Simultaneous Pose and Non-Rigid Shape with Particle Dynamics

Antonio Agudo[1] and Francesc Moreno-Noguer[2]
[1]Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Spain.
[2]Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, Spain.

Recovering the 3D shape of a deforming object and the camera motion from monocular video is an inherently ill-posed problem and it normally requires prior knowledge on the scene structure and camera motion. Most Non-Rigid Structure from Motion approaches solve this problem using statistical priors to model the global deformable structure as a linear combination of low-rank bases of either shapes [5, 8, 16] or 3D point trajectories [3, 12]. This is typically used with additional smoothness constraints that further disambiguate the problem [6, 10]. Yet, while low-rank methods can effectively encode global deformations, they cannot, in general, handle non-linear motion patterns and strong local deformations. Piecewise strategies [14, 15] allow recovering larger deformations, although their performance highly depends on having overlapping features in neighboring patches, which can be hard to obtain in practice. In any event, these previous approaches batch process all frames of the sequence at once, after video capture, preventing them from being used on-line and in real-time applications.

While sequential solutions exist for the rigid case [9], sequential estimation of deformable objects based only on the measurements up to that moment remains a challenging and unsolved problem. Only recently, this problem has been also handled in a sequential manner, which represents an even more complex scenario because of the intrinsic strong ill-posedness of the problem [1, 2, 11]. While this is a promising direction, these methods still focus on global models only valid for relatively small deformations [11] or continuous surfaces [1, 2].

An alternative to statistical and low-rank approaches is to directly model the physical laws that locally govern object kinematics. Drawing inspiration from computer graphics [13], there have been several attempts at using these models for tracking non-rigid motion [7] and human activities [4]. Unfortunately, these methods are usually focused to specific types of motion, and their underlying laws rely on non-linear relations complex to optimize.

In this paper, we propose a *sequential solution* to simultaneously estimate camera pose and non-rigid 3D shape from a monocular video. In contrast to most existing approaches that rely on global representations of the shape, we model the object at a *local level*, as an ensemble of particles, each ruled by the linear equation of the Newton's second law of motion to constrain their motion. So, the deformable object is modeled as a system of individual particles and represent global deformation by locally modeling the underlying physical laws that govern each of the particles, according to a constant velocity model with acting forces. This dynamic model is incorporated into a bundle adjustment framework, in combination with simple regularization components that ensure temporal and spatial consistency of the estimated shape and camera poses. The resulting approach is both efficient and robust to several artifacts such as noisy and missing data or sudden camera motions, while it does not require any training data at all. Our method can handle different types of motions such as articulated, isometric, stretchable and abrupt deformations.

We show successful non-rigid 3D reconstructions results in a high variety of real video sequences (such as those depicted in Fig. 1), including articulated and non-rigid motion, both for continuous and discontinuous shapes. Our system is shown to perform comparable to competing batch, computationally expensive, methods and shows remarkable improvement with respect to the sequential ones.

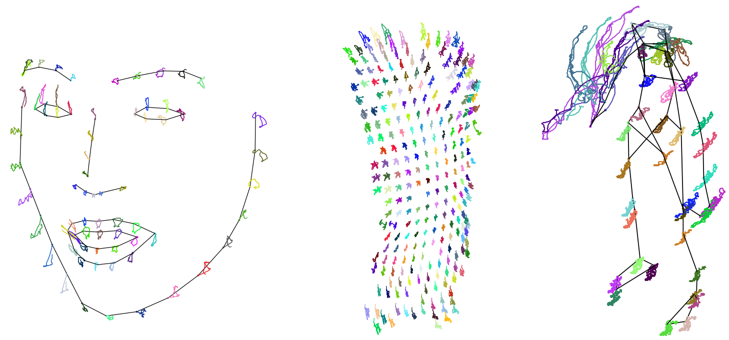

Figure 1: 3D Reconstruction using our physically-inspired velocity model for different types of deformations: face, human torso and articulated motion. Each line represents the per point non-rigid motion detected by our algorithm. Best viewed in color.

[1] A. Agudo, B. Calvo, and J. M. M. Montiel. Finite element based sequential bayesian non-rigid structure from motion. In *CVPR*, 2012.

[2] A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In *CVPR*, 2014.

[3] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Non-rigid structure from motion in trajectory space. In *NIPS*, 2008.

[4] M. Brubaker, L. Sigal, and D. Fleet. Estimating contact dynamics. In *ICCV*, 2009.

[5] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure from motion factorization. In *CVPR*, 2012.

[6] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *CVPR*, 2013.

[7] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *TPAMI*, 15(6):580–591, 1993.

[8] F. Moreno-Noguer and J. M. Porta. Probabilistic simultaneous pose and non-rigid shape recovery. In *CVPR*, 2011.

[9] R. Newcome and A. J. Davison. Live dense reconstruction with a single moving camera. In *CVPR*, 2010.

[10] M. Paladini, A. Del Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito. Factorization for non-rigid and articulated structure using metric projections. In *CVPR*, 2009.

[11] M. Paladini, A. Bartoli, and L. Agapito. Sequential non rigid structure from motion with the 3D implicit low rank shape model. In *ECCV*, 2010.

[12] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh. 3D reconstruction of a moving point from a series of 2D projections. In *ECCV*, 2010.

[13] Z. Popovic and A. Witkin. Physically based motion transformations. In *ACM SIGGRAPH*, 1999.

[14] C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *CVPR*, 2011.

[15] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, 2010.

[16] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors. *TPAMI*, 30(5):878–892, 2008.

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