as virtual try-on or teleconferencing. It can also be used during character prototyping in computer animation. 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. Most of existing works [2, 3, 4, 5, 6] focus on human-body reconstruction and recent advances have made significant progress in this area. However, the problem itself is naturally ambiguous, given limited input and occlusion. Although applying a predefined prior can alleviate this ambiguity, it is still insufficient in several cases, especially when a part of the body is occluded by clothing, or when the pose direction is perpendicular to the camera viewing plane. For example, when the human is walking towards the camera, it can be difficult to distinguish the difference between a standing vs. walking pose using a direct front-view image, while a side-view image could be more informative of the posture.

Moreover, recovery of garment properties, especially for physical material estimation, has been under-explored due to the complexity and the diversity of cloth dynamics and coupled interaction with a human body. Estimating worn garment materials using RGB frames of a video is highly challenging. Image features are often sparse, containing many noisy signals regarding the fabric materials worn on the body. An effective way to amplify useful signals is to estimate garment geometry from images as a by-product. However, this is already an open problem far from being solved due to several reasons. First, garments have highly dynamical geometry that is not easy to capture and model. Previous works on garment modeling [7, 8, 9] and estimation [10, 11, 12] often propose solutions on one single type of garment, mostly t-shirts. Although the methods are also applicable to other garments, lack of generalization in capturing different garment geometries present a considerable barrier for real-world applications: users can only choose one of few pre-trained garment types and are not able to import new ones easily. Second, accurate estimation is often hindered by camera projection and human body occlusion. For example, the human-body estimation network may disagree with the garment reconstruction network in skeleton orientation due to the projection ambiguity (*e.g.* an arm is posed forward vs. backward), resulting in prediction misalignment. Therefore, without a general garment representation and an accurate geometry estimation, it is very difficult to regress the fabric materials solely from garment image sequences.

1.2 Differentiable Simulation for Material Optimization

Differentiable physics simulation is a powerful family of techniques that applies gradient-based methods to learning and control of physical systems [13, 14, 15, 16, 17]. It can enable optimization for control, and can also be integrated into neural network frameworks for performing complex tasks. My work focuses on cloth simulation, which relates to applications in robotics, computer vision, and computer graphics [8, 18, 19, 20, 21, 22, 23]. My goal is to enable differentiable cloth simulation, which can provide a unified approach to a variety of inverse problems for 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 [13]) 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 [13]. In my simple example with cloth, the number of pairwise constraints would be at least $289 \times 512 = 140K$ 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.

1.3 Simulation-Based Virtual Try-On

Drape prediction systems fall mainly in two categories: physics-based simulation and learning-based garment generation. Significant progress has been achieved in visual simulation of cloth over the past decades [24, 25, 26, 27]. 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. Given recent advances in manycore and cloud computing, parallel computing has emerged as a possible alternative to achieve the desired runtime performance. Parallelization is a popular, practical way to achieve performance improvement. With a highly scalable parallelization scheme, physics-based simulation can be accelerated by orders of magnitude, enabling fast user feedback in virtual try-on.

As an alternative, learning-based cloth draping is also one of the key components in virtual try-on systems. With the help of a well-trained draping network, virtual try-on can predict quickly and accurately how garments look and fit on a body. Previously mentioned cloth simulation typically is prohibitively slow, while image-based try-on [28, 29] does not provide fit-accurate information. Apart from its use to avoid the dressing room, fast garment draping can also be a critical component in interactive character prototyping for a wide range of applications, like teleconferencing, computer animations, special effects or computer games.

1.4 Thesis Statement

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

To support this thesis statement, I present five novel algorithms:

1. A learning-based *shape-aware* human body mesh reconstruction for both pose and shape estimation,

2. A differentiable simulation algorithm for fabric material optimization,
3. A joint estimation framework for estimating human body and apparels through a close-loop iterative optimization,
4. A time-domain parallelization algorithm to accelerate the simulation performance in distributed systems, and
5. A dynamics-inspired learning framework to directly predict the cloth draping on a wide range of body shapes.

1.5 Main Results

This dissertation presents five methods to support each component of the thesis statement. I list the main results obtained within these methods below:

1.5.1 Shape-Aware Human Reconstruction Using Multi-View Images

In Chapter 2, I propose a scalable neural network framework to reconstruct the 3D mesh of a human body from multi-view images, in the subspace of the SMPL model [30]. Use of multi-view images can significantly reduce the projection ambiguity of the problem, increasing the reconstruction accuracy of the 3D human body under clothing. The key contributions of this work include:

- 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 to account for the ambiguity of the 3D human body (geometry) reconstruction problem from 2D images, achieving improved estimation results.

- A large simulated dataset, including *clothed* human bodies and the corresponding ground-truth parameters, to enhance the reconstruction accuracy, especially in shape estimation, where no ground-truth or supervision is provided in the real-world dataset.

- Accurate *shape* recovery *under occlusion of garments* by (a) providing the corresponding supervision and (b) deepening the model using the multi-view framework.

1.5.2 Differentiable Simulation for Material Optimization

In Chapter 3, I propose a differentiable cloth simulator that can be embedded as a layer in deep neural networks (DNN) for estimating fabric material parameters. Differentiable simulation provides an effective, robust framework for modeling cloth dynamics, self-collisions, and contacts in DNN. Due to the high dimensionality of the dynamical system in modeling cloth, traditional gradient computation for collision response can become impractical. To address this problem, I propose to compute the gradient directly using QR decomposition of a much smaller matrix. The key contributions of this work include:

- A dynamic collision detection scheme to reduce collision dimensionality.

- A novel gradient computation method of collision response using implicit differentiation.
- An optimized backpropagation algorithm using QR decomposition.

1.5.3 Joint Estimation of Human and Garment from Video

In Chapter 4, I propose a network model that uses video input to jointly estimate the human body, the garment shape, as well as fabric materials of the garment dressed on a human. During the estimation, I use a closed-loop optimization structure to share information between tasks and feed the learned garment features for temporal estimation of garment material type. The key contributions include:

- The first end-to-end neural network that recovers fabric material(s) of a garment from one single RGB video;
- A two-level auto-encoder for learning the latent space of garments through multi-scale feature coupling;
- Joint estimation of human body and apparels through a close-loop iterative optimization;
- The first garment prediction model that can account for arbitrary topologies and uses a feedback loop to the body estimation for prediction consistency.

1.5.4 Time-Domain Parallelization for Accelerating Cloth Simulation

In Chapter 5, I propose a novel time-domain parallelization technique that makes use of the two-level mesh representation to resolve the time-dependency issue and develop a practical algorithm to smooth the state transition from the corresponding coarse to fine meshes. A load estimation and a load balancing technique used in online partitioning are also proposed to maximize the performance acceleration. The key contributions of this work include:

- A time-domain parallelization algorithm supporting *adaptive meshes* with minimal communication overhead;
- Load estimation and load balancing techniques that maximize the overall performance acceleration;
- A practical state transitioning algorithm between low- and high-resolution simulations to recover details and ensure the visual quality of the simulated sequences.

1.5.5 Dynamics-Inspired Garment Draping Prediction

In Chapter 6, I propose a novel learning framework for draping prediction that can incorporate arbitrary loss functions at runtime composed of three key components. First, a physics-inspired supervision on a novel neural network. Second, an unsupervised optimization process coupled to the physics of individual garments at runtime. Finally, self-correction of the network based on the samples optimized in the previous stage. The key contributions of this work include:

- A semi-supervised framework that enables easy integration of constraints into the deep learning model.
- Introduction of novel loss functions that encode geometric, physical, design, and tailoring constraints.
- A novel encoder/decoder network that effectively captures global and local features from the input and dynamically aggregates neighborhood information.
- A new self-correcting method based on data augmentation that enables both more accurate predictions and reduced data preparation time.

1.6 Outline of Dissertation

The subsequent chapters of this dissertation are organized as follows.

Chapter 2 introduces a new shape-aware human body estimation method that uses multi-view input for accurate reconstruction.

Chapter 3 presents my differentiable cloth simulation method that can compute the gradients efficiently and achieve faster optimization convergence than gradient-free methods.

Chapter 4 demonstrates a novel joint learning model that simultaneously predict human body shapes, garment geometry, and its fabric material in an end-to-end network using temporal garment geometry features represented by latent codes.

Chapter 5 presents my time-domain parallelization algorithm for accelerated cloth simulation.

Chapter 6 describes a semi-supervised learning framework that integrates multiple geometric and physics constraints into learning garment draping on various bodies.