Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Quantifying Vector Field Distances Based on a Reduced Stream Line Representations

Lars Gräning, Thomas Ramsay

2013

Preprint:

This is an accepted article published in Int. Conf. on Intelligent Data Engeneering and Automated Learning (IDEAL). The final authenticated version is available online at: https://doi.org/[DOI not available]

Quantifying Flow Field Distances Based on a Compact Streamline Representation

Lars Graening[1] and Thomas Ramsay[2]

[1] Honda Research Institute Europe,
Carl-Legien-Straße 30, 63073 OffenbachMain, Germany
lars.graening@honda-ri.de
[2] Honda R&D Americas, Inc, Raymond, OH 43067
tramsay@oh.hra.com

Abstract. In many engineering domains like aerospace, vehicle or engine design the analysis of flow fields, acquired from computational fluid dynamics (CFD) simulations, can reveal important insights on the behavior of the simulated objects. However, the huge amount of flow data produced by each simulation complicates the data processing and limits the application of computational tools for flow analysis. Thus, an a priori transformation of the flow data into a compact low dimensional representation is desired. This paper introduces a new procedure for transforming flow field data into a compact streamline based representation. Wherein, streamlines with negligible information contribution are removed from the representation. The reduced set of streamlines defines the basis for a subsequent quantification of flow field distances. Experimental studies show that the distances calculated based on the compact representation well approximate the distances of the uncompressed flow field with a significant drop in memory consumption.

Key words: Flow field distance, Flow features, Compact flow field representation, CFD

1 Introduction

The computational simulation and modeling of fluid dynamics has become a corner stone for a huge variety of applications, e.g. related to aerospace, climate modeling, vehicle or engine design. The analysis and comparison of the acquired flow field data can provide relevant domain information related to the behavior of the technical system under consideration, e.g. revealing important interactions between design and flow features [6]. The comparison of two flow fields requires a proper similarity measure being defined, which can e.g. be adopted for discovering flow field categories [1] or to compare simulated against ideal flow, as for the design of combustion engines [3]. However, means for the computational evaluation of flow field similarities are rarely being used in practice, rather modern visualization techniques [5] are applied to support the qualitative evaluation of flow field similarities through visual inspection. Quantitative methods require

that explicit considerations about the flow field representation and the similarity measure are being made. Xu et al. [9] have studied probabilistic similarity measures based on featured flow fields. The measures take feature data from the entire simulation domain into account. For practical applications, this requires a big data storage and is linked to high computational costs for the similarity evaluation. In this paper a new compact streamline vise flow field representation is introduced which is derived from an automatic streamline reduction process. The reduced flow field representation is the basis for the definition of a probabilistic similarity measure.

2 Flow Field Representation

Flow fields, as a result of computational fluid dynamics simulations (CFD), define properties of the fluid or air flow in interaction with solid objects. Beside pressure, density and temperature, the velocity vector field defines the speed of the air or fluid flow at fixed positions in the Euclidean space. Considering steady state flow, the vector field is formally defined as a function that maps each point in the Euclidean space to a velocity vector, which remains constant over time,

$$v(p) : (x, y, z)^T \mapsto (v_x, v_y, v_z)^T.$$

Tools for computational flow simulation model the flow physics with the Navier Stokes equations. In order to numerically solve these partial differential equations the simulation area is subdivided into discrete cells. Especially for high fidelity simulations the cell sizes need to be kept small, resulting in a huge velocity vector field representing the detailed flow of the entire simulation domain. The storage of all the flow data is often impractical.

Local Flow Features The orientation and magnitude of the velocity vector does not always provide all relevant information about the flow at a fixed point p. Therefore, additional features need to be derived. Wherein, the calculation of the features are always problem and application dependent. Following Xu et al. [9], features representing higher order properties derived from the gradient of the velocity vector field have been considered in this paper. The local gradient in a 3D velocity vector field is defined by an asymmetric 3D tensor:

$$T(p, v) = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} = \begin{pmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} & \frac{\partial v_x}{\partial z} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} & \frac{\partial v_y}{\partial z} \\ \frac{\partial v_z}{\partial x} & \frac{\partial v_z}{\partial y} & \frac{\partial v_z}{\partial z} \end{pmatrix} = S + A, \tag{1}$$

which can be decomposed into its symmetric and asymmetric components, S and A:

$$S = \begin{pmatrix} \frac{\partial v_x}{\partial x} & \frac{1}{2}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right) & \frac{1}{2}\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right) & \frac{\partial v_y}{\partial y} & \frac{1}{2}\left(\frac{\partial v_z}{\partial y} + \frac{\partial v_y}{\partial z}\right) \\ \frac{1}{2}\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial v_z}{\partial y} + \frac{\partial v_y}{\partial z}\right) & \frac{\partial v_z}{\partial z} \end{pmatrix} = \begin{pmatrix} \varepsilon_1 & \frac{1}{2}\theta_3 & \frac{1}{2}\theta_2 \\ \frac{1}{2}\theta_3 & \varepsilon_2 & \frac{1}{2}\theta_1 \\ \frac{1}{2}\theta_2 & \frac{1}{2}\theta_1 & \varepsilon_3 \end{pmatrix}, \tag{2}$$

$$A = \begin{pmatrix} 0 & -\frac{1}{2}\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) & \frac{1}{2}\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) & 0 & -\frac{1}{2}\left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) \\ -\frac{1}{2}\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}, \tag{3}$$

where S comprises isotropic scaling and rotation information, and A anisotropic information about the stretching of the velocity vectors in the field, based on which the following scalar features can be calculated:

1. *Vector magnitude*: $M_v = |(v_x, v_y, v_z)^T|$,
2. *Tensor magnitude*: $M_t = \sqrt{\frac{1}{2}\sum_i^3 \sum_j^3 T_{i,j}^2}$,
3. *Dilatation magnitude*: $M_d = \sqrt{\sum_i^3 \varepsilon_i^2}$,
4. *Magnitude of shear strain rate*: $M_s = \sqrt{\sum_i^3 \theta_i^2}$,
5. *Magnitude of vorticity*: $M_\omega = \sqrt{\sum_i^3 \omega_i^2}$.

Given those five features we define the feature space as:

$$f(p, v) : p, v \mapsto (M_v, M_t, M_d, M_s, M_\omega)^T.$$

The defined feature vector is considered to comprise all problem relevant information within the simulated flow domain. If this is not the case any subsequent analysis and modeling step might fail to reveal desired information.

Featured Streamlines Streamlines are one of the most important flow field visualization techniques see e.g. [10]. Streamlines describe the trajectories (tangential to the velocity vector) that particles would travel through the flow starting from an initial seeding position. Streamlines are represented by discrete sample points along the trajectory, where the distance between adjacent sample points is defined by the density of the CFD mesh. Assigning the feature vector to each sample point along the streamline, a *featured streamline* is defined as:

$$l(u) : u \mapsto p, v, (M_v, M_t, M_d, M_s, M_\omega)^T,$$

where u defines the position of the sampling point along the streamline l. A set of featured streamlines is denoted as $L = \{l_i\}, i \in [1, N_l]$, with N_l stream lines.

Beside the purpose of visualization, the transformation of the flow field into a set of streamlines can be applied for data reduction, where the entire flow and feature data is condensed into information along individual trajectories. Without a priori knowledge about the flow field, streamlines are seeded uniformly at the inlet. In order to capture all relevant flow phenomena, often a huge number of streamlines is required, what complicates the subsequent visual investigation of the flow and goes along with an increase in memory consumption.

3 Automatic Streamline Reduction

A program for automatic streamline reduction has been developed to reduce the amount of flow data to be processed by humans or computer programs. Given a large set of streamlines, the algorithm targets to remove individual streamlines from the set which contribute least to the overall information content. Thus, targeting the creation of an efficient streamline based flow field representation with a minimum number of streamlines. Therefore, streamlines are ranked according to their information contribution, which is quantified based on the formulation of the Shanon entropy [8]. Given discrete feature data for each streamline, the expected entropy accumulated over all features defines the considered measure of relevance, formally defined as:

$$\overline{H}_i = E(H(l_i)) = -\frac{1}{N_f} \sum_{i}^{N_f} \sum_{j}^{N_b^i} p(\hat{f}_i^j) \log p(\hat{f}_i^j),$$

with N_f defining the number of features, N_b^i defining the number of discrete states and $\hat{f}_i^j$ defining the discrete feature value of feature i, e.g. the velocity magnitude M_v. The probability $p(\hat{f}_i^j)$ denotes the likelihood of a discrete feature value to be measured along the considered streamline. On the one hand, if all discrete feature values are equally probable, the feature values are uniformly distributed and the entropy gets maximal. On the other hand, if $p(\hat{f}_i^j) = 1.0$, the feature value along the streamline does not change within a certain boundary. In the latter case the value of the Shannon entropy vanishes. Under the assumption that streamlines with none or small feature value variations are of less relevance and capture least information about the flow field, a heuristic has been defined that removes streamlines with smallest $\overline{H}_i$. Ranking all streamlines according to their expected entropy, a subset of streamlines L' can be derived which captures the majority of the overall flow feature information. The information content of the reduced subset is defined by the relative expected entropy:

$$IC(L') = \frac{\sum_j^{N_k} \overline{H}_j}{\sum_i^{N_l} \overline{H}_i}, \tag{4}$$

where N_k defines the number of streamlines in the reduced set. For an automatic streamline reduction procedure a threshold λ needs to be applied to $IC(L')$. The threshold guarantees that at least λ percent of the flow information remain in the subset.

4 Flow Field Distance

Applying techniques for featured streamline extraction and reduction, the quantification of the distance between two flow fields is rephrased to the quantification of the distance between sets of reduced featured streamline representations. Given that the distance $d_{m,n}$ between two corresponding streamlines m and n

of two representations A and B are given, the overall distance has been defined based on the accumulation of all $d_{m,n}$:

$$D(L'_A, L'_B) = \sum_{k}^{N_A} d_{m,n}, \tag{5}$$

with N_A defining the number of streamlines of representation A and l_m denoting the corresponding streamline to l_n. The calculation of the overall distance based on reduced streamline sets requires to detect corresponding streamlines between two representations, as well as the definition of an appropriate streamline wise distance measure $d_{m,n}$.

Identification of Corresponding Streamlines The task of identifying corresponding streamlines is to assign a streamline of one flow field representation A to its comparable streamline of flow field B. This paper attempts the problem by assuming that corresponding streamlines are those which origin at the same or a near by seeding point. Given the initial point p^0_m of one streamline $l_m \in L^A$ with $u = 0$, the corresponding streamline in B is defined by the Euclidean closest point p^0_n over all streamline seeding positions in L^B, with $n \in [1, N_B]$. If the minimal distance is within a vicinity r the streamline l_n is said to be the corresponding streamline of l_m. The vicinity r is defined as the minimal distance between the seeding positions of neighboring streamlines considering all streamlines in A and B. If the minimal distance is not within the vicinity r a corresponding streamline to l_m is said to not exist.

Streamline Wise Distance Given N_f different features and a related weight factor w_i, the streamline wise distance in its general form is denoted as:

$$d_{m,n} = \sum_{i}^{N_f} w_i \cdot d(P^m_i, P^n_i)$$

with $\sum_i^{N_f} w_i = 1.0$, and P^m_i, P^n_i defining the discrete feature distribution for feature i with respect to streamline l_m and the corresponding streamline l_n, respectively. In this paper four different probabilistic distance measures $d(P^m_i, P^n_i)$ have been considered to evaluate the distance between discrete feature distributions. The measures can be categorized into *bin-wise distances*: L^1 and χ^2 [9], and *cross-bin distances*: Earth Mover's Distance EMD [7] and the quadratic χ distance QC [4]. In the case that no corresponding streamline to l_m has been found, $d(P^m_i, P^n_i) = 0.0$.

5 Results

The flow distance quantification based on a reduced featured streamline representation has been evaluated given a flow field test data set, which covers velocity and feature data from the simulation of different polyhedral objects.

5.1 Experimental Setup

Various geometrical objects have been generated by modifying the baseline surface mesh of an icosahedron object. First, a subdivision algorithms[3] has been applied to smoothen the surface of the icosahedron, so that the object more and more morphs into a spherical shape, see Fig. 1. Second, objects with different level of subdivision have been rotated along the z-axis. The objects are denoted as ICO_{α}^{SL}, where $SL \in [0, 3]$ defines the subdivision level and $\alpha \in \{0, 10, \ldots, 90\}$ the rotation angle. In order to derive comparable flow fields, all objects have been constrained to a frontal area of $3m^2$. For the simulation of the flow around the

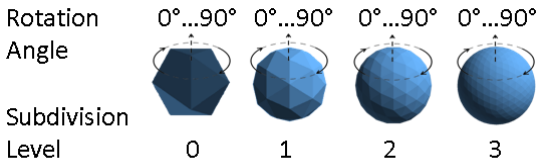

Fig. 1. Modified geometries for the test dataset have been generated by rotation and by subdividing the surface mesh of the icosahedron.

polyhedral objects, each object has been positioned onto a ground plane with a distance of $48.0m$ from the inlet. In order to numerically solve the Navier stokes equations the vicinity around the object has been discretized into a volume mesh with around $350,000$ mesh cells using *snappyHexMesh*[4]. The flow has been modeled at a speed of $110km/h$. To solve the partial differential equations the simpleFOAM tool from *openFOAM*[5] has been applied. The velocity field linked to each object has been extracted from the time averaged flow. The initial set of streamlines has been uniformly seeded from a plain at the position $x = -48.0m$. The flow features along the individual streamlines have been calculated based on finite differences.

5.2 Entropy based streamline reduction

In the first experiment a set of 450 streamlines has been generated for each of the 40 objects. By applying entropy based streamline reduction, first the relation between the ratio of the number of streamlines in the subsets and its related information content, (Eq. 4) have been under investigation. The streamline subsets are constraint to cover at least λ times the expected entropy of the entire streamline set. For different values of λ from 0.11 to 0.95 the related number of streamlines of the reduced subsets are depicted in Fig. 2 a). Each box denotes the statistics on the number of remaining streamlines over all different objects.

[3] http://brainder.org/, [2]
[4] http://www.openfoam.org/docs/user/snappyHexMesh.php
[5] http://www.openfoam.com/

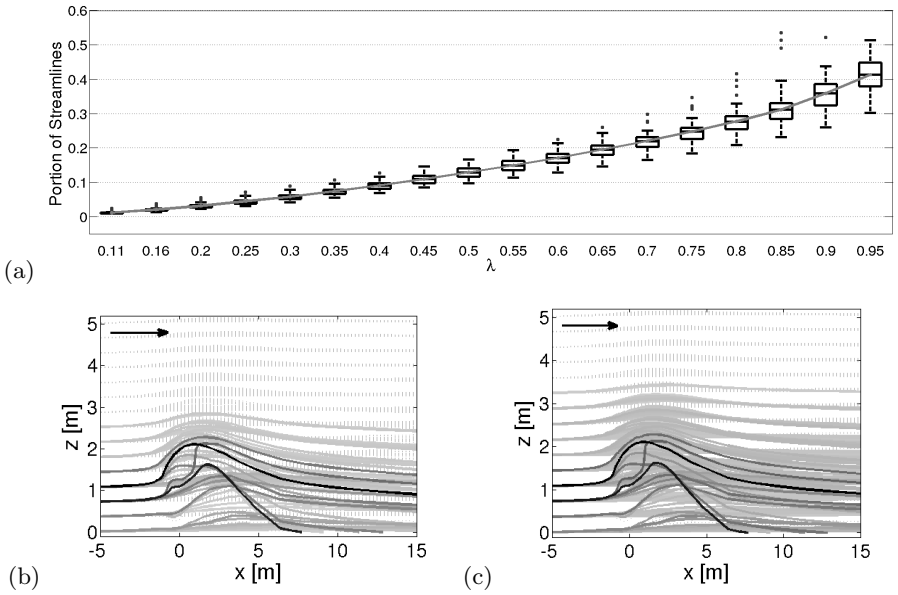

Fig. 2. a) Illustration of the portion of streamlines and the percentage of the related information kept after the streamline reduction step. Visualization of the statistics over all polyhedral objects in the test data set. The illustrations at the bottom show the reduced streamline sets regarding object Ico_0^0 with b) $\lambda = 0.50$ and c) $\lambda = 0.90$ (projected onto the x-z-axis).

As an example, with $\lambda = 0.95$ around 95% of the information about the flow features can be preserved by reducing the number of streamlines by more than 50%. The filtered 50% of the streamlines contribute only 5% to the expected entropy of the entire streamline data set. The results show that the suggested procedure for automatic streamline reduction can lead to a significant data reduction, reducing the memory consumption of the flow field representation.

Figs. 2 b) and c) depict two examples with $\lambda = 0.5$ and $\lambda = 0.9$, illustrating the reduced streamline data sets concerning object Ico_0^0. Solid lines show the most informative streamlines, with color proportional to the respective information content. Dotted lines highlight streamlines which have been removed from the dataset. It shows that the streamline ranking can simplify the visualization, by hiding all uninformative streamlines.

5.3 Flow Field Distance

In the following experiments it has been investigated how the distances between flow fields, which have been calculated based on the reduced streamlines, compare with the distances evaluated based on the complete streamline set. For reference, for each object three sets of uniform seeded streamlines with 450, 200 and 50 lines have been generated. The distances calculated based on these

three representations have been compared against the distances of three different compact streamline representations, with $\lambda \in \{0.9, 0.7, 0.5\}$. All in all, for each object six different representations have been generated. Related to each representation, the distances between the streamline data of all possible object combinations have been calculated according to Eq. 5, using either of the four distance measures (L^1, χ^2, EMD, $QC^B_{0.5}$).

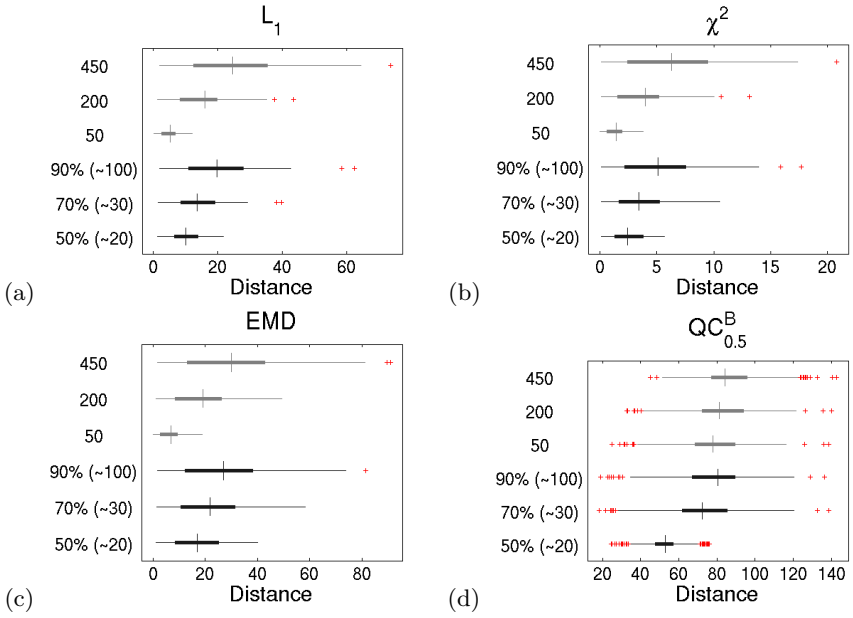

Fig. 3. Comparison of the flow field distances using either of the representation with fixed number of streamlines 450, 200 and 50 (gray) or using the compact streamline representation with $\lambda \in \{0.9, 0.7, 0.5\}$ (black).

The results of the experiments are depicted in Figs. 3 a) to d). Each bar denoting the distribution of the flow field distances of all object combinations, with all in all 1440 distance values. From all the different representation, it is assumed that the one with 450 uniform seeded streamlines most accurately represents the distances between the different flow fields. A decrease in the mean value over all flow field distances is expected to decrease the likelihood to correctly discriminate the flow fields. Considering the representations with uniform seeding, with 450, 200 and 50 streamlines, a reduction in the number of streamlines leads to a drop in the median distance value. Especially with only 50 streamlines a correct discrimination of the flow fields seems to be hardly possible. This effect can be similarly observed for all four distance measures.

In contrast, using the reduced streamline representation (reduced from 450 uniform seeded streamlines), the reference distance distribution can be almost

maintained even for low number of streamlines, independent of the distance measure used. As an example, consider Fig. 3 c), where the EMD has been used, it can be concluded that only around 100 streamlines are sufficient, with $\lambda = 0.5$, in order to maintain the distance distribution of the reference representation with 450 streamlines. This holds for almost any distance measure, except using the $QC^B_{0.5}$ distance. However, it can be concluded that around two third of the streamlines can be removed from the flow data, while maintaining almost the same distance values between flow fields.

6 Conclusion

In this paper a new attempt for deriving a reduced streamline wise flow field representation has been introduced, where streamlines are reduced based on the evaluation of its expected flow feature entropy. It has been demonstrated that the suggested procedure can lead to a significant reduction in memory consumption for storing the flow data, while maintaining the majority of the flow information. The quantification and analysis of the distance values between different flow fields, resulting from the simulation of different polyhedral objects, has turned out that the compact representation can almost maintain the same distances as a uniformly seeded streamline representation with a huge number of streamlines.

For consolidating the usability of such compact representations and the related technique for evaluating the streamline wise flow field distance, the same procedure will be applied to the simulation data of various passenger cars in the future. The resulting distances can be utilized then to visualize or automatically retrieve passenger car categories. Furthermore, the streamline wise distance measure needs to be extended to time-varying flow data.

7 Acknowledgement

The authors would like to thank Stefan Menzel and Sebastian Schmitt from the Honda Research Institute Europe GmbH for their valuable comments and advice.

References

1. S. Depardon, J. Lasserre, L. Brizzi, and J. Bore. Automated topology classification method for instantaneous velocity fields. *Experiments in Fluids*, 42:697–710, 2007.
2. H. Kenner. *Geodesic math and how to use it.* Dome series. University of California, 1976.
3. R. S. Laramee, D. Weiskopf, J. Schneider, and H. Hauser. Investigating swirl and tumble flow with a comparison of visualization techniques. In *Proceedings IEEE Visualization04*, pages 51–58. IEEE Computer Society, 2004.
4. O. Pele and M. Werman. The quadratic-chi histogram distance family. In *Proceedings of the 11th European conference on Computer vision: Part II*, ECCV'10, pages 749–762, Berlin, Heidelberg, 2010. Springer-Verlag.

5. F.H. Post, B. Vrolijk, H. Hauser, R.S. Laramee, and H. Doleisch. Feature extraction and visualization of flow fields. *Eurographics 2002 State-of-the-Art Reports*, pages 69–100, 2002.
6. M. Rath and L. Graening. Modeling design and flow feature interactions for automotive synthesis. In *Int. Conf. on Intelligent Data Engeneering and Automated Learning (IDEAL)*. Springer, 2011.
7. Y. Rubner, C. Tomasi, and L. J. Guibas. The earth movers distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
8. C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379423, July 1948.
9. L. Xu, H. Q. Dinh, P. Mordohai, and T. Ramsay. Detecting patterns in vector fields. In *AIAA Aerospace Sciences Meeting*, Orlando, Florida, USA, 2011.
10. X. Ye, D. Kao, and A. Pang. Strategy for seeding 3d streamlines. In *Visualization, 2005. VIS 05. IEEE*, pages 471–478, 2005.