Visualizing Statistical Complexity in 3D Turbulent Flows using a Robust Entropy Calculation Method

Arne Präger
Leipzig University
Augustusplatz 10
04109, Leipzig, Germany
praeger@informatik.uni-leipzig.de

Baldwin Nsonga
Leipzig University
Augustusplatz 10
04109, Leipzig, Germany
nsonga@informatik.uni-leipzig.de

Gerik Scheuermann
Leipzig University
Augustusplatz 10
04109, Leipzig, Germany
scheuermann@informatik.uni-leipzig.de

ABSTRACT

Highly resolved flow simulation data is becoming more common. These simulations frequently feature high turbulence with complex flow patterns. Finding these regions often requires expert knowledge, and in more complex cases, flow patterns of interest may remain hidden. The concept of statistical complexity was shown to be suitable to indicate regions of interest within flow data with limited prior knowledge. One way to determine the statistical complexity of flow fields is via the local vector field entropy and the correlation. In this work, we improve the method for calculating the Shannon entropy in vector fields. To this end, we introduce a robust entropy computation that takes the scale of the corresponding regions into account. The improved method uses a novel way to determine the distributions required for the entropy calculation and is applicable to unstructured domains. We validate our method with analytic flow fields and apply it to fluid simulation data, visualizing the results via volume rendering. This work shows the applicability of our technique to highlight regions of interest in turbulent flows.

Keywords
flow visualization, data exploration, information theory, statistical complexity, turbulence visualization

1 INTRODUCTION

Three-dimensional flow simulations have become increasingly complex [TLMF21]. Feature detection methods (e.g., vortex identification) and 2D cuts with color mappings of physical properties, e.g., velocity or pressure, are most common for visual analysis [BCP+12, ZH15]. Even though these visualizations are helpful, they require expert knowledge of specific configurations to understand which flow features in which regions are of interest and how to detect them with appropriate parameters.

Our goal is to provide a tool to assist the researcher in identifying regions of interest in individual time steps, which the researcher can further investigate purposefully. To this end, we apply the concept of statistical complexity based on the Shannon entropy [NT12] of vector directions which Xu et al. [XLS10] successfully applied to flow fields. Even though the Shannon entropy could be applied directly, Arbona et al. [ABMP14] argue that complete chaos and complete order should have similar value to the analysis. In contrast, entropy exhibits minimal values for complete order and maximum values for complete chaos.

Our primary focus is improving on the Shannon entropy computation. After providing a background on statistical complexity and Shannon entropy, we discuss the interpretation of statistical complexity in flow fields in general and provide arguments for the usefulness of this type of analysis for turbulent flows. We then discuss the methodology which distinguishes itself from the common approach by: (I) applying a neighborhood scheme suitable for unstructured domains, (II) utilizing the concept of information gain to set neighborhood sizes for the entropy computation dynamically, and (III) introducing a binning scheme to reduce overestimated complexity values. We validate our method with analytic datasets and apply it to a Direct Numerical Simulation (DNS) dataset. The results are visualized via volume rendering.

2 RELATED WORK

In the context of environmental sciences, a recent survey conducted by Bujack and Middel [BM20] gives a broad overview of various flow visualization techniques.

Geometric approaches based on the visualization of integral structures, e.g., streamlines or path-

https://www.doi.org/10.24132/CSRN.3201.5

lines, were discussed by McLoughlin [MLP+09]. Feature-based visualization approaches aim to detect and represent specific flow patterns like vortices [Hal05, JWM88, JH95], shock waves [WXWH13] ,or splat events [NNF+20, NSG+20].

Methods based on vector field topology visualize the topological skeleton consisting of critical points and segment the flow domain into regions of similar behavior [HH89, HH91, LHZP07]. Lagrangian coherent structures (LCS) can be interpreted as an unsteady analogy to flow topology and were extensively studied [Hal15, HS11, HY00].

As flow visualizations often depend on the chosen reference frame, a body of research concerned with this issue was published in recent years [BPKB14, GGT17, RG20, WGS05, WGS07].

Statistical complexity was introduced by Crutchfield and Young [CY89] and later refined by Lopez-Ruiz et al. [LRMC95] who separated the influences of entropy and structure.

Jaenicke et al. [JWSK07] have shown the usefulness of visualizing the statistical complexity computed via cellular automata in multifield data. They have also proven that statistical complexity can automatically detect flow features [JBTS08, JS10]. Arbona et al. [ABMP14] propose a method to apply the concept of statistical complexity to 3D vector data based on the product of Shannon entropy and local correlation.

Xu et al. [XLS10] applied the Shannon entropy to vector fields. They base the entropy at a point on the variance of the orientations of the vectors located in its neighborhood. Xu et al. [XLS10] did not directly visualize this information but used it to seed streamlines in information-rich regions instead. Wang et al. [WYM08] determine the entropy of dataset chunks to customize the level of detail used for their visualization. Furuya and Itoh [FI08], as well as Lee et al. [LMSC11], applied information-theoretic techniques to select a small number of streamlines out of a set optimally representing the original data. Tao et al. [TMWS12] and Ma et al. [MWW+14] used entropy to create automated camera cruises through possibly significant flow regions based on their entropy. An overview of how information-theoretic techniques can be utilized for visualization can be found in the book by Wang and Chen [CFV+16].

3 BACKGROUND

This section provides a background on entropy, statistical complexity, and how these concepts can be applied to flow fields.

3.1 Entropy of Velocity Fields

The Shannon entropy [NT12] describes the information content of a single variable. Its calculation is based on the distribution of the values of said variable. For any given variable X with the possible outcomes $\{x_1,..,x_n\}$ the Shannon entropy is:

$$H(X) = -\sum_{i=1}^{N} p(x_i)\log_2(p(x_i)), \tag{1}$$

where $p(x_i)$ denotes the probability with which the variable X takes the value x_i. $H(X)$ becomes minimal if the variable always assumes the same value and becomes maximal when all potential outcomes share the same possibility.

Xu et al. [XLS10] adapted the concept of entropy for vector fields. The entropy of single points is based on the variance of the orientations of vectors in their neighborhood. They do not specify in detail how a neighborhood is defined. As they use a manually set natural number as a parameter, we assume neighborhoods are grid-based. To generate the required distribution, they use vector orientations. They discretize a sphere and evaluate which vector orientation corresponds to which segment of the sphere. Note that the segments correspond to the bins of a histogram. To avoid bias, they utilize the technique proposed by Leopardi [Leo06] dividing the sphere surface into equally sized regions of similar shape. This technique divides the sphere into almost circular top and bottom areas as well as multiple rings with several quadrangular segments.

3.2 Statistical Complexity of Velocity Field

Arbona et al. [ABMP14] expanded upon the concept of vector field entropy with their definition of statistical complexity by taking local correlation into account. In their study, Arbona et al. [ABMP14] aim to determine the complexity at multiple scales, which they accomplish by utilizing meshes of varying sizes for the computation of the complexity field. They then compute the correlation between the mean velocity within a cell and its neighboring cells. The statistical complexity is then the product of the correlation and the entropy acquired using the method from Xu et al. [XLS10]. Their results show that complex structures become more visible when visualizing statistical complexity instead of the vector field entropy. This work and the velocity vector entropy computation by Xu et al. [XLS10] are the foundations of our research.

3.3 Information in Turbulent Flows

Visualizing entropy and statistical complexity gives insights into a flow with minimal prior knowledge. Xu et al. [XLS10] state that the entropy of a vector field should be suitable to identify critical points in non-converging flows. They show that it can also emphasize structures like regions near separation lines. This property is a result of the inherent nature of entropy. Low entropy values represent that the orientations of vectors in a neighborhood only vary slightly. Thus low entropy regions depict the more laminar parts of a flow. With rising variations in the flow directions, the corresponding entropy values climb as well. Turbulent flow fields, expected to result in high entropy values, exhibit structure depending on the chosen configuration and reference frame. High entropy values, emphasizing regions of high disorder, are not sufficient to visualize potential regions of interest, as highly turbulent regions will always be favored. On the other hand, statistical complexity gives more nuance to the visualization, reducing the impact of the pure disorder. As a result, using statistical complexity, we aim to visualize regions containing critical points and other complex flow patterns like coherent structures.

4 METHOD

We improve the computation of the entropy [XLS10] by making it more robust and applicable to irregular grids as well as utilizing dynamically computed neighborhood sizes. We adapt the method from Arbona et al. [ABMP14] from a cell-based to a point-based approach.

Our method takes a velocity field $\mathbf{v}(\mathbf{x},t)$ as input and computes the statistical complexity:

$$C(\mathbf{x},t) = H(\mathbf{x},t)D(\mathbf{x},t), \qquad (2)$$

where $H(\mathbf{x},t)$ is the entropy and $D(\mathbf{x},t)$ is the velocity correlation at a point $\mathbf{x} \in G$ in the flow domain $G \subset \mathbb{R}^3$ at time $t \in \mathbb{R}$.

4.1 Neighborhood Definition

The computation of $H(\mathbf{x},t)$ and $D(\mathbf{x},t)$ requires a neighborhood definition. As stated before, we assume that the neighborhoods applied by Xu et al. [XLS10] depend on the grid of the flow domain. Neighborhoods based on grids can introduce a directional bias when the spacings between points are not equal along the axes. Furthermore a grid based approach may not suitable for non-uniform grids.

Our goal is to find a neighborhood definition free of directional bias which is applicable to any grid type. The neighbors should preferably be distributed equally in space so that the results of the entropy calculation are uninfluenced by the orientation of the field. In the following when we mention dodecahedra we always refer to convex, regular deodecahedra.

We base our neighborhood definition on dodecahedron vertices as they are equally distributed on the surface of a sphere and thus share the same distance to the point in its center. Even though all platonic solids share this property, we use the dodecahedron as it has the most vertices of them, leading to a fine resolution of the neighborhood. We define the neighborhood $n_d(\mathbf{x}) \subset G$ as the union of the vertices of d equidistant layers of dodecahedra (cf. Figure 1). The faces of neighboring dodecahedra layers are parallel. Note that, the number of points contained in the neighborhood is $|n_d(\mathbf{x})| = 20d$, as a single dodecahedron contains 20 vertices. We obtain a neighborhood consisting of points with uniform radial and angular spacing.

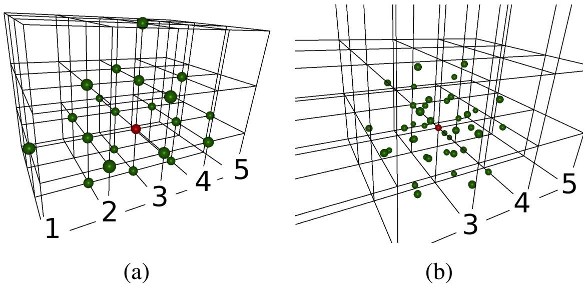


Figure 1: Neighborhoods of size two in green around the red point for (a) the old and (b) the new neighborhood definition

The neighborhood vectors used for the computation are then acquired via interpolation. To prevent aliasing due to undersampling, we respected the Nyquist frequency and chose the distance between two neighboring dodecahedra layers to be 0.4 of the minimal cell edge distance within the domain. Note that the density of neighborhood points decreases as the distance to $\mathbf{x}$ increases. This is justifiable as it creates differences between points in small areas of turbulence surrounded by a laminar field and points in small laminar areas surrounded by a turbulent field. The first case should result in higher entropy values which our method captures correctly.

4.2 Binning Strategy

As mentioned before Xu et al. [XLS10] segment a sphere into equal-sized areas using Leopardis algorithm [Leo06]. This introduces a bias, as flows in the direction of the poles are more robust to small variations in orientation than their orthogonal counterparts. Binning, in general, can lead to vectors with small variances in orientation being assigned to different bins. Especially in laminar flows, small variation can lead to an overestimated statistical complexity. An example of the artifacts resulting from this is shown in Section 5.1.

To avoid bias introduced through the shape of the segment bins, we used the triangular faces of an icosahedron as the base for our segmentation, as they all share the same shape and size. Each of these faces can be subdivided into four equilateral triangles of the same area. The user can set the number of subdivisions $s \in \mathbb{N}$. Using this subdivision strategy, we can provide different consistent binning resolutions as shown in Figure 2.

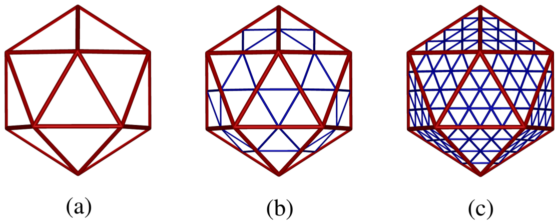


Figure 2: Icosahedral binning scheme using no (a), one (b) and two (c) subdivisions

To solve the issue of overestimated complexity values for small orientation variances, we apply a randomized rotation to the binning icosahedron. We then repeat the entropy calculation multiple times with different rotation angles. The number of iterations k can be set by the users. As the entropy of the velocity direction should be invariant to rotation, we can assume that high values result from overestimation. Therefore, after k iterations, we store the minimal value. The quality of the results increases in proportion to the chosen number k as artifacts resulting from small variances are not robust against variations of the binning. Note that the random rotations introduce noise. To reduce the noise without further increasing k, we repeat the same number of iterations for points with a higher statistical complexity value than all their edge neighbors. The effects of our binning technique are discussed in Section 5.

4.3 Dynamic Neighborhoods

Choosing appropriate neighborhoods is challenging, as two different regions might require a different number of dodecahedron layers d to highlight regions of interest of varying scales. To solve this, we introduce a local dynamic neighborhood size utilizing relative entropy. The general concept is that we only need to add additional layers to the neighborhood if their vectors contain new information. Beginning with the second layer, we calculate the relative entropy [KL51] between the current orientation distribution of the neighborhood and the distribution containing the vectors of the next layer. We define the distributions as follows.

For each triangular bin, we compute the number of neighbors $|\hat{n}_d(\mathbf{x},t)|$ where the corresponding (scaled) vectors intersect the triangle before and after adding another dodecahedron layer. This is divided by the number of all neighbors $|n_d(\mathbf{x},t)|$, resulting in:

$$p_i(\mathbf{x},t) = \frac{|\hat{n}_d(\mathbf{x},t)|}{|n_d(\mathbf{x},t)|} \quad p_i^+(\mathbf{x},t) = \frac{|\hat{n}_{d+1}(\mathbf{x},t)|}{|n_{d+1}(\mathbf{x},t)|}, \tag{3}$$

where $p_i(\mathbf{x},t)$ is the estimate of the probability for neighbors of point $\mathbf{x}$ belonging to bin i. $p_i^+(\mathbf{x},t)$ denotes the estimate of the probability after adding the additional dodecahedron layer. As $p_i^+(\mathbf{x},t) = 0$ implies $p_i(\mathbf{x},t) = 0$, absolute continuity is provided, enabling us to apply relative entropy:

$$KL(\mathbf{x},t) = \sum_{i=1}^{N} p_i(\mathbf{x},t)\log_2 \frac{p_i(\mathbf{x},t)}{p_i^+(\mathbf{x},t)}, \tag{4}$$

where $N = 20 * 4^s$ is the number of bins resulting from s subdivisions of the 20 icosahedron faces (cf. Figure 2). High relative entropy values indicate more significant differences between the two distributions. If $KL(\mathbf{x},t)$ is higher than a user defined threshold r, we repeat the process by setting $p_i(\mathbf{x},t)$ to $p_i^+(\mathbf{x},t)$ and adding the next layer.

4.4 Entropy and Correlation

We can now compute the entropy as follows:

$$H(\mathbf{x},t) = -\sum_{i=1}^{N} p_i(\mathbf{x},t)\log_2 p_i(\mathbf{x},t), \tag{5}$$

where $p_i(\mathbf{x},t)$ is the distribution after the relative entropy has fallen below the threshold r and $N = 20 * 4^s$ is the number of bins. The correlation is computed by adapting the cell-based strategy from Arbona [ABMP14] to our point based approach:

$$D(\mathbf{x},t) = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbf{v}(\mathbf{x},t)\mathbf{v}(\mathbf{x}_i,t)}{\mathbf{v}(\mathbf{x},t)^2 + \mathbf{v}(\mathbf{x}_i,t)^2} + \frac{1}{2}, \tag{6}$$

where $\mathbf{x}_i \in n(\mathbf{x},t)$ is the i^{th} neighbor vertex position in the neighborhood of point $\mathbf{x}$. Note that close to boundaries, neighborhood points can be out of bounds and have to be excluded from the calculation. To prevent the correlation values from being influenced by a varying number of neighborhood points, we set N to $20d$.

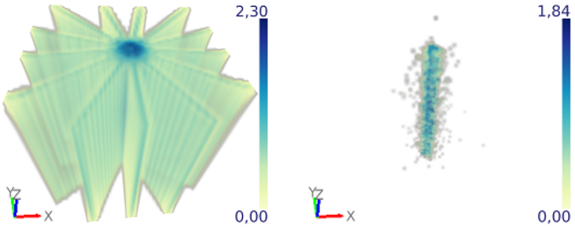
We can now compute the statistical complexity using Equation 2. Note that the results of the presented method are invariant to the magnitude of the velocity vectors.

4.5 Visualization

Developing a novel and suitable visualization for the regions of interest is not within the scope of this work. As the statistical complexity $C(\mathbf{x},t)$ is a scalar field, several established visualization techniques are available. For an in-depth interactive analysis of the results, we apply volume rendering [DCH88]. This is a GPU-based ray casting technique that renders volumes with different transparencies based on their scalar values. We explore the range of the results and emphasize values exhibiting structure. It is to note that the utilized colormap can be adjusted in real-time. An example of the resulting visualization is shown in Figure 3.

5 EVALUATION

In this section, we present the results of our evaluation. First, we show that our method produces the expected



(a) Previous binning approach (b) New binning approach

Figure 3: A comparison of the statistical complexities of $\mathbf{v}_{\mathrm{rot}}(\mathbf{x})$ for $r = 0.01$, $s = 1$ and $k = 5$

results when applying it to analytic test cases. Then we conduct a parameter study to provide suggestions for finding suitable input parameters. Finally, we investigate the performance of our method.

5.1 Application to Analytic Fields

To evaluate our method, we conducted experiments on analytical 3D data sets. We construct the steady fields:

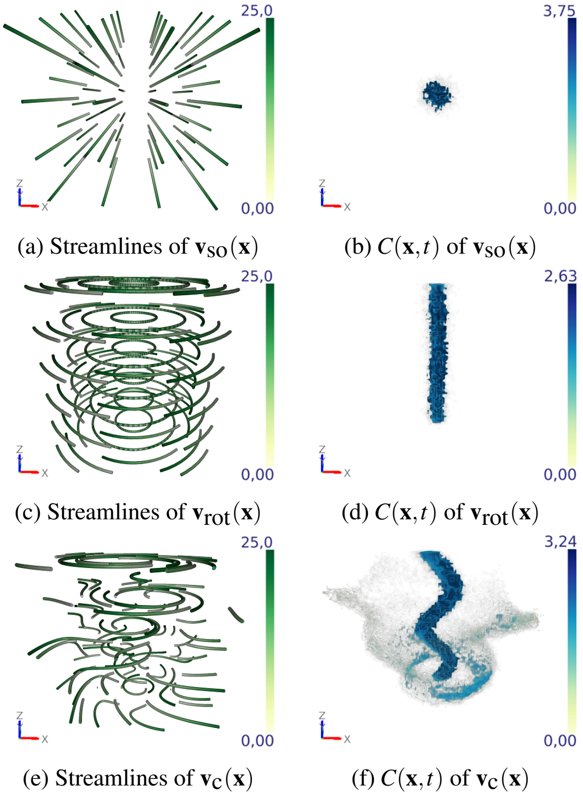
$$\mathbf{v}_{\mathrm{so}}(\mathbf{x}) = [x,y,z]^T \quad \mathbf{v}_{\mathrm{rot}}(\mathbf{x}) = [y,-x,0]^T, \tag{7}$$

where $\mathbf{v}_{\mathrm{so}}(\mathbf{x})$ is a source and $\mathbf{v}_{\mathrm{rot}}(\mathbf{x})$ is a rotation. Additionally we constructed a steady Crawfis tornado [Cra03] we denote as $\mathbf{v}_{\mathrm{C}}(\mathbf{x})$. The values were then sampled on a $50 \times 50 \times 50$ grid within the domain $[-24.5, 24.5] \times [-24.5, 24.5] \times [-24.5, 24.5]$.

Figure 3 shows a comparison between the classical binning method and our binning scheme applied to $\mathbf{v}_{\mathrm{rot}}(\mathbf{x})$. The left image demonstrates that the classic binning can produce artifacts. This is consistent with the results shown in Figure 3 in [XLS10]. These artifacts, however, are not robust against the binning scheme and are reduced. Note that, as a downside, noise is introduced. This is a clear improvement, as the noise in the right figure is clearly identifiable, whereas the artifacts in the left figure could imply structure in an unknown dataset. We discuss the effect of the number of iterations in the following section.

Figure 4 shows volume rendering visualizations of the results of our statistical complexity approach in comparison to streamlines seeded uniformly in the field. As expected, the center of the source field and the rotation center of $\mathbf{v}_{\mathrm{rot}}(\mathbf{x})$ are the regions with the highest statistical complexity. The rotation center of the Crawfis tornado is also clearly visible. Note that the mantle-like surface around the tornado is inherent to the dataset and also appears when applying vortex identification methods.

Through these experiments, it becomes clear that visualizing the statistical complexity is a useful method to highlight critical points, vortex cores, and other potential regions of interest.



(a) Streamlines of $\mathbf{v}_{\mathrm{so}}(\mathbf{x})$ (b) $C(\mathbf{x},t)$ of $\mathbf{v}_{\mathrm{so}}(\mathbf{x})$

(c) Streamlines of $\mathbf{v}_{\mathrm{rot}}(\mathbf{x})$ (d) $C(\mathbf{x},t)$ of $\mathbf{v}_{\mathrm{rot}}(\mathbf{x})$

(e) Streamlines of $\mathbf{v}_{\mathrm{C}}(\mathbf{x})$ (f) $C(\mathbf{x},t)$ of $\mathbf{v}_{\mathrm{C}}(\mathbf{x})$

Figure 4: A comparison of the visualization of streamlines and the statistical complexities $C(\mathbf{x},t)$ of analytic datasets for parameters $r = 0.01$, $s = 1$ and $k = 5$

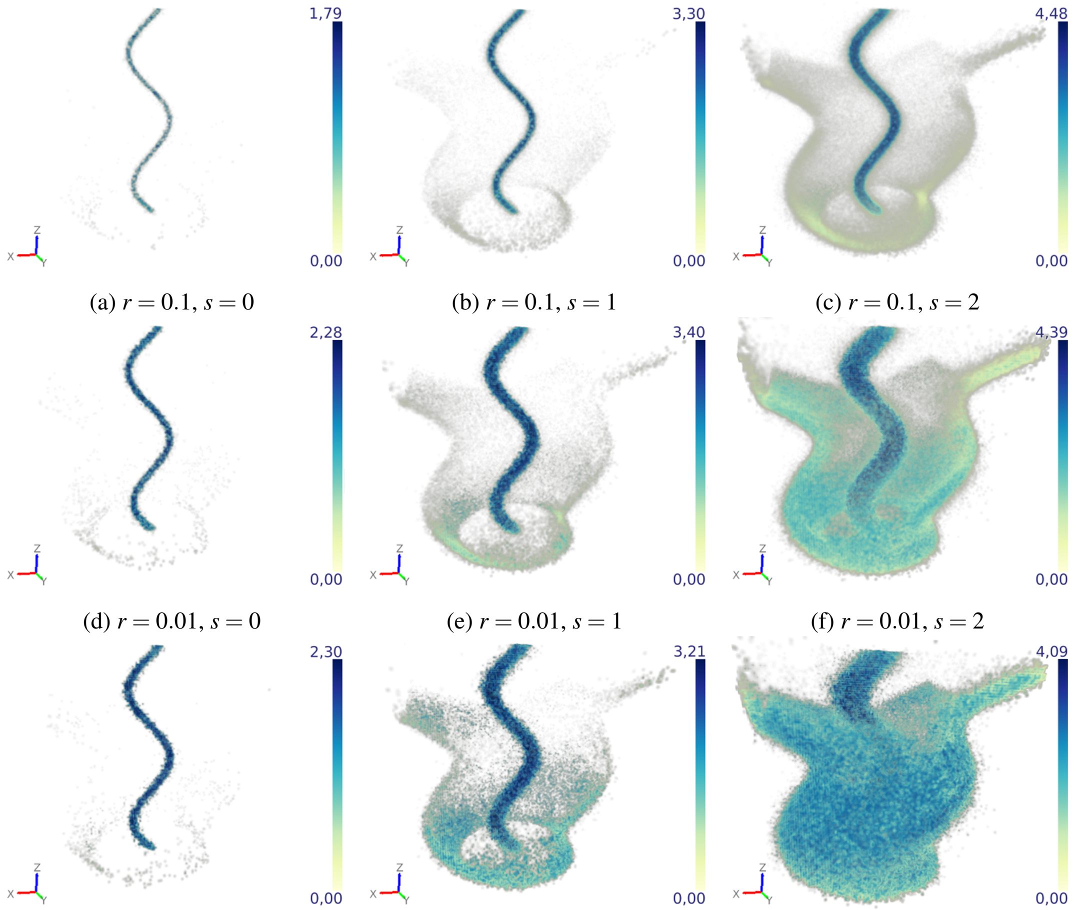
5.2 Parameter Study

5.2.1 *Information Gain and Bin Resolution*

The most impacting parameters in our method are the relative entropy threshold utilized during the dynamic neighborhood size computation and the number of subdivisions of the icosahedron faces used for binning. In this experiment, we compared the results of our algorithm for the Crawfis tornado defined on a $100 \times 100 \times 100$ grid for combinations of these parameters. We apply five iterations ($k = 5$) to provide a robust visualization. For this analysis, we base the evaluation on the quality on the following criteria: (1) the highest statistical complexity values lie exclusively in the vortex core line, and (2) medium values should be located around the vortex core and on the mantle shaped surface (cf. Figure 4f).

The results in Table 1 show that a relative entropy threshold of $r = 0.1$ is too high since the information gained through adding a new layer is rarely enough to surpass it. This leads to small neighborhoods which may not be able to capture the behavior of the flow correctly (cf. Table 1). Figure 5 (first row) shows that no clear distinction between regions with different statistical complexities is possible and our criteria are not met.

A relative entropy threshold of 0.01 delivers the desired results as it allows the neighborhoods to grow to viable



(a) $r = 0.1$, $s = 0$ (b) $r = 0.1$, $s = 1$ (c) $r = 0.1$, $s = 2$

(d) $r = 0.01$, $s = 0$ (e) $r = 0.01$, $s = 1$ (f) $r = 0.01$, $s = 2$

(g) $r = 0.001$, $s = 0$ (h) $r = 0.001$, $s = 1$ (i) $r = 0.001$, $s = 2$

Figure 5: Visualization of the statistical complexity on a $100 \times 100 \times 100$ Crawfis tornado dataset with $k = 5$ and varying r and s

$[d_{min}, d_{max}]$	$s = 0$	$s = 1$	$s = 2$
$r = 0.1$	$[2,4]$	$[2,6]$	$[2,9]$
$r = 0.01$	$[2,13]$	$[3,21]$	$[2,30]$
$r = 0.001$	$[2,34]$	$[2,52]$	$[2,88]$

Table 1: Neighborhood size d intervals resulting from specific subdivision numbers s and information gain thresholds r on the Crawfis tornado dataset (the table corresponds to Figure 5)

sizes. Subdivision of $s = 1,2$ seems sufficient as the difference between areas of different statistical complexity values becomes clearly visible.

For a threshold of 0.001 the neighborhood sizes become larger, accentuating large-scale features. Increasing the subdivision intensifies the phenomenon further (cf. Table 1).

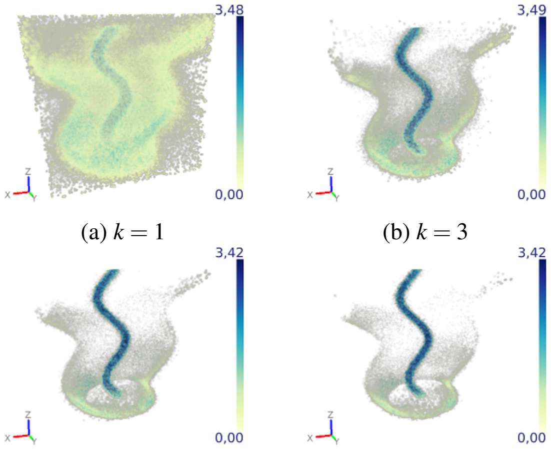
As a strategy for parameter selection, we recommend exploring the parameter space of the threshold r, starting with small values, e.g., $r \approx 0.001$. Choosing a small initial value reduces the risk of false negatives. By increasing the threshold, structures may persist, informing the user of potential regions of interest.

Subdividing the surface of the icosahedron one and two times is both viable. While the first approach shortens the calculation time, the second one should perform better for stronger turbulences as it can capture finer details of the behavior of a flow.

5.2.2 *Number of Iterations*

We conducted another experiment to determine a recommendation for the number of iterations needed to create reliable results. We used the $100 \times 100 \times 100$ Crawfis tornado. The relative entropy threshold was set to 0.01, and we subdivided the icosahedron one time since these parameter settings have produced satisfactory results in the previous test.

Figure 6 displays the statistical complexity field after a different amount of iterations. The visual noise is drastically reduced after three iterations, but some is still present. After five iterations, most of the higher statistical complexity values are near the vortex core and the mantle surface. Increasing the number of iterations



(a) $k = 1$ (b) $k = 3$

(c) $k = 5$ (d) $k = 7$

Figure 6: Results of the parameter study on a 100 × 100 × 100 Crawfis tornado dataset with $r = 0.01$ and $s = 1$ for varying k

further still improves the quality and reliability of the results, albeit decreasingly less.

In order to evaluate the relationship between the statistical complexity and the number of iterations, we computed

$$\delta_k = \frac{1}{N}\sum_{i=1}^{N} |C_k(\mathbf{x_i},t) - C_{k+1}(\mathbf{x_i},t)|, \quad (8)$$

where N is the number of grid points and $\mathbf{x_i} \in G$ is a grid point. As depicted in Figure 7 we can observe that the δ_k values decrease and approach zero when k grows. Thus when using our method there always is a tradeoff between the runtime of the algorithm and the robustness of the results. In practical applications, five to seven iterations should be sufficient as the quality of the results only increases marginally afterward.

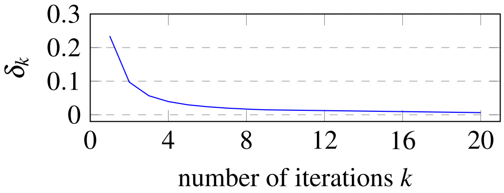


Figure 7: Relationship between δ_k and k

5.3 Performance

In this section, we measure the scalability of our method. The test was conducted on a Linux system with Ubuntu 20.04.3 LTS, 32 GB of RAM and a 32x Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz. Since the calculation of the statistical complexity is performed individually for every grid point, we can utilize OpenMP for parallelization. It is to note that our method is still not fully optimized. We tried to avoid unnecessary calculations but making the algorithm run on GPUs could introduce a drastic speedup as one of the most frequently used operations is a test for triangle intersections.

During our experiment, we measured the runtimes of our method for its application on Crawfis tornados defined on $n \times n \times n$ structured grids with $n = 50, 100, 150, 200$. We set the relative entropy threshold to 0.01, subdivided the icosahedron faces used for the binning one time, and let the algorithm run through three iterations. As shown in Figure 8 a linear increase in the number of grid points leads to a linear increase in runtime.

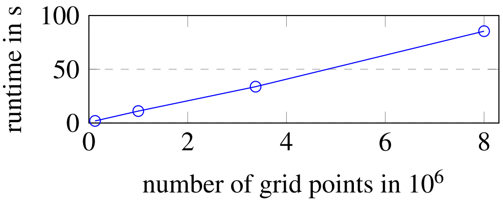


Figure 8: Relationship between runtime and grid points

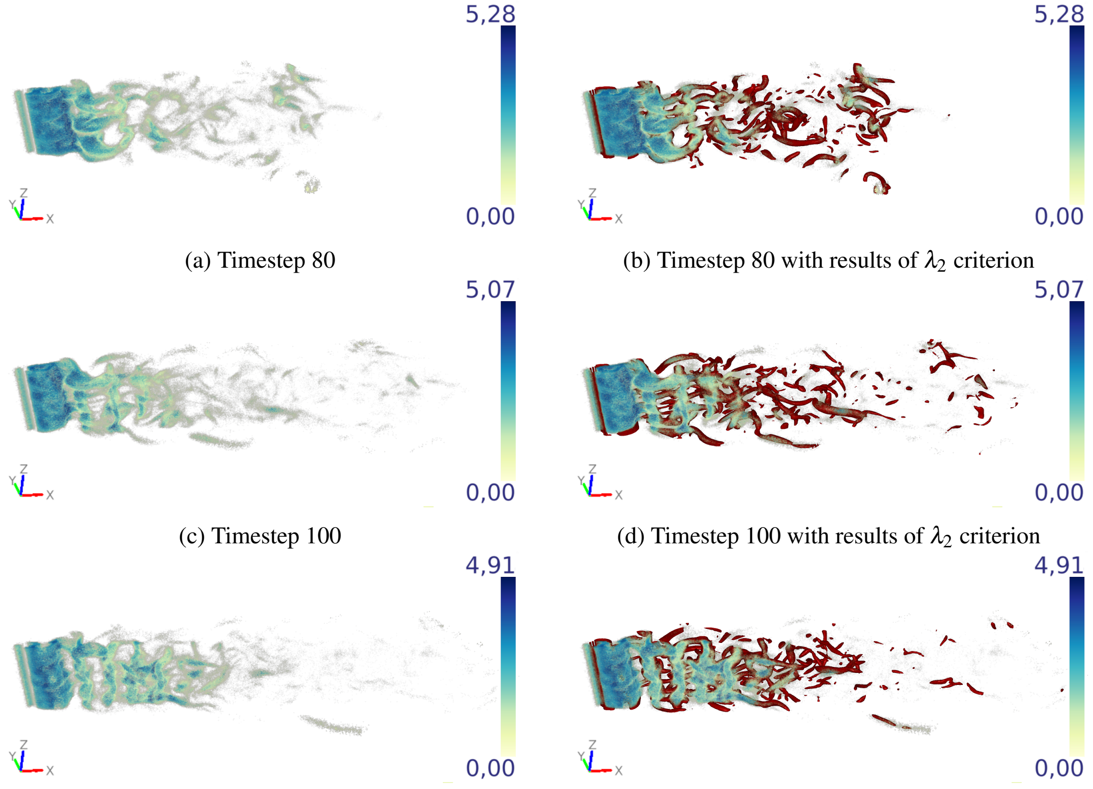
We repeated the the same tests for crawfis tornados defined on $n \times n \times n$ unstructured domains with $n = 50, 100$. The runtimes were increased by a factor of 6.32 and 7.88, respectively, compared to the structured grids.

6 APPLICATION TO VON KÁRMÁN VORTEX STREET

We applied our method to a turbulent DNS dataset of a von Kármán vortex street as it is a well-studied configuration. The dataset was simulated with the Gerris flow solver [Pop04] and made publicly available by the ETH Zürich [BRG19]. The dataset contains a constant flow that hits a half-cylinder, leading to vortex shedding. The flow is defined on a 640 × 240 × 80 grid and has a Reynolds number of 6400. The dataset contains 151 timesteps depicting the build-up of the flow. At around timestep 90, the flow starts to span the whole domain.

We computed the statistical complexities of three timesteps in the later stages of the flows development. The regions with the highest complexities can be found directly around the cylinder. Behind the cylinder, structures of high statistical complexity emerge in regular intervals on alternating sides. With advancing time, more of these structures form as the existing ones move with the general direction of the flow akin to the vortices of a von Kármán vortex street.

We also compared the results of our method with the vortices detected by the λ_2 vortex criterion. A clear relation is visible between the regions with the highest statistical complexity values and the λ_2 vortices in the first third of the field behind the cylinder. Further behind the cylinder, some vortices coincide with regions of middle to lower complexity values, whose magnitude



(a) Timestep 80

(b) Timestep 80 with results of λ_2 criterion

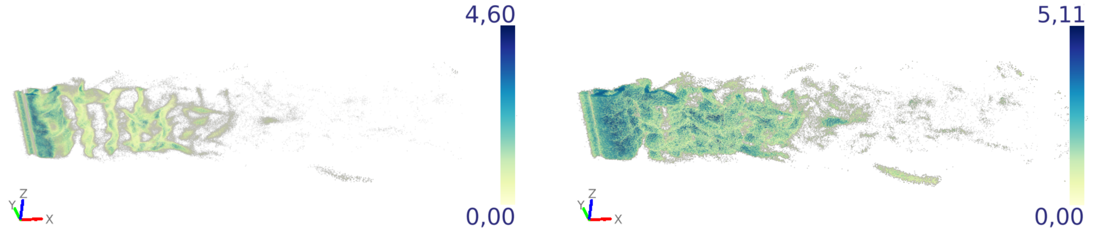
(c) Timestep 100

(d) Timestep 100 with results of λ_2 criterion

(e) Timestep 120

(f) Timestep 120 with results of λ_2 criterion

Figure 9: Visualization of the statistical complexities for $r = 0.01$, $s = 2$ and $k = 5$ of multiple timesteps of the von Kármán vortex street [BRG19, Pop04] compared with results of λ_2 vortex criterion [JH95] displayed as red isosurfaces of the value $\lambda_2 = -15$



(a) $r = 0.1$

(b) $r = 0.001$

Figure 10: Visualization of the statistical complexities for $s = 2$ and $k = 5$ of timestep 120 of the von Kármán vortex street [BRG19, Pop04] for different r

is probably diminished through them lying in areas of high turbulence. The statistical complexity, in general, behaves as expected and captures the behavior of the flow.

We also ran some test on how the statistical complexity behaves for different relative entropy thresholds r. We did not change the degree of subdivision s of the binning icosahedron as we deem 320 bins to be an appropriate number to capture the behavior of a turbulent flow. The number of algorithm iterations was set to $k = 5$ since the results of Section 5.2 indicate that increasing this number further has no significant influence on the quality of the visualization. As seen in Figure 10, changing r has the same influence on the statistical complexity values as we observed in Section 5.2. For high relative entropy thresholds r, the regions with the highest complexities shrink and finer structures are hardly visible anymore. Lower r-values lead to bigger regions of high complexity.

7 CONCLUSIONS

This paper presented an improvement of the statistical complexity calculation for three-dimensional flow fields by making the entropy computation more robust. We achieved this by presenting a new neighborhood definition applicable to any grid type, determining dynamic neighborhood sizes based on the concept of information gain, and introducing a novel binning approach to reduce overestimated complexity values and artifacts. By applying our method to analytical datasets, we evaluated our approach and showed that the statistical complexity is suited to identify critical points and other regions of interest. We have further shown that the direct visualization of the improved statistical complexity via volume rendering gives meaningful insights into the behavior of a flow for DNS data. Thus the proposed method provides an additional tool for the visual analysis of such data.

In the future, we want to apply our method to a more complex DNS dataset and evaluate the applicability of our method to the expert domain. In addition, we will develop a visualization technique suitable for the visualization of statistical complexity, taking into account the neighborhood sizes. This should enable perceiving regions of interest of different scales. Lastly, we will conduct research on the utilization of statistical complexity in different reference frames and especially investigate how to make our method translation invariant.

8 ACKNOWLEDGMENTS

This work was partially funded by the German Federal Ministry of Education and Research within the project Competence Center for Scalable Data Services and Solutions (ScaDS) Dresden/Leipzig (BMBF 01IS14014B) and by the Development Bank of Saxony (SAB) under project number 100335729.

9 REFERENCES

[ABMP14] A Arbona, C Bona, B Miñano, and A Plastino. Statistical complexity measures as telltale of relevant scales in emergent dynamics of spatial systems. *Physica A: Statistical Mechanics and its Applications*, 410:1–8, 2014.

[BCP+12] Andrea Brambilla, Robert Carnecky, Robert Peikert, Ivan Viola, and Helwig Hauser. Illustrative flow visualization: State of the art, trends and challenges. *Visibility-oriented Visualization Design for Flow Illustration*, 2012.

[BM20] Roxana Bujack and Ariane Middel. State of the art in flow visualization in the environmental sciences. *Environmental Earth Sciences*, 79(2):65, 2020.

[BPKB14] H. Bhatia, V. Pascucci, R. M. Kirby, and P.-T. Bremer. Extracting features from time-dependent vector fields using internal reference frames. In *Proceedings of the 16th Eurographics Conference on Visualization*, EuroVis '14, page 21–30, Goslar, DEU, 2014. Eurographics Association.

[BRG19] Irene Baeza Rojo and Tobias Günther. Vector field topology of time-dependent flows in a steady reference frame. *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE Scientific Visualization)*, 2019.

[CFV+16] Min Chen, Miquel Feixas, Ivan Viola, Anton Bardera, Han-Wei Shen, and Mateu Sbert. *Information theory tools for visualization*. CRC Press, 2016.

[Cra03] Roger Crawfis. Tornado data set generator, 2003.

[CY89] James P Crutchfield and Karl Young. Inferring statistical complexity. *Physical review letters*, 63(2):105, 1989.

[DCH88] Robert A Drebin, Loren Carpenter, and Pat Hanrahan. Volume rendering. *ACM Siggraph Computer Graphics*, 22(4):65–74, 1988.

[FI08] Shiho Furuya and Takayuki Itoh. A streamline selection technique for integrated scalar and vector visualization. *Vis Š08: IEEE Visualization Poster Session*, 2(4), 2008.

[GGT17] Tobias Günther, Markus Gross, and Holger Theisel. Generic objective vortices for flow visualization. *ACM Trans. Graph.*, 36(4):141:1–141:11, July 2017.

[Hal05] G. Haller. An objective definition of a vortex. *Journal of Fluid Mechanics*, 525:1–26, 2005.

[Hal15] George Haller. Lagrangian coherent structures. *Annual Review of Fluid Mechanics*, 47:137–162, 2015.

[HH89] J. Helman and L. Hesselink. Representation and display of vector field topology in fluid flow data sets. *Computer*, 22(8):27–36, 1989.

[HH91] J.L. Helman and L. Hesselink. Visualizing vector field topology in fluid flows. *IEEE Computer Graphics and Applications*, 11(3):36–46, 1991.

[HS11] George Haller and Themistoklis Sapsis. Lagrangian coherent structures and the smallest finite-time lyapunov exponent. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(2):023115, 2011.

[HY00] G. Haller and G. Yuan. Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D: Nonlinear Phenomena*, 147(3–4):352–370, 2000.

[JBTS08] Heike Jänicke, Michael Böttinger, Xavier Tricoche, and Gerik Scheuermann. Automatic detection and visualization of distinctive structures in 3d unsteady multi-fields. In *Computer Graphics Forum*, volume 27, pages 767–774. Wiley Online Library, 2008.

[JH95] Jinhee Jeong and Fazle Hussain. On the identification of a vortex. *Journal of fluid mechanics*, 285:69–94, 1995.

[JS10] Heike Jänicke and Gerik Scheuermann. Towards automatic feature-based visualization. In *Dagstuhl Follow-Ups*, volume 1. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2010.

[JWM88] Hunt' JCR, A Wray, and P Moin. Eddies, stream, and convergence zones in turbulent flows. *Center for turbulence research report CTR-S88*, pages 193–208, 1988.

[JWSK07] Heike Janicke, Alexander Wiebel, Gerik Scheuermann, and Wolfgang Kollmann. Multifield visualization using local statistical complexity. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1384–1391, 2007.

[KL51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951.

[Leo06] Paul Leopardi. A partition of the unit sphere into regions of equal area and small diameter. *Electronic Transactions on Numerical Analysis*, 25(12):309–327, 2006.

[LHZP07] Robert S. Laramee, Helwig Hauser, Lingxiao Zhao, and Frits H. Post. Topology-based flow visualization, the state of the art. In Helwig Hauser, Hans Hagen, and Holger Theisel, editors, *Topology-based Methods in Visualization*, pages 1–19, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[LMSC11] Teng-Yok Lee, Oleg Mishchenko, Han-Wei Shen, and Roger Crawfis. View point evaluation and streamline filtering for flow visualization. In *2011 IEEE Pacific Visualization Symposium*, pages 83–90. IEEE, 2011.

[LRMC95] Ricardo Lopez-Ruiz, Héctor L Mancini, and Xavier Calbet. A statistical measure of complexity. *Physics letters A*, 209(5-6):321–326, 1995.

[MLP+09] Tony McLoughlin, Robert S. Laramee, Ronald Peikert, Frits H. Post, and Min Chen. Over Two Decades of Integration-Based, Geometric Flow Visualization. In M. Pauly and G. Greiner, editors, *Eurographics 2009 - State of the Art Reports*. The Eurographics Association, 2009.

[MWW+14] Jun Ma, James Walker, Chaoli Wang, Scott Kuhl, and Ching Kuang Shene. Flowtour: An automatic guide for exploring internal flow features. In *2014 IEEE Pacific Visualization Symposium*, pages 25–32. IEEE, 2014.

[NNF+20] Baldwin Nsonga, Martin Niemann, Jochen Fröhlich, Joachim Staib, Stefan Gumhold, and Gerik Scheuermann. Detection and visualization of splat and antisplat events in turbulent flows. *IEEE Transactions on Visualization and Computer Graphics*, 26(11):3147–3162, 2020.

[NSG+20] B. Nsonga, G. Scheuermann, S. Gumhold, J. Ventosa-Molina, D. Koschichow, and J. Fröhlich. Analysis of the near-wall flow in a turbine cascade by splat visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):719–728, 2020.

[NT12] Ikujiro Nonaka and Hirotaka Takeuchi. *Die Organisation des Wissens: Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen*. Campus Verlag, 2012.

[Pop04] S. Popinet. Free computational fluid dynamics. *ClusterWorld*, 2(6), 2004.

[RG20] Irene Baeza Rojo and Tobias Günther. Vector field topology of time-dependent flows in a steady reference frame. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):280–290, 2020.

[TLMF21] Silvio Tschisgale, Bastian Löhrer, Richard Meller, and Jochen Fröhlich. Large eddy simulation of the fluid–structure interaction in an abstracted aquatic canopy consisting of flexible blades. *Journal of Fluid Mechanics*, 916, 2021.

[TMWS12] Jun Tao, Jun Ma, Chaoli Wang, and Ching-Kuang Shene. A unified approach to streamline selection and viewpoint selection for 3d flow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):393–406, 2012.

[WGS05] Alexander Wiebel, Christoph Garth, and Gerik Scheuermann. Localized Flow Analysis of 2D and 3D Vector Fields. In Ken Brodlie, David Duke, and Ken Joy, editors, *EUROVIS 2005: Eurographics / IEEE VGTC Symposium on Visualization*. The Eurographics Association, 2005.

[WGS07] Alexander Wiebel, Christoph Garth, and Gerik Scheuermann. Computation of localized flow for steady and unsteady vector fields and its applications. *IEEE transactions on visualization and computer graphics*, 13:641–51, 07 2007.

[WXWH13] Ziniu Wu, Yizhe Xu, Wenbin Wang, and Ruifeng Hu. Review of shock wave detection method in cfd post-processing. *Chinese Journal of Aeronautics*, 26(3):501–513, 2013.

[WYM08] Chaoli Wang, Hongfeng Yu, and Kwan-Liu Ma. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1547–1554, 2008.

[XLS10] Lijie Xu, Teng-Yok Lee, and Han-Wei Shen. An information-theoretic framework for flow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1216–1224, 2010.

[ZH15] Liang Zhou and Charles D Hansen. A survey of colormaps in visualization. *IEEE transactions on visualization and computer graphics*, 22(8):2051–2069, 2015.