Machine learning for digital try-on: Challenges and progress

Junbang Liang[1] (✉), **Ming C. Lin**[1]

Abstract Digital try-on systems for e-commerce have the potential to change people's lives and provide notable economic benefits. However, their development is limited by practical constraints, such as accurate sizing of the body and realism of demonstrations. We enumerate three open challenges remaining for a complete and easy-to-use try-on system that recent advances in machine learning make increasingly tractable. For each, we describe the problem, introduce state-of-the-art approaches, and provide future directions.

Keywords machine learning; digital try-on; garment modeling; human body estimation; material modeling

1 Introduction

E-commerce has grown at a rapid pace in recent years. Consumers today are more likely to shop online than to visit a retail store. The situation is much more complicated, however, when it comes to buying clothes. People need to know how a garment fits on them, how it looks, and how it feels. Digital try-on systems can potentially satisfy these needs, providing a direct visual impression, and possibly customized clothes sizing as well. Therefore, it has drawn much attention as an attractive alternative to improve the user experience and popularize online fashion shopping.

However, the technology is still far from practical, easy-to-use, and adequate to replace physical try-on. Currently, most try-on systems rely on either image-editing, copy-pasting, or template demonstrations,

1 University of Maryland, College Park, MD 20785, USA. E-mail: J. Liang, liangjb@cs.umd.edu (✉); M. C. Lin, lin@cs.umd.edu.

while the ultimate goal is a fast and realistic try-on system adaptive to each customer's body. There is still a substantial technological gap between modeling and demonstrating garment fitting in the digital and real worlds, including fast and realistic demonstration, accurate modeling of human body and garments, faithful modeling of garment material, and lossless transformation of garments between virtual and physical worlds.

In this paper, we present some open research issues that contribute to this technological gap, including:
1. accurate estimation of human shapes and sizes using consumer devices,
2. faithful recovery of garment materials via (online) images, and
3. ease of design and manipulation of sewing patterns and garment pieces by end-users.

Although traditional methods have made important progress on these under-constrained problems, learning-based approaches have shown tremendous potential to make a notable impact. Compared to traditional methods, machine-learning algorithms are usually much faster since training and optimization are performed offline. They are also good at generalizing to unseen images without the need for tedious data pre-processing. While extensive research exists on 2D image learning, machine learning of highly variable 3D human body shapes is still far from mature, which is the reason why the open issues described above remain elusive.

For each problem listed above, we motivate its importance, provide a problem description, and present state-of-the-art approaches with potential for improvements. We believe that solutions to these challenging problems will lead to significant advances in digital try-on, as well as other areas of e-commerce.

TSINGHUA UNIVERSITY PRESS | Springer

2　Open problems

In this section we first introduce three major challenges that limit digital try-on technology from being widely adopted and accepted by shoppers. There are several reasons why shoppers still prefer physical try-on. Firstly, consumers are unsure if what they buy online will fit them well. Although general sizing systems exist, their lack of consistency and standardization across different brands and garment materials can often make it difficult to size clothes, especially for persons with non-standard body shapes and proportion. Accurate estimation of human body shape is the key to successful digital try-on. Secondly, fabric is usually a key consideration when shopping for clothes. Different fabric affects how garments look and fit, how consumers would wear them, and whether or not they would buy them. However, the correspondence between the actual material and its digital representation are not well understood. It is also challenging to acquire a full fabric digital model from real-world examples.

For the customers, appearance is as critical as other factors. There are two approaches to displaying garments: 2D image-based, and 3D mesh-based with photo-realistic rendering. They have different advantages and drawbacks, but both need a large garment database for support. While creating a 3D garment takes considerable effort, 2D images often suffer from a lack of variation and are much more difficult to customize. In either case, the try-on system needs a user-friendly design and manipulation backend. Last, but not least, a fast and realistic animation of the garments in motion along with body movements greatly improves the user experience. Although it is not so critical as other factors, it would effectively reduce the perceptual gap between the real and digital worlds for online shopping. Previous work has proposed using cloud computing to improve the animation speed, but there is still a notable technology gap for high-quality, interactive 3D animation of clothes.

3　Human shape estimation

As noted, accurate human shape estimation is key to enabling digital try-on. 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 and augmented environments. Yet, it remains a challenging and popular topic of interest. While direct 3D body scanning can provide excellent and accurate results, its adoption is somewhat limited by the required specialized hardware. RGB images are widely available for input to digital try-on and can be easily captured using commodity mobile devices. Although purely image-based try-on methods have been proposed [1], learning-based 3D body estimation is more widely applicable in that the 3D body can be articulated and so re-posed and re-targeted.

We define the human-body reconstruction problem informally as, given one or more RGB images, to estimate the human body geometry and size, and output (preferably) a 3D humanoid mesh. Traditional algorithms often formulate it as an optimization problem, in which the silhouette difference is a major part of the objective function [2]. Therefore, these methods either require the human to wear tight clothes, or alternatively relax the target function to be unilateral on uncovered body parts [3], or to point correspondences [4]. The use of machine learning methods in this problem has led to significant advances. Firstly, it has moved the algorithm from online to offline, significantly reducing response time. Second, by using a parametric human model [5], one can easily construct a regression network for the parameters while the losses needed can also be inferred from them. While early works proposed network models for only 2D/3D body skeletons [6–8], more recent works have introduced techniques to perform regression for the entire human body—either using a parametric human model [9, 10] or a voxel-based representation [11–13]. As annotations in most real-world datasets contain only joint positions, the learning process has been refined in various ways [14–17]. The current state of the art is the recent work by Ref. [18] [1]. It emphasizes shape learning, while many other works often focus on body-joint losses, but neglect the effect of body shapes.

The key contribution of Ref. [18] is a multi-view, multi-stage framework to address ambiguity caused by camera projection (see Fig. 1). Their model performs several stages of error correction. Each of the image inputs is passed on step by step; at each step, a shared-

---

① Liang and Lin's data and code are available at `https://gamma.umd.edu/researchdirections/virtualtryon/humanmultiview`
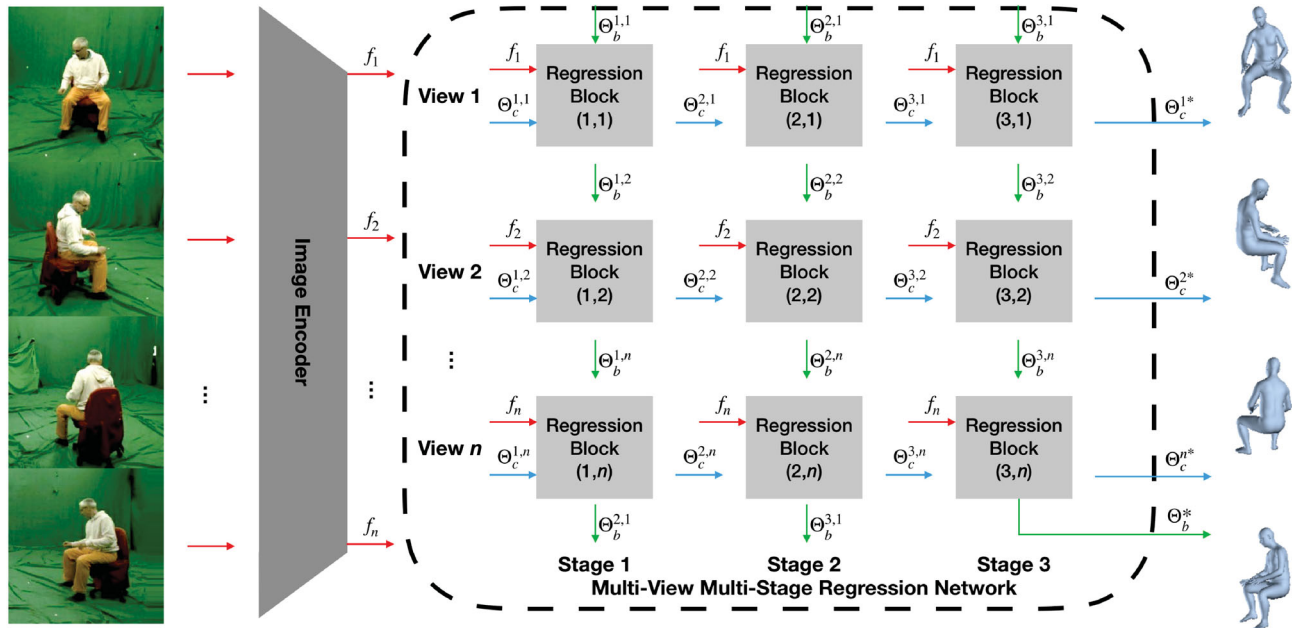
Fig. 1 Network structure from Ref. [18]. By using an iterative value correction structure, visual information from different views is effectively integrated to provide a unified human shape. Reproduced with permission from Ref. [18], © The Author(s) 2019.

parameter prediction block computes the correction based on the image feature and the input guesses. The camera and the human body parameters are estimated at the same time, projecting the predicted 3D joints back to 2D for loss computation. The estimated pose and shape parameters are shared among all views, while each view maintains its own camera calibration and global orientation. Their proposed framework uses a recurrent structure, making it a universal model applicable to any number of views. At the same time, it couples shareable information across different views so that the human body pose and shape are optimized using image features from all views. Unlike static multi-view CNNs which have a fixed number of inputs, they make use of the RNN-like structure in a cyclic form to accept any number of views, and prevent the gradient vanishing by predicting corrective values instead of updating parameters in each regression block.

Experiments have shown that, after training, this model can form a single view image, provide equally good pose estimation as the state of the art, and provide considerably improved pose estimation when using multi-view inputs, leading to better shape estimation across all datasets. An example is demonstrated in Fig. 2. Moreover, a physically-based synthetic data generation pipeline is introduced to enrich the training data, which is very helpful for

shape estimation and regularization in cases that traditional datasets do not capture. While synthetic data improves the diversity of human bodies with ground-truth parameters, a larger garment dataset and a more convenient registration process are needed to minimize the performance gap between real-world images and synthetic data. In addition, other variables such as hair, skin color, and 3D backgrounds are subtle elements that can influence the perceived realism of the synthetic data at the higher expense of a more complex data generation pipeline. With the recent progress in image style transfer using GAN, a

(a) Input image   (b) Recovered body

Fig. 2 Prediction results using the state of the art [18]. The model captures the shape of the human body by learning from synthetic data. The recovered legs and chest are close to those of the person in the image. Reproduced with permission from Ref. [18], © The Author(s) 2019.

promising direction is to transfer the synthetic result to more realistic images to further improve the result.

4 Garment material modeling

4.1 Introduction

Garment material plays an important role in digital try-on systems. Physical recreation of the fabric not only gives a compelling visual simulation of the cloth, but also affects how the garment feels and fits on the body. However, fabric modeling is a challenging task: the appearance and physical properties of the garment are determined not only by the type of materials the clothes are made of, but also by sewing and weave. Thus, researchers often focus on the physical behaviour, rather than the underlying semantic primitives.

Hence, we state the garment material modeling problem as follows. Given a sufficient amount of data, model the material's physical behavior and physical properties, so that visual effects the same as or similar to those of the real material can be reproduced by a computer. This has two implications: firstly, we need to define a physical model of the material, and then we must estimate the parameters in the model.

There are many ways to model clothes, including spring–mass systems and finite elements. The latter is the most popular model since it can produce realistic results. While one can use isotropic properties such as Young's modulus and Poisson ratio, an anisotropic model is a better choice since it can support different behaviors caused by the weave of the material.

4.2 Learning-based estimation

While traditional optimization methods [19] often take a long time to compute material parameters, machine-learning methods can make predictions in real time by a simple feed-forward operation, which is more useful in applications that need fast feedback, such as garment prototyping. The state-of-the-art model from Yang et al. [20] [1] uses CNNs combined with LSTM to recover material parameters from videos. To constrain both the input and solution space, they choose one of the materials as a basis; the material sub-space is constructed by multiplying this material basis with a positive coefficient. To construct an optimal material parameter sub-space, a

---

material parameter sensitivity analysis is conducted to examine the sensitivity of the material parameters κ with respect to the amount of deformation $D(\kappa)$. Physically based cloth simulations are used to generate a much larger number of data samples within these sub-spaces, which would otherwise be difficult or time-consuming to capture. The cloth meshes are generated through physically-based simulation, and then rendered as 2D images with a randomly assigned texture. Using the data samples, they combine the image signal feature extraction method, a CNN, with the temporal sequence learning method, LSTM, to learn the mapping from visual appearance to material. As shown in Fig. 3, the CNN layer is used to extract both low- and high-level visual features, while the LSTM layer focuses on learning the mapping between the material properties of the cloth and its consequent movement.

They demonstrated the proposed framework with the application of "material cloning". With the trained deep neural network model being able to capture the cloth motion (Fig. 4), the material type can be inferred from a video recording of the motion of the cloth in a fairly small amount of time. The recovered material type can be "cloned" onto another piece of cloth or garment as shown in Fig. 5.

In this work, the videos contain only a single piece of cloth which does not interact with any other object. While this is not applicable to all real-world scenarios, this method provides new insights into addressing this challenging problem. A natural extension would be to learn from videos of clothing directly interacting with the human body, under varying lighting conditions and partial occlusion.

4.3 Optimization using differentiable physics

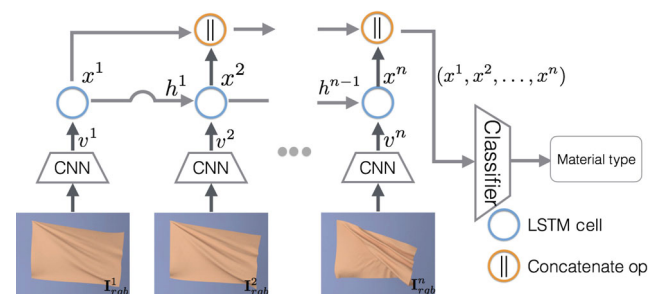Another approach to modeling the fabric is to measure geometric differences directly during parameter

Fig. 3 Network model from Ref. [20]. The material is modeled by learning motion patterns of image features given by CNNs. Reproduced with permission from Ref. [20], © The Author(s) 2017.
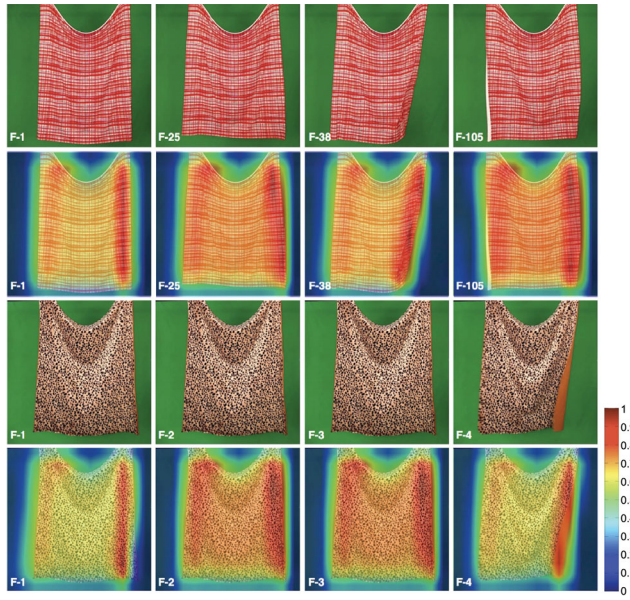

Fig. 4 Learned CNN conv5-layer activation visualization from Ref. [20]. Experiments show that the trained model is able to capture moving parts of the cloth even in an unseen video. Reproduced with permission from Ref. [20], © The Author(s) 2017.
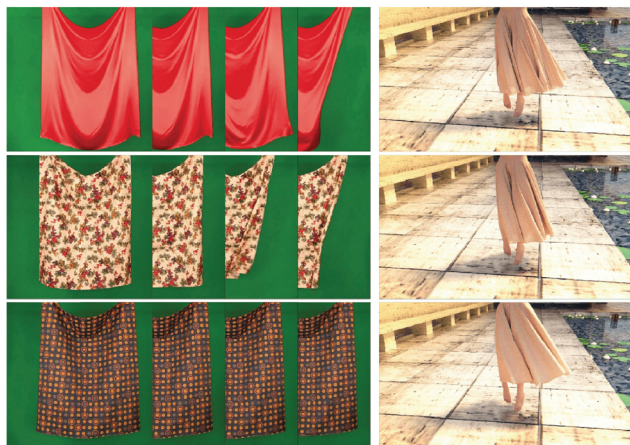

Fig. 5 Yang et al. [20] modeled clothes materials in input videos (left), and applied those materials to a simulated skirt (right). Reproduced with permission from Ref. [20], © The Author(s) 2017.

optimization. Assuming that the environment is known to the system, computation of the estimated motion and its gradient with respect to the material parameters can be achieved using differentiable simulation. A typical usage of differentiable simulation is motion control (see Fig. 6), where the difference to the target is measured and the loss backpropagated to the network. Similar processes can be applied to material parameter estimation as well. By measuring the distance to the target as the loss and computing corresponding gradients, either in pixel space or in 3D space, the material parameters

can be learned or optimized to achieve the desired cloth motion or visual effect. Recent differentiable physics work covers rigid bodies [22, 23], cloth [24], and particle-grid systems [25, 26]. The state-of-the-art is Ref. [24] [①], which proposes a method for differentiable cloth simulation. It is the first work to tackle a high dimensional simulation problem and to propose a general differentiable collision handling algorithm. Later, a follow-up work [21] extended the algorithm to be applicable to coupled dynamics with rigid bodies. Overall, they follow the computational flow of the common approach to cloth simulation: discretization using the finite element method, integration using an implicit Euler method, and collision response on impact zones. They use implicit differentiation in the linear solver and optimization in order to compute the gradient with respect to the input parameters. The discontinuity introduced by collision response is negligible because the discontinuous states constitute a zero-measure set. During backpropagation in the optimization, gradient values can be directly computed after QR decomposition of the constraint matrix. Their pipeline contains several techniques that can be employed in other differentiable simulations.

4.3.1 Derivatives of the physical solution

In modern simulation algorithms, an implicit Euler method is often used for stable integration results. Thus the mass matrix M often includes the Jacobian of the forces, and is denoted as $\hat{M}$ to indicate this difference. A linear solver is needed to compute the acceleration since it is time-consuming to compute $\hat{M}^{-1}$. Implicit differentiation is used to compute the gradients of the linear solution. Given an equation $\hat{M}a = f$ with a solution z and propagated gradient $\partial\mathcal{L}/\partial a|_{a=z}$, where $\mathcal{L}$ is the task-specific loss function, implicit differentiation is used to derive the gradients. We refer readers to the original paper [24] for more details.

4.3.2 Derivatives of the collision response

A general approach using LCP to integrate collision constraints into physics simulations has been proposed, but constructing a static LCP is often impractical in cloth simulation due to the high dimensionality. Collisions and contacts which happen at each step are very sparse compared to the complete

---

① Liang et al.'s data and code are available at `https://gamma.umd.edu/researchdirections/virtualtryon/differentiablecloth`
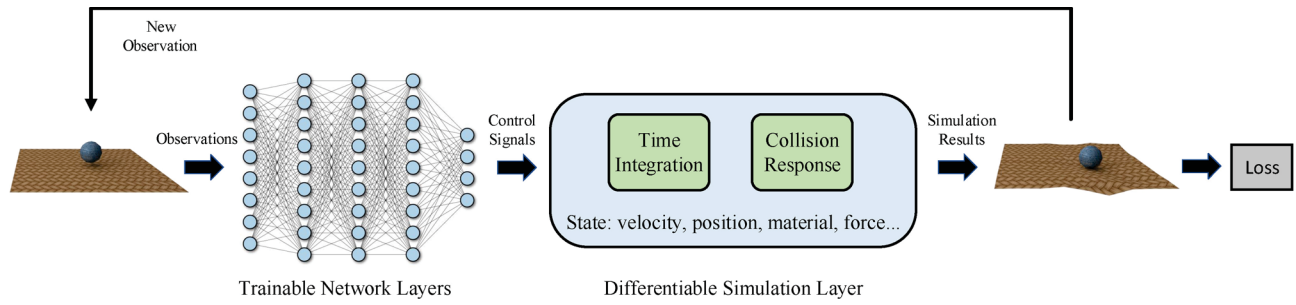
Fig. 6 Differentiable simulation embedding example from Ref. [21]. The loss can be backpropagated through the physics simulator to the neural network, enabling learning tasks such as material modeling and motion control.

data. Therefore, a dynamic approach is used that incorporates collision detection and response.

Collision handling in their implementation is based on impact zone optimization. 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, a QP problem is developed to determine the constraints. Since the signed distance function is linear in $\boldsymbol{x}$, the optimization takes a quadratic form, as shown originally in Ref. [24]:

$$\underset{\boldsymbol{z}}{\text{minimize}} \quad \frac{1}{2}(\boldsymbol{z}-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{z}-\boldsymbol{x}),$$
$$\text{subject to} \quad \boldsymbol{G}\boldsymbol{z}+\boldsymbol{h} \leqslant \boldsymbol{0}$$

where $\boldsymbol{W}$ is a constant diagonal weight matrix related to the mass of each vertex, and $\boldsymbol{G}$ and $\boldsymbol{h}$ are constraint parameters. The numbers of variables and constraints are n and m, i.e. $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{h} \in \mathbb{R}^m$, and $\boldsymbol{G} \in \mathbb{R}^{m \times n}$. Note that this optimization problem has inputs $\boldsymbol{x}$, $\boldsymbol{G}$, and $\boldsymbol{h}$, and output $\boldsymbol{z}$. The goal here is to derive $\partial \mathcal{L}/\partial \boldsymbol{x}$, $\partial \mathcal{L}/\partial \boldsymbol{G}$, and $\partial \mathcal{L}/\partial \boldsymbol{h}$ given $\partial \mathcal{L}/\partial \boldsymbol{z}$, where $\mathcal{L}$ is the loss function.

When computing the gradient using implicit differentiation, the dimensionality of the linear system can be very high. Their key observation here is that $n \gg m > \text{rank}(\boldsymbol{G})$, since one contact often involves 4 vertices (thus 12 variables) and some contacts may be linearly dependent (e.g., multiple adjacent collision pairs). They minimize the size of the linear equation based on the QR decomposition of $\boldsymbol{G}$, which is the key to accelerating backpropagation of high dimensional QP problems.

One of their experiments shows its ability to optimize material parameters from observation. The scene features a piece of cloth hanging under gravity and subjected to a constant wind force. The material model consists of three parts: density d, stretching stiffness $\boldsymbol{S}$, and bending stiffness $\boldsymbol{B}$. The stretching stiffness quantifies the reaction force when the cloth is stretched; the bending stiffness models how easily the cloth can be bent and folded. Table 1 shows results. They achieve a much smaller error in most measurements in comparison to the baselines; the linear part of the stiffness matrix is modeled well. With the computed gradient using their model, one can effectively optimize the unknown parameters that dominate cloth movement to fit the observed data.

In follow-up work, Qiao et al. extended the differentiable simulation pipeline to couple with rigid body dynamics, formulated using generalized coordinates:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \boldsymbol{q} \\ \dot{\boldsymbol{q}} \end{pmatrix} = \begin{pmatrix} \dot{\boldsymbol{q}} \\ \ddot{\boldsymbol{q}} \end{pmatrix} = \begin{pmatrix} \dot{\boldsymbol{q}} \\ \boldsymbol{M}^{-1}f(\boldsymbol{q},\dot{\boldsymbol{q}}) \end{pmatrix}$$

and update the optimization formulation for collision response accordingly (see Ref. [21] for details):

Table 1 Material parameter estimation results from Ref. [24]. Their proposed method runs faster than L-BFGS. Values of material parameters are Frobenius norms of the difference normalized by the Frobenius norm of the target. Values of the simulated result are the average pairwise vertex distances normalized by the size of the cloth. The gradient-based method yields much smaller errors than the baselines

Method	Runtime (sec/step/iter)	Density error (%)	Linear stretching stiffness error (%)	Bending stiffness error (%)	Simulation error (%)
Baseline	—	68 ± 46	160 ± 119	**70 ± 42**	12 ± 3.0
L-BFGS	2.89 ± 0.02	4.2 ± 5.6	72 ± 90	**70 ± 43**	4.9 ± 3.3
Liang et al. [24]	**2.03 ± 0.06**	**1.8 ± 2.0**	**45 ± 41**	77 ± 36	**1.6 ± 1.4**

$$\operatorname*{minimize}_{\boldsymbol{q'}} \quad \frac{1}{2}(\boldsymbol{q}-\boldsymbol{q'})^{\mathrm{T}}\hat{\boldsymbol{M}}(\boldsymbol{q}-\boldsymbol{q'})$$
$$\text{subject to} \quad \boldsymbol{G}f(\boldsymbol{q'})+\boldsymbol{h}\leqslant\boldsymbol{0}$$

Due to the inclusion of rigid bodies, the constraints used in the optimization are no longer linear. When computing gradients, they linearize the constraints around a neighborhood as an approximation to enable QR decomposition for acceleration as previously mentioned.

5 Garment modeling and design

Realistic apparel model generation has become increasingly popular, due to the rapid changes in fashion trends and the growing need for garment models in different applications such as virtual try-on. It is already used even for state-of-the-art interactive apparel design systems [27]. Application requirements mean that it is important to have a general cloth model that can represent a diverse set of garments. However, there are many challenges in automatic garment model generation. Firstly, garments usually have different types of topology, especially for fashion apparel, that makes it difficult to design a universal pipeline. Moreover, it is often not straightforward for general garments design to be retargeted onto another body shape, making customization difficult.

Previous work has addressed this problem to some extent. Huang et al. [28] proposed a realistic 3D garment generation algorithm based on front and back image sketches, but it cannot readily retarget generated garments to other body shapes. Wang et al. [29] proposed an algorithm which can conveniently perform retargeting, but permits limited topology like T-shirts or skirts. There is no recent work that addresses these two problems at the same time.

We introduce a learning-based parametric generative model to overcome the above difficulties, given garment sewing patterns and human body shapes as input. One possible approach would be to compute a *displacement image* on the U–V space of the human body as a unified representation of the garment mesh. Different topology and sizes of the garment are represented by different values in the image. The 2D displacement image, as the representation of the 3D garment mesh data, can then be fed into a conditional generative adversarial network (cGAN) for latent space learning. The 2D representation for the garment mesh can transfer the irregular 3D mesh data to regular image data where a traditional CNN can easily learn. It can also extract relative geometric information with respect to the human body, enabling garment retargeting to a different person.

6 Conclusions

Although virtual reality and digital try-on have excellent potential and are rapidly developing, there remain open problems before online try-on systems can be widely adopted. We have listed three major challenges, all of which can be addressed or further improved using machine learning algorithms. For garment material prediction, state-of-the-art methods are still limited in that the training data is highly constrained: the scenario contains only a piece of cloth floating in the wind. To improve its applicability to daily tasks, it is necessary to focus on solving the problem on a more diverse set of inputs. Predicting the material from a garment on a fixed human body could be a good start, before generalizing to arbitrary human motions and predicting multiple garments on the same body. In the area of human shape estimation, it would be interesting to learn how external constraints could improve estimation accuracy. For example, the shape and size of the garment are hard constraints to which the predicted body should conform. While optimization-based methods can integrate these constraints fairly easily, doing so remains elusive for learning-based approaches. One possibility is to jointly estimate body and garment together and introduce an intersection loss. This approach would require a new solution to the open problem of unified deep garment representation, if we do not want to train one model for every garment type, which could be even more challenging. We believe that substantial breakthroughs in digital try-on are achievable with more investigation in these directions.

TSINGHUA UNIVERSITY PRESS Springer

References

[1] Zheng, Z. H.; Zhang, H. T.; Zhang, F. L.; Mu, T. J. Image-based clothes changing system. *Computational Visual Media* Vol. 3, No. 4, 337–347, 2017.

[2] Dibra, E.; Jain, H.; Öztireli, C.; Ziegler, R.; Gross, M. HS-Nets: Estimating human body shape from silhouettes with convolutional neural networks. In: Proceedings of the 4th International Conference on 3D Vision, 108–117, 2016.

[3] Bălan, A. O.; Black, M. J. The naked truth: Estimating body shape under clothing. In: *Computer Vision – ECCV 2008. Lecture Notes in Computer Science, Vol. 5303.* Forsyth, D.; Torr, P.; Zisserman, A. Eds. Springer Berlin, 15–29, 2008.

[4] Lassner, C.; Romero, J.; Kiefel, M.; Bogo, F.; Black, M. J.; Gehler, P. V. Unite the people: Closing the loop between 3D and 2D human representations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4704–4713, 2017.

[5] Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M. J. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics* Vol. 34, No. 6, Article No. 248, 2015.

[6] Wei, S.-E.; Ramakrishna, V.; Kanade, T.; Sheikh, Y. Convolutional pose machines. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 4724–4732, 2016.

[7] Cao, Z.; Simon, T.; Wei, S.; Sheikh, Y. Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1302–1310, 2017.

[8] Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H.-P.; Xu, W.; Casas, D.; Theobalt, C. VNect: Realtime 3D human pose estimation with a single RGB camera. *ACM Transactions on Graphics* Vol. 36, No. 4, Article No. 44, 2017.

[9] Alldieck, T.; Magnor, M.; Xu, W.; Theobalt, C.; Pons-Moll, G. Video based reconstruction of 3D people models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 8387–8397, 2018.

[10] Kanazawa, A.; Black, M. J.; Jacobs, D. W.; Malik, J. End-to-end recovery of human shape and pose. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7122–7131, 2018.

[11] Varol, G.; Ceylan, D.; Russell, B.; Yang, J.; Yumer, E.; Laptev, I. Bodynet: Volumetric inference of 3D human body shapes. In: Proceedings of the European Conference on Computer Vision, 20–36, 2018.

[12] Zheng, Z.; Yu, T.; Wei, Y.; Dai, Q.; Liu, Y. Deephuman: 3D human reconstruction from a single image. In: Proceedings of the IEEE International Conference on Computer Vision, 7739–7749, 2019.

[13] Saito, S.; Huang, Z.; Natsume, R.; Morishima, S.; Kanazawa, A.; Li, H. PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization. In: Proceedings of the IEEE International Conference on Computer Vision, 2304–2314, 2019.

[14] Xu, Y.; Zhu, S.-C.; Tung, T. Denserac: Joint 3D pose and shape estimation by dense render-and-compare. In: Proceedings of the IEEE International Conference on Computer Vision, 7760–7770, 2019.

[15] Smith, D.; Loper, M.; Hu, X.; Mavroidis, P.; Romero, J. FACSIMILE: Fast and accurate scans from an image in less than a second. In: Proceedings of the IEEE International Conference on Computer Vision, 5329–5338, 2019.

[16] Alldieck, T.; Magnor, M.; Bhatnagar, B. L.; Theobalt, C.; Pons-Moll, G. Learning to reconstruct people in clothing from a single RGB camera. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1175–1186, 2019.

[17] Kolotouros, N.; Pavlakos, G.; Black, M. J.; Daniilidis, K. Learning to reconstruct 3D human pose and shape via modelfitting in the loop. In: Proceedings of the IEEE International Conference on Computer Vision, 2252–2261, 2019.

[18] Liang, J.; Lin, M. C. Shape-aware human pose and shape reconstruction using multi-view images. In: Proceedings of the IEEE International Conference on Computer Vision, 4352–4362, 2019.

[19] Yang, S.; Pan, Z. R.; Amert, T.; Wang, K.; Yu, L. C.; Berg, T.; Lin, M. C. Physics-inspired garment recovery from a single-view image. *ACM Transactions on Graphics* Vol. 37, No. 5, Article No. 170, 2018.

[20] Yang, S.; Liang, J.; Lin, M. C.; Learning-based cloth material recovery from video. In: Proceedings of the IEEE International Conference on Computer Vision, 4383–4393, 2017.

[21] Qiao, Y. L.; Liang, J. B.; Koltun, V.; Lin, M. C. Scalable differentiable physics for learning and control. *arXiv preprint* arXiv:2007.02168, 2020.

[22] De Avila Belbute-Peres, F.; Smith, K. A.; Allen, K.; Tenenbaum, J.; Kolter, J. Z. End-to-end differentiable physics for learning and control. In: Proceedings of the Advances in Neural Information Processing Systems, 2018.

[23] Degrave, J.; Hermans, M.; Dambre, J.; Wyffels, F. A differentiable physics engine for deep learning in robotics. *Frontiers in Neurorobotics* Vol. 13, 6, 2019.

[24] Liang, J.; Lin, M.; Koltun, V. Differentiable cloth simulation for inverse problems. In: Proceedings of the 33rd Conference on Neural Information Processing Systems, 2019.

[25] Hu, Y.; Liu, J.; Spielberg, A.; Tenenbaum, J. B.; Freeman, W. T.; Wu, J.; Rus, D.; Matusik, W. ChainQueen: A real-time differentiable physical simulator for soft robotics. In: Proceedings of the International Conference on Robotics and Automation, 6265–6271, 2019.

[26] Hu, Y. M.; Anderson, L.; Li, T. M.; Sun, Q.; Carr, N.; Ragan-Kelley, J.; Durand, F. DiffTaichi: Differentiable programming for physical simulation. *arXiv preprint* arXiv:1910.00935, 2019.

[27] Liu, K. X.; Zeng, X. Y.; Bruniaux, P.; Tao, X. Y.; Yao, X. F.; Li, V.; Wang, J. 3D interactive garment pattern-making technology. *Computer-Aided Design* Vol. 104, 113–124, 2018.

[28] Huang, P.; Yao, J.; Zhao, H. Automatic realistic 3D garment generation based on two images. In: Proceedings of the International Conference on Virtual Reality and Visualization, 250–257, 2016.

[29] Wang, T. Y.; Ceylan, D.; Popović, J.; Mitra, N. J. Learning a shared shape space for multimodal garment design. *ACM Transactions on Graphics* Vol. 37, No. 6, Article No. 203, 2018.

Junbang Liang is a 4th-year Ph.D. student in the University of Maryland, College Park. He received his B.E. degree from Tsinghua University in 2016, and his M.S. degree from the University of North Carolina in 2018. His research interests are physics-based cloth simulation, computer vision, and machine learning.

Ming C. Lin is a Distinguished University Professor and Elizabeth Stevinson Iribe Chair of Computer Science at the University of Maryland College Park and John R. and Louise S. Parker Distinguished Professor Emerita of Computer Science at the University of North Carolina, Chapel Hill. She obtained her B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley. She is a Fellow of ACM, IEEE, and Eurographics, and a member of ACM SIGGRAPH Academy.

TSINGHUA UNIVERSITY PRESS Springer