A Novel Method for Tracking Tensor-based Regions of Interest in Large-Scale, Spatially-Dense Turbulent Combustion Data

Timothy Luciani*
Dept. of Computer Science
University of Pittsburgh

Adrian Maries
Dept. of Computer Science
University of Pittsburgh

Hoang Tran
Dept. of Mathematics
University of Pittsburgh

Mehdi B. Nik
Center for Simulation & Modeling
University of Pittsburgh

Levent Yilmaz
Center for Simulation & Modeling
University of Pittsburgh

G. Elisabeta Marai
Dept. of Computer Science
University of Pittsburgh

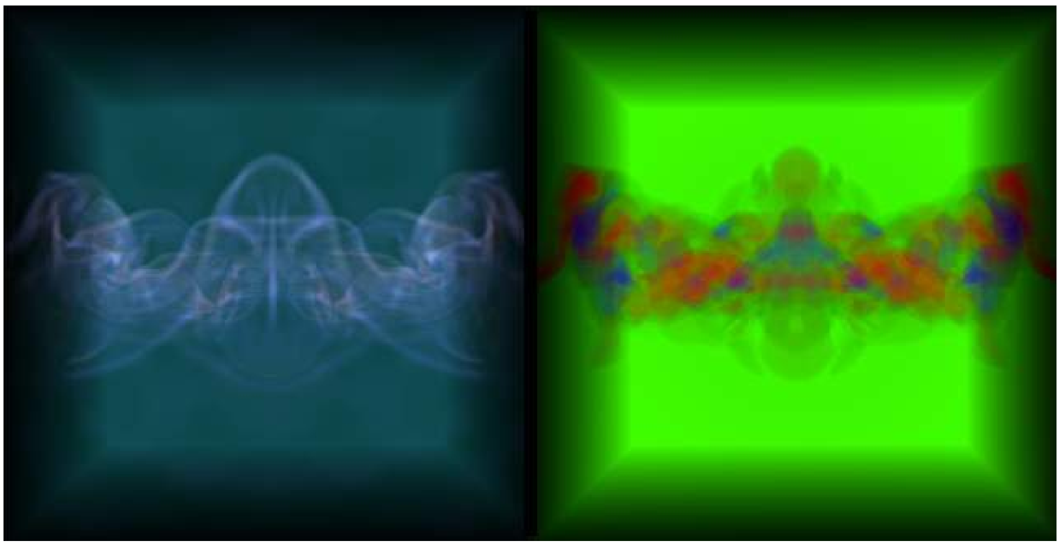


Figure 1: Tensor-based region of interest tracking for a mixing layer dataset. On the left is a volume rendering of the strain tensor trace (divergence). On the right is a volume rendering of three tensor clusters. The main regions of interest follow the tensor trace almost exactly.

1 Introduction

In visualization of tensor data associated with turbulent combustion simulations, the datasets we have to process and visualize are typically enormous in size and continue to grow exponentially. The large amount of data may severely affect the manipulating speed, thus posing a major challenge to interactive visualization [2]. Even worse, tensor datasets tend to be very dense leading to clutter and occlusion problems. To address these issues, feature extraction is an emerging method. Typically, only a small percentage of data is of interest, thus making the effective visualization of very large datasets possible. Feature extraction also helps the users highlight and focus on regions of interest.

We introduce an approach for the segmentation, visualization and tracking of regions of interest in large scale tensor field datasets generated by computational turbulent combustion simulations. We use canopy clustering followed by a K-means algorithm to partition and cluster the tensor field components. The resulting clusters are tracked through multiple timesteps. Interactive, hardware-accelerated volume renderings [4] are generated using the cluster indices. Results on two rich datasets show this approach can assist in the visual analysis of combustion tensor fields.

A great deal of research has been conducted on the problems of feature extraction and tracking [5], as well as in characterizing, detecting and visualizing regions of interest [1]. The method we present herein is novel in that it integrates machine learning with visualization for extracting and clustering regions of interest. It is thus a promising approach to apply to very large flow datasets.

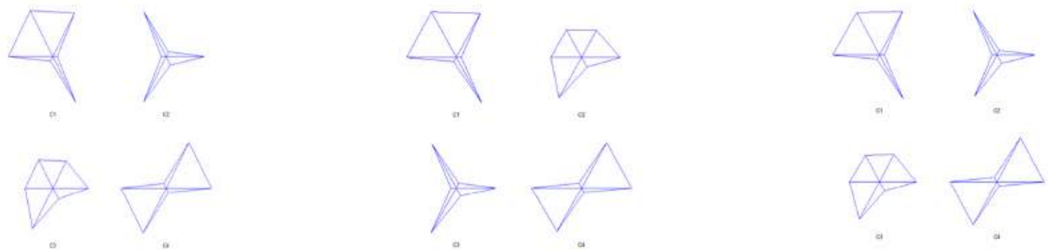
*e-mail: tbl8@pitt.edu

2 Methods

Clustering analysis is used to group data points that are similar to one another. Clustering was performed on the 6 distinct values of the strain-rate tensor, which is a symmetric quantity (Eq. 1):

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{1}$$

where x_i and u_i for $i = 1,2,3$ are the Cartesian components of position and velocity, respectively; and S_{ij} is the strain-rate tensor.



(a) timestep 70 (b) timestep 75 (c) timestep 80

Figure 2: Star-plot glyphs corresponding to the cluster centroids for a four-cluster segmentation (left to right: centroids at timesteps 70, 75, and 80). The different signatures of the cluster centroids make possible the consistent labeling and thus tracking of clusters over time.

Due to the density of the data, a pre-clustering step was performed through canopy clustering [3] to obtain good starting cluster centers. Once the centers were found, k-means was run, this time limiting the number of iterations to 50. Star-plots were generated around the cluster centers to facilitate the consistent labeling of clusters across multiple timesteps. The star-plot representation was selected by the domain expert as being the most intuitive representation of the tensor. (Figure 2) Alternative designs had explored a variety of glyph representations including ellipsoids and superquadrics [2]. Finally, a ray-based volume rendering technique was employed with a user-defined transfer function. This allowed for the presentation of interactive linked views to the user. By design, the comparison between regions of interest and the actual tensor fields could thus be observed simultaneously. The first volume showed the divergence of the dataset. The second volume displayed the regions of interest, rendered by assigning each cluster an individual value and setting the transfer function monochromatically to each.

3 Results

Datasets. The first dataset, the temporal mixing layer, is a simple configuration where two streams of fuel and oxidizer flow over and against each other. The flow speeds are adjusted for a low Reynolds number yielding a narrow range of length scales, and this configuration can be easily tackled with Direct Numerical Solution (DNS) and then used as a benchmark. The data covers a grid of size 193 grid points in two Cartesian directions and 194 in the other (approx. 8M grid points); single timestep. The second dataset, the shocklet, has a similar configuration, but is significantly larger (194 x 577, approx. 21M points), and varies along the temporal dimension (t = 0:600; 12,900 time steps.) Through time, the flow is going through pairing and exhibits 3D effects ("shocklet"). Flow field variables such as Mach number, divergence of velocity and gradients of density, temperature and pressure change sharply across the shocklet surface.

Performance. Clustering was performed on a quad-core 3.33 GHz Intel i5 CPU machine with 16 GB of RAM running Windows 7. On average, the data took between 15 to 20 minutes to generate 3 and 4 clusters. The application was then tested on a machine with two NVIDIA GeForce GTX 550Ti 1024MB GPUs and 16 GB of RAM running Windows 7 as well as a Macbook Pro with an NVIDIA GeForce GT 650M 1024MB GPU and 8GB of RAM. For both datasets, both machines achieved frame rates of 125 frames per second, allowing the user to interact with the volume renderings in real-time.

Feedback. The goal for the mixing layer dataset was to see if the clustering can provide insight into the structure of the flow. We provided a senior combustion researcher with a 3D volume rendering of the divergence of the tensor (sum of components on the main diagonal, indicates fluid density changes), and a volume rendering of the 3-group clustering (Fig. 1). When asked for an evaluation, he noted that the clusters coincided surprisingly well with the interesting regions of the flow.

The goal for the second dataset was to see if the distinct tensor field regions have a clear relationship with the shock region. Figure 2 shows the starplot glyphs corresponding to the cluster centroids for the four-cluster segmentation, steps 70, 75 and 80. Two senior combustion researchers analyzed the corresponding volume rendering output and concluded that the tensor clusters did not correlate with the regions where the shock is. This indicated to the combustion researchers that the tensor field did not modify significantly in the shock region, which was considered an interesting finding.

4 Discussion and Conclusion

The goal of this project was to examine the potential of using cluster analysis on tensor field data generated by turbulent combustion simulations. Specifically, our aim was firstly to see if tensor field clustering and rendering could give researchers insights into the structure of the flow through a volume. Secondly, we wanted to investigate if cluster analysis would allow combustion researchers to detect the positive or negative correlation of the tensor field with a specific feature of high-speed flow, namely the region where the speed becomes supersonic. The answer to both questions is affirmative.

In conclusion, we have introduced an approach for the segmentation, visualization and tracking of regions of interest in large scale tensor field datasets generated by computational turbulent combustion simulations. The approach is novel in that it integrates *machine learning* – canopy and k-means clustering – with *visualization* – interactive volume rendering – to extract, cluster, and track regions of interest in the tensor field. Our evaluation on two rich combustion datasets shows this approach can assist in the visual analysis of the combustion tensor field.

References

[1] K.-L. Ma, J. V. Rosendale, and W. Vermeer. 3d shock wave visualization on unstructured grids. *Proceedings of the 1996 Symposium on Volume Visualization*, pages 87–94, 1996.

[2] A. Maries, M. A. Haque, S. Yilmaz, M. Nik, and G. Marai. *Interactive Exploration of Stress Tensors Used in Computational Turbulent Combustion*. New Developments in the Visualization and Processing of Tensor Fields, D. Laidlaw and A. Villanova (editors), Springer, 2011.

[3] A. McCallum, K. Nigam, and L. Ungar. Efficient clustering of high-dimensional data sets with application to reference matching. *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining KDD 00*, pages 169–178, 2000.

[4] NVIDIA, Inc. *The CUDA API Reference Manual*, February 2011.

[5] F. H. Post, B. Vrolijk, H. Hauser, R. S. Laramee, and H. Doleisch. The state of the art in flow visualization: Feature extraction and tracking. *Computer Graphics Forum*, 22(4):775–792, 2003.