VIRTUAL 3D TRIAL ROOM

Anusha A G
Dept of Computer Science and Engineering
K S Institute of Technology
Bengaluru, Karnataka, India
anushashrike1999@gmail.com

Deepika S H
Dept of Computer Science and Engineering
K S Institute of Technology
Bengaluru, Karnataka, India
deepikahande06@gmail.com

Divya Yashaswi Kanney
Dept of Computer Science and Engineering
K S Institute of Technology
Bengaluru, Karnataka, India
disney.kstew@gmail.com

Priyanka Hanni
Dept of Computer Science and Engineering
K S Institute of Technology
Bengaluru, Karnataka, India
hannipriyanka@gmail.com

Kushal Kumar B N
Dept of Computer Science and Engineering
K S Institute of Technology
Bengaluru, Karnataka, India
kushalkumarbn@ksit.edu.in

Abstract

Presenting "Virtual 3D trial room", an application which uses CNN to produce a 3D model of a person and allows them to try on clothes virtually. Based on the parameters taken from the monocular video provided by the user as input, the 3D model is constructed. Dataset is used to train Machine Learning models to get almost accuracy during the construction of the 3D model. This created 3D model can try on clothes from the digital wardrobe which has clothes which have been converted to 3D models. Using MGN algorithm the perfect size of cloth for the individual can be predicted. The virtual 3d trial room has a wide range of applications; it can be used with the already existing online clothes shopping website to give users a better shopping experience. It would be extremely helpful during this COVID-19 pandemic and could be used later on as well.

Keywords: 3D model, digital wardrobe, monocular video, CNN, DeepCut.

Introduction

During this pandemic timeline, citizens felt obliged to violate the norms and protocols to keep up with their basic needs. These essentials include clothes. Following the guidelines to avoid human contact, buying convenient, fashionable and comfortable clothes is a challenge. Online shopping for clothes which was considered a solution has various issues of its own, such as users not getting a proper look and feel of the clothes being purchased. This issue could be solved by the use of a virtual 3D trial room. Virtual 3D trial rooms would involve a 3D model of the user and the user would be able to try on clothes virtually.

A personalized realistic, 3D model of a human is created. This model should

comprise the person-specific static geometry of the body, hair and clothing, alongside a coherent surface texture. This model is used to virtually try on clothes [2].

A common way to acquire such models is with a scanner which is not cost efficient, numerous researches have been made to capture body shape in a more practical setup. One such practical way would be capturing a video from a single monocular RGB camera [2].

The 3D model is constructed using a 3D generative model called SMPL(Skinned Multi-Person Linear) and CNN called DeepCut.[1] All this is done from a few frames of input video. The input video consists of a person standing in a T pose and rotating 360 degree, making sure the person is seen from all the sides. After the model creation the user is allowed to try clothes from the digital wardrobe.

Literature survey

A] Video Based Reconstruction of 3D People Models

This paper describes a method to obtain accurate 3D body models and texture of arbitrary people from a single, monocular video in which a person is moving. Based on a parametric body model, the paper presents a robust processing pipeline to infer 3D model shapes including clothed people with 4.5mm reconstruction accuracy. The approach used is the transformation of dynamic body pose into a canonical frame of reference. This is done using only an RGB camera.

There are three steps involved in this method:

1) pose reconstruction
2) consensus shape estimation
3) frame refinement and texture map generation.

B] Learning to Reconstruct People in Clothing from a Single RGB Camera - Octopus

Octopus, a learning-based model to infer the personalized 3D shape of people from a few frames (1-8) of a monocular video in which the person is moving with a reconstruction accuracy of 4 to 5mm, while being orders of magnitude faster than previous methods. From semantic segmentation images, Octopus model reconstructs a 3D shape, including the parameters of SMPL plus clothing and hair in 10 seconds or less. The model achieves fast and accurate predictions based on two key design choices. First, by predicting shape in a canonical T-pose space, the network learns to encode the images of the person into pose invariant latent codes, where the information is fused. Second, based on the observation that feed-forward predictions are fast but do not always align with the input images, it predicts using both, bottom-up and top-down streams (hybrid method) allowing information to flow in both directions. Learning relies only on synthetic 3D data. Once learned, Octopus can take a variable number of frames as input, and is able to reconstruct shapes even from a single image with an accuracy of 5mm.

C] Multi-Garment Net: Learning to Dress 3D People from Images

Multi-Garment Network (MGN), a method to predict body shape and clothing, layered on top of the SMPL model from a few frames of a video. This model allows to predict garment geometry, relate it to the

body shape, and transfer it to new body shapes and poses.

Methodology

The system has 2 steps.
1] To create a 3D model of the user from the input video.
2] Dress any body shape in arbitrary poses using the available 3D digital wardrobe.

The 3D model creation has three stages. First, it analyses the video for a few seconds long of someone moving with T-pose, preferably turning 360°- to show all sides and for each frame creates a silhouette separating the person from the background. Based on machine learning techniques in which computers learn a task from many examples it roughly estimates the 3D body shape and location of joints using SMPL. In the second stage, it combines information about the T-posed people into one, more accurate model. Finally, in the third stage, it applies colour and texture to the model based on recorded hair, clothing, and skin [7].

In the next step, the 3D model is analysed to separate and extract the underlying body geometry, motion component and the clothing as separate geometric layers. MGN model is used to distinguish the clothes and the body geometry and to dress any body shape in arbitrary poses. A publicly available digital wardrobe, the MGN model [3], and code is used to dress SMPL obtained 3D models with the garments.

Conclusion

Even though a lot of research is being done in the field of AR & VR for construction of 3D models, no specific application has been dedicated to virtually trying on clothes for online fashion shopping. Our application can be integrated with dedicated online cloth shopping websites to provide a platform to try-on clothes virtually. Hence providing better experience to the users while shopping online.

Reference

[1] B. Bhatnagar, G. Tiwari, C. Theobalt and G. Pons-Moll, "Multi-Garment Net: Learning to Dress 3D People From Images," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 5419-5429, doi: 10.1109/ICCV.2019.00552.

[2] T. Alldieck, M. Magnor, W. Xu, C. Theobalt and G. Pons-Moll, "Video Based Reconstruction of 3D People Models," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 8387-8397, doi: 10.1109/CVPR.2018.00875.

[3] Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, Gerard Pons-Moll "Learning to Reconstruct People in Clothing from a Single RGB Camera", 2019 IEEE Conference on Computer Vision and Pattern Recognition, Germany.

[4] 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 Conf. on Computer Vision. Springer International Publishing, 2016.

[5] Jinlong Yang, Jean-Sebastien Franco, Franck Hetroy- ´ Wheeler, and Stefanie Wuhrer. Analyzing clothing layer deformation statistics

of 3d human motions. In European Conf. on Computer Vision, pages 237–253, 2018

[6] W. Xu, A. Chatterjee, M. Zollhoefer, H. Rhodin, D. Mehta, H.-P. Seidel, and C. Theobalt. Monoperfcap: Human performance capture from monocular video. In ACM Transactions on Graphics, 2018

[7] Matthew Hutson, “Watch artificial intelligence create a 3D model of a person—from just a few seconds of video”, AAAS Journal Apr. 3,2018. [online]. Available: https://www.sciencemag.org/news