Deformable Surface Dynamics Modeling

Tony Tung[1,a)]

Abstract: Deformable surfaces are surfaces that undergo non-rigid deformations. They can represent soft tissues (e.g., human body, face, organ, cloth), and can be modeled by sequences of 3D surface meshes. Thanks to progresses in sensing technologies, 3D reconstruction of deformable surfaces from real-world visual and spatial information can nowadays be achieved with high accuracy (i.e., below 0.5 cm) and in reasonable time. For example, 3D video (i.e., sequence of full 3D models) representing live human performances or daily activities (e.g., dance or yoga) can be obtained using various techniques (e.g., multiview stereo [1], [2], [3] or depth data fusion [4]). However, the complexity and constant change in geometry and topology of these objects pose challenges for applying traditional vision algorithms to the sequences. Over the past few years, new algorithms have been proposed for vision tasks such as registration [5], [6], [7], [8], segmentation [9], and categorization [10] of such data. Among these tasks, the *microscopic* categorization of surfaces is of specific interest to us, because it is critical to surveillance applications, such as detecting organ anomalous behavior, skin deformation, leaks in tanks at power plant, or assessing fabric quality.
Deformable surfaces representing real-world objects (e.g., such as humans) can be assumed as a stream of temporally continuous and indefinitely varying 3D geometrical data that possess certain temporal statistics. For example, clothing made of soft fabrics worn by a human in motion usually exhibit different surface variations compared to bare skin. However, as opposed to dynamic textures [11], [12], [13], visual appearance-based methods cannot be used directly to capture the complex behavior of deformable surface, as surface texture does not carry spatial information and can have poor quality (e.g., due to multiview stereo reconstruction artifacts). On the other hand, although actual sensing devices and capture systems usually provide data contaminated with noise, the accuracy is already sufficient for research purpose and many applications. Moreover it is reasonable to assume that sensing systems will continue to improve very quickly. Hence, we propose to characterize deformable surfaces using a geometry-dynamics-based approach that relies on intrinsic surface properties [14], [15]. We process as follows: 1) intrinsic surface features represented by continuous scalar values (e.g., Koenderink shape indices [16]) are tracked over time and produce time series at each surface point, 2) these observed local geometrical deformations are modeled using switching linear dynamical systems (LDS) that capture spatiotemporal variation dynamics, and 3) classification is performed using a discriminative model based on temporal distribution of LDS sets. Particularly, we introduce a timing-based low-level descriptor to capture local spatiotemporal variations, and the bag-of-timings (BoT) paradigm to further handle challenging deformable surface classification and segmentation with respect to different level of rigidities.
Our experiments on synthesized and real-world datasets with ground truth showcase classification and segmentation robustness. In particular, we use public datasets of 3D video representing human performances which are challenging, of good quality, and popular in CV and CG communities (i.e., data from MIT [17], Univ. Surrey [18], and INRIA [6]).

References

[1] Matsuyama, T., Wu, X., Takai, T. and Nobuhara, S.: Real-Time 3D Shape Reconstruction, Dynamic 3D Mesh Deformation, and High Fidelity Visualization for 3D Video, *CVIU*, Vol. 96, No. 3, pp. 393–434 (2004).
[2] Franco, J., Menier, C., Boyer, E. and Raffin, B.: A Distributed Approach for Real-Time 3D Modeling, *CVPR Workshop* (2004).
[3] Matsuyama, T., Nobuhara, S., Takai, T. and Tung, T.: 3D Video and its Applications, *Springer* (2012).
[4] Jiang, H., Liu, H., Tan, P., Zhang, G. and Bao, H.: 3D reconstruction of dynamic scenes with multiple handheld cameras, *ECCV* (2012).
[5] de Aguiar, E., Theobalt, C., Stoll, C. and Seidel, H.-P.: Marker-less deformable mesh tracking for human shape and motion capture, *CVPR* (2007).
[6] Cagniart, C., Boyer, E. and Ilic, S.: Probabilistic Deformable Surface Tracking From Multiple Videos, *ECCV* (2010).
[7] Tung, T. and T.Matsuyama: Dynamic surface matching by geodesic mapping for 3D animation transfer, *CVPR* (2010).
[8] Allain, B., Franco, J.-S., Boyer, E. and Tung, T.: On Mean Pose and Variability of 3D Deformable Models, *ECCV* (2014).
[9] Franco, J. and Boyer, E.: Learning Temporally Consistent Rigidities, *CVPR* (2011).
[10] Tung, T. and T.Matsuyama: Topology Dictionary for 3D Video Understanding, *PAMI* (2012).
[11] Doretto, G., Chiuso, A., Wu, Y. and Soatto, S.: Dynamic textures, *IJCV*, Vol. 51, No. 2, pp. 91–109 (2003).
[12] Chan, A. B. and Vasconcelos, N.: Mixtures of dynamic textures, *ICCV* (2005).
[13] Ravichandran, A., Chaudhry, R. and Vidal, R.: View-Invariant Dynamic Texture Recognition using a Bag of Dynamical Systems, *CVPR* (2009).
[14] Tung, T. and T.Matsuyama: Intrinsic Characterization of Dynamic Surfaces, *CVPR* (2013).
[15] Tung, T. and T.Matsuyama: Timing-Based Local Descriptor for Dynamic Surfaces, *CVPR* (2014).
[16] Koenderink, J. and van Doorn, A.: Surface shape and curvature scales, *Image and Vision Computing* (1992).
[17] de Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.-P. and Thrun, S.: Performance capture from sparse multi-view video, *ACM Trans. Graphics*, Vol. 27, No. 3 (2008).
[18] Starck, J. and Hilton, A.: Surface Capture for Performance-Based Animation, *IEEE CGA* (2007).

[1] Kyoto University, Graduate School of Informatics, Kyoto 606–8501, Japan
[a)] http://www.tonytung.org

