Inferring 3D shapes from a single viewpoint is an essential human vision feature extremely difficult for computer vision machines. Despite the advances in the field of 3D human reconstruction, most research has concentrated only on unclothed bodies and faces, but modelling and recovering garments have remained notoriously tricky.

In this paper we adapt the LGGAN [1] architecture to predict garment UV map from the input image. We also introduce 3D loss functions to improve the surface quality.

Introduction

- UV-based reconstruction of 3D garments from a single RGB image
- Inferring 3D shapes from a single viewpoint is an essential human vision feature extremely difficult for computer vision machines.
- Despite the advances in the field of 3D human reconstruction, most research has concentrated only on unclothed bodies and faces, but modelling and recovering garments have remained notoriously tricky.
- In this paper we adapt the LGGAN [1] architecture to predict garment UV map from the input image. We also introduce 3D loss functions to improve the surface quality.

UV Maps

- Garments are registered on top of SMPL [2] mesh to have homogeneous topology at both training and inference time.
- The UV coordinates are discrete points, thus UV maps have empty gaps between vertices. We use image inpainting techniques to estimate the values of the empty spaces.
- We use displacement UV maps that store garment vertices as an offset over the estimated SMPL body vertices.
- We create UV maps for the garment mesh, garment semantic segments and body mesh.

Biography

- Garments are registered on top of SMPL [2] mesh to have homogeneous topology at both training and inference time.
- The UV coordinates are discrete points, thus UV maps have empty gaps between vertices. We use image inpainting techniques to estimate the values of the empty spaces.
- We use displacement UV maps that store garment vertices as an offset over the estimated SMPL body vertices.
- We create UV maps for the garment mesh, garment semantic segments and body mesh.

Class-Specific Local Generation Network

- LGGAN is conditioned on body UV map and garment UV semantic segments,
- Global decoder G_g is responsible to predict low frequency details (overall shape),
- Local decoder G_l is responsible to learn specific dynamics w.r.t. each garment class,
- The amount of aggregation of global and local information is controlled by weight decoder G_w through softmax,
- Class-specific discriminative feature learning branch classifies masked features to the target garment class,
- Discriminator D receives the conditioning image (body UV map) and predicted garment UV map to distinguish real data from fake.

Generator Loss

- L1 loss on the reconstructed UV map,
- L2 loss on the surface normals based on nearest neighbor matching,
- Laplacian smoothing regularization,
- Edge length regularization.

RGB to UV Map Translation - LGGAN Architecture

- LGGAN is conditioned on body UV map and garment UV semantic segments,
- Global decoder G_g is responsible to predict low frequency details (overall shape),
- Local decoder G_l is responsible to learn specific dynamics w.r.t. each garment class,
- The amount of aggregation of global and local information is controlled by weight decoder G_w through softmax,
- Class-specific discriminative feature learning branch classifies masked features to the target garment class,
- Discriminator D receives the conditioning image (body UV map) and predicted garment UV map to distinguish real data from fake.

Comparison with SMPLicit [4]

- We use surface-to-surface (S2S) error [5], an extension of the chamfer distance (CD), that computes the distance based on the nearest faces rather than nearest vertices.
- As S2S evaluates the results based on mesh format, during evaluation we unwrap UV map representations into 3D meshes.

CLOTH3D++ Dataset [3]

- We use CLOTH3D++, the first large-scale synthetic dataset of 3D clothed human sequences.
- Garments are simulated on top of thousands of different human pose sequences and body shapes, generating realistic cloth dynamics.
- It has over 2 million 3D samples with a large variety of garment types, topology, shape, size, tightness and fabric.

Evaluation Metric

- We use surface-to-surface (S2S) error [5], an extension of the chamfer distance (CD), that computes the distance based on the nearest faces rather than nearest vertices.
- As S2S evaluates the results based on mesh format, during evaluation we unwrap UV map representations into 3D meshes.