Suivi topologique rapide et robuste par appariement de Wasserstein augmenté

Lifted Wasserstein Matcher for Fast and Robust Topology Tracking

Maxime Soler, Mélanie Plainchault, Bruno Conche, Julien Tierny

This talk presents a robust and efficient method for tracking topological features in time-varying scalar data. Structures are tracked based on the optimal matching between persistence diagrams with respect to the Wasserstein metric. This fundamentally relies on solving the assignment problem, a special case of optimal transport, for all consecutive timesteps. Our approach relies on two main contributions. First, we revisit the seminal assignment algorithm by Kuhn and Munkres which we specifically adapt to the problem of matching persistence diagrams in an efficient way. Second, we propose an extension of the Wasserstein metric that significantly improves the geometrical stability of the matching of domain-embedded persistence pairs. We show that this geometrical lifting has the additional positive side-effect of improving the assignment matrix sparsity and therefore computing time. The global framework computes persistence diagrams and finds optimal matchings in parallel for every consecutive timestep. Critical trajectories are constructed by associating successively matched persistence pairs over time. Merging and splitting events are detected with a geometrical threshold in a post-processing stage. Extensive experiments on real-life datasets show that our matching approach is up to two orders of magnitude faster than the seminal Munkres algorithm. Moreover, compared to a modern approximation method, our approach provides competitive runtimes while guaranteeing exact results. We demonstrate the utility of our global framework by extracting critical point trajectories from various time-varying datasets and compare it to the existing methods based on associated overlaps of volumes. Robustness to noise and temporal resolution downsampling is empirically demonstrated.

Index Terms—opological Data Analysis, Optimal Transport, Feature Trackingopological Data Analysis, Optimal Transport, Feature TrackingT

1 Introduction

Performing feature extraction and object tracking is an important topic in scientific visualization, for it is key to understanding time-varying data. Specifically, it allows to detect and track the evolution of regions of interest over time, which is central to many scientific domains, such as combustion [2], aerodynamics [6], oceanography [12] or meteorology [18]. With the increasing power of computational resources and resolution of acquiring devices, efficient methods are needed to enable the analysis of large datasets.

The emergence of new paradigms for scientific simulation, such as *in-situ* and *in-transit* [1], [9], [11], [13], [17], clearly exhibits the ambition to reach toward exascale computing [4] in the forthcoming years. In this context, as both spatial and temporal resolutions of acquired or simulated datasets keep on increasing, understanding the evolution of features of interest throughout time proves challenging.

Topological data analysis has been used in the last decades as a robust and reliable setting for hierarchically defining features in scalar data [5]. In particular, its successful application to time-varying data [3], [14] makes it a prime candidate for tracking. Both topological analysis and feature tracking have been applied *in-situ* [8], [19], which demonstrates their interest in the context of large-scale data. Nonetheless, major bottlenecks of state-of-the art topology tracking methods are still the high required computation cost as well as the need for high temporal resolution.

In this talk, we propose a novel feature-tracking framework, which correlates topological features in

- *Maxime Soler: Total SA and Sorbonne Université E-mail:soler.maxime total.com*
- *Mélanie Plainchault: Total SA E-mail: melanie.plainchault@total.com*
- *Bruno Conche: Total SA E-mail: bruno.conche@total.com*
- *Julien Tierny:CNRS and Sorbonne Université E-mail: julien.tierny@sorbonne-universite.fr*

time-varying data in an efficient and meaningful way. It is the first approach, to the best of our knowledge, combining the setting of topological data analysis with optimal transport for the problem of feature tracking. More precisely, the key idea is to use combinatorial optimization for matching topological structures (namely, *persistence diagrams*) according to a fine-tuned metric. After exposing our formal setting, we introduce an extension of the exact assignment algorithm by Kuhn and Munkres [7], [10] that we adapt in an efficient way to the case of persistence diagrams. We highlight the issues raised by the classical Wasserstein metric between diagrams, and propose a robust *lifted* metric that overcomes these limitations. We then present the detailed tracking framework. Extensive experiments demonstrate the utility of our approach.

2 Contributions

This talk will present the following new contributions (further described in [15]:

1) **Approach:** We present a sound and original framework, which is the first combining topology and transportation for feature tracking, comparing favorably to other state-of-the-art approaches, both in terms of speed and robustness.
2) **Metric:** We extend traditional topological metrics, for the needs of time-varying feature tracking, notably enhancing geometrical stability and computing time.
3) **Algorithm:** We extend the assignment method by Kuhn and Munkres to solve the problem of persistence matchings in a fast and exact way, taking advantage of our metric.
4) **Implementation:** We provide a lightweight VTK-based C++ implementation of our approach for reproduction purposes.

References

[1] J. C. Bennett, H. Abbasi, P. T. Bremer, R. Grout, A. Gyulassy, T. Jin, S. Klasky, H. Kolla, M. Parashar, V. Pascucci, P. Pebay, D. Thompson, H. Yu, F. Zhang, and J. Chen. Combining in-situ and in-transit processing to enable extreme-scale scientific analysis. In *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference for*, pages 1–9, Nov 2012.

[2] P. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. Interactive exploration and analysis of large scale simulations using topology-based data segmentation. *IEEE TVCG*, 17(9):1307–1324, 2011.

[3] P. T. Bremer, G. Weber, V. Pascucci, M. Day, and J. Bell. Analyzing and tracking burning structures in lean premixed hydrogen flames. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):248–260, March 2010.

[4] D. A. S. C. A. Committee. Synergistic challenges in data-intensive science and exascale computing. Technical report, DoE Advanced Scientific Computing Advisory Committee, Data Sub-committee, 2013.

[5] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2009.

[6] S. M. Hannon and J. A. Thomson. Aircraft wake vortex detection and measurement with pulsed solid-state coherent laser radar. *Journal of Modern Optics*, 41(11):2175–2196, 1994.

[7] H. W. Kuhn and B. Yaw. The hungarian method for the assignment problem. *Naval Res. Logist. Quart*, pages 83–97, 1955.

[8] A. Landge, V. Pascucci, A. Gyulassy, J. Bennett, H. Kolla, J. Chen, and T. Bremer. In-situ feature extraction of large scale combustion simulations using segmented merge trees. In *SuperComputing*, 2014.

[9] K. Moreland, R. Oldfield, P. Marion, S. Jourdain, N. Podhorszki, V. Vishwanath, N. Fabian, C. Docan, M. Parashar, M. Hereld, M. E. Papka, and S. Klasky. Examples of in transit visualization. In *Proceedings of the 2Nd International Workshop on Petascal Data Analytics: Challenges and Opportunities*, PDAC '11, pages 1–6, New York, NY, USA, 2011. ACM.

[10] J. Munkres. Algorithms for the assignment and transportation problems, 1957.

[11] M. Rasquin, P. Marion, V. Vishwanath, B. Matthews, M. Hereld, K. Jansen, R. Loy, A. Bauer, M. Zhou, O. Sahni, J. Fu, N. Liu, C. Carothers, M. Shephard, M. Papka, K. Kumaran, and B. Geveci. Electronic poster: Co-visualization of full data and in situ data extracts from unstructured grid cfd at 160k cores. In *Proceedings of the 2011 Companion on High Performance Computing Networking, Storage and Analysis Companion*, SC '11 Companion, pages 103–104, New York, NY, USA, 2011. ACM.

[12] T. Ringler, M. Petersen, R. L. Higdon, D. Jacobsen, P. W. Jones, and M. Maltrud. A multi-resolution approach to global ocean modeling. *Ocean Modelling*, 69:211 – 232, 2013.

[13] M. Rivi, L. Calori, G. Muscianisi, and V. Slavnic. In-situ visualization: State-of-the-art and some use cases. 2011.

[14] B. S. Sohn and C. Bajaj. Time-varying contour topology. *IEEE TVCG*, 12(1):14–25, 2006.

[15] M. Soler, M. Plainchault, B. Conche, and J. Tierny. Lifted wasserstein matcher for fast and robust topology tracking. In *IEEE Symposium on Large Data Analysis and Visualization*, 2018.

[16] J. Tierny and V. Pascucci. Generalized topological simplification of scalar fields on surfaces. *IEEE TVCG*, 18(12):2005–2013, 2012.

[17] H. Yu, C. Wang, R. W. Grout, J. H. Chen, and K. L. Ma. In situ visualization for large-scale combustion simulations. *IEEE Computer Graphics and Applications*, 30(3):45–57, May 2010.

[18] F. Zhang, E. Fiorelli, and N. E. Leonard. Exploring scalar fields using multiple sensor platforms: Tracking level curves. In *2007 46th IEEE Conference on Decision and Control*, pages 3579–3584, Dec 2007.

[19] F. Zhang, S. Lasluisa, T. Jin, I. Rodero, H. Bui, and M. Parashar. In-situ feature-based objects tracking for large-scale scientific simulations. In *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, pages 736–740, Nov 2012.

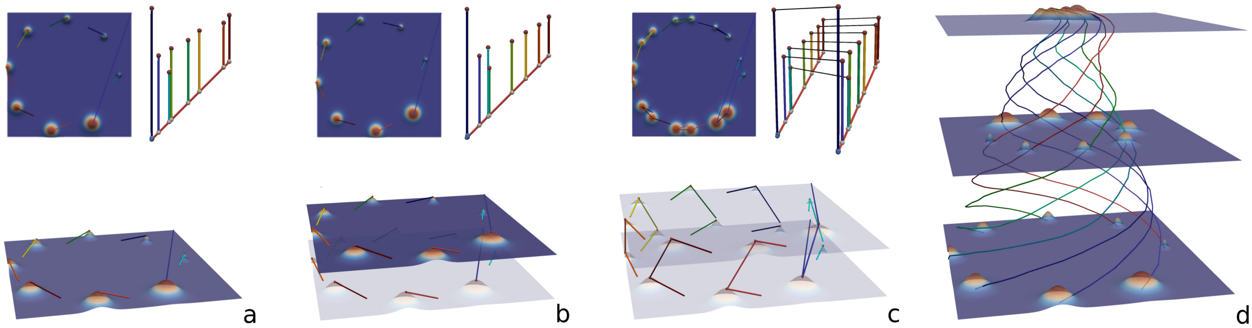

Fig. 1. Overview of our tracking approach on a dataset consisting of eight whirling gaussians: persistence diagram computations for two consecutive timesteps (a) and (b); matching of persistence pairs of two timesteps (c), propagation of matchings and construction of a trajectory (d).

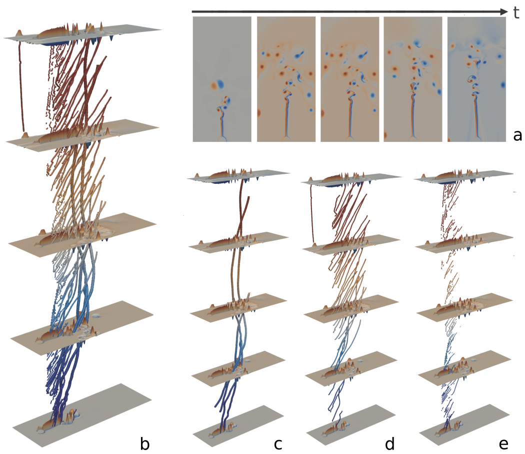

Fig. 2. Boussinesq flow generated by a heated cylinder (a). Feature tracking is performed (b) on the fluid vorticity. Some vortices exist over a long period of time (c), as others vanish more rapidly (d), sometimes akin to noise (e). Feature trajectories can easily be filtered from their lifespan.

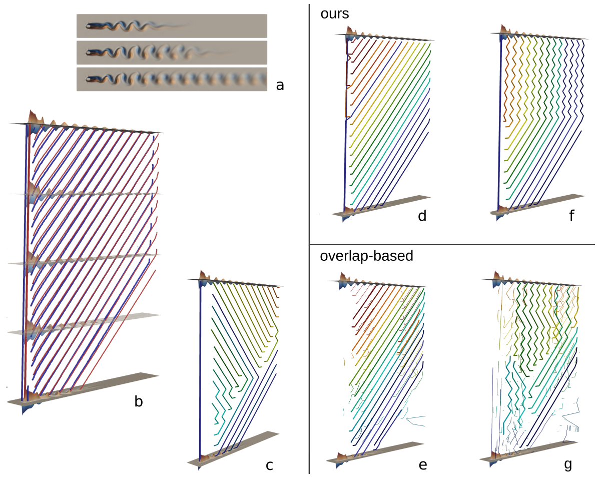

Fig. 3. Simulated von Kármán vortex street (a), on which minima and maxima of the vorticity are tracked with our approach and 1% persistence filtering (b). Only taking the geometry and scalar value into account while doing the matchings (i.e. completely ignoring the *birth* in the lifted metric), is not sufficient to correctly track features (c). Maxima only are tracked considering 1 frame every 5 timesteps (d). With the same temporal resolution, the overlap-based approach (e) does capture small trajectories corresponding to noise, displayed with thinner lines, that have to be filtered for instance using topological simplification [16]. Considering 1 frame every 7 timesteps (f) still yields correct trajectories up to the point where, every other frame, optimal matchings for the metric are between a feature and the preceding one, due to features traveling fast. The overlap approach (g) is less stable in this case as it extracts erroneous trajectories from the very first stages of the simulation to the end.