Entropy Guided Visualization And Analysis Of Multivariate Spatio-Temporal Data Generated By Physically Based Simulation

Selcuk Sumengen*
SAP Innovation Center, Istanbul, Turkey

Ekrem Serin†
Sabanci University, Istanbul, Turkey

Selim Balcisoy‡
Sabanci University, Istanbul, Turkey

Abstract

Flow fields produced by physically based simulations are subsets of multivariate spatio-temporal data, and have been in interest of many researchers for visualization, since the data complexity makes it difficult to extract representative views for the interpretation of fluid behavior. We utilize Information Theory to find entropy maps for vector flow fields, and use entropy maps to aid visualization and analysis of the flow fields. Our major contribution is to use Principal Component Analyses (PCA) to find a projection that has the maximal directional variation in polar coordinates for each sampling window in order to generate histograms according to the projected 3D vector field, producing results with fewer artifacts than the traditional methods. Entropy guided visualization of different data sets are presented to evaluate proposed method for the generation of entropy maps. High entropy regions and coherent directional components of the flow fields are visible without cluttering to reveal fluid behavior in rendered images. In addition to using data sets those are available for research purposes, we developed a fluid simulation framework using Smoothed Particle Hydrodynamics.

1 Introduction

Time-varying multivariate spatio-temporal data sets are produced by physically based simulations of natural phenomenons, however fluid simulations producing flow fields are the most common ones that are frequently used for practical applications. We experiment on every stage beginning from the simulation to the visualization, and introduce improvements on certain tasks until proposing a novel method for histogram generation to calculate entropy and aid visualization. Ultimate goal of this research is finding methods and techniques to assist analysis and visualization of multivariate data sets using Information Theoretical approaches.

2 Visualization & Analysis Of Multivariate Spatio-Temporal Data

Although the flow fields are only a subset of multivariate spatio-temporal data, majority of the data studied in this category are provided from the simulations that are producing flow fields. A detailed overview of methods for integration based geometric flow visualization are presented by McLoughlin et al. in their paper[2]. Pobitzer et al. [3] published a state of art report on topology based flow visualization for unsteady flow. We briefly overview direct and integration based visualization methods for flow fields with applications to the sample data obtained from our SPH simulations as well as Hurricane Isabel data produced by Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF) [1]. In addition to the direct approaches for detecting salient features, we employ Entropy function used in Information Theory with a histogram based method similar to Xu et al. [5]. We improve their method with a modification to take vector magnitudes into account in addition to the directions for diversity. A new method using singular value decomposition and principal components is proposed afterwards.

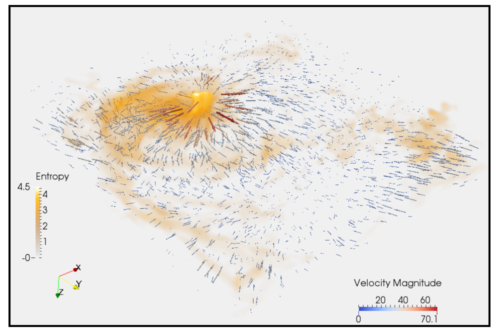

Figure 1: Hurricane Isabel data is rendered using color coding for the entropy values and arrow glyph for the velocity vector where the entropy value is below the threshold to reveal fluid behavior.

Direct methods aim to present data without generation of new attributes. Those methods are usually applied in a simplistic manner, and easy to implement. In color coding, usually vector magnitudes are visualized and the direction information is omitted, or each component rendered separately. It's also possible to map each directional component of the vectors to a color value in RGB color space. Arrow glyph methods are able to represent directions. However, arrow glyph technique is vulnerable to cluttering and occlusion. The majority of integration based geometry extraction methods are originated from streamlines. Pathlines are the particle trajectories a mesh-less particle takes in the flow field in time, and streaklines are the curves connecting the particles which are seeded from the same location. All those methods require choosing spatial locations to start integration, and defining intervals in spatial or time domain. They are also vulnerable to occlusion and cluttering based the seeding locations and integration intervals.

3 Entropy Guided Visualization of Multivariate Spatio-Temporal Data

Information Theory is a very powerful tool to quantify information content of the data used for flow visualization as well as determining salient features in 3D data fields. Although the use of entropy to highlight the important features is practiced in scientific visualization[4, 5], our approach differs in evaluating the vector directions and magnitudes while creating histograms to calculate entropy.

Probability distribution function $p(x_i)$ should be defined to calculate an entropy field. It's trivial to calculate $p(x)$ when the variable is an integer with enough samples in the subspace. When a variable is a floating point, a threshold needed to place the samples into the same bucket. Treating each distinct value separately would cause sparsity. The general approach is placing the buckets uniformly, casting the floating point numbers to the integers. For

*e-mail:selcuk.sumengen@sap.com
†e-mail:eserin@su.sabanciuniv.edu
‡e-mail:balcisoy@sabanciuniv.edu

the vector fields, common approach is to convert the vectors to angle/magnitude pairs, the magnitudes are ignored or entropy fields are calculated separately to join them later with interpolation techniques. Our method differs in terms of histogram generation, and we propose two different approaches to utilize angles and magnitudes together before and after entropy calculation.

3.1 Histogram Generation for 3D Vector Fields

Xu et al. [5] partition the range of vectors represented by polar angles; Θ, $0 < \Theta <= 2\pi$ for 2D vectors. For 3D vectors, unit sphere is decomposed into 360 patches to use the cones that connect the center to the patches, and vectors are quantized ignoring magnitudes.

Our first approach is to rearrange the buckets to group the vectors into the same bucket depending on the directions and magnitudes. Instead of dividing into patches, we put equidistant points on the unit sphere. We also scale the sphere according to the global distribution of magnitudes and use the points on two spheres to represent the buckets.

Our final approach to generate histograms for 3D vectors is using Principal Component Analysis (PCA). We use Singular Value Decomposition (SVD) to find a new coordinate system for each window, in which the projection of 3D vectors onto the new XY plane produces highest directional variation among all possible projections onto any plane. We use polar coordinate transformations to calculate Θ for angles, r for radiuses and keep the transformed z values. After applying entropy calculation for each scalar value, results in Figure 2 are pretty consistent with the previous methods. In addition, our entropy maps produce distinctive regional boundaries, not interfered with lines due to histogram discretization. The artifacts due to discretization are still present, but they're not forming boundaries as straight lines. They are curvilinear due to rotating coordinate frames of the histogram bins.

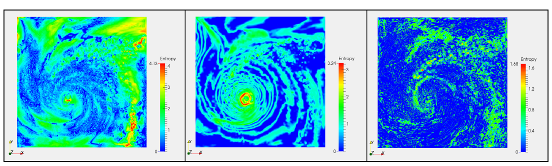

Figure 2: Entropy calculated with our method on Hurricane Isabel dataset; with angle of direction on the projected plane with the magnitudes, and z coordinate after projection.

Second entropy map is calculated from r for radiuses of the vectors, and reveals important information about vector magnitudes which can also be used in analysis as a whole, or averaging with the directional entropy field as an existing approach. Third entropy map is produced with z values corresponds to the components with less directional variation. Low entropy values observed in the map verifies this assumption.

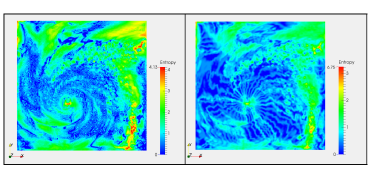

Figure 3: The directional entropy calculated after utilizing PCA is given on the left, and the directional entropy calculated using a unit sphere for discretization is given on the right.

3.2 Results and Discussion

Information Theory is utilized to calculate entropy for the purpose of aiding visualization and analysis of the flow fields. High entropy regions indicate the areas of high directional variation in vector flow field. Low entropy field is an indicator of order, it reveals the areas where the directions are coherent. To demonstrate the utilization of entropy fields generated, direct rendering methods are merged to visualize SPH simulation and Hurricane Isabel data sets. In Figures 4 and 1, volume rendering is used for the entropy field, and due to the transfer function correlated with the entropy, low entropy regions are transparent. High entropy regions are chaotic in terms of directions, so we utilized arrow glyph method on low entropy regions to reveal the flow behavior without cluttering.

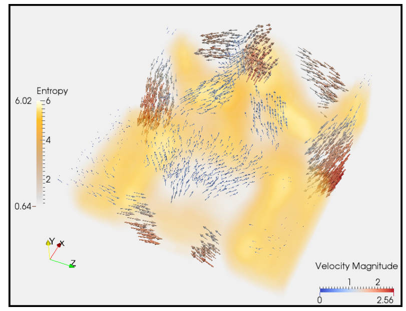

Figure 4: SPH simulation data is rendered using color coding for the entropy values and arrow glyph for the velocity vector where the entropy value is below the threshold to reveal fluid behavior.

Since SPH simulation data in figure 4 has high variation in three dimensions, our method produces more readable and informative visualizations than the direct rendering techniques. Hurricane Isabel data rendered in Figure 1 also reveals the center of vortex and the directional information around it. Our approach produces a representative view of the whole field.

4 CONCLUSION

We presented a framework for visualization of multivariate spatio-temporal data sets generated by physically based simulation. PCA is used for generating histograms of 3D vector fields with polar coordinate transformation. Our method is less prone to discretization errors than the previous methods grouping 3D vectors into buckets with fixed geometry in 3D space, since dimensional reduction allows us to use less number of buckets oriented in space according to the variation of data, and avoids sparsity. After projection, entropy fields are generated from the vector magnitudes, and z-coordinates in addition to the entropy field from the directional component. We also performed entropy guided visualization of the flow fields for demonstration. Regions of interest can be defined according to the entropy ranges, and the entropy field itself can reveal the characteristics and behavior of underlying flow field.

REFERENCES

[1] C. B. T. S. Bill Kuo, Wei Wang and D. Middleton. Hurricane isabel data produced by the weather research and forecast (wrf) model, courtesy of ncar and the u.s. national science foundation (nsf)., 2004.

[2] T. McLoughlin, R. S. Laramee, R. Peikert, F. H. Post, and M. Chen. Over two decades of integration-based, geometric flow visualization. *Computer Graphics Forum*, 29(6):1807–1829, 2010.

[3] A. Pobitzer, R. Peikert, R. Fuchs, B. Schindler, A. Kuhn, H. Theisel, K. Matkovi, and H. Hauser. The state of the art in topology-based visualization of unsteady flow. *Computer Graphics Forum*, 30(6):1789–1811, 2011.

[4] C. Wang and H.-W. Shen. Information theory in scientific visualization. *Entropy*, 13(1):254–273, 2011.

[5] L. Xu, T.-Y. Lee, and H.-W. Shen. An information-theoretic framework for flow visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1216 –1224, nov.-dec. 2010.