

Unsteady Flow Visualization via Physics based Pathline Exploration

Duong B. Nguyen*
University of Houston

Lei Zhang†
University of Houston
Rodolfo Ostilla Monico‡
University of Houston

Robert S. Laramée§
Swansea University
Guoning Chen||
University of Houston

David Thompson§
Mississippi State University

ABSTRACT

This work proposes to analyze the time-dependent characteristics of the physical attributes measured along pathlines derived from unsteady flows, which can be represented as a series of time activity curves (TAC). A new TAC-based unsteady flow visualization and analysis framework is proposed. The center of this framework is a new event-based distance metric (EDM) that compares the similarity of two TACs, from which a new spatio-temporal, hierarchical clustering of pathlines based on their physical attributes and an attribute-based pathline exploration are proposed. These techniques are integrated into a visual analytics system, which has been applied to a number of unsteady flow in 2D and 3D to demonstrate its utility.

Keywords: Flow visualization, Time activity curves, Clustering

1 INTRODUCTION

Vector field visualization is a ubiquitous technique that is employed to study a wide range of dynamical systems. Many effective approaches have been developed to visualize such complex data [6, 11, 17, 19]. Among these techniques, the geometric-based approaches [6, 14] are commonly applied due to their intuitive representation of flow behavior. Examples of geometric-based techniques include various integral curve/surface based representations and integral curve clustering that concentrate on the geometric characteristics of the flow (e.g., the shape or curvature of integral curves).

Previous work and limitations. However, the geometric representation of the flow need not provide sufficient representation of the underlying physical characteristics. For example, vector field topology [10], an abstract representation of the geometric characteristics of steady flow, only encodes hyperbolic features, such as fixed points [18, 23] and hyperbolic periodic orbits [3, 22, 27] in the flow. Other physically relevant information, such as vortices, shearing, etc., are not always captured [31]. Similarly, clustering methods that select integral curves to ensure sufficient spatial coverage and to reduce cluttering typically do not consider physical importance. Second, a geometric representation may not intuitively reveal the physical behavior of the flow, as shown by Zhang et al. [32]. In Figure 1(a), the red and purple pathlines are both flat but have rather different physical characteristics (Figure 1(b)). To incorporate physics into the visualization of unsteady flow, Zhang et al. [32] introduced a Lagrangian accumulation framework that can be used to characterize integral curves by inspecting their respective overall attribute behaviors (i.e., each integral curve is assigned a value by accumulating the values of an attribute of interest along the curve). That technique was inspired by the pathline attributes introduced by Shi et al. [20]. In the meantime, Lee et al. proposed a visualization framework to analyze time-varying data sets with a time activity curve (TAC) based distance field [12], which is used to

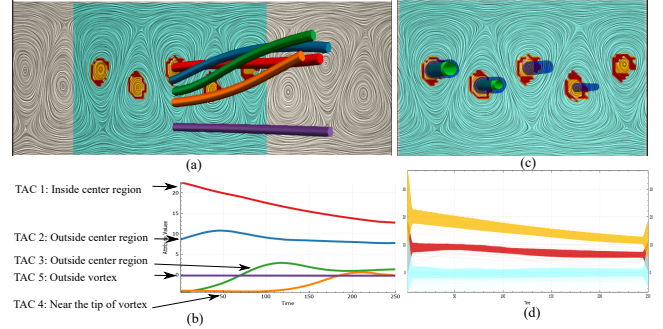


Figure 1: Clustering result of the 2D cylinder flow using attribute Q reveals a three-layered configuration of the vortex system ($p_c = 100$). Pathlines sampled along a line passing the center of a vortex (a) and their corresponding TACs (b). The TAC (red) of the pathline seeded in the vortex center decreases monotonically over time, indicating the diffusion of the concentrated vorticity, which gradually increases the vorticity in the outer regions of the shedded vortices (TACs 2-4). TAC 5 corresponds to the pathline seeded outside of vortex region, which exhibits stable characteristic. (c) The two iso-contours (blue and green) with Q values of 17 and 24, respectively, cannot fully capture the vortex configuration. (d) TAC profiles of our clustering results.

highlight features. TACs have also been studied in other scientific visualization problems [7, 9, 25, 28–30]. The existing research looks at the overall (or global) characteristics of attributes, which may not capture all of the relevant characteristics, e.g., the diffusion of Q during the advection of vortices over time (Figure 1), due to the suppression of local information.

Our contributions. To incorporate more detailed physics into the analysis and visualization of unsteady flow, we propose a novel visual analysis framework based on the temporal behavior of local, physical attributes of interest measured along individual pathlines. Similar to Lee et al. [12], we refer to the temporal profile of the attribute along a pathline as time activity curve (TAC), specifically a Lagrangian TAC. Unlike Shi et al. [20] and other similar methods that also compute the pathline attributes, our framework takes into account the arbitrary movement of the observer. That is, the physical properties are computed from the new instantaneous vector fields after applying the optimal reference frames, which achieves a better alignment between physical features and the geometric representation of the flow, as demonstrated in [8]. The benefits of analyzing the flow behavior based on TACs are two-fold. First, they are 1D plots that are independent of the flow dimension (i.e., applicable to both 2D and 3D data). Second, the geometric characteristics of TACs (e.g., ascending, descending, peaks, valleys, etc.) reveal the interaction of flow particles with physical features over time. This enables us to explain the geometric characteristics of the corresponding pathlines and vice versa.

By utilizing the advantages of analyzing TACs, We introduce a new spatio-temporal, hierarchical clustering of pathlines based on their respective TACs and a TAC-based pathline selection and exploration. Central to these operations is a comprehensive distance metric for the comparison of two TACs, which we refer to as an Event-based Distance Metric (EDM) that incorporates the correlation of pair-wise TACs and the spatio-temporal Euclidean distances between events (Section 3.1). We develop a visual explo-

*e-mail: dbnguyen11@uh.edu

†e-mail: zhanglei.later@gmail.com

‡e-mail: R.S.Laramée@swansea.ac.uk

§e-mail: dst@cavs.msstate.edu

¶e-mail: rostilla@central.uh.edu

||e-mail: chengu@cs.uh.edu

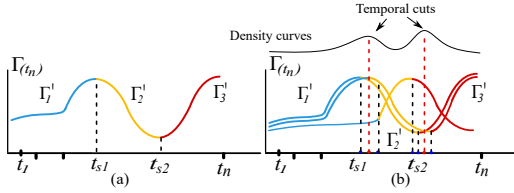


Figure 2: (a) An example TAC with three events. t_{s1} and t_{s2} are two extrema. (b) An example of identifying temporal cuts for temporal clustering. Cuts are selected at the maxima of the density curve.

ration system that integrates the aforementioned TAC-based analysis and exploration techniques with a modified edge-bundling visualization of TAC clusters. We have applied our TAC-based exploration system to a number of 2D and 3D unsteady flows. Our framework effectively reveals the two-layer configuration of a vortex and its decay over time in vortex shedding (Figure 1), which is difficult to reveal via conventional methods. We also facilitate interpretation of the temporal behavior of a 2D vortex ring, including its interaction with a wall and its breakdown.

2 VECTOR FIELD AND TAC BACKGROUND

Consider a spatio-temporal domain $\mathbb{D} = \mathbb{M} \times \mathbb{T}$ where $\mathbb{M} \subset \mathbb{R}^d$ is a d -manifold ($d = 2, 3$) and $\mathbb{T} \subset \mathbb{R}$, a general vector field can be expressed as an ordinary differential equation (ODE) $\dot{\mathbf{x}} = V(\mathbf{x}, t)$. For an *unsteady* (or *time-dependent*) vector field $V(\mathbf{x}, t)$, the trajectory of a particle starting at $\mathbf{x}_0$ and at time t_0 is called a *pathline*, denoted by $\mathbf{x}_{\mathbf{x}_0, t_0}(t) = \mathbf{x}_0 + \int_{t_0}^t V(\mathbf{x}_{\mathbf{x}_0, t_0}(\tau), \tau) d\tau$. Given a steady vector field $\mathbf{v}$, its spatial gradient $\nabla_{\mathbf{x}} \mathbf{v}$ is referred to as its *Jacobian*, denoted by $\mathbf{J}$. From V and $\mathbf{J}$, a number of local physical attributes of the flow can be derived [2], including kinetic energy, vorticity, λ_2 , Q , and local shear rate, which are used in this work.

Definition of TAC. Given a local attribute $\mathbf{A}$, a Lagrangian TAC along a pathline $\mathcal{C}$ of a particle, seeded at $\mathbf{x}$ at time t , can be expressed as $\Gamma_{\mathbf{A}, \mathcal{C}}[i] = \mathbf{A}(\mathcal{C}(\mathbf{x}, t_i), t_i) | i = 1, 2, \dots, n$, where $t_1, t_2, \dots, t_n$ are the sample times within the time window $T \subset \mathbb{T}$ and $\mathcal{C}(\mathbf{x}, t_i)$ is the location of $\mathbf{x}$ on the pathline $\mathcal{C}$ at time t_i .

For simplicity, we denote a time activity curve as $\Gamma = \{\Gamma[i] | i = 1, 2, \dots, n\}$ where $\Gamma[i]$ is the local attribute value at time t_i . Figure 2(a) illustrates a TAC where the x axis indicates time indexes and the y axis shows the local attribute values at the corresponding times. The length of a TAC is the number of values in the TAC, indicating the lifespan of the corresponding particle.

TAC Decomposition. To better describe the behavior of a TAC and measure the difference between TACs (Section 3.1), we perform a decomposition. Given a TAC Γ (Figure 2(a)), the decomposition subdivides Γ into a sequence of contiguous, non-overlapping sub-TACs, Γ'_i , i.e., $\Gamma = \langle \Gamma'_1, \Gamma'_2, \dots, \Gamma'_z \rangle$ (e.g., $\Gamma = \langle \Gamma'_1, \Gamma'_2, \Gamma'_3 \rangle$ in Figure 2(b)). Each sub-TAC consists of exactly one primitive trend, which can be either stable, ascending or descending, and each pair of two neighboring sub-TACs have differing primitive trends. The position that splits two neighboring sub-TACs is called a split point t_s (e.g., the extrema t_{s1} and t_{s2} in Figure 2(b)). These split points will be used for the later temporal clustering (Figure 2(b)). In our implementation, we simply apply 1D Morse decomposition [5] to generate a number of sub-TACs for each TAC.

3 TAC-BASED FLOW EXPLORATION FRAMEWORK

Our framework starts with computing densely placed pathlines in the flow. Depending on the attributes of interest, the corresponding TACs are derived from the obtained pathlines and segmented into multiple time intervals. Next, a hierarchical clustering is performed based on the characteristics of the entire TACs (Section 3.2). Based on the global clustering result, a hierarchical temporal clustering of TACs is performed to capture the level-of-detail characterization of their temporal behavior (Section 3.2.1).

3.1 TAC-based Distance Metrics

To assist the spatio-temporal clustering of TACs, we first describe our distance metric for TACs. To accurately compare the difference

in the characteristics of two TACs, the distance metric must take both the temporal trends and the magnitude of TACs into account. The traditional distance metrics, such as the Euclidean distance and the Pearson correlation coefficient, concentrate on either the trend or the magnitude of the TACs and cannot satisfy our needs. Another metric for measuring the similarity of two time series is Dynamic Time Warping (DTW) [12]. DTW considers both shift and deformation of the time series. However, the time stamp for each sample in the TACs has specific meaning, which requires us to align the TACs based on the time stamps, making DTW not suitable in our cases.

To address the limitations of the existing metrics, we introduce a new distance metric, called the Event-based Distance Metric (EDM). Based on the TAC decomposition discussed earlier, we define the distance between two TACs Γ_1 and Γ_2 as the similarity of the corresponding events detected from Γ_1 and Γ_2 , i.e., $\Gamma_1 = \langle \Gamma'_{11}, \Gamma'_{12}, \dots, \Gamma'_{1k} \rangle$ and $\Gamma_2 = \langle \Gamma'_{21}, \Gamma'_{22}, \dots, \Gamma'_{2k} \rangle$. Our Event-based Distance Metric (EDM) distance is defined as follows:

$$D_{edm}(\Gamma_1, \Gamma_2) = \sum_{i=1}^k (1 + P_c \times D_{corr}(\Gamma'_{1i}, \Gamma'_{2i})) \times D_e(\Gamma'_{1i}, \Gamma'_{2i}) \quad (1)$$

$$D_{corr}(\Gamma_1, \Gamma_2) = 0.5 - \frac{cov(\Gamma_1, \Gamma_2)}{2\sigma_{\Gamma_1}\sigma_{\Gamma_2}} \quad (2)$$

where cov is the covariance and σ_{Γ} is the standard deviation of Γ .

$$D_e(\Gamma_1, \Gamma_2) = \sqrt{\sum_{i=1}^n (\Gamma_1[i] - \Gamma_2[i])^2} \quad (3)$$

In the above definition, $D_e(\Gamma_1, \Gamma_2)$ represents the Euclidean spatial distance between sub-TACs Γ'_{1i} and Γ'_{2i} . $D_e(\Gamma_1, \Gamma_2)$ addresses the challenges where TACs exhibit similar trends but with different magnitudes. $D_{corr}(\Gamma_1, \Gamma_2)$ measures the global correlation between TACs Γ_1 and Γ_2 . If $D_{corr}(\Gamma_1, \Gamma_2) = 1$, then $(1 + P_c \times D_{corr}(\Gamma'_{1i}, \Gamma'_{2i})) = 2$, which means Γ_1 and Γ_2 have opposite trends. In contrast, if $D_{corr}(\Gamma_1, \Gamma_2) = 0$, Γ_1 and Γ_2 have the same trend and the value of the first term is 1. $D_{corr}(\Gamma_1, \Gamma_2)$ aims to resolve the ambiguity where Γ_1 and Γ_2 have similar Euclidean distance, but different correlation distance relative to Γ_{base} . P_c represents a user-assigned importance for the spatial difference and the global correlation. The higher value of P_c , the more weight given to the global correlation. By default, we set $P_c = 1$.

We use multiplication instead of addition to combine D_e and D_{corr} in Eq. (1) due to the relation between the two terms, and their value range difference. The first term is equal to 1 when the two TACs contain similar trends. In this case, the distance between two TACs is completely based on the second term (i.e., Euclidean distance). If the two TACs have opposite trends, then the distance between the TACs is expected to be large. By multiplying, we magnify the second term by a maximum of two when the TACs have inverse trends. EDM requires linear time to compute, and thus, can be applied to large data sets.

Note that EDM requires that Γ_1 and Γ_2 have the same number of sub-TACs, and the lengths of two corresponding sub-TACs Γ'_{1i} , Γ'_{2i} are equal. Consider $\Gamma_1 = \langle \Gamma'_{11}, \Gamma'_{12}, \dots, \Gamma'_{1z} \rangle$ and $\Gamma_2 = \langle \Gamma'_{21}, \Gamma'_{22}, \dots, \Gamma'_{2k} \rangle$, $z > k$. To ensure Γ_1 and Γ_2 have equal length, we apply the split points of Γ_1 to Γ_2 as it enables to capture the detailed correlation of the two TACs in a finer temporal partitioning.

3.2 TAC-based Clustering

Global clustering. In order to provide different levels of detail for flow behavior w.r.t. the local attributes, we perform the clustering of TACs using the new distance metric over *all* temporal samples, coupled with the popular agglomerative hierarchical clustering (AHC). The linkage type used in this work is the complete linkage since it is better for finding compact clusters of approximately equal diameter [4]. To reduce the traditional cubic time complexity, we implement the parallel, locally-ordered AHC proposed by Walter et al. [1], which runs in sub-quadratic time.

3.2.1 TAC-based Temporal AHC

Two TACs that belong to two clusters globally may possess local segments having similar behavior. See Figure 2(b) for an example.

To capture this detail, we propose a hierarchical clustering algorithm in the temporal dimension, i.e., a temporal AHC.

Time interval segmentation of TACs. To study TACs in a level-of-detail fashion, we apply time interval segmentation to a group of TACs. The time intervals that segment TACs should preserve TAC characteristics. In other words, one primitive trend of a TAC should not be segmented into two time intervals, which causes fragmentation. As described in Section 2, each TAC can be decomposed into a number of sub-TACs, corresponding to a number of temporal segments, as shown in Figure 2(b). However, for a group of TACs, it is not guaranteed that the segment split points are identical at time. To address this, we utilize a 1D Gaussian kernel density estimation (KDE). Specifically, we first identify the split points for each TAC. Let $x_1, x_2, \dots, x_N$ be the numbers of split points at the time location $i = 1, 2, \dots, N$. We estimate a density curve (Figure 2 (b)) using the univariate fixed bandwidth kernel estimator [21] defined as: $f(x) = \frac{1}{NH} \sum_{i=1}^N K(\frac{x-x_i}{H})$, where $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the Gaussian kernel, and H is the width of the kernel. The optimal cuts are selected as the local maxima of the estimated density curve (Figure 2 (b)). Selection of the value of H is important in KDE as it can make the density estimate smoother or noisier. In our experiment, we found that setting H to 1 enables us to capture a reasonable number of cuts for the later temporal clustering.

The temporal cuts obtained from the above KDE segment the entire time period T into a number of intervals of varying length, referred to as $T = (T_1, T_2, \dots, T_m)$. In this way, all TACs are segmented by these cuts which attempt to preserve the most common characteristics of all TACs (Figure 2(b)).

Temporal hierarchical clustering After performing the temporal partitioning and obtaining the local time intervals, we now perform AHC within each interval. We apply the distance metric proposed in Eq. (1) for temporal clustering.

Assigning the cluster number for each time interval is difficult because, on one hand, the number of time intervals in time-hierarchical clustering varies, while on the other hand, the cluster distances in different time intervals may be different. To show the consistent changes across time intervals, the same treatment needs to be applied uniformly. Therefore, we use the distance threshold ϵ for the global clustering to guide the clustering within individual time intervals. Specifically, the distance threshold ϵ_i for time interval T_i is determined by the time range of the interval, i.e., $\epsilon_i = \frac{|T_i|}{|T|} \epsilon$. In this way, it is foreseeable that there are more clusters generated in the time intervals where the TACs behave more diversely, i.e., when TACs have larger dissimilarities.

The goal of temporal clustering is to build up a hierarchical tree of the input m time intervals obtained in the previous temporal partitioning, i.e., m leaf nodes of the tree, so that the level-of-detail of a TAC's behavior can be observed in the temporal dimension. In contrast to the spatial hierarchical clustering, in which any two clusters can be selected for a merging operation, in temporal hierarchical clustering only two clusters that are contiguous in time can be merged together, which makes the merging operation simpler. In our implementation, starting from the initial m leaf nodes, a distance array $D \in \mathbb{R}^{(m-1)}$ is created. Each entry indicates the dissimilarity after a pair of consecutive time intervals are merged into one. $D[i]$ can be computed as follows.

$$D[i] = \eta(T_i) + \eta(T_{i+1}) - \eta(T_i \cup T_{i+1}) \quad (4)$$

where $\eta(T_k)$ is the average pairwise dissimilarity within a time interval T_k , $T_i \cup T_{i+1}$ is the new time interval obtained by merging T_i and T_{i+1} .

$$\eta(T_k) = \sqrt{\frac{\sum_{i=1}^{|T_k|} \sum_{j=i}^{|T_k|} (D_{edm}(\Gamma_i, \Gamma_j))^2}{|T_k|(|T_k| - 1)}} \quad (5)$$

$\eta(T_k)$ reflects the compactness of the TACs in the time interval T_k . The larger the value of $\eta(T_k)$, the further the TACs in T_k are located from the centroid.

In spatial AHC, the two clusters with the smallest distance are selected for merging. Similarly, in temporal hierarchical cluster-

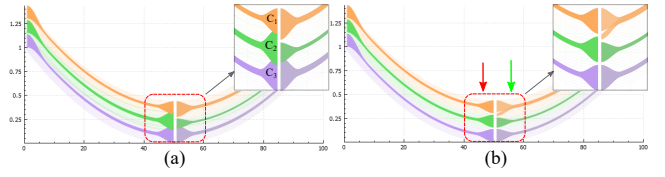


Figure 3: Visualization of transition between time intervals. (a) Original edge-bundling visualization (b) Our modified version. Magnified views show the transition between two time intervals.

ing, the two time intervals with the smallest dissimilarity changes are merged together first. In other words, time interval T_k and its neighboring time interval T_{k+1} that satisfies $D[k] \leq D[i], \forall i \leq M$, are first merged together to generate a new time interval $T_k + T_{k+1}$ and then removed from the node list. Consequently, a new $m-2$ dimension distance array $D_{(m-2)}$ is generated with the remaining $m-1$ nodes. The above merge process is iterated until only one time interval, i.e., the entire time period, remains as the root of the temporal hierarchical tree. The height of the temporal hierarchical tree built on m time intervals is $m-1$. On the i^{th} level of the tree, i.e., the height is i , there are $m-i$ time intervals.

3.3 TAC-based Flow Visualization System

We integrate the above clustering technique for TACs and their exploration into a visualization system. Our system consists of a number of linked views, including (1) a spatial view for the user to select region of interest in the space for exploration and to visualize the analysis result in the flow domain, (2) a TAC view that shows the clustering result of TACs and supports the user interaction, and (3) an abstract view that shows the hierarchical tree of the clusters obtained using AHC. We recommend the readers to watch the accompanying video for more details about the interface and user interactions of our system.

Visualize TAC clusters. To visualize TAC clusters, we adapt the edge bundling technique for parallel coordinate plot visualization by Palmas et al. [16] with the adjustment at the tail and head of each cluster (Figure 3(a)) to reduce visual clutter and achieve color consistency across different time intervals.

To address the visual overlap at two ends of edge bundles (Figure 3(a)), we offset proportionally to clusters' size, whose heads or tails are overlapping. As illustrated by the red arrow in Figure 3(b), the minimum value of C_2 at the tail end is increased and the maximum value of C_3 at the tail end is decreased, eliminating the overlapping between C_2 and C_3 while preserving the relative range size simultaneously. Removing overlapping at the tail of T_{k-1} makes the boundaries of source clusters clear. To fully resolve the connections among time intervals, we visualize both main and minor sources at the head of a cluster. From Figure 3(b), we can easily ascertain the transition of clusters between two time intervals.

4 APPLICATIONS OF TAC-BASED EXPLORATION

2D flow behind a cylinder. We first apply our clustering to a 2D simulation of the flow behind a cylinder with a Reynolds number of 160 [26]. The simulation covers a subset of the spatio-temporal domain, $[-0.5, 7.5] \times [-0.5, 0.5] \times [15, 23]$, where the vortex shedding is fully formed. According to the domain experts, the core region of a vortex in this flow has a motion close to that of a rigid body rotation, which helps to preserve the shape of the vortex. However, the concentrated vorticity in the vortex cores will diffuse due to viscosity (i.e., friction) and the absence of an external forces to maintain the rotation [13]. The diffused vorticity will reach the outer layer of vortices where it will interact with vorticity from other vortices, thus losing the coherent character. The overall structure of the vortices is stable due to interleaving and somewhat symmetric configuration of the counter-rotating vortices.

We choose the first 250 time steps of this simulation and use uniform sampling in a spatial resolution of 1200×150 to compute pathlines and measure the attributes along them. As demonstrated in Figure 1(a), our framework identifies three regions using the TACs of the Q attribute without significant user intervention: the

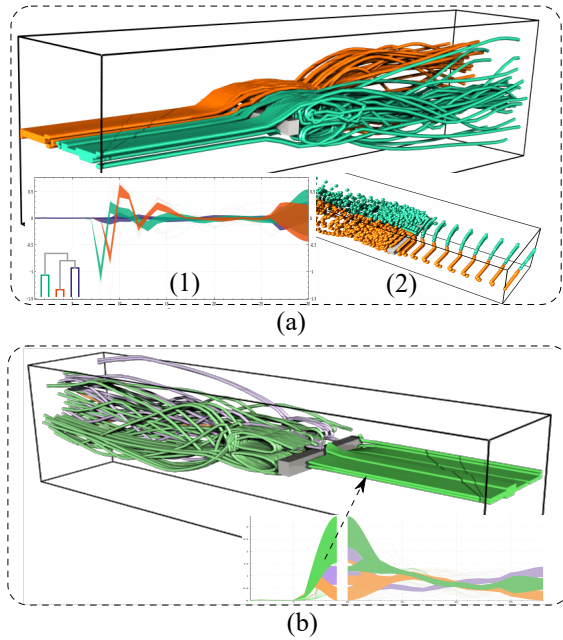


Figure 4: Clustering result of the 3D Cylinder flow based on attribute Q ($p_c = 1$). (a) Global clustering result. The TAC profiles corresponding to the two pathline groups at the center location show a shifting, indicating particles in these two regions exhibit vortex shedding at different times. (b) Temporal clustering result. The pathlines sampled at shearing layers ($Q < 0$) exhibit different behaviors in the second time interval (i.e., splitting into three clusters)

viscous vortex core where the vorticity is concentrated, the outer layer of the vortices where vorticity diffuses and grows and the region outside of the vortices where the flow is almost irrotational. In addition, the TACs’ visualization in Figure 1(d) informatively characterizes the attribute behaviors within different flow regions. Specifically, the decay of the rotational momentum of the vortex core is clearly depicted by the monotonic decrease of the orange TAC. In contrast, the traditional iso-contouring (or iso-surfacing in the space-time domain) has a difficult time depicting this configuration. For instance, Figure 1 (c) shows two iso-surfaces computed with two different Q values. Due to the decrease of the Q value from left to right in space, the selected thresholds may not lead to iso-surfaces that depict the behavior of vortices in the far right of the flow, whose Q values may be similar to other regions without a vortex.

To study the detailed behavior of vortices, we sample 5 pathlines along a line passing through the center of a vortex (Figure 1 (a-b)). Clearly, we see three different types of TAC behaviors: (1) the decaying of the Q value over time along the core (the red TAC/pathline); (2) the increasing and shifting of the peak Q values of the TACs corresponding to the pathlines seeded at locations gradually moving away from the vortex core (blue, green, and orange TACs/pathlines); and (3) a flat TAC/pathline (purple). Among these TACs, the behavior of the TACs in group (2) is interesting. On the one hand, one can see the correlation of the peak locations of these TACs with the changing direction (or turning) of their respective pathlines. On the other hand, the shifting of the peaks indicates the propagation of the rotational momentum outwardly from the vortex core. The increase from negative Q values to positive Q values for the green and orange TACs also associate the shearing layer ($Q < 0$) with its corresponding vortex region. Such an integrated exploration cannot be easily achieved with other methods alone.

2D vortex ring. The second 2D data set simulates a vortex ring hitting a wall with a Reynolds number of 2000. During the interaction, the vortex ring approaches the wall and causes a boundary layer to appear. As the vortex slides against the wall, the boundary layer becomes unstable and is lifted up as a secondary vortex, which in turns lifts up the primary vortex. Our temporal segmentation

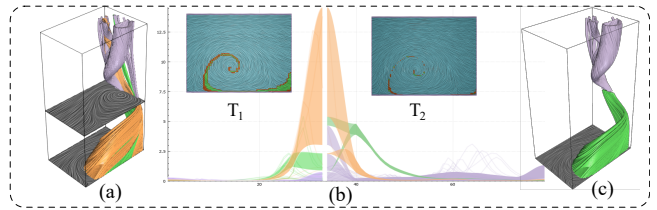


Figure 5: Temporal clustering of the vortex ring ($p_c = 1$) using the *vorticity* attribute. (a) Pathline clusters in two time intervals. (b) TAC clusters. The two insets show the spatial segmentation at two representative times within the two time intervals. (c) One group of pathlines (green) involved in two vortices (purple) after hitting the wall.

Table 1: Performance of AHC clustering on four datasets

Simulations	The Number of TAC	Time steps	Running Time
Flow behind Cylinder 2D [26]	20000	500	48.6s
Vortex Ring [15]	16384	80	12.4s
Flow behind Cylinder 3D [24]	30720	102	30.5s

results using the *vorticity* attribute are shown in Figure 5. It shows that our method can detect the moment when the vortex impacts the wall, and automatically generates a temporal cut at that time. The clustering results reveal a small curly region (colored in orange) whose attributes along the pathlines are large at the temporal cut. These pathlines represent particles that either enter one of the two vortices or move to the outermost layer and exhibit a decrease in velocity to zero.

3D flow behind a cylinder. Figure 4 shows the clustering results using TACs of attribute Q for the 3D flow behind a cylinder [24]. Considering the transitional nature of this flow, we select a seeding plane near the left boundary (i.e., $X = -11$) with 64×48 uniform samples. Pathlines that leave the domain earlier are discarded. The result with our method (Figure 4 (a)) reveals a temporal shifting in the Q profiles of the two groups of pathlines that possess similar geometry characteristics, which is not known before. The detailed view of the particle advection over time in (a.2) indicates that the two groups of particles travel in different speed, which explains the shifting. Figure 4 (b) show the temporal behavior of pathlines seeded at the shearing layer ($Q < 0$) of the flow.

Clustering performance: All numerical experiments are carried out on a PC with an Intel Core i7-3537U CPU and 128GB RAM with a NVIDIA Quadro 4000 graphic card. The most time consuming task in the system is the AHC clustering. The detailed average running time of AHC clustering on three unsteady flow simulations is reported in Table 1.

5 CONCLUSION

We propose an interactive visualization framework for the analysis and exploration of unsteady flow based on TACs measured along individual pathlines. To compare TACs behaviors for clustering, we introduce the Event-based Distance metric (EDM) to calculate the dissimilarity of TACs based on their events. This distance metric enables the development of a hierarchical, spatio-temporal clustering of pathlines based on their TAC behaviors. We integrate this clustering framework along with an improved edge-bundling visualization into an interactive visual exploration system to aid the understanding of unsteady flows. Our framework has been evaluated on multiple unsteady flow simulations, and helps domain experts analyze vortex structure.

There are a few limitations of this work. First, our temporal cut estimation strategy applies the same cut for all TACs, which may not be ideal. Second, we have yet to study the dependency of the temporal clustering on the quality of the global spatial clustering, and how the noise in the data affects the clustering quality. Finally, our AHC computation is not optimal and cannot handle large datasets. In the future, we plan to address these limitations.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable feedback. This research was supported by NSF IIS 1553329.

REFERENCES

- [1] M. K. B. Walter, K. Bala and K. Pingali. Fast Agglomerative Clustering for Rendering. *IEEE Symposium on Interactive Ray Tracing*, 2008.
- [2] M. Berenkoub, R. O. Monico, R. S. Laramée, and G. Chen. Visual analysis of spatio-temporal relations of pairwise attributes in unsteady flow. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1246–1256, 2019.
- [3] G. Chen, K. Mischaikow, R. S. Laramée, P. Pilarczyk, and E. Zhang. Vector field editing and periodic orbit extraction using Morse decomposition. *IEEE Transactions on Visualization and Computer Graphics*, 13(4):769–785, Jul./Aug. 2007.
- [4] D. Defays. An efficient algorithm for a complete link method. *The Computer Journal*, 20(4):364–366, 1977. doi: 10.1093/comjnl/20.4.364
- [5] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical morse complexes for piecewise linear 2-manifolds. In *SCG '01: Proceedings of the Seventeenth annual Symposium on Computational Geometry*, pp. 70–79. ACM, New York, NY, USA, 2001. doi: 10.1145/378583.378626
- [6] M. Edmunds, R. S. Laramée, G. Chen, N. Max, E. Zhang, and C. Ware. Surface-based flow visualization. *Computers & Graphics*, 36(8):974–990, 2012.
- [7] Z. Fang, T. Möller, G. Hamarneh, and A. Celler. Visualization and exploration of time-varying medical image data sets. In *Proceedings of Graphics Interface 2007, GI '07*, pp. 281–288. ACM, New York, NY, USA, 2007. doi: 10.1145/1268517.1268563
- [8] T. Gunther and H. Theisel. Hyper-objective vortices. *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [9] H. Guo, R. Renaut, K. Chen, and E. Reiman. Clustering huge data sets for parametric pet imaging. *BioSystems*, 71:8192, 2003.
- [10] J. L. Helman and L. Hesselink. Representation and display of vector field topology in fluid flow data sets. *IEEE Computer*, 22(8):27–36, August 1989.
- [11] R. Laramée, H. Hauser, L. Zhao, and F. H. Post. Topology based flow visualization: the state of the art. In *Topology-Based Methods in Visualization (Proceedings of Topo-in-Vis 2005)*, Mathematics and Visualization, pp. 1–19. Springer, 2007.
- [12] T.-Y. Lee and H.-W. Shen. Visualizing time-varying features with tac-based distance fields. In *Visualization Symposium, 2009. PacificVis '09. IEEE Pacific*, pp. 1–8. IEEE, 2009.
- [13] J. G. Leishman. *Principles of Helicopter Aerodynamics*. Cambridge University Press, New York, NY, USA, December 2016.
- [14] T. McLoughlin, R. S. Laramée, R. Peikert, F. H. Post, and M. Chen. Over two decades of integration-based, geometric flow visualization. In *Computer Graphics Forum*, vol. 29, pp. 1807–1829. Wiley Online Library, 2010.
- [15] P. Orlandi and R. Verzicco. Vortex rings impinging on walls: axisymmetric and three-dimensional simulations. *Journal of Fluid Mechanics*, 256:615646, 1993. doi: 10.1017/S0022112093002903
- [16] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. P. Seidel, and T. Weinkauff. An edge-bundling layout for interactive parallel coordinates. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific*, pp. 57–64. IEEE, 2014.
- [17] A. Pobitzer, R. Peikert, R. Fuchs, B. Schindler, A. Kuhn, H. Theisel, K. Matkovic, and H. Hauser. The state of the art in topology-based visualization of unsteady flow. *Computer Graphics Forum*, 30(6):1789–1811, September 2011.
- [18] K. Polthier and E. Preuß. Identifying vector fields singularities using a discrete hodge decomposition. In *Mathematical Visualization III*, pp. 112–134. Ed: H.C. Hege, K. Polthier, 2003.
- [19] T. Salzbrunn, T. Wischgoll, H. Jänicke, and G. Scheuermann. The state of the art in flow visualization: Partition-based techniques. In H. Hauser, S. Strassburger, and H. Theisel, eds., *In Simulation and Visualization 2008 Proceedings*, pp. 75–92. SCS Publishing House, 2008.
- [20] K. Shi, H. Theisel, H. Hauser, T. Weinkauff, K. Matkovic, H.-C. Hege, and H.-P. Seidel. Path line attributes - an information visualization approach to analyzing the dynamic behavior of 3D time-dependent flow fields. In H.-C. Hege, K. Polthier, and G. Scheuermann, eds., *Topology-Based Methods in Visualization II*, Mathematics and Visualization, pp. 75–88. Springer, Grimma, Germany, 2009. doi: 10.1007/978-3-540-88606-8_6
- [21] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.
- [22] H. Theisel, T. Weinkauff, and H.-P. Seidel. Grid-independent detection of closed stream lines in 2D vector fields. In *Proceedings of the Conference on Vision, Modeling and Visualization 2004 (VMV 04)*, pp. 421–428, Nov. 2004.
- [23] X. Tricoche, G. Scheuermann, and H. Hagen. Continuous topology simplification of planar vector fields. In *Proceedings of IEEE Visualization 2001*, pp. 159–166, 2001.
- [24] W. von Funck, T. Weinkauff, H. Theisel, and H.-P. Seidel. Smoke surfaces: An interactive flow visualization technique inspired by real-world flow experiments. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1396–1403, November - December 2008.
- [25] C. Wang, H. Yu, and K.-L. Ma. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1547–1554, 2008.
- [26] T. Weinkauff and H. Theisel. Streak lines as tangent curves of a derived vector field. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization 2010)*, 16(6):1225–1234, November - December 2010.
- [27] T. Wischgoll and G. Scheuermann. Detection and visualization of closed streamlines in planar fields. *IEEE Transactions on Visualization and Computer Graphics*, 7(2):165–172, 2001.
- [28] K.-P. Wong, D. Feng, S. Meikle, and M. Fulham. Segmentation of dynamic pet images using cluster analysis. *IEEE Nuclear Science Symposium*, 3:18/126 – 18/130, 200.
- [29] J. Woodring and H.-W. Shen. Multiscale time activity data exploration via temporal clustering visualization spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 15(1):123–137, 2009.
- [30] J. Woodring and H.-W. Shen. Semi-automatic time-series transfer functions via temporal clustering and sequencing. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization, EuroVis'09*, pp. 791–798. The Eurographs Association & John Wiley & Sons, Ltd., Chichester, UK, 2009. doi: 10.1111/j.1467-8659.2009.01472.x
- [31] L. Zhang, R. S. Laramée, D. Thompson, A. Sescu, and G. Chen. Compute and visualize discontinuity among neighboring integral curves of 2d vector fields. In *Topological Methods in Data Analysis and Visualization*, pp. 187–203. Springer, 2015.
- [32] L. Zhang, D. Nguyen, D. Thompson, R. S. Laramée, and G. Chen. Enhanced Vector Field Visualization via Lagrangian Accumulation. *Computer and Graphics (special issue of CAD/Graphics)*, pp. 224–234, 2017.