Boundary-aided Human Body Shape and Pose Estimation from a Single Image for Garment Design and Manufacture

Zongyi Xu and Qianni Zhang

Queen Mary University of London, London, UK

Abstract

Current virtual clothing design applications mainly use predefined virtual avatars which are created by professionals. The models are unrealistic as they lack the personalised body shapes and the simulation of human body muscle and soft tissue. To address this problem, we firstly fit the state-of-the-art parametric 3D human body model, SMPL, to 2D joints and boundary of the human body which are detected using CNN methods automatically. Considering the scenario of virtual dressing where people are usually in stable poses, we define a stable pose prior from CMU motion capture (mocap) dataset for further improving accuracy of pose estimation. Accurate estimation of human body shape and poses provides manufacturers and designers with more comprehensive human body measurements, which put a step forwards clothing design and manufacture through Internet.

1. Introduction

2D images are the most convenient data source for acquiring 3D model for real people in the scenario of realistic virtual dressing because the most commonly available device in our daily life is our smart phones. We acquire our realistic human body avatar from a single 2D image by fitting the 3D parametric human body model, SMPL [LMR*15] to the image to estimate human body shapes and poses. We take the SMPLify [BKL*16] method as a base method and go beyond it to exploit boundary information of images to constrain the deformation of SMPL. Thus, we fit SMPL to the image with the constraints of both 2D joints as well as boundary and show that the boundary information improves the accuracy of estimated shapes significantly. Like other works, we use CMU mocap dataset to build human pose prior. In the scenario of virtual dressing people usually stand still or make slow movements. Given that CMU mocap dataset provides a wide range of motion data, we further improve the accuracy of pose estimation by extracting stable poses from CMU mocap data, providing stable pose priors to our work. Our accurate human body shape and poses though single 2D image allows for personalised garment design and manufacture.

2. Boundary Aided Human Body Shape and Pose Estimation

Build stable pose prior In the scenario of virtual clothing design, people commonly stand or move slowly in front camera. The pose variance is limited. As CMU dataset covers various human poses presented in daily life and sports for 144 subjects. A general pose prior cannot describe some specific poses accurately. Experiments show that the results of SMPLify present bent knees or stoop for the stable pose of "Stand". In order to provide more accurate pose prior for our case, we firstly extract the stable poses from CMU dataset. We define the stable poses to be those that change slightly in a short period of time. For each frame, we calculate the pose difference between its neighboring frames:

$$err = \frac{\sum_{k=-step}^{step} norm(\theta_i - \theta_{i+k})}{2 \times step} \tag{1}$$

Here, we set step to be 1 and θ is the pose parameter of the motion in each frame and when *err* is smaller than *threshold*, we regard the pose as stable poses. In our experiment, the threshold is set to be 0.001. Mosh [LMB14] is applied to calculate the pose parameters θ for each frame of stable poses, which captures motion and shape from sparse markers provided by CMU mocap data. With stable poses, Gaussian Mixture Model (GMM) is used to describe the pose prior for our work.

Full body estimation We take SMPL as our human body representation. Given the detected 2D joints and boundary of images, the full body estimation is formulated as :

$$E(\beta,\theta) = E_M(\beta,\theta) + E_b(\beta,\theta;K,U) \tag{2}$$

The $E_M(\beta,\theta)$ is the estimated human body model only relying on 2D joints J_{est} by :

$$E_M(\beta,\theta) = E_J(\beta,\theta;K,J_{est}) + \lambda_\theta E_{S\theta}(\theta) + \lambda_\alpha E_\alpha(\theta) + \lambda_\beta E_\beta(\beta) \tag{3}$$

where E_J is the data term which penalizes the distance between estimated 2D joints of images J_{est} and the corresponding projected SMPL joints. $E_\beta(\beta)$ is shape prior learned from the SMPL body shape training set. $E_{S\theta}(\theta)$ and $E_\alpha(\theta)$ are pose prior which are learned from precomputed stable poses. Here, $E_{S\theta}(\theta)$ can favor probable stable poses over unstable ones. After we trained our stable pose prior, $E_{S\theta}(\theta)$ is defined as the negative logarithm of a sum.

DOI: 10.2312/egp.20181022

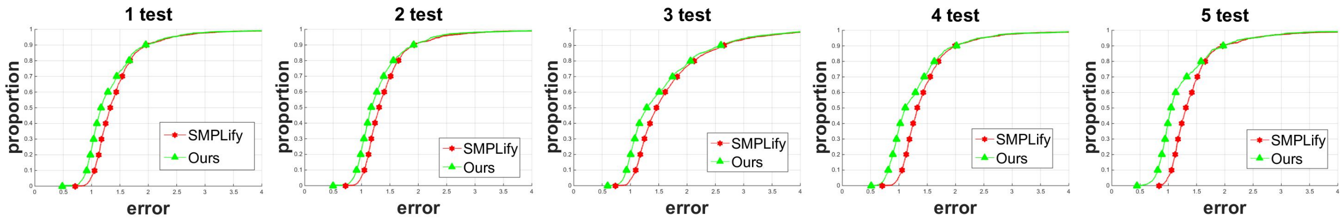

Figure 1: *The quantitative comparison of our method with SMPLify.*

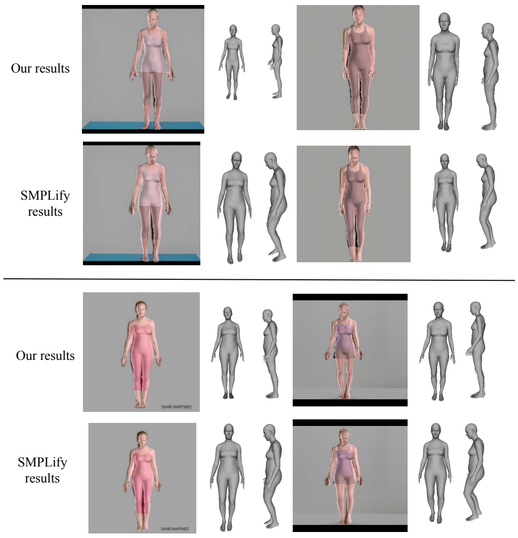

Figure 2: *The qualitative comparison of our method with SMPLify.*

Similar to [BKL*16], we approximate the sum in the mixture of Gaussian by a max operator:

$$E_{S\theta}(\theta) = -log\sum_{j}(g_j)\mathcal{N}(\theta;\mu_{\theta_j}\Sigma_{\theta_j}) \quad (4)$$

where μ_{θ_j} and Σ_{θ_j} are trained with our stable poses. Boundary information is very important to enlarge or shrink the model to make the final estimated human body shape similar to the real person. In Eq. 2, we define the cost function of boundary as:

$$E_b(\beta,\theta;K,U) = \sum_{i}^{N}||(B_i - U_i(\Pi_K(M(\beta,\theta))))||^2 \quad (5)$$

where B_i is the ith point on the boundary of images, $\Pi(\cdot)$ is the project function and $U_i(\cdot)$ is the corresponding points of B_i on the boundary of projected model. We minimize Eq. 2 and Eq.3 using *Powell*'s dogleg method, using OpenDR† and Chumpy‡. Optimization for a single image takes around 1 minute on a common desktop machine with 16 GB RAM and 4 cores.

† https://github.com/mattloper/opendr/wiki

‡ https://github.com/mattloper/chumpy

3. Results and Conclusion

We perform 5-fold cross validation for each class to compare our method against SMPLify in terms of pose estimation on 42797 stable poses in Fig. 1. We calculate the Euclidean error between the ground truth pose p_{gt} and the estimated pose p_{est} for each round over different proportions of stable pose dataset. The accuracy of our methods is higher than SMPLify in all rounds. Given the 2D joints of images, there would be cases that several different 3D joints have same 2D projection. Thus, the pose prior trained with wide range of poses cannot prevent this case. One million pose data is used in SMPLify to train the GMM model while we only use 30000 pose data to achieve better results. Visualized comparison is given in Fig. 2. Obviously, our estimated results "stand" more straight than SMPLify due to the stable pose prior. Seen from the side, our results are more consistent with what images show while SMPLify results tend to bend knees. This is because only 2D joints of a single image cannot provide enough constraints on pose estimation. The introduction of stable pose prior manages to favor possible stable poses over impossible ones.

In this poster, we propose to provide a stable pose prior and add boundary constraints for more accurate estimation of human body from a single image. We show that the pose estimation of proposed approach is more accurate than the state-of-the-art with only one percent of the training set of SMPLify and boundary information improves the accuracy of shape estimation significantly.

4. Acknowledgement

We thank Naureen Mahmood for information about SMPL pose data; Tuanfeng Y. Wang for technical discussion. Funding is provided by China Scholarship Council (CSC).

References

[BKL*16] 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 *European Conference on Computer Vision* (2016), Springer, pp. 561–578. 1, 2

[LMB14] Loper M., Mahmood N., Black M. J.: Mosh: Motion and shape capture from sparse markers. *ACM Transactions on Graphics (TOG) 33*, 6 (2014), 220. 1

[LMR*15] Loper M., Mahmood N., Romero J., Pons-Moll G., Black M. J.: Smpl: A skinned multi-person linear model. *ACM Transactions on Graphics (TOG) 34*, 6 (2015), 248. 1