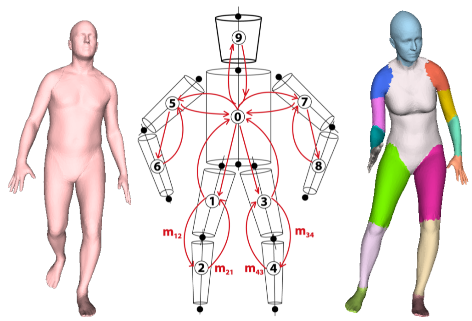
The Stitched Puppet: A Graphical Model of 3D Human Shape and Pose

Silvia Zuffi[1,2], Michael J. Black[1]
[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany. [2]ITC - Consiglio Nazionale delle Ricerche, Milan, Italy.



(a) (b) (c)

Figure 1: **3D Body Models.** (a) A SCAPE body model [2] realistically represents 3D body shape and pose using a single high-dimensional state space. (b) A graphical body model composed of geometric primitives connected by pairwise potentials [5]. (c) The **stitched puppet (SP)** model has the realism of (a) and the graphical structure of (b). Each body part is described by its own low-dimensional state space and the parts are connected via pairwise potentials that "stitch" the parts together.

Inference of human body shape and pose from images, depth data, 3D scans, and sparse markers is of great interest. There are two main classes of 3D body models in use: highly realistic models like SCAPE (Fig. 1(a)) [2], which use a high dimensional state space that combines shape and pose parameters, making inference computationally challenging; or part-based models like loose-limbed people (Fig. 1(b)) [5], which are advantageous for inference but do not make it possible to recover body shape and do not match well to image evidence. In this paper we propose a new human body model, named stitched puppet (SP), that offers the best features of both approaches in that it is both part-based and highly realistic (Fig. 1(c)).

The SP model is learned from a detailed 3D body model based on SCAPE [2]. Each body part is represented by a mean shape and two subspaces of shape deformations learned using principal component analysis (PCA), independently accounting for variations in intrinsic body shape and pose-dependent shape deformations. These shape variations allow SP to capture and fit a wide range of human body shapes in different poses (Fig. 2).

Figure 2: **Example SP bodies.** Several female and male bodies generated using SP. Note the realism of the 3D shapes.

As with other part-based models, the parts form a graph with pairwise potentials between nodes in the graph. The SP potentials represent a "stitching cost" that penalizes parts that do not fit properly together in 3D to form a coherent shape. Unlike the SCAPE model, parts can move away from each other but with some cost. This ability of parts to separate and then be stitched back together is an important property that is exploited during inference to better explore the space of solutions (Fig. 3).

We apply SP to two challenging problems involving estimating human

Figure 3: **Alignment of SP to FAUST with D-PMP [4].** (left) D-PMP particles corresponding to independent body parts. (middle) Best set of particles (light blue) and 3D scan data (red) after several iterations of the algorithm. (right) Solution.

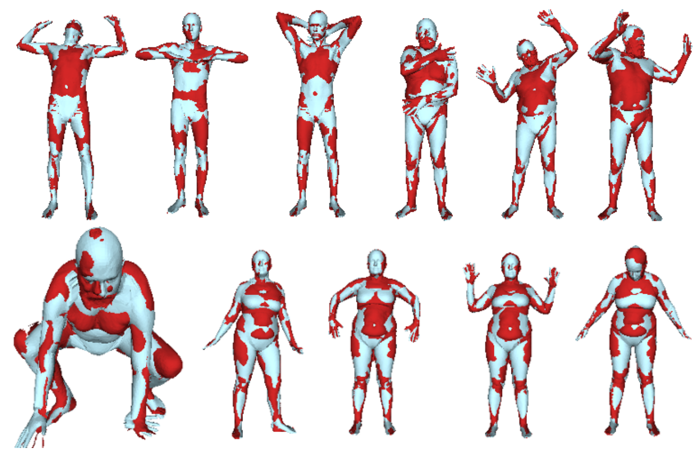


Figure 4: **Alignment on FAUST**. We show the test scan in red and SP in light blue.

shape and pose from 3D data. The first is the FAUST mesh alignment challenge [3], where ours is the first method to successfully align all 3D meshes. The second is the fit of SP to noisy and low-resolution visual hull data. To align SP to 3D data we minimize an energy composed by a stitching term and a data term. Inference for 3D pose and body shape is performed using a recently proposed iterative particle-based algorithm for *maximum-a-posteriori* inference in graphical models with pairwise potentials, the D-PMP algorithm [4] (Fig. 3). Figure 4 shows examples of fully automatic alignment on FAUST. Note how we can accurately estimate pose and body shape.

The model and code are available for research purposes [1].

[1] `http://stitch.is.tue.mpg.de`.
[2] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *ACM Trans. Graph.*, 24(3):408–416, 2005.
[3] F. Bogo, J. Romero, M. Loper, and M. J. Black. FAUST: Dataset and evaluation for 3D mesh registration. *CVPR*, pp. 3794–3801, 2014.
[4] J. Pacheco, S. Zuffi, M. J. Black, and E. Sudderth. Preserving modes and messages via diverse particle selection. *ICML*, 32(1):1152–1160, 2014.
[5] L. Sigal, M. Isard, H. Haussecker, and M. J. Black. Loose-limbed people: Estimating 3D human pose and motion using non-parametric belief propagation. *IJCV*, 98:15–48, 2011.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.