Real-time Image-based Virtual Try-on with Measurement Garment

Toby Chong
The University of Tokyo, Hongo
Tokyo, Japan

I-Chao Shen
The University of Tokyo, Hongo
Tokyo, Japan

Yinfei Qian
The University of Tokyo, Hongo
Tokyo, Japan

Nobuyuki Umetani
The University of Tokyo, Hongo
Tokyo, Japan

Takeo Igarashi
The University of Tokyo, Hongo
Tokyo, Japan

Figure 1: Our real-time system allows the user to virtually try-on various garments using different interactions, such as stretching the body and pulling the garment. The user first wears the color-coded measurement garment (top row), and our system synthesizes the target garments with detailed wrinkles according to different motions.

ABSTRACT

Virtual try-on technology that replaces a customer's wearing with arbitrary garments can significantly improve the online cloth shopping experience [1, 2, 4]. In this work, we present a real-time image-based virtual try-on system composed of two parts, i.e., photo-realistic clothed person image synthesis for the customers to experience the try-on result and garment capturing for the retailers to capture the rich deformations of the target garment. Distinguished from previous image-based virtual try-on works, we formulate the problem as a supervised image-to-image translation problem using a measurement garment, and we capture the training data with a custom actuated mannequin.

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; **Computer vision**; • **Human-centered computing** → **Human computer interaction (HCI)**.

KEYWORDS

Virtual Try-on, Deep image synthesis

ACM Reference Format:
Toby Chong, I-Chao Shen, Yinfei Qian, Nobuyuki Umetani, and Takeo Igarashi. 2021. Real-time Image-based Virtual Try-on with Measurement Garment. In *ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia (SIGGRAPH ASIA '21), December 14–17, 2021, virtual.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

SIGGRAPH ASIA '21, December 14–17, 2021, virtual

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

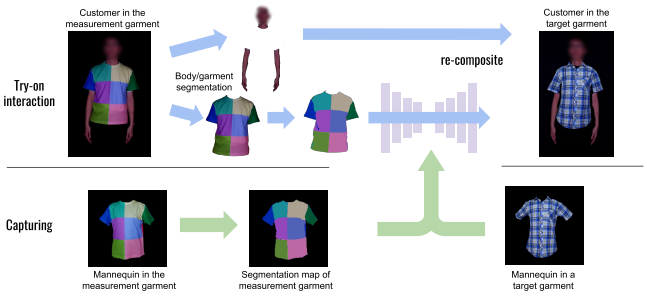

Figure 2: Overview of our system.

1 OUR METHOD

To achieve realistic real-time virtual try-on and to support common interactions during try-on, we propose a novel clothing person image synthesis method and a per garment capturing method (Figure 2). The clothed person synthesis module then learns the mapping between the captured paired data using an image-to-image network. The garment capturing module captures RGBD images of the actuated mannequin wearing a garment (including the measurement garment and the target garments). It produces the paired data between the measurement and the target garment under the same body sizes and poses.

Synthesis and try-on. When a customer comes into the fitting room, the customer first wears the measurement garment and stands in front of the video camera. Then the customer chooses a target garment among those captured by the retailer and she can move in front of the camera and change poses during the virtual try-on. Our synthesis module takes the images of measurement and target garment as training data, and we train an image-to-image translation network for each measurement garment and target garment pair. This system does not require retraining the image-to-image translation network for each customer. Generating the try-on result contains three stages: our system first extracts the garment from the input camera feed, and used the extracted garment as the input to the image-to-image translation model to infer the resulting target garment. Finally, the inferred target garment image from the image-to-image translation model is then re-composed back into the original frame.

Garment capturing. To use our system, the retailer first needs to capture the measurement garment using the actuated mannequin in a capturing studio. After capturing the measurement garment, the retailer can capture arbitrary target garments by repeating the same capturing process. And then, they can use the captured paired data to train a image-to-image translation network for each target garment separately. We designed the basic body shape of our actuated mannequin that approximates the average male body shapes in a commonly used parametric body shape model SMPL [3]. To enable the mannequin to pose for a large variety of motions commonly found during physical try-out sessions in a physical store, we equipped it with nine degrees-of-freedoms, including independent arm movement for both arms, four adjustable body region, and rotation around the center (Figure 3). To improve the robustness of the translation network used in [1] over different lighting condition,

Figure 3: Actuated mannequin DoFs. (a) The torso of our actuated mannequin is segmented into four adjustable regions (illustrated in red arrows), with three front-facing actuators and one side-facing actuator. (b) (c) And each upper arm is actuated with two servo motors, allowing movement along the yaw and pitch axes.

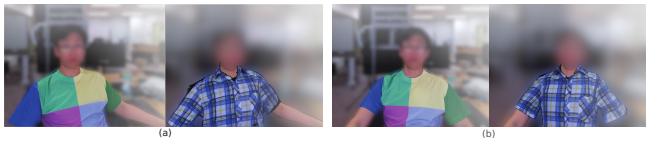

Figure 4: Two example frames where we change the measurement garment into a shirt during a video conference.

we further capture both the measurement garment and the target garments under different light conditions.

2 RESULTS AND APPLICATIONS

To demonstrate the effectiveness of our system, we show the virtual try-on results of various target garment in Figure 1. On the top row of Figure 1, we show the users wearing the measurement garment and perform different common poses during try-on. In the bottom row, we show the virtual try-on results synthesized by our method. Our try-on results are with (i) detailed wrinkles,(ii) preserved textures on the garment, and (iii) clear body parts.

2.1 Cloth changing for video conference

In recent days, we often have professional remote video conferences with colleagues due to pandemic. Our method can virtually change the upper body clothing without physically changing to formal clothing such as shirt. Please watch the supplemental video for the result.

3 ACKNOWLEDGEMENT

[to be added]

REFERENCES

[1] anonymous. 2021. Per Garment Capture and Synthesis for Real-time Virtual Try-on. In *Proceedings of the 34rd Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA.

[2] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. 2018. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7543–7552.

[3] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. 2015. SMPL: A skinned multi-person linear model. *ACM transactions on graphics (TOG)* 34, 6 (2015), 248.

[4] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, and Ping Luo. 2020. Towards Photo-Realistic Virtual Try-On by Adaptively Generating-Preserving Image Content. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.