Computers & Graphics (2017)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/results-in-physics

Metric-based Curve Clustering and Feature Extraction in Flow Visualization

Lieyu Shi[a], Guoning Chen[a]

[a]*Department of Computer Science, University of Houston, Houston, TX, USA, 77204*

ARTICLE INFO

Article history:
Received 24 April 2017
Received in final form 15 June 2017

Keywords: Metric Design, Curve Clustering, Feature Extraction, Flow Visualization

ABSTRACT

We design a number of distance metrics with linear-complexity based on the geometric and statistic properties of 3D curves, and apply them to classify streamlines and the trajectories of particles for flow visualization. The results show that our geometric metrics are more effective than existing metrics (especially spatial-based metrics), enabling the extraction of clusters and representative curves with more insightful and rich information about the flow behaviors.

1. Introduction

Visualizing integral curves or particle trajectories derived from the high-dimensional flow data are important for understanding the physical behaviors of the flows, as they provide an intuitive geometry proxy to depict flow behaviors. Various integral curve generation and visualization techniques have been proposed [1]. Among these techniques, integral curve clustering aims to classify densely placed integral curves into a small number of clusters, from each of which a representative curve is selected to generate a simplified representation of the flow.

Problem Description Designing an appropriate distance metric for the measurement of the similarity between integral curves is not trivial, as it may not only affect the quality of the clustering results but also determine the scalability of the clustering technique. For instance, one common strategy applied by many distance (or similarity) metrics for the comparison of integral curves is to decompose the individual integral curves into segments based on their geometric characteristics (e.g., curvature). Then, the similarity of two integral curves is computed based on the pairs of closest points on the two curves and their corresponding segments. This strategy requires an expensive neighborhood search to determine the closest pairs for the similarity computation, whose complexity is at least $O(NM^2)$ (M is the number of curves and N is the average number of points of each curve), indicating that it does not scale well to large-scale data (e.g., particle trajectory data that consists of hundreds of thousands of curves). In addition, current metrics are sensitive to the parameter selection and user-specified thresholds (e.g., threshold for curve decomposition), making the clustering results unreliable.

Our Contributions To address the above challenge, we propose a number of new distance metrics that can measure the similarity among integral curves or particle trajectories (from particle-based data) without requiring the complicated decomposition. Based on their characteristics, our new metrics can be classified into either spatial-based, geometric-based or statistic-based metrics. Since the similarity computation is performed directly on the entire curves, the complexity of the similarity computation is in general linear. To our best knowledge, we are the first to introduce and design the direct and scalable metrics for particle trajectory clustering and feature extraction for particle-based flow data. We compare our new metrics with the state-of-the-art metric proposed in [2] on both streamline and particle trajectory clustering and find that our geometric-based metrics can select streamlines that represent meaningful features (e.g. vortices), while the metrics in the previous work may not. We also apply the information theoretical framework [3] to assess the generated clustering results and show that our metrics are as good as the best metric from [2] in streamline clustering, while they significantly outperform the other metrics in clustering particle trajectories. Furthermore, we provide an informal analysis on the properties of our new metrics, which we believe can be considered as a guideline (Table 1 in Supplemental Material for the selection and design of suitable distance metrics for the needs of various applications.

2. Related Work

Measuring distance (or similarity) between curves is a fundamental and critical task in various data visualization and exploration tasks, including integral curve clustering [4] and retrieval [5, 6, 7], DTI fiber bundling and exploration [8, 9, 10, 11, 12], and geographical movement study [13, 14]. A large number of distance (or similarity) metrics have been proposed to support the above wide variety of tasks.

Spatial-based proximity Among these many metrics, the spatial-based similarity metrics are widely used for the purposes of curve clustering and bundling. These metrics usually take into account both the point-wise Euclidean distance and the similarity of local geometric characteristics around pairs of closest points. For instance, Chen et al. [15] proposed a distance metric based on Euclidean distance of streamline segments, and Zhang et al. [10] defined it as distance between closest point pairs. Note that the above proximity can also be defined as either the average distance between these pairs of closest points, the smallest distance among these pairs of points [8], the thresholded average distance [11] or the weighted and normalized sum of minimum distance [12]. However, one limitation of these spatial-based metrics is that curves that exhibit similar geometric characteristics but spatially far away will not be considered similar. In addition, searching closest points for the similarity computation often result in high overhead in practice if the number of curves for clustering is large.

Spatial-independent proximity To overcome the above issue of the spatial-based metrics, non-spatial based (or spatial-independent) metrics have been proposed lately. Lee et al. [13] proposed a metric that linearly combines the perpendicular distance, parallel distance and angle distance of the line segments on two curves. McLoughlin et al. [16] introduced the streamline signatures accounting for the curvature, torsion and tortuosity in streamline similarity measurement. However, two streamlines that are not similar may still have similar signatures. To address that, Lu et al. [5] proposed a distribution-based metric that uses dynamic time warping (DTW) ([17]) of 2D histograms mapped from streamline segments and applied it to streamline pattern search. Li et al. [18, 6] defined distance between streamlines as distance between bag-of-feature vectors. Recently, Wang et al. [19, 7] presented a transformation invariant similarity calculation from the symmetric Chamfer distance among closest points, which is effective for segment-based streamline pattern search but sensitive to over-segmentation and suffering from the fixed segmentation of streamlines that may not be able to represent all patterns. The metrics mentioned above often have complexity of $O(N^2)$ for the similarity computation of a single curve as shown in [20], making them not scalable to large-scale simulation data.

In contrast to the above spatial-based and spatial-independent metrics, our metrics do not require 1) the search of nearest points, 2) the decomposition of the input curves, and 3) parameter tuning, making them more efficient to compute with a linear time-complexity.

3. Metric Designing

In this section, we describe our strategy of designing and performing comparison on a number of new distance metrics for flow curve clustering and extraction.

Pre-processing A curve or trajectory in flow data is represented by a series of line segments connecting a sequence of $3D$ vertices. In order to make streamlines equal-size for point-wise distance metrics described in Section 3.1, we fill streamlines of short lengths (at least 2 points) with their last position so that they all have N points (N is the length of the longest streamline in the data) as [2] did. However, this pre-processing is not needed for particle-based flow trajectories.

Based on aspects a metric is measuring, we classify metrics into three groups, i.e., spatial-based, geometric-based and statistic-base metrics. Among these, one spatial-based metric, i.e., the fraction distance metric, $\mathbf{L}_{Frac}$, is adapted from the machine learning community [21] to our flow visualization problem, while all the geometric-based and statistic-based metrics are new metrics proposed in this work. Due to the limitation of the space, we focus on the proposed geometric-based and statistic-based metrics and leave the detailed discussions of other metrics to the supplemental document.

3.1. Geometric-based Metrics

The intuition for our geometric-based metrics is that, 3D curves are composed of line segments, hence, we could incorporate line segment similarity into overall similarity measurements between two curves, similar to the strategy employed by Chen et al. [15]. These metrics on geometric similarity are also translation-free (spatially independent), length-free, scaling-free, and satisfy triangle-inequality. Based on our designing objective (i.e. capturing the most geometrically interesting behavior of a 3D curve), our geometric-based metrics would **favor those curves with higher tortuosity and curvature**.

1. Geometric piece-wise intersection metric $\mathbf{L}_{GPW}$*: Motivated by the fact that two equal-size curves are morphologically similar if their line segments are parallel to each other, we propose a novel geometric piece-wise intersection metric $\mathbf{L}_{GPW}$ as below

$$\mathbf{L}_{GPW}(\mathbf{x},\mathbf{y}) = \frac{1}{N-1}\sum_{i=1}^{N-1} \arccos \frac{\mathbf{p}(\mathbf{x}_i)\cdot\mathbf{p}(\mathbf{y}_i)}{\parallel \mathbf{p}(\mathbf{x}_i)\parallel \cdot \parallel \mathbf{p}(\mathbf{y}_i)\parallel} \quad (1)$$

where $\mathbf{p}(\mathbf{x}_i)$ and $\mathbf{p}(\mathbf{y}_j)$ are the corresponding line segments of $\mathbf{x}$ and $\mathbf{y}$. $\mathbf{L}_{GPW}$ measures the average parallelism between two curves (see Figure 1 in Supplemental Materials for an illustration), and it is translation-free, spatially independent and scaling-free.

2. Signed-angle piece-wise intersection metric $\mathbf{L}_{SAPW}$: Different from $\mathbf{L}_{GPW}$ defined above, $\mathbf{L}_{SAPW}$ measures signed angles based on a pre-defined orientation as

$$\mathbf{L}_{SAPW}(\mathbf{x}, \mathbf{y}) = \left| \frac{1}{N-1} \sum_{i=1}^{N-1} \Phi(\mathbf{p}(\mathbf{x}_i), \mathbf{p}(\mathbf{y}_i)) \right| \quad (2)$$

$$\Phi(\mathbf{p}(\mathbf{x}_i), \mathbf{p}(\mathbf{y}_i)) = \begin{cases} \arccos \frac{\mathbf{p}(\mathbf{x}_i)\cdot\mathbf{p}(\mathbf{y}_i)}{\|\mathbf{p}(\mathbf{x}_i)\|\cdot\|\mathbf{p}(\mathbf{y}_i)\|} & \mathbf{n}_i \cdot \mathbf{n} > 0 \\ -\arccos \frac{\mathbf{p}(\mathbf{x}_i)\cdot\mathbf{p}(\mathbf{y}_i)}{\|\mathbf{p}(\mathbf{x}_i)\|\cdot\|\mathbf{p}(\mathbf{y}_i)\|} & \mathbf{n}_i \cdot \mathbf{n} < 0 \end{cases} \quad (3)$$

$$\mathbf{n}_i = \mathbf{p}(\mathbf{x}_i) \times \mathbf{p}(\mathbf{y}_i) \quad (4)$$

where $\mathbf{n}$ is a fixed normal direction which in our implementation is defined as the normal direction of first pair of the line segments on the two curves. Based on this normal direction, we can distinguish 3D relative counter-clockwise and clockwise rotation. Intuitively, $\mathbf{L}_{SAPW}$ is able to group those trivial and unimportant curves if they are similar in shape despite of symmetric rotation (Figure 3 of Supplemental Materials), and conserve highly tortuous curves of great interesting features. This metric is both translation-invariant and symmetry-invariant.

3. Geometric piece-wise metric with deviation $\mathbf{L}_{GPWD}$*: Similar to $\mathbf{L}_{GPW}$ defined in Eq. (1), $\mathbf{L}_{GPWD}$ only adds the deviation of piece-wise angles as a rotation-free factor, hence, preserves more rotation-free property than $\mathbf{L}_{GPW}$.

$$\mathbf{L}_{GPWD}(\mathbf{x}, \mathbf{y}) = \mathbf{L}_{GPW}(\mathbf{x}, \mathbf{y}) \cdot \sigma \quad (5)$$

where σ is the standard deviation of sequences of $\{\Phi(\mathbf{p}(\mathbf{x}_i), \mathbf{p}(\mathbf{y}_i))\}_{i=1}^{N-1}$ as $\Phi(\cdot,\cdot)$ is defined in Formula (3).

3.2. Statistic-based Metrics

Since each curve can be regarded as a single-variate or multi-variate Gaussian distribution by the law of large numbers in [22], Bhattacharyya metric [23] can be potentially used to measure distance between two curves. Statistic-based metrics derived from Bhattacharyya metric does not need pair-wise comparison, hence the two curves need not be exactly the same size. Despite performance improved by matrix computation, statistic metrics will work best with large enough N, though. Detailed information on statistics-based metric designing is provided in the supplemental document.

4. Metric Analysis

Three important properties are usually examined for a mathematical metrics, including the 1) homogeneity – measures whether a metric is scaling-free or not, 2) triangle-inequality, and 3) definiteness. Besides these three properties, there are also symmetry and non-negativeness property for a mathematical metric, which is satisfied naturally when designing all our metrics, thus, we leave them out from our discussion. Understanding what properties a metric possesses helps select an appropriate metric to reveal the desired characteristics in the 3D curves of interest. We provide a brief description on the properties of the metrics designed in Section 3 in Table 1 of the supplemental document as a reference.

5. Experiments and Evaluation

In this section, we report our experiments and evaluation of the proposed metrics via a number of 3D flow datasets. In addition to the visual comparison and inspection for the metric efficacy evaluation, we also resort to the information theoretical framework [3] to quantitatively assess the metric performance based on the generated clustering results. We compare our results with the-state-of-art principal component analysis (PCA) based streamline clustering result by Ferstl et al. [2], which is the first direct metric for streamline clustering to our best knowledge.

Flow Dataset The flow datasets used in our experiment include the flow behind a square cylinder and two position-based-fluid (PBF) simulation datasets, i.e. dam-breaking and two-half-merging scenarios. The flow behind cylinder data contains 9226 streamlines with 16 symmetric vortex structures. The two PBF simulations [24] were performed with $128K$ and $300K$ particles, respectively. We extracted their trajectories within 250 frames. Figure 6 in Supplemental Materials illustrates two simulations and particle-based trajectories in the simulation data.

Clustering Technique Considering time and memory overhead given the size of the datasets we are working on, we chose K-means clustering to test our new metrics on the above flow datasets.

Evaluation of Results We would visually compare results and check whether important features (e.g., vortices) or overall representative structures of the flow, are extracted. Besides , we also applied the information entropy [3] to quantitatively measure the effectiveness of these metrics.

$$\begin{array}{c} H(X) = \sum_{x_i \in X} p(x_i) \log_2 p(x_i) \\ p(x_i) = \frac{C(x_i)}{\sum_{i=1}^{n} C(x_i)} \end{array} \quad (6)$$

where X is a random variable with a sequence of possible outcomes x,$x \in \{x_1, x_2, \cdots, x_n\}$, and $C(x_i)$ is the number of curves in the i_{th} cluster.

5.1. Results on Streamline Datasets

To visualize the data, we select the closest and the furthest streamlines to the centroid of each cluster, respectively. The reason of not using the centroid streamline–an average curve of all streamlines in the same cluster, is that the centroid streamlines may self-intersect, which does not convey the correct characteristics of streamlines in the steady flows.

Flow behind cylinder data

Figure 1 shows the identified representative streamlines from the clustering results computed using different metrics for the flow behind cylinder data. From these results, we see that our new geometric metrics clearly outperform the other metrics, as they successfully preserve the vortex structures based on visual inspection. Among the three geometric metrics we proposed (i.e., $\mathbf{L}_{GPW}$ (row 5), $\mathbf{L}_{SAPW}$ (row 6) and $\mathbf{L}_{GPWD}$ (row 7)), $\mathbf{L}_{SAPW}$ provides the best representation with streamlines closest to the cluster centroids (middle column), because it preserves more vortex structures. However, when inspecting the representative streamlines furthest away from their centroids,

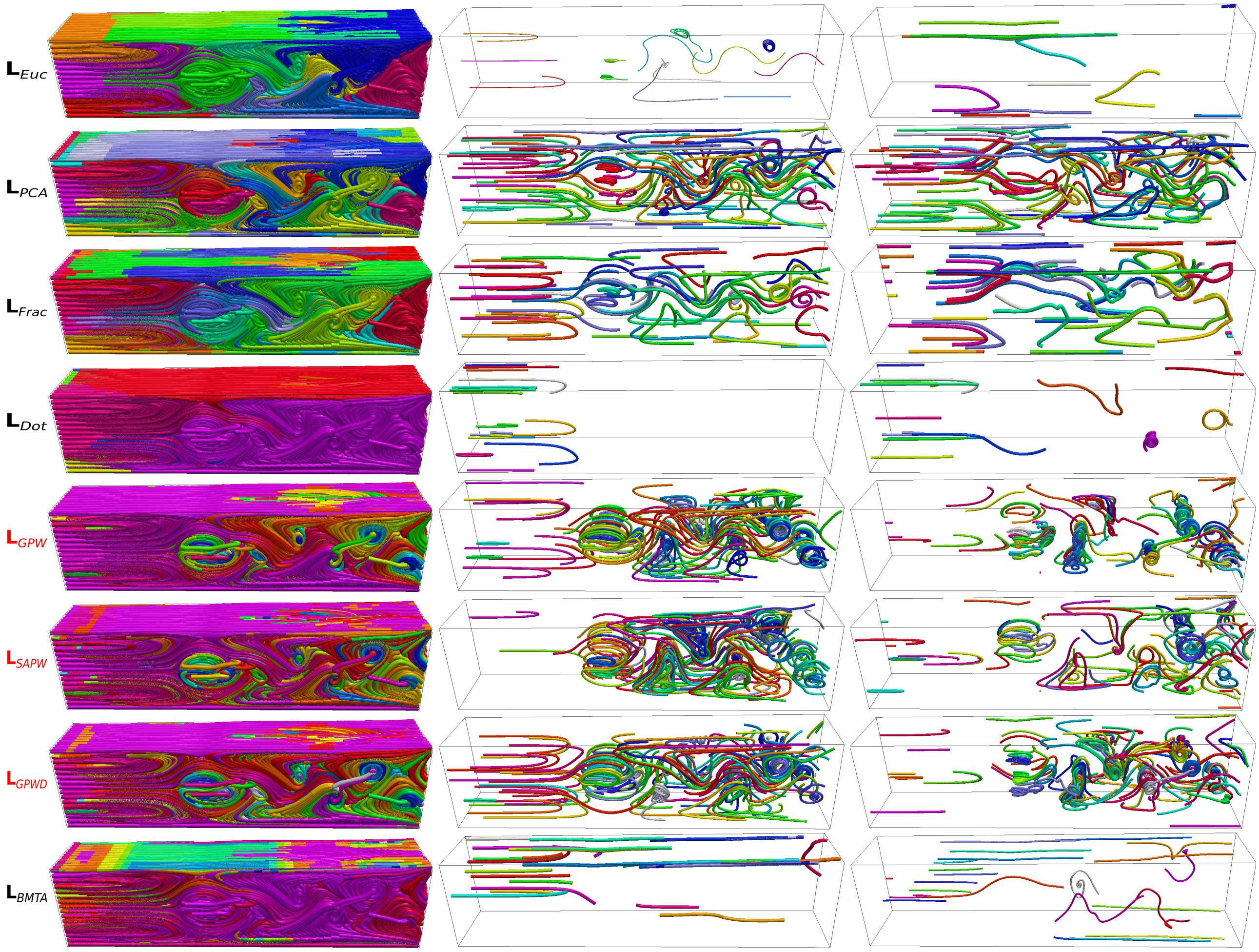


Fig. 1. Metric tested on the flow behind cylinder with the K-means clustering. Several metrics cannot extract more than 10 clusters, so we leave them out. From top to bottom rows, the results corresponding to L_{Euc}, L_{PCA}, L_{Frac}, L_{Dot}, L_{GPW}, L_{SAPW}, L_{GPWD} and L_{BMTA}, respectively. From left to right of each row, grouping streamlines, closest streamlines to centroid, and farthest streamlines to centroid are shown, respectively. Streamlines are color coded based on which clusters they belong to. From these results, we see that three geometric metrics, i.e., L_{GPW} (row 5), L_{SAPW} (row 6) and L_{GPWD} (row 7) (highlighted by red title) are able to extract more vortex structures, while spatial metrics, L_{PCA} (row 2) and L_{Frac} (row 3) result in streamlines that depict the overall structure other than vortices of flow domain.

we see that $\mathbf{L}_{GPW}$ and $\mathbf{L}_{GPWD}$ generate better visual representation than $\mathbf{L}_{SAPW}$. This may be in part due to the fact that the standard deviation in $\mathbf{L}_{GPWD}$ as a multiplication increases the distance of those highly-tortuous streamlines (i.e. having large standard deviation) to their centroids. Therefore, those interesting streamlines may be identified as the furthest streamlines from the centroids.

In the meantime, $\mathbf{L}_{PCA}$ is better in hierarchically presenting the overall structures of flow domain, while it fails to identify streamlines with high curvature and tortuosity. Also, our three geometric-based metrics have close ranking to the $\mathbf{L}_{PCA}$ in the entropy analysis.

To conclude, our geometric metrics, i.e., $\mathbf{L}_{GPW}$, $\mathbf{L}_{SAPW}$ and $\mathbf{L}_{GPWD}$, are robust and effective to extract vortex-like features with high curvature and torturosity. Their entropies are very close to the clustering result as in [2].

5.2. Results on Particle-based Trajectories

The trajectories of particle-based simulations represent spatial positions of particles over time. Due to very large number of particles involved in simulations, directly visualizing these trajectories cannot provide any interpretable information. To address this, we apply the clustering technique on various distance metrics to study some common characteristics in the trend of those particles (i.e., either spatially, geometrically or statistically).

Unlike our experiments on streamline clustering, we also extracted centroid trajectories in addition to the closest and furthest curves because the centroid trajectories may provide the general flow information for particles with similar trajectories without worrying about the self-intersecting, which is a nature characteristic for particle trajectories.

PBF Dam Breaking

The first particle-based dataset for our metric evaluation is dam-breaking simulation [24]. $128K$ trajectories with 250 points for each trajectory are used. The clustering results and their corresponding representative trajectories are partially shown in Figure 3 (see full results in Figure 8 of Supplemental Materials). Besides trivial trajectories either too short but

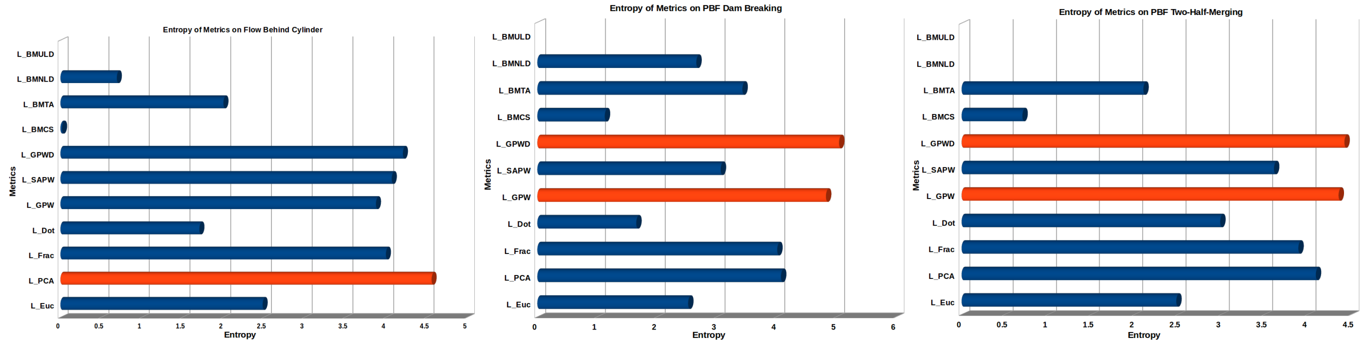


Fig. 2. Entropy values for metrics on the flow-behind-cylinder dataset(left), pbf dam-breaking experiment(middle) and pbf two-half-merging(right). Highest metric entropies are highlighted as red. For flow-behind cylinder, $\mathbf{L}_{PCA}$ is ranked the highest. For pbf dam-breaking, $\mathbf{L}_{GPWD}$ and $\mathbf{L}_{GPW}$ are ranked the first and second, respectively. For pbf two-half-merging, $\mathbf{L}_{GPWD}$ and $\mathbf{L}_{GPW}$ are ranked top two.

tortuous or winding along the wall, we focus more on regions with trajectories stretching highest on the front and back left corners for this simulation data. From the representative trajectories that are closest to the centroids (second column), we see that $\mathbf{L}_{PCA}$, $\mathbf{L}_{Frac}$, $\mathbf{L}_{GPW}$, and $\mathbf{L}_{GPWD}$, $\mathbf{L}_{BMTA}$ and $\mathbf{L}_{BMNLD}$ are able to preserve these two important trajectories. For centroids, $\mathbf{L}_{PCA}$, $\mathbf{L}_{Frac}$ and $\mathbf{L}_{GPW}$ work better. However, we notice that the leftmost centroid trajectory exits the flow domain using $\mathbf{L}_{PCA}$, which is not physically plausible. Overall, we see that $\mathbf{L}_{GPW}$ performs the best in extracting features while preserving the overall structures of the flow of this simulation. From this experiment, we can conclude that our geometric-based metrics can encode richer details than the other metrics and are able to preserve features of importance. Our entropy analysis (Figure 2) also supports our conclusion, in which two geometric metrics, $\mathbf{L}_{GPW}$, and $\mathbf{L}_{GPWD}$ (highlighted in red) outperform other metrics with the highest entropy values.

We also provide analysis on another particle trajectories of the two-half-merging simulation in Supplemental Material, and both visual inspection and entropy value favor our geometric metrics.

Remark: Although results on statistics-based metrics do not extract features or patterns as accurately as the geometric-based metrics, we believe the statistic-based metrics can be useful in the clustering of integral curves derived from uncertain flow, which we plan to explore further in the future.

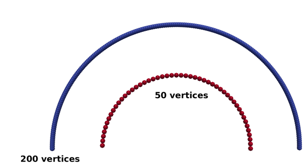
6. Conclusion

In this work, we proposed a number of geometric-based and statistic-based metrics with linear time-complexity for $3D$ curve comparison combined with standard clustering techniques for various curve-based exploration tasks. We analyze the properties of these metrics and evaluate their effectiveness by combining them with the conventional K-means clustering to a number of curve-based datasets derived from various fluid simulations and with varying sizes. Our analysis and experiments show that our geometric-based metrics are able to identify curves that convey interesting flow features (e.g. vortical behaviors) in most cases. In particular, our geometric-based metrics can help identify representative trajectories that possess richer details than the other metrics. We also perform the quantitative measurement of the clustering results based on the information theoretical framework and discover that when dealing with the streamline-based data, our geometric metrics have comparable performance to those spatial-based metrics, while our geometric metrics outperform the others in the processing of the particle trajectory data.

Our new geometric metrics are intuitive and simple, and they do not require complex pre-processing, hence, can be applied to large-scale data processing

Limitations and future work

Our geometric metrics for curves are based on piecewise parallelism, and hence requires equal-size of curves. Currently, we fill in the missing positions of those short streamlines using their end positions to satisfy the equal-size requirement. However, it's difficult to accurately predict and measure visually similar curves with extremely different sizes (e.g., the two curves in the inset The statistic-based metrics have the potential to address this. However, their current clustering results are not as good as the geometric-based metrics, which requires to further investigation.



Clustering strategy is also a critical problem that needs to be improved in the future. In particular, we would like to employ the hierarchical clustering to provide a level-of-detail representation of the flow. However, its computational overhead is quite high. Also, particle trajectories from the particle-based simulations usually convey more physical information than purely geometric information, e.g., vorticity, shearing and acceleration, which may be incorporated into the future design of metrics in hope of producing a more physically meaningful representation of the flow data.

Finally, in this work we directly applied information entropy [3] to quantitatively evaluate the clustering results, which may not conceptually intuitive enough. We would like to exploit a more accurate and reasonable standard for quantitative comparisons of clustering in the future work.

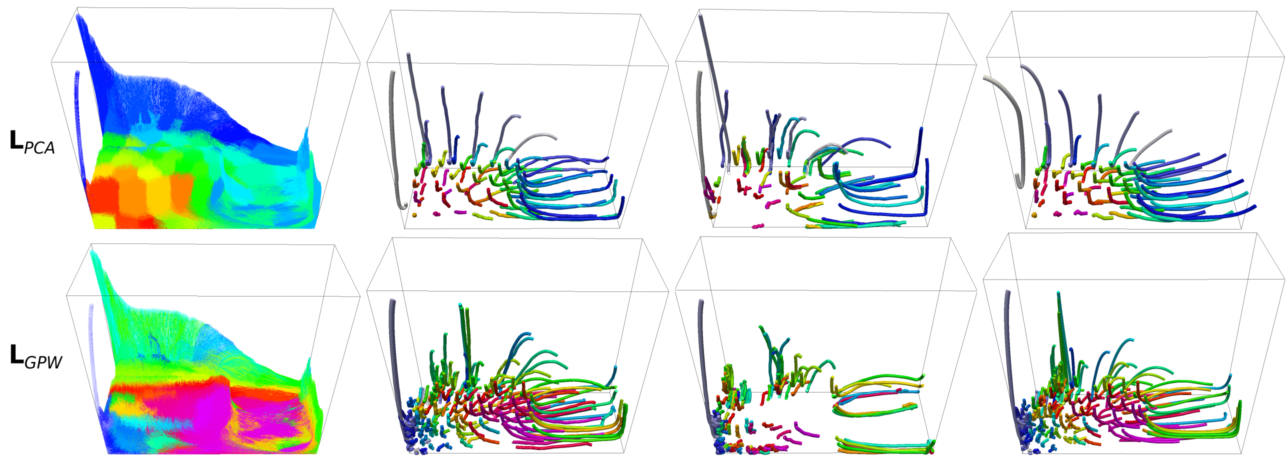


Fig. 3. Metric evaluation on the dam-breaking data with the K-means clustering.From top to bottom, results with L_{PCA} and L_{GPW} are shown, respectively. Left to right of each row, grouping trajectory, trajectories closest, furthest away from centroid and centroid of each cluster are shown, respectively. In all three representative trajectory identification, L_{GPW} works the best with the latter having more details. The other metrics have more or less omitted trajectories of importance.

Acknowledgments

This work is supported by US NSF IIS-1553329. We wish to thank Lei Zhang for providing streamline datasets. We would also like to thank the anonymous reviewers for their constructive comments.

References

[1] McLoughlin, T, Laramee, RS, Peikert, R, Post, FH, Chen, M. Over two decades of integration-based, geometric flow visualization. Computer Graphics Forum 2010;29(6):1807–1829. URL: `http://dx.doi.org/10.1111/j.1467-8659.2010.01650.x`. doi:`10.1111/j.1467-8659.2010.01650.x`.

[2] Ferstl, F, Brger, K, Westermann, R. Streamline variability plots for characterizing the uncertainty in vector field ensembles. IEEE Transactions on Visualization and Computer Graphics 2016;22:767–776.

[3] Xu, L, Lee, TY, Shen, HW. An information-theoretic framework for flow visualization. IEEE Transactions on Visualization and Computer Graphics 2010;16(6):1216–1224.

[4] Yu, H, Wang, C, Shene, CK, Chen, JH. Hierarchical streamline bundles 2012;18(8):1353–1367.

[5] Lu, K, Chaudhuri, A, Lee, TY, Shen, HW, Wong, PC. Exploring vector fields with distribution-based streamline analysis. In: 2013 IEEE Pacific Visualization Symposium (PacificVis). 2013, p. 257–264. doi:`10.1109/PacificVis.2013.6596153`.

[6] Li, Y, Wang, C, Shene, CK. Extracting flow features via supervised streamline segmentation. Computers & Graphics 2015;52:79–92.

[7] Wang, Z, Esturo, JM, Seidel, HP, Weinkauf, T. Stream linebased pattern search in flows. Computer Graphics Forum 2016;:n/a–n/aURL: `http://dx.doi.org/10.1111/cgf.12990`. doi:`10.1111/cgf.12990`.

[8] Corouge, I, Gouttard, S, Gerig, G. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro. 2004, p. 344–347.

[9] Moberts, B, Vilanova, A, van Wijk, JJ. Evaluation of fiber clustering methods for diffusion tensor imaging. In: VIS 05. IEEE Visualization, 2005. 2005, p. 65–72. doi:`10.1109/VISUAL.2005.1532779`.

[10] Zhang, S, Correia, S, Laidlaw, D. Identifying white-matter fiber bundles in DTI data using an automated proximity-based fiber-clustering method. IEEE Trans Vis Comput Graph 2008;14(5):1044–1053.

[11] Chen, W, Zhang, S, Correia, S, Ebert, DS. Abstractive representation and exploration of hierarchically clustered diffusion tensor fiber tracts. Computer Graphics Forum 2008;27(3):1071–1078. URL: `http://dx.doi.org/10.1111/j.1467-8659.200i.01244.x`. doi:`10.1111/j.1467-8659.2008.01244.x`.

[12] Jianu, R, Demiralp, C, Laidlaw, D. Exploring 3d DTI fiber tracts with linked 2d representations. IEEE Transactions on Visualization and Computer Graphics 2009;15(6):1449–1456.

[13] Lee, JG, Han, J, Whang, KY. Trajectory clustering: A partition-and-group framework. In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. SIGMOD '07; New York, NY, USA: ACM. ISBN 978-1-59593-686-8; 2007, p. 593–604. URL: `http://doi.acm.org/10.1145/1247480.1247546`. doi:`10.1145/1247480.1247546`.

[14] Demšar, U, Buchin, K, Cagnacci, F, Safi, K, Speckmann, B, Van de Weghe, N, et al. Analysis and visualisation of movement: an interdisciplinary review. Movement ecology 2015;3(1):5.

[15] Chen, Y, Cohen, JD, Krolik, J. Similarity-guided streamline placement with error evaluation. IEEE Trans Vis Comput Graph 2007;13(6):1448–1455.

[16] McLoughlin, T, Jones, MW, Laramee, RS, Malki, R, Masters, I, Hansen, CD. Similarity measures for enhancing interactive streamline seeding. IEEE Transactions on Visualization and Computer Graphics 2013;19(8):1342–1353. URL: `http://dx.doi.org/10.1109/TVCG.2012.150`. doi:`10.1109/TVCG.2012.150`.

[17] Berndt, DJ, Clifford, J. Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. AAAIWS'94; AAAI Press; 1994, p. 359–370. URL: `http://dl.acm.org/citation.cfm?id=3000850.3000887`.

[18] Li, Y, Wang, C, Shene, CK. Streamline similarity analysis using bag-of-features. In: IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics; 2013,.

[19] Wang, Z, Esturo, JM, Seidel, HP, Weinkauf, T. Pattern Search in Flows based on Similarity of Stream Line Segments. In: Bender, J, Kuijper, A, von Landesberger, T, Theisel, H, Urban, P, editors. Vision, Modeling & Visualization. The Eurographics Association. ISBN 978-3-905674-74-3; 2014,doi:`10.2312/vmv.20141272`.

[20] Ratanamahatana, CA, Keogh, E. Everything you know about dynamic time warping is wrong. 2004,.

[21] On the Surprising Behavior of Distance Metrics in High Dimensional Spaces. ICDT '01; 2001. ISBN 3-540-41456-8.

[22] Dinov, ID, Christou, N, Gould, R. Law of large numbers: the theory, applications and technology-based education. J Stat Educ 2009;17:1–19.

[23] Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society 1943;35:99–109.

[24] Macklin, M, Müller, M. Position based fluids. ACM Transactions on Graphics (TOG)-SIGGRAPH 2013 2013;32.