An Efficient Volumetric Framework for Shape Tracking

Benjamin Allain, Jean-Sébastien Franco, Edmond Boyer
Inria Grenoble Rhône-Alpes - LJK, Grenoble Universities, France.

Motivation Recovering 3D shape motion using visual information is an important problem with many applications in computer vision and computer graphics, among other domains. Most existing approaches rely on surface-based strategies, where surface models are fit to visual surface observations. While numerically plausible, this paradigm ignores the fact that the observed surfaces often delimit volumetric shapes, for which deformations are constrained by the volume inside the shape. Consequently, surface-based strategies can fail when the observations define several feasible surfaces, whereas volumetric considerations are more restrictive with respect to the admissible solutions. In this work, we investigate a novel volumetric shape parametrization to track shapes over temporal sequences.

Contributions

CVT In constrast to Eulerian grid discretizations of the observation space, such as voxels, we consider general shape tesselations yielding more convenient cell decompositions, in particular the Centroidal Voronoi Tesselations (CVT). In a CVT, since each Voronoi site coincides with its cell centroid, cell shapes and connectivity are quasi-regular.

With this shape representation, we devise a tracking method that exploits volumetric information, both for the data term evaluating observation conformity, and for expressing deformation constraints that enforce prior assumptions on motion.

Deformation Model We extend to the volumetric case a state-of-the-art surface tracking approach [1, 2] where a template shape is decomposed into samples (here cell centroids) which are then clustered into uniformly sized patches (see Fig. 1). Each patch may move rigidly, and neighbor patches are softly constrained to remain in the same relative pose (with respect to each other) as in a reference pose.

Volumetric Observation Model While surface fitting methods often rely on surface normal orientation to match observed points to the model's points, it is not applicable in the case of volumetric fitting, where no normal can be defined. Our approach introduces a volumetric fitting term, based on the assumption that the distance between an interior point and the shape surface is preserved across time. Our volumetric fitting term is designed to deal with shapes reconstructed from multiple views, in particular visual hulls, where the shape surface is usually an overestimate of the true surface.

Model and Inference The tracking problem is formulated as a MAP over a probabilistic generative model. Inference is performed with the Expectation Maximization algorithm, where latent variables represent the associations between the template cells and the observed shape cells.

Experiments We validate our approach on a dataset containing meshes reconstructed from images, and marker-based motion capture ground truth[3]. The latter is used as ground truth to evaluate the tracking consistency. We also measure silhouette reprojection error on several datasets.

Experiments demonstrate similar or improved precisions over state-of-the-art methods, as well as improved robustness, a critical issue when tracking sequentially over time frames. Experiments notably show improved volume stability of tracking, and more realistic behaviour when tracking bending objects, as the volumetric tracking solution proposed prevents the typical overfolding seen in existing surface-based tracking approaches.

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

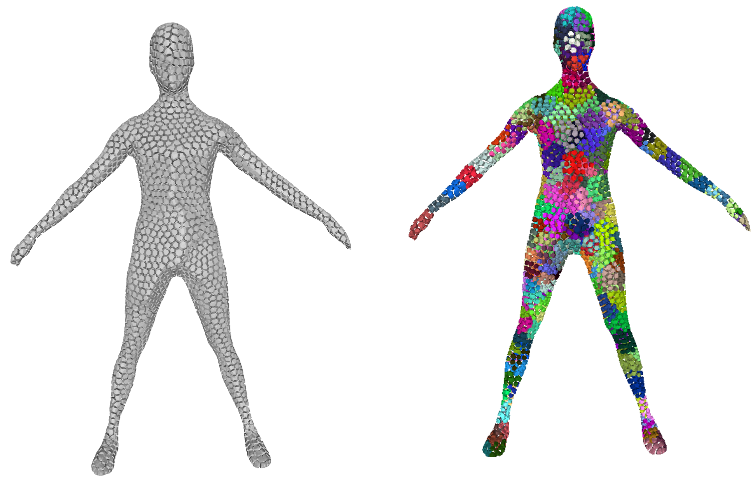

Figure 1: Left: CVT of the template shape. Right: clustering of cells into patches.

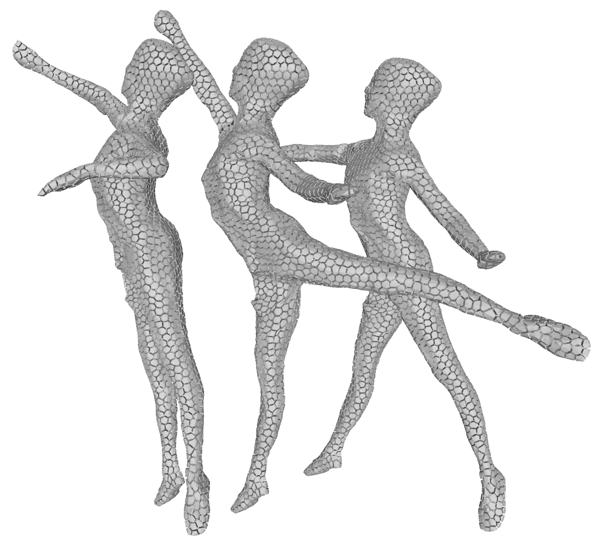

Figure 2: Tracking results for three poses of a dancing sequence.

[1] B. Allain, J.-S. Franco, E. Boyer, and T. Tung. On mean pose and variability of 3d deformable models. *ECCV*, 2014.

[2] C. Cagniart, E. Boyer, and S. Ilic. Probabilistic deformable surface tracking from multiple videos. *ECCV*, 2010.

[3] Y. Liu, J. Gall, C. Stoll, Q. Dai, H.-P. Seidel, and C. Theobalt. Markerless motion capture of multiple characters using multi-view image segmentation. *PAMI*, 2013.