VIRTUAL 3D TRIAL ROOM

Anusha A G[1], Deepika S H[2], Divya Yashaswi Kanney[3], H Priyanaka [4] , Kushal Kumar B N[5]

[1] Student, Department of Computer Science and Engineering K S Institute of Technology , Karnataka, India
[2] Student, Department of Computer Science and Engineering K S Institute of Technology , Karnataka, India
[3] Student, Department of Computer Science and Engineering K S Institute of Technology , Karnataka, India
[4] Student,Department of Computer Science and Engineering K S Institute of Technology , Karnataka, India
[5] Assistant professor, Department of Computer Science and Engineering K S Institute of Technology , Karnataka, India

ABSTRACT

Shopping online can be a tedious process as the user has no way to see how the clothes would look on them. This problem can be solved by creating a 3D model to do the same. 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.
Presenting "Virtual 3D trial room", an application which uses Convolutional Neural Network (CNN) and OpenCV to produce a 3D model of a person and allows them to try on clothes virtually. The monocular video of a person turning around 360 degrees is provided as input by the user, out of this 8 frames are collected to create the 3D model. 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 the Multi-Garment Network (MGN) algorithm the look and feel of the cloth for the individual can be estimated. 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 to promote social distancing and quarantine, this could also be used later on as well. The process would involve various steps such as separating the person from the background, estimating the joint location, creating the texture of the person, etc.

Keyword : *3D model, digital wardrobe, monocular video, CNN*

---

1. INTRODUCTION

During this pandemic people weren't allowed to go outside due to which people had to shop online for everything including clothes. It was particularly inconvenient as there is no way to know how the clothes would fit. This issue could be solved by the use of Virtual 3D trial room wherein a 3D model of the person is created using a few frames taken from the video of the user turning around 360 degrees. New clothes would then be displaced on the created 3D model to check the fitting of the clothes.

2. PROBLEM STATEMENT

Due to the pandemic, everyone has taken to shopping online for their essentials. These essentials include clothing, and going outside to buy clothes hinders the social distancing health measure. Shopping online for clothes is difficult as there is no way to check the fitting and comfort of the clothes. Hence, a virtual 3D trial room allowing the user to upload a video of themselves along with few other inputs such as height and weight resulting in a robust, texture-rich 3D human model with clothes tried on would be helpful.

3. LITERATURE SURVEY

- MGN: *Bhatnagar et al. 2019.* [1] developed the first model which is capable of inferring the human body and layered garments on top as separate meshes from images directly. Then the clothes are displaced on the model.
- CAPE: *Q. Ma et al. 2020.* [2] developed a model that factorizes clothed human bodies into two parts: the minimally-clothed body, and a clothing layer represented as displacements from the body which enables the model to naturally extend SMPL to a class of clothing types by treating clothing as an additional additive shape term. This generative model does so to add realistic clothing deformations and wrinkles for variable poses.
- TailorNet: *Patel et al. 2020.* [10] is a model developed over a mixture of Neural Networks learned from Physics based simulations. It learns deformations as displacements to a garment template in a canonical pose, while the articulated motion is driven by skinning.
- SIZER: *Tiwari et al. 2020.* [2] developed SIZER dataset containing 100 subjects wearing 10 garment classes, SizerNet, a model that learns from real scans to predict how clothing drapes on the body as a function of size and ParserNet, a model that maps a single mesh registration into a multi-layered representation of clothing without the need for segmentation or non-linear optimization.
- SMPLicit: *Corona et al. 2021.* [3] is a topology-aware generative model that separates multiple cloth layers by retrieving complex garment geometries under different body poses.

3.METHODOLOGY AND RESULTS

3.1 Human parsing / Semantic Part Segmentation on instance level

This is done on the frames taken from the video. Semantic segmentation is used for obtaining human silhouette.[4]
PGN recasts instance-level human parsing as two sub-tasks that may be learned and refined together through a single network:
1) semantic part segmentation to assign each pixel as a human part (e.g., face, arms);
2) instance-aware edge detection to group semantic parts into unique person instances.
As a result, the shared intermediate representation would be capable of describing fine-grained pieces as well as inferring instance belongings for each part. Finally, during inference, a basic instance partition technique is used to obtain final results.

Fig-1: PGN Segmentation

3.2 Pose estimation

It involves detecting key point location that describes the object. Human pose estimation can be done by detecting various landmarks such as neck, nose, eyes, hands, legs, etc to generate the skeleton.[6]
Pose estimate entails determining the location of key points that describe the item. Various markers such as the neck, nose, eyes, hands, and legs can be used to estimate human position.
There are three phases involved here: [7]

- Localization of important points: One branch of the convolutional neural network is in charge of predicting all of the key points and assigning a confidence value to each one. This is referred to as a confidence map.

- Another branch of the network predicts a 2D vector field connecting joints and limbs, which predicts key point association data.
- Greedy Inference: Greedy inference is used to connect all of the keypoints.

Fig-2: Keypoint estimation (JSON file)

3.3 Texture stitching

Texturing the models' surfaces is an important step in the reconstruction of 3D models. Texture mapping is used on the surface of a 3D human model to provide a continuous, integrated, and smooth texture. This is done as an image stitching optimization.
The following are the steps involved:

- Texture Completion
- Segmentation completion
- Displacement map prediction

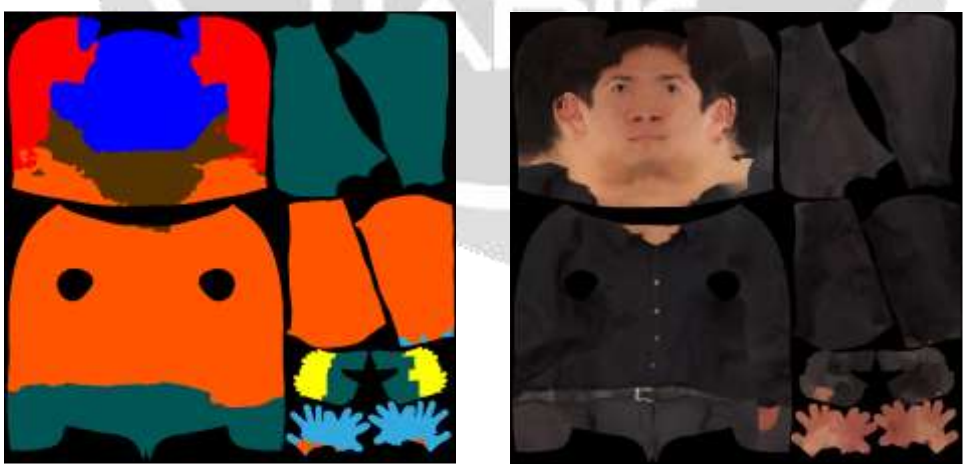

Fig-3: Texture stitching

3.4 Creating 3D shape/ model

Inputs such as frames, human silhouette, texture maps, estimated pose are taken as input and a mesh is created which is further used to create the 3D model. For this purpose we extend the SMPL model with per-vertex displacements/deformations where garments are obtained separately instead of treating the complete as single surface geometry.[3,7]

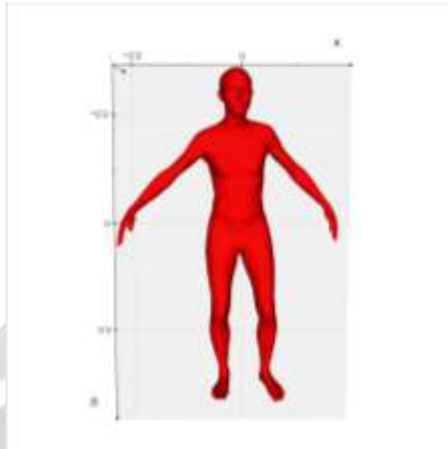

Fig-4: Skinned 3D model

3.5 Separating clothes from 3D model (MGN)

MGN's formula is used to extract garments from the images of a source and then the inferred 3D garments are used to dress arbitrary human bodies in various poses from SMPL shape subjects. [1]

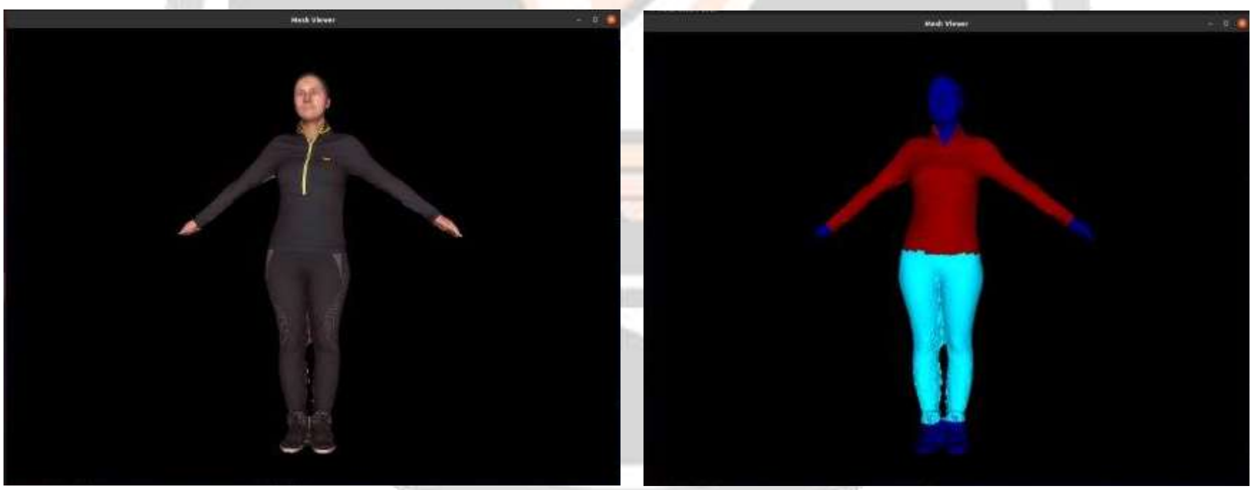

Fig-5: Garments separated from the perso

3.6 Clothing

In this step the 3D model is taken and clothes from the digital wardrobe is placed on the person to see how the clothes look.[1]

Fig-6: Dressed SMPL model

4. CONCLUSIONS

In conclusion, contactless virtual trials of clothes with the least amount of return rates can be achieved using this application. It is a good step towards 3D model visualization for online shopping. A video taken from a smartphone with an RGB camera can be uploaded to the system to create a 3D model of the person. Multiple frames are collected from the uploaded video for segmentation purposes, detecting key points, etc., to create a 3D model of the person using SMPL. This created 3D model is then dressed using the clothes from an existing dataset.

This project creates a 3D model of a person and allows them to see how the clothes would look on the person,

Future enhancements:

The future enhancement for the project would be to create a more realistic human model, predict the proper fitting of the clothes for the 3D model and ultimately create an API of the project, so it can be used in clothing websites, thus allowing the users to make a more informed choice.

6. REFERENCES

[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), 2019, pp. 5419-5429, doi: 10.1109/ICCV.2019.00552.

[2] Tiwari, G., Bhatnagar, B.L., Tung, T., Pons-Moll, G.: **Sizer: A dataset and model for parsing 3d clothing and learning size sensitive 3d clothing**. In: European Conference on Computer Vision (ECCV). Springer (aug 2020)

[3] Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M.J.: **Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image**. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) European Conf. on Computer Vision. Springer International Publishing (2016)

[4] Gong, K., Liang, X., Li, Y., Chen, Y., Yang, M., & Lin, L. (2018). **Instance-level Human Parsing via Part Grouping Network**. ArXiv, abs/1808.00157.

[5] G. Pavlakos et al., "**Expressive Body Capture: 3D Hands, Face, and Body From a Single Image,**" 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10967-10977, doi: 10.1109/CVPR.2019.01123.

[6] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "**OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields**," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2929257.

[7] Alldieck, T., Magnor, M., Xu, W., Theobalt, C., Pons-Moll, G.: **Detailed human avatars from monocular video**. In: International Conference on 3D Vision (3DV) (sep 2018)

[8] B. Jiang, J. Zhang, Y. Hong, J. Luo, L. Liu, H. Bao. (2020). **BCNet: Learning Body and Cloth Shape from A Single Image.** arXiv:2004.00214v2 Available at: https://arxiv.org/pdf/2004.00214.pdf

[9] T. Alldieck, M. Magnor, B. L. Bhatnagar, C. Theobalt, and G. Pons-Moll. **Learning to Reconstruct People in Clothing from a Single RGB Camera**. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019

[10] C. Patel, Z. Liao, G. Pons-Moll. **TailorNet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style**. arXiv:2011.14619 Available at: https://arxiv.org/pdf/2011.14619.pdf