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.

2D images are the most convenient data source for acquiring 3D model for real people in the scenario of realistic virtual dressing.

Objectives

- We acquire our realistic human body avatar from single 2D image
- To estimate human body shape and poses more accurately
- To put a step forwards clothing design and manufacture through Internet

Methodology

- Build stable pose prior
  - In the scenario of virtual clothing design, people commonly stand or move slowly in front camera.
  - 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:
    \[ err = \frac{\sum_{k=-\text{step}}^{\text{step}} \text{norm}(\theta_1 - \theta_{1+k})}{2 \times \text{step}} < \text{threshold} \]

- 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 :
    \[ E(\beta, \theta) = E_M(\beta, \theta) + E_b(\beta, \theta; K, U) \]
  - The \( E_M(\beta, \theta) \) is the estimated human body model only relying on 2D joints \( f_{\text{est}} \) by:
    \[ E_M(\beta, \theta) = E_f(\beta, \theta; K, f_{\text{est}}) + \lambda_\beta E_{\Sigma}(\theta) + \lambda_\alpha E_{\alpha}(\theta) + \lambda_\beta E_{\beta}(\beta) \]
  - where \( E_f \) is the data term which penalizes the distance between estimated 2D joints of images \( f_{\text{est}} \) and the corresponding projected SMPL joints. \( E_{\beta}(\beta) \) is shape prior. \( E_{\Sigma} \) and \( E_{\alpha} \) are pose prior which are learned from precomputed stable poses. Here, \( E_{\Sigma} \) can favor probable stable poses over unstable ones.
    \[ E_{\Sigma}(\theta) = -\log \sum_j (g_j N(\theta; \mu_{\theta j}, \Sigma_{\theta j})) \]
  - where \( \mu_{\theta j} \) and \( \Sigma_{\theta j} \) are trained with our stable poses.

Boundary term:

\[ E_b(\beta, \theta; K, U) = \sum_{i=1}^{N} ||(B_i - U_i(\Pi_K(M(\beta, \theta))))||^2 \]

where \( B_i \) is the \( i_{th} \) point on the boundary of images, \( \Pi_K(\cdot) \) is the project function and \( U_i \) is the corresponding points of \( B_i \) on the boundary of projected model.

Results

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

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

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).