SMPL: A Skinned Multi-Person Linear Model

The authors present a learned model of human body shape and pose dependent shape variation that is more accurate than previous models and is compatible with existing graphics pipelines. The Skinned Multi-Person Linear model (SMPL) is a skinned vertex based model that accurately represents a wide variety of body shapes in natural human poses. The parameters of the model are learned from data including the rest pose template, blend weights, pose-dependent blend shapes, identity-dependent blend shapes, and a regressor from vertices to joint locations.

The goal is to automatically learn a model of the body that is both realistic and compatible with existing graphics software. To that end, they describe a "Skinned Multi-Person Linear" (SMPL) model of the human body that can realistically represent a wide range of human body shapes, can be posed with natural pose-dependent deformations, exhibits soft-tissue dynamics, is efficient to animate, and is compatible with existing rendering engines.

The formulation admits an objective function that penalizes the pervertex disparities between registered meshes and their model, enabling training from data. To learn how people, deform with pose, they use 1786 high-resolution 3D scans of different subjects in a wide variety of poses. They align the template mesh to each scan to create a training set. They optimize the blend weights, pose-dependent blend shapes, the mean template shape (rest pose), and a regressor from shape to joint locations to minimize the vertex error of the model on the training set. This joint regressor predicts the location of the joints as a function of the body shape and is critical to animating realistic pose-dependent deformations for any body shape.

All parameters are automatically estimated from the aligned scans.

The SMPL model decomposes body shape into identity-dependent shape and non-rigid pose-dependent shape, and take a vertex-based skinning approach that uses corrective blend shapes. A single blend shape is represented as a vector of concatenated vertex offsets. They begin with an artist created mesh with 6890 vertices and 23 joints. The mesh has the same topology for men and women, spatially varying resolution, a clean quad structure, a segmentation into parts, initial blend weights, and a skeletal rig.

Because their model decomposes shape and pose, they train these separately, simplifying optimization, and there are separate models for men and women.

The model is working because the good quality training data is important. Here they use thousands of high-quality registered template meshes. Importantly, the pose training data spans a range of body shapes enabling us to learn a good predictor of joint locations. Additionally, training all the parameters (template shape, blend weights, joint regressor, shape/pose/dynamic blend shapes) to minimize vertex reconstruction error is important to obtain a good model. Here the simplicity of the model is an advantage as it enables training everything with large amounts of data.

SMPL uses standard skinning equations and defines body shape and pose blend shapes that modify the base mesh. They train the model on thousands of aligned scans of different people in different poses. The form of the model makes it possible to learn the parameters from large amounts of data while directly minimizing vertex reconstruction error. Specifically, they learn the rest template, joint regressor, body shape model, pose blend shapes, and dynamic blend shapes.