DeepGarment

3D Garment Shape Estimation from a Single Image

R.Danerek, E.Dibra, C.Oztireli, R.Ziegler, M.Gross
ETH Zurich, Vizrt Switzerland

Problem:

Recover the 3D shape of a garment from a single image

Related Work

Structure-from-motion-based techniques

modify and extend the standard SfM (Structure from motion) setup to estimate the shape of the garment

Rely on special markers depicted on the garment

Multi-view camera setup

Related Work

Shape-from-shading-based techniques

model garment from a single image, utilizing statistical human body models and having the user outline the silhouette of the garment and set the body pose

User-specified garment outlines

assuming symmetry in order to model the garment from the back as well

run-time in order of minutes

inability to handle self occlusions of the character

Related Work

Data-driven techniques

focus on estimating the naked human body shape from images and model clothing as an “offset” from the naked body shape

Related Work

Estimate 3D garment shape from a single image

assume depth is know

restricted mannequin poses and given cloth panels

considerable manual interaction

map wrinkle patterns and segmented garments to cloth parameters and materials through fitting

DeepGarment

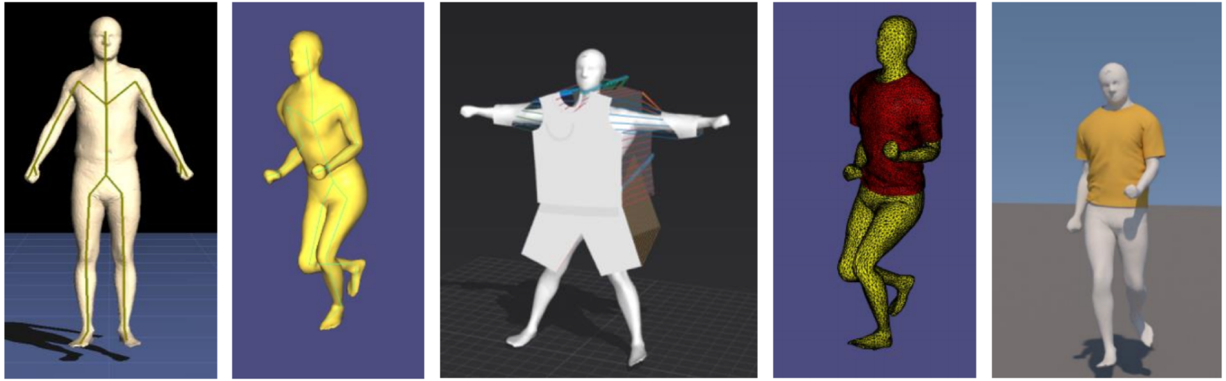
Data-driven technique leveraging deep learning technique

Goal: obtain the correct global 3D shape, possibly with plausible high-frequency deformations (such as wrinkles and folds)

1. Data Generation
2. 3D Garment Shape Reconstruction with CNN
3. Interpenetration Handling

Data Generation

Database: pairs of renderings and the corresponding 3D garment shapes



(a) (b) (c) (d) (e)

Data Generation

Human Model Creation : picked 10 meshes generated by a statistical human shape model

Human Model Animation: 3D skeletons extracted from motion capture dataset. Attach the skeletons to human shapes with an automatic method that computes skinning weights. Animate by applying Dual Quaternion Skinning

Garment Design: Marvelous Designer

Garment Simulation: ARCSim physically based simulation

Rendering

3D Garment Shape Reconstruction with CNN

Learn the deformation (simply represented as the vertex displacement) from a reference mesh (either a garment mesh or a body mesh, depending on the application) with respect to image observations using a CNN

Given an input image, the system masks the garment, and feeds it as an input to a specialized CNN, trained end-to-end to regress to 3D garment vertex deformations or offsets from a template mesh or human body

Mesh Deformation Representation

2 different representations for garment deformation

Both are based on the idea of vertex displacements, with different reference meshes

Garment-from-Garment Shape Representation: displacements from a reference garment mesh which is dressed on a character in a T-pose

Garment-from-Body Shape Representation: offset from a body mesh

Displacement of 3D coordinates

PCA analysis to reduce the dimension

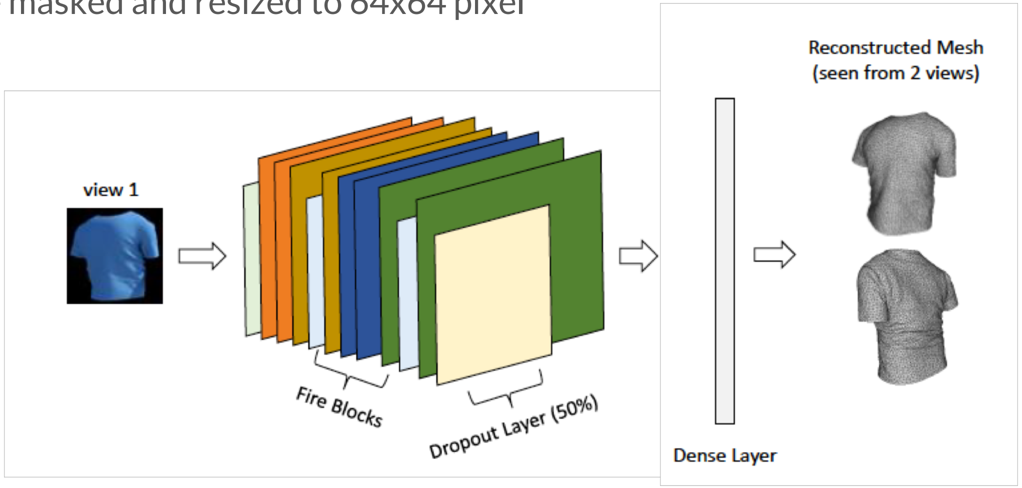
Segmentation

Segment out the garment from the background

An accurate segmentation is obtained by assuming a background model, utilizing Gaussian Mixture Models to learn a background model, and finally segmenting with graphcuts

Single-View Architecture

Input images are masked and resized to 64x64 pixel

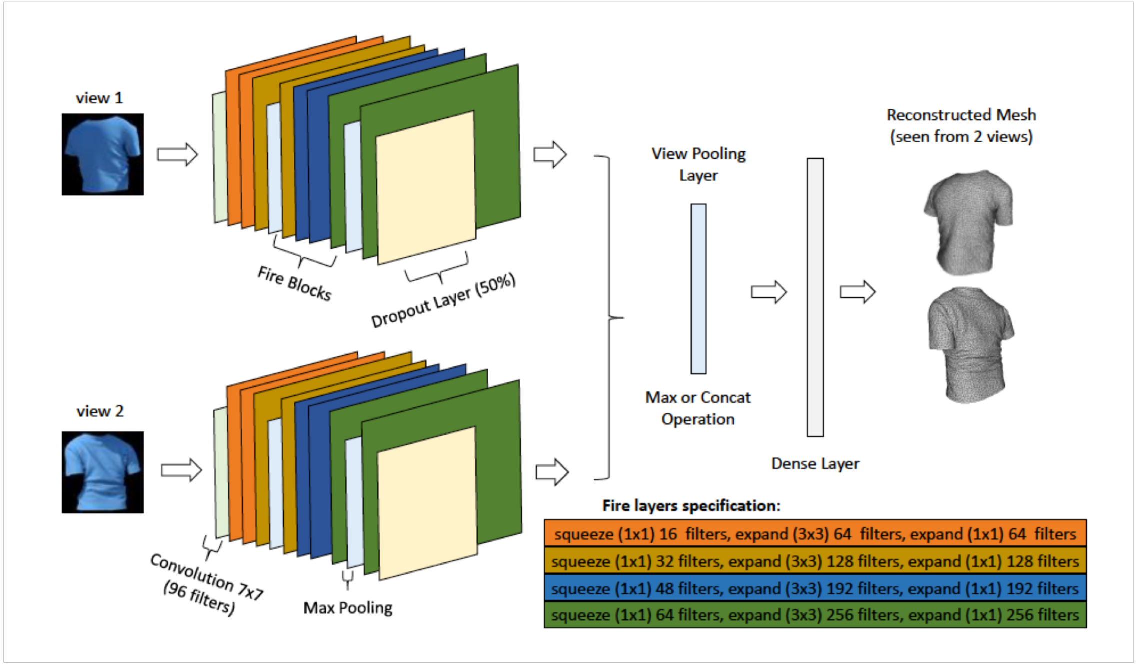


Two-View Architecture

Utilize the information from the multi-view input to produce more accurate results

Combine information coming from multiple views at a later stage by separately training two similar CNN-s on each view, and then concatenating the outputs of the last convolutional layer of each CNN through a view-pooling layer that performs either a max or a concatenation operation

Two-View Architecture



Loss Layer

Loss function: measures the error between the parameters of the estimated and the ground truth meshes

Compute the mean squared error over vertex positions

$$L_{full}(Y^P, Y^{GT}) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i^P - Y_i^{GT}\right\|^2$$

Loss Layer

Loss function: measures the error between the parameters of the estimated and the ground truth meshes

Regress to PCA component coefficient, use weighted mean squared error function

$$L_{PCA}(Y^P, Y^{GT}) = \frac{1}{l}\sum_{i=1}^{n} w_i \left| Y_i^P - Y_i^{GT} \right|$$

w_i is the PCA variance ratio corresponding to the i-th principal component, and l is the number of component

Loss Layer

In order to capture the curvature better and in turn the folds and wrinkles, integrating normal estimations through an additional term in loss

At each training iteration, compute the normal of the estimated vertices and compare them to the ground truth normal computed on the ground truth garment meshes

$$L^*(Y^P, Y^{GT}) = \frac{1}{n}\sum_{i=1}^{n} \left\| Y_i^P - Y_i^{GT} \right\|^2 - \lambda \left[k\left(N_i^P\right)^T N_i^{GT} \right]^3$$

Interpenetration Handling

Estimated mesh does not fit the body perfectly but some vertices may be placed inside it

Employ a least squares energy minimization to push the interpenetrating vertices out of the body mesh.

$$E_B(y) = pc(y) + \lambda_s s(y) + \lambda_d d(y)$$

Interpenetration Handling

$pc(y)$: interpenetration term. push the interpenetrating vertices out of the body mesh

$$pc(y) = \sum_{(i,j)\in C \wedge i \in P} \left\| \varepsilon + {\vec{n}_{b_j}}^T (\vec{v}_i - \vec{b}_j) \right\|^2$$

A garment vertex $\vec{v}_i$ is located inside the body if ${\vec{n}_{b_j}}^T (\vec{v}_i - \vec{b}_j) < 0$, where $\vec{n}_{b_j}$ is the normal of the body vertex $\vec{b}_j$

ε is set to small negative number to ensure the garment vertices are moved safely out of the body

Interpenetration Handling

$s(y)$: smoothness term. Make sure the vertices are being moved smoothly with respect to their neighbors

Prevent the final solution from having undesirable spikes in place of the interpenetrating vertices which are being moved out of the body

$$s(y) = \sum_{i \in V} \left\| \left(\vec{v}_i - \tilde{\vec{v}}_i\right) - \frac{1}{|B_i|} \sum_{j \in B_i} (\vec{v}_j - \tilde{\vec{v}}_j) \right\|^2$$

Interpenetration Handling

$d(y)$: damping term. The positions of the vertices have not changed very much from the input mesh

$$d(y) = \sum_{i \in V} \left\| (\vec{v}_i - \tilde{\vec{v}}_i) \right\|^2$$

Results

100000 T-shirt meshes on 7 male bodies of various shapes creating a geometry dataset

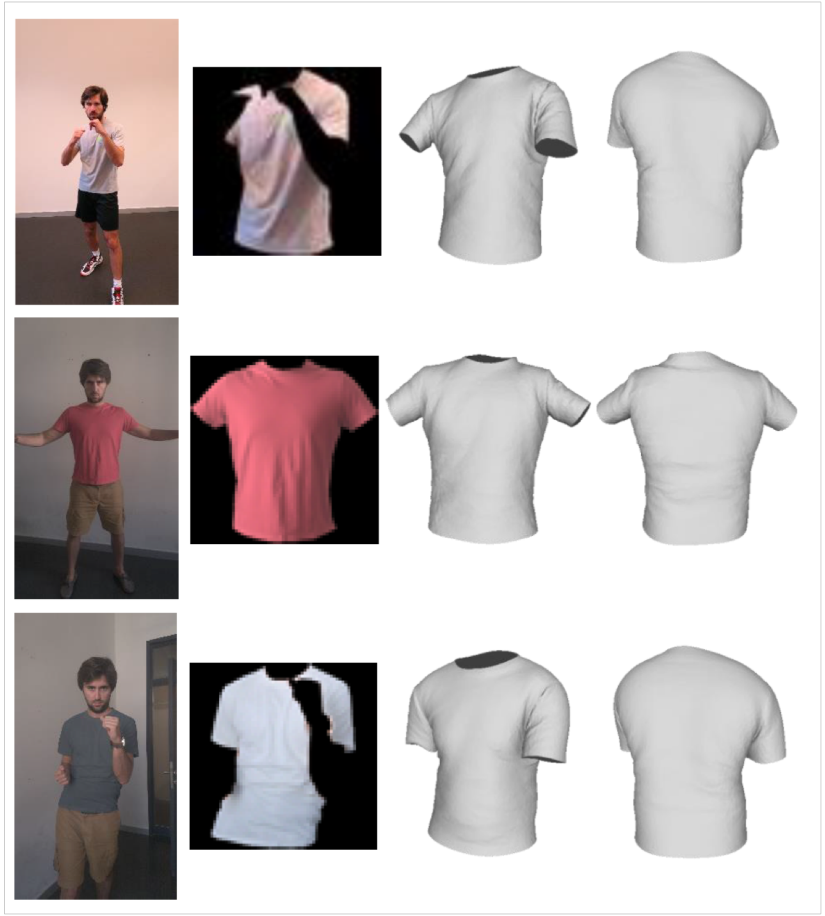
15000 dress meshes

Dataset consists of purely synthetic images, hence the training has never seen a real image, but is still able to capture plausible low-frequency deformations on real data

Results

Results

Results



Results

Conclusions

- “Garment-from-Body” formulation is better for dressing characters
- “Garment-from-Garment” is more suitable for reconstructing the garment only
- The multi-view architectures achieve superior performance in comparison to the single view models on both datasets

Limitations

- Introduce the artifacts during data generation
- Small input images prevent the method from capturing some of the wrinkles
- Need to train a new model for every new clothing type since the reference garment mesh may vary

Contributions

- An end-to-end 3D garment shape estimation algorithm. The algorithm automatically extracts 3D shape from a single image captured with an **uncontrolled setup** that depicts a dynamic state of a garment at interactive rates

- A regressor based on convolutional neural networks (CNN-s) combined with statistical priors and a specialized loss function for garment shape estimation

Thank You!