MACHINE LEARNING FOR DIGITAL TRY-ON

Junbang Liang **Ming C. Lin**
University of Maryland, College Park

ABSTRACT

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. With recent advances in machine learning, these challenging problems become increasingly more tractable. We list a set of three open problems towards a complete and easy-to-use try-on system that can be enabled by recent advances in machine learning. For each of them, we define the problem, introduce state-of-the-art approaches, and provide future directions. A digital try-on system enabled by machine learning techniques can further enhance the consumer's E-shopping experience and provide notable economic benefits to the society.

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 (Giovanni et al., 2012). 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.

In this paper, we present the key open research issues that contribute to the gap. They include:

1. Accurate estimation of human shapes and sizes through consumer devices;
2. Faithful recovery of garment materials via (online) images;
3. Ease of design and manipulation on sewing patterns and garment pieces by end-users.

Although traditional methods have made some progresses on these under-constrained problems, learning-based approaches have shown tremendous potential in making notable impact. For each problem listed above, we motivate its importance, give a problem definition, and present state-of-the-art approaches with potential improvements. We believe that the solutions to these challenging problems will lead to significant advances in digital try-on, as well as other areas of E-commerce.

2 OPEN PROBLEMS

In this section we first introduce the three major challenges that hamper digital try-on from being widely adopted and accepted by customers. 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 a full 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 mainstream ways of showing 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 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 suffice the needs. Last, but not least, a fast and vivid animation of the garments in motion along with the body movement will greatly improve the user experience. Although it is not so critical as other factors, it would effectively reduce the perceptual gap between the real world and the digital one for online shopping. Previous work (Liang & Lin, 2018) used cloud computing to improve the animation speed, but there is still a notable technology gap towards high-quality, interactive 3D animation of clothes.

2.1 HUMAN SHAPE ESTIMATION

As mentioned above, an accurate human shape estimation algorithm is the key to enabling the functionality of 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/augmented environments. Yet, it remains a challenging and popular topic of interest. While direct 3D body scanning can provide excellent and sufficiently accurate results, its adoption is somewhat limited by the required specialized hardware. To this end, it is preferred to use RGB images as input rather than expensive scanned meshes when applied to digital try-on. The problem can be defined informally as, given one or a set of RGB images, estimate the human body sizes and output preferably a 3D humanoid mesh. Traditional algorithms often formulate it as an optimization problem, and compute the silhouette difference as its major part of the target function (Dibra et al., 2016). Therefore, these methods either require the human to wear tight clothes, or alternatively relax the target function to be unilateral on uncovered body parts (Bălan & Black, 2008) or to point correspondences (Lassner et al., 2017). The current state of the art is the recent work by Liang & Lin (2019) that emphasizes on shape learning, while many other works often focus on body-joint losses, but neglect the effect of body shapes.

Their key contribution is a multi-view multi-stage framework to address the ambiguity issue caused by camera projection. Their model is iteratively run for several stages of error correction. Inside each stage, the multi-view image input is passed on one at a time. At each step, the shared-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 camera calibration and the global rotation. Their proposed framework 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. Different from static multi-view CNNs which have to fix the number of inputs, they make use of the RNN-like structure in a cyclic form to accept any number of views, and avoid the gradient vanishing by predicting the corrective values instead of the updated parameters at each regression block.

(a) The input image. (b) Liang & Lin (2019).

Figure 1: Prediction results of the state of the art. The model can capture the shape of the human body by learning from synthetic data. The recovered legs and chest are close to the person in the image.

Experiments have shown that this trained model can provide equally good pose estimation as the 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. Moreover, a physically-based synthetic data generation pipeline is introduced to enrich the training data, which is very helpful for shape estimation and regularization of end effectors 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 promising direction is to transfer the synthetic result to more realistic images to further improve the learning result.

2.2 Garment Material Cloning

Garment material plays an important role in digital try-on systems. Physical recreation of the fabric not only gives a compelling visual simulacrum of the cloth, but also affects how the garment feels and fits on the body. However, modeling material is a challenging task: the visual effect and physical interaction of the garment is determined not only by the type of matter it is made of, but also the way of sewing and yarning. Because of this, researchers often focus on the physical properties it behaves, rather than the underlying semantic primitives.

Following such assumption, we model the garment material cloning problem as below. Given a sufficient amount of data, model its physical behaviors and compute the corresponding properties, such that the same or similar visual effect can be reproduced on computers. It has two implication: first we need to give a physical model of the material, next we estimate the parameters in the model.

There are many works to model clothes, including spring-mass systems (Provot et al., 1995) and finite elements (Etzmuß et al.). Finite elements method is the most popular model since it can produce realistic results. While one can use isotropic properties such as Yang's Modulus and Poisson Ratio, anisotropic model is the better choice since it can support different behaviors caused by yarning. One of the state-of-the-art model is from Wang et al. (2011) where they use a piece-wise linear model to account for different forces in different magnitudes of deformation. Building upon it, Yang et al. (2017) used CNN combined with LSTM to recover the material parameters from videos.

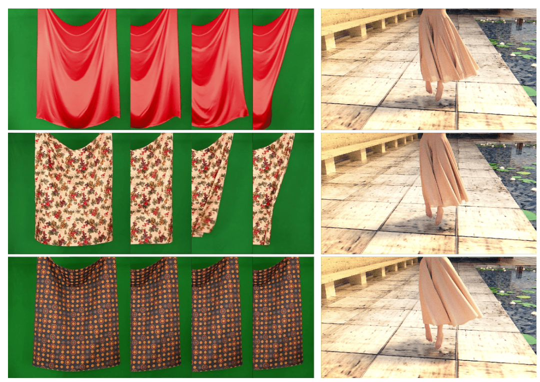

Figure 2: Yang et al. (2017) predicted the material type of the cloth in input videos from the left and cloned those material on to the skirt. The simulated skirt are shown on the right.

To constrain both the input and solution space, they choose one of the material as the basis and 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 that would otherwise be difficult or time-consuming to capture. The cloth meshes are generated through physically based simulation, then rendered to 2D images with a randomly assigned texture. With the data samples, they combine the image signal feature extraction method, CNN, with the temporal sequence learning method, LSTM, to learn the mapping from visual "appearance" to "material". 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 sequential movement.

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

2.3 Garment Modeling and Design

Realistic apparel model generation has become increasingly popular, due to the rapid change of the fashion trend and the growing need for garment model in different applications such as virtual try-on. For the application requirements, 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. First, garments usually have different types of topology, especially for fashion apparel, that makes it difficult to design a universal generation pipeline. Moreover, it is often not straightforward for the general garments design to be retargeted onto another body shape, making it difficult for customization.

Some previous work has addressed this problem to some extent. Huang et al. (2016) proposed an realistic 3D garment generation algorithm based on front and back image sketches, but it cannot retarget the generated garments to other body shape easily. Wang et al. (2018) proposed an algorithm which can do retargeting conveniently, but have limited topology like T-shirt or skirt. As such, there is no recent work that addresses these two problems at the same time.

Generally, we prefer 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 UV space of the human body as an 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 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 the relative geometry information with respect to the human body, enabling garment retargeting to a different human body.

3 Conclusion

Although Virtual Reality and digital try-on demonstrate significant potential and are rapidly developing, there still open problems to address before the online try-on systems can be widely adopted. We listed three major challenges, all of which can solved or improved by machine learning algorithms. For each of these challenges, we motivate its importance, briefly describe the current research progress, and point to possible future directions. We believe there will be substantial breakthroughs if more investigation is carried out towards these directions.

References

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

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*, pp. 108–117. IEEE, 2016.

Olaf Etzmuß, Michael Keckeisen, and Wolfgang Straßer. A fast finite element solution for cloth modelling. In *11th Pacific Conference on Computer Graphics and Applications, 2003. Proceedings.* IEEE.

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*, pp. 55–65. Springer, 2012.

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.

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, pp. 3, 2017.

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

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*, pp. 4352–4362, 2019.

Xavier Provot et al. Deformation constraints in a mass-spring model to describe rigid cloth behaviour. In *Graphics interface*, pp. 147–147. Canadian Information Processing Society, 1995.

Huamin Wang, James F O'Brien, and Ravi Ramamoorthi. Data-driven elastic models for cloth: modeling and measurement. In *ACM Transactions on Graphics (TOG)*, volume 30, pp. 71. ACM, 2011.

Tuanfeng Y Wang, Duygu Ceylan, Jovan Popović, and Niloy J Mitra. Learning a shared shape space for multimodal garment design. In *SIGGRAPH Asia 2018 Technical Papers*, pp. 203. ACM, 2018.

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*, pp. 4383–4393, 2017.