EUROGRAPHICS 2017 / L. Barthe and B. Benes
(Guest Editors)

Volume 36 (2017), Number 2

DeepGarment : 3D Garment Shape Estimation from a Single Image - Supplement

R. Daněřek[1,2,⋆], E. Dibra[1,2,⋆], C. Öztireli[1], R. Ziegler[2], M. Gross[1]

[1] Computer Graphics Laboratory, ETH Zürich
[2] Vizrt Switzerland
⋆ These authors contributed equally

1. Introduction

This supplementary material contains information which is additional to the original paper. It includes detailed specification of both single and multi-view neural net architecture we use, additional evaluations, experiments and further discussion of limitations of our technique.

2. Technique specification

2.1. Full Specification of the Architecture

Here we include the full detail of both single and two-view architectures, including the expansion on Fire layers (as proposed in [IMA*16]). The architectures are captured in Figure 1. Here we expand on the choice of architecture. In the beginning of the project we have experimented with much bigger CNNs such as Deep Residual Nets. However, due the size of such networks and the heavy experimentation needed, we decided to opt for SqueezeNet, which delivers comparable performance, is much faster to train and in our experience converged more reliably. For this reason, we encourage the use of SqueezeNet for similar problems. Despite that, it is definitely possible to optimize the results further with deeper architectures.

3. Additional Results

In this section we show a potential application of our method as a fast approximation of physically based clothing simulation. Furthermore, we perform a simple quantitative experiment on real data.

Fast Garment Simulations A potential application of our method is speeding-up of physically based cloth simulation (PBS). PBS is known to be computationally expensive. Therefore, one could simulate very coarse clothing and use the resulting rendered images in our technique. As long as the garments used for training and testing are similar in terms of their geometries and materials, this image-based method works well without having to go through any scene-level 3D information, or requiring correspondences between the underlying body, skeleton, or garment meshes. The performance gain is dependent on the resolution of the high-res training and the low-res input meshes. In our experiments, we opted for a low-resolution mesh of 398 vertices instead of the original 6065, resulting in a speed-up of about 8x, allowing the simulation to run in real time. We demonstrate the quality of the resulting garments in Fig.2.

3.1. Pixel Overlap of the Reconstruction

Performing any meaningful quantitative evaluation on real-life data is not easy because there are no suitable datasets available and creating one is a very challenging task. To the best of our knowledge, such an accurate dynamic clothing capture system is not available. The consumer-level depth cameras are insufficient as they cannot capture high frequency details.

Therefore, to provide at least some evaluation on the real data we perform the following experiment. We render the reconstructed mesh of the garment in the correct orientation. Then we compare the image masks, which are the mask of the input image which was fed to our neural net model and the mask of the rendering of the estimated mesh in the correct orientation. Please note that the results of this experiment should be taken only as an informative lower bound as our rendering of the final estimation does not contain any occluders such as arms. Furthermore, the experiment does not measure the quality of the reconstructed 3D mesh as a whole. Wrinkles and other finer deformations are not considered by this metric. We report the average pixel overlap for single view and two view T-shirt dataset and also for the single view dress dataset. The averages are 87.9%, 89.4% and 91.4%.

4. Further Discussion

We have presented a novel approach to clothing shape estimation. The advantages of our technique are discussed in the paper. We are very well aware that it brings along many limitations which we will discuss next. We aim to prove our idea correct instead of deploying a ready-to-be-used, production-level solution for a specific purpose. However, we humbly believe our study to be a pioneering one with the potential to spark a fruitful line of research in this area. The number of use cases and the simplicity of the method has a very big potential for a vast amount of applications. Nevertheless,

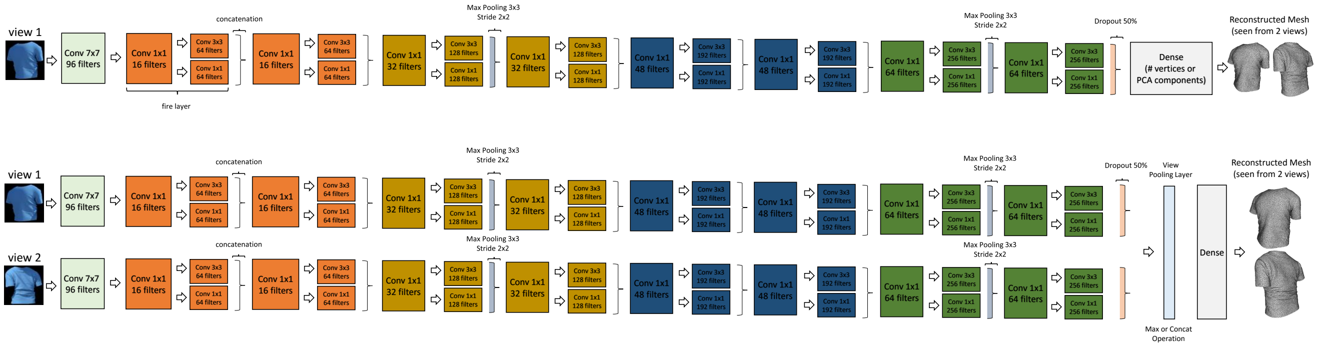

Figure 1: ***Top:*** *Single-view architecture.* ***Bottom:*** *Two-view architecture.*

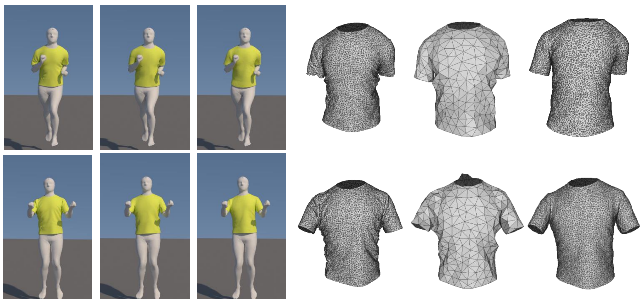

Figure 2: *From left to right: simulation with a high resolution mesh, with the corresponding low resolution mesh, and the estimated high resolution mesh with our method. For each case, we also show the raw mesh. The input to our method is the rendering of the simulation with the low resolution mesh.*

in order to be broadly usable much more research has to be put into this direction. This paper shall provide an initial setup, which can be explored much further in the future.

Garment Generality One of the limitations of our technique is that it has to be trained for every type of clothing as each reference garment mesh may vary dramatically in resolution based on its shape complexity.

Impact of the Dataset Generation on Performance Since our approach is data driven, the performance is only as good as the training dataset. There are many factors that come into play when creating such a dataset such as the artistic quality of the reference garment design, the resolution of the mesh, the versatility of the motion capture database and the characters animated by it, the correct behavior of the physically based simulation and last but not least the realism of the renderings (illumination, reflectance, texture, occluders etc.). Please note, that our data generation pipeline is not perfect and as such does not create a perfect dataset. Despite that, we are able to estimate the rough shape of the garments quite accurately. The closer the input image is to the training dataset, the better estimations we get and with very similar images (the synthetic test set), we show that we are even capable of recovering the high frequency detail. We believe that most of the flaws in the estimation could be lifted by creating an ultimate dataset. The creation of such a dataset, however, is another challenging engineering task. Another interesting line of research would be applying Generative Adversarial Nets [GPAM*14] which could help to create bigger variance in our training data. These relatively new approaches of generative and discriminative nets have not been used for clothing so far.

Temporal Smoothness In our supplementary material we have included the performance of our technique on a couple of real video sequences. Although the results are already temporally smooth overall, they can still exhibit certain artifacts or discontinuities. Please note that the focus of the paper was purely on single and two-view images. Some of the artifacts of the estimation in the videos can be caused by either an artifact in the segmentation (the importance of silhouettes was discussed in the paper) or the fact that certain poses were not present in the training set. The most visible artifacts in the video sequences, however, are temporal discontinuities. Please note that we do not enforce temporal smoothness in any way as we perform the estimation for each frame independently. However, we are confident that the results can be further improved by applying RNNs. We leave this as an interesting topic for future work.

Dependency on Segmentation It is true, that our technique requires quality segmentation of the garments on the input. Utilizing the recent deep learning based works on image segmentation such as [LSD14] could eventually alleviate the need for supervised segmentation in an uncontrolled environment altogether.

References

[GPAM*14] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y.: Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, Ghahramani Z., Welling M., Cortes C., Lawrence N. D., Weinberger K. Q., (Eds.). Curran Associates, Inc., 2014, pp. 2672–2680. 2

[IMA*16] IANDOLA F. N., MOSKEWICZ M. W., ASHRAF K., HAN S., DALLY W. J., KEUTZER K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR abs/1602.07360* (2016). 1

[LSD14] LONG J., SHELHAMER E., DARRELL T.: Fully convolutional networks for semantic segmentation. *CoRR abs/1411.4038* (2014). 2