Department: Head
Editor: Name, xxxx@email

Information-Theoretic Exploration of Multivariate Time-Varying Image Databases

Humayra Tasnim
Department of Computer Science, UNM

Soumya Dutta
Information Sciences, LANL

Terece L. Turton
Information Sciences, LANL

David H. Rogers
Information Sciences, LANL

Melanie E. Moses
Department of Computer Science, UNM

***Abstract*—Modern scientific simulations produce very large datasets, making interactive exploration of such data computationally prohibitive. An increasingly common data reduction technique is to store visualizations and other data extracts in a database. The Cinema project is one such approach, storing visualizations in an image database for post hoc exploration and interactive image-based analysis. This work focuses on developing efficient algorithms that can quantify various types of multivariate dependencies existing within multi-variable datasets. It applies specific mutual information measures for the quantification of salient regions from multivariate image data. Using such information measures, the opacity of the images is modulated so that the salient regions are automatically highlighted and the domain scientists can interactively explore the most relevant regions for scientific discovery.**

IMAGE-BASED data reduction techniques have emerged as one of the viable solutions to minimize the size of the stored data so that it can be analyzed and visualized interactively post hoc by the application scientists [1]. Storing large-scale three-dimensional multivariate simulation

Published by the IEEE Computer Society

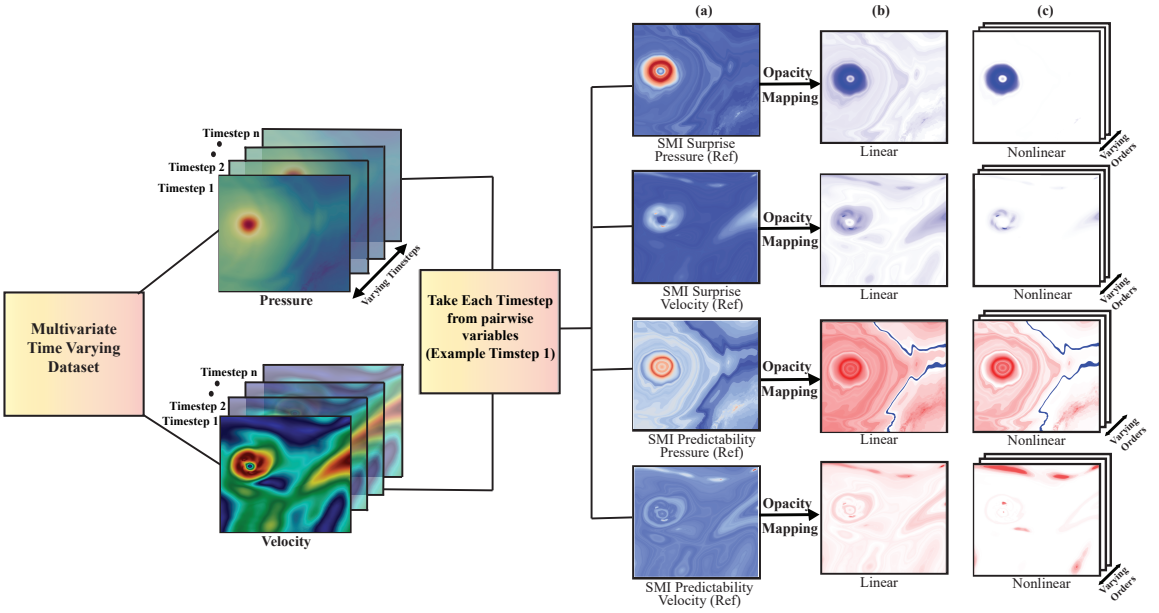

Figure 1. An illustrative diagram of our workflow. Here we have chosen the variables pressure and velocity from the Hurricane Isabel dataset to demonstrate the steps in our technique. Specific mutual information (SMI) measures: Surprise and Predictability are applied on the variable pairs and corresponding images are shown in column (a). After modulating the opacity using linear and nonlinear mapping functions, images with salient regions are analyzed as shown in columns (b) and (c) respectively.

datasets in the form of an indexed image database, called a Cinema Database[1] [2], facilitates exploration of the large-scale scientific data efficiently without overwhelming the users. These Cinema databases are ideally generated in situ, i.e., when the simulation is running on the supercomputer and the data is not yet moved to the disks. Instead of keeping the raw data, Cinema databases are stored onto disk as a proxy for the data, capturing various types of visualizations of the data. Later during offline analysis, the Cinema databases can be explored interactively to analyze the data in the image space. The success of this approach has been shown in many application domains [2], [3].

Even though Cinema databases result in a significant amount of data reduction, such databases still consist of multiple variables, timesteps, visualization parameters, etc. Hence, efficient image-based data analysis and visualization algorithms are necessary to find salient data features automatically so that the domain experts do not have to manually explore them. This problem becomes more challenging when the experts want to analyze features in the multivariate spatiotemporal domain to study their interaction pattern. In many scientific applications, variables collectively show association/dissociation relationships and such properties are often correlated to a physical phenomenon in the data. For example, in hurricane simulation data, low-pressure and low-velocity regions are characterized as the hurricane eye, indicating the strength of the storm. Therefore, multivariate analysis techniques are essential to efficiently detect association/dissociation relationships in image databases. Ideally, these relationships should be visually incorporated to the image database to support further interactive exploration for new scientific discovery.

In this work, we propose an information-theoretic analysis framework that works on multivariate time-varying Cinema databases and performs automatic identification of salient regions given a pair of variables. The technique uses *specific mutual information measures* (SMI) that are a decomposition of traditional mutual information so that the information content of specific data values can be quantified. Each SMI measure

[1]https://cinemascience.github.io

captures a unique multivariate property of the data. Using the strength of these SMI measures, the opacity of the images is modulated during visual analysis so that the important spatial regions are highlighted automatically and the users can quickly focus on them while exploring the Cinema databases. The analysis results are presented interactively using a web-based visual-analytics tool, CinemaView[2], which allows side-by-side interactive comparison of analysis results. The efficacy of the proposed framework is demonstrated by applying it to scientific simulation datasets from weather and combustion sciences.

The contributions of our work are twofold:

- We propose a new technique to perform automatic feature analysis in multivariate time-varying scientific data. Our image-based representations of the 3D spatiotemporal data helps reducing the overhead of the analysis significantly.
- We propose an information-theoretic opacity mapping technique to highlight the statistically salient regions in the data considering pairs of variables.

RELATED WORKS

In this section, we present a comparative discussion of the existing related works and indicate how our work is different. Information theory [4] have been used successfully for solving problems across many computational domains[5], [6]. Instead of using traditional mutual information, the use of various decomposition of mutual information, called specific mutual information (SMI), have gained significant attraction in recent years. By applying SMI, Bramon et al. showed that multi-modal 3D medical datasets can be fused into a single dataset [7]. In another work, Bramon et al. used mutual information to design color transfer function for medical data [8]. To analyze uncertainty of isosurfaces in scientific 3D data, Biswas et al. [9] used SMI and Dutta et al. extended this work into time-varying domain [10]. In contrast to the above works, in this work, we have focused on 2D image-based databases, generated from multivariate time-varying simulations, where our primary focus is to use SMI to automatically first detect the statistically salient regions considering images from variable pairs and then use the SMI values at each pixel location to define opacity values so that the salient regions are automatically highlighted. These images will be ideally generated during the simulation run, i.e., in situ, and as these simulations can have many variables and hundreds to thousands of time steps, we believe that our approach can significantly accelerate the multivariate analysis for the domain scientists by providing them an image-based time-varying summary of simulation variable interactions where the salient regions are automatically highlighted.

[2]https://github.com/cinemascience/cinema_view

OVERVIEW

Our aim is to develop an interactive analysis technique to enable scientists to explore salient regions in time-varying multivariate datasets. The images in the Cinema database are derived from three dimensional simulation data for each variable over multiple timesteps. To study the relationship among multiple variables, we use *specific mutual information* (SMI) to provide information about a target variable based on the knowledge of a specific scalar value of another reference variable. We employ two SMI measures to explore multivariate interaction between variable pairs and use the SMI values to design opacity mapping for the images to highlight statistically salient regions automatically. A workflow of the proposed framework is presented in Figure 1.

INFORMATION-DRIVEN FRAMEWORK FOR MULTIVARIATE FEATURE EXPLORATION

Cinema Database and Image Format

To generate the Cinema database images, 2D slice rendering is applied to the 3D scalar valued variables. Instead of applying a transfer function via a colormap and storing the RGB valued images, we use perspective projection on the 2D slice of the 3D data so that each pixel stores the corresponding value of scalar data [3]. Such images are called *float images* and are stored using standard PNG format. This also allows us to compute the SMI measures directly using the raw data values rather than data distorted by an underlying colormap. A colormap can then be applied post hoc. In Figure 2, we show examples of

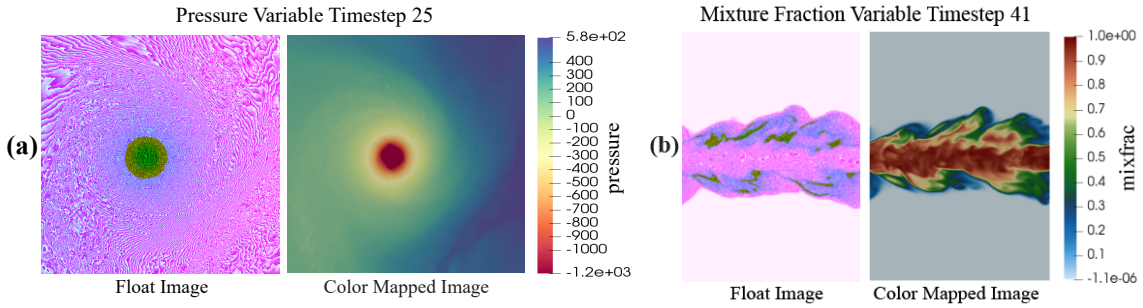

Figure 2. Visualization of float and colored images. 2(a) presents float image and the corresponding colored image using the colorbar shown on right for the *pressure* variable from the Hurricane Isabel dataset. 2(b) presents an example of the *mixture fraction* variable from the Turbulent combustion dataset.

the float images and corresponding color mapped images that are used in this work.

Specific Mutual Information Measures

The key factor in this work is determining the degree of association among the different variables in order to identify and highlight salient regions. Because scientific data often has nonlinear dependencies between variables, any correlation analysis technique must handle nonlinear cases. There are several correlation analysis techniques available for measuring variable relationship. Mutual Information (MI) is one of the well-known measures to quantify the mutual correlation between two variables. MI's ability to capture nonlinear dependency between variables makes it a better choice than a more typical approach such as Pearson's correlation. Mutual information quantifies the total amount of information overlap between two variables, i.e., if we observe a certain variable, then MI tells us how much uncertainty has been reduced regarding the information of another variable. Given two random variables X and Y, MI $I(X,Y)$ is formally defined as:

$$\mathrm{I}(X,Y) = \sum_{y\in Y}\sum_{x\in X} \mathrm{p}(x,y)\log\frac{\mathrm{p}(x,y)}{\mathrm{p(x)\,p(y)}} \qquad (1)$$

where $\mathrm{p}(x)$ and $\mathrm{p}(y)$ are the probabilities of occurrence of values x for X and y for Y respectively and $\mathrm{p}(x,y)$ is the joint probability of occurrence of values x and y together.

MI quantifies the total association or disassociation between two variables and provides a single value. Since we aim to extract salient regions, we need a measure that can provide us with information related to individual scalar values. Traditional MI can be further decomposed into specific mutual information (SMI) measures to quantify individual data values' contribution towards such association or disassociation. For specific scalar values $x \in X$, SMI computes the information content of x when another variable Y is observed. In this case, X is called the reference variable and Y is called the target variable. Knowledge about the scalar values in the reference variable can increase knowledge about the target variable. This increase in information or decrease in uncertainty helps in identifying important regions in the float-image data. MI can be decomposed in multiple ways to obtain several SMI measures and we focus on two such SMI measures, Surprise and Predictability, [7], [11] for finding different types of multivariate characteristics between variable pairs.

SMI measure Surprise: $I_1(x;Y)$

The *Surprise* measure quantifies the change in the information content in the occurrences of the target variable after observing individual scalar values of the reference variable. This measure has the potential of providing information which would seem improbable otherwise, hence the name surprise [7], [11]. The regions where data values have higher surprise values can be informative. For two random variables X and Y, surprise is denoted as I_1 and presented as:

$$\mathrm{I}_1(x;Y) = \sum_{y\in Y} \mathrm{p}(y|x)\log\frac{\mathrm{p}(y|x)}{\mathrm{p(y)}} \qquad (2)$$

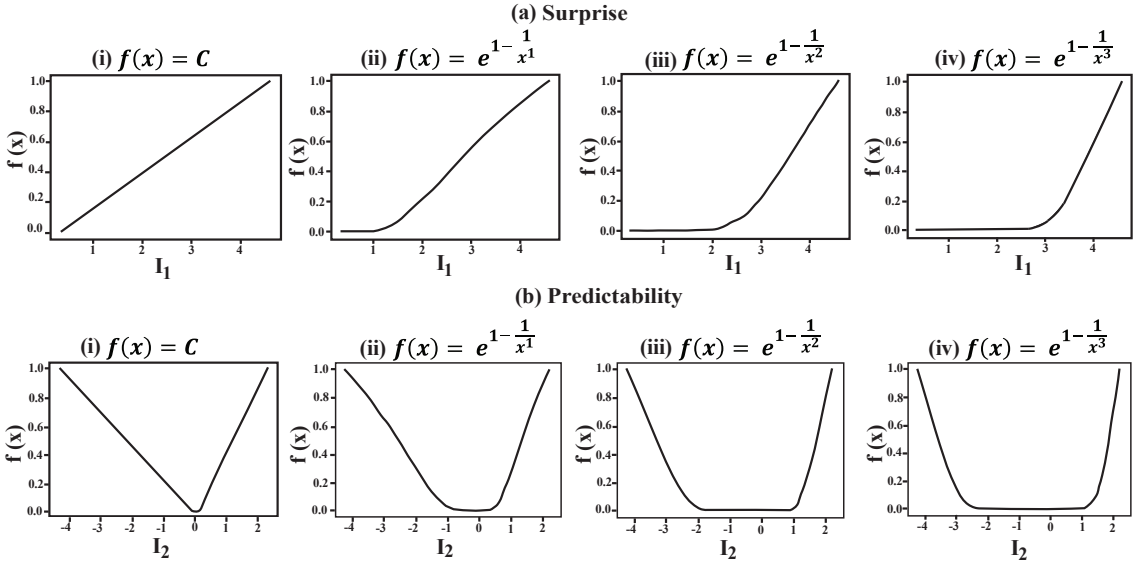

Figure 3. Function plots of the opacity mapping for modulating transparency in the images. Upper row 3(a), presents plots from SMI measure surprise (I_1) and lower row 3(b), presents plots from SMI measure predictability (I_2). Column (i) represents linear mapping and columns (ii), (iii) and (iv) represent increasing order of nonlinear mapping. x-axis of the plots shows the values from the SMI measure and y-axis shows the mapped values from the corresponding functions.

where $x \in X$ is the reference variable and $y \in Y$ is the target variable. $\mathrm{p}(y)$ is the probabilities of occurrence of values y for Y and $\mathrm{p}(y|x)$ is the conditional probabilities of values y given values x. Surprise is always positive as it is the distance between $\mathrm{p}(y|x)$ and $\mathrm{p}(y)$. A high $I_1(x;Y)$ implies that after observing the reference variable x, some low probability values of $y \in Y$ have become more probable. This surprising element is potentially informative for our analysis.

SMI measure Predictability: $I_2(x;Y)$

The *Predictability* measure provides us with the amount of increase/decrease in uncertainty about the target variable after observing the reference variable [7], [11]. This quantification of the uncertainty change helps to identify statistically significant regions in the images. Predictability is denoted as I_2 and can be computed as:

$$\mathrm{I}_2(x;Y) = -\sum_{y\in Y} p(y)\log p(y) + \sum_{y\in Y} p(y|x)\log p(y|x) \tag{3}$$

where $x \in X$ is the reference variable and $y \in Y$ is the target variable. $\mathrm{p}(y)$ is the probabilities of occurrence of values y for Y and $\mathrm{p}(y|x)$ is the conditional probabilities values y given values x. Based on the amount of information increase and decrease, I_2 can be both positive and negative. A high positive $I_2(x;Y)$ value indicates that the uncertainty of target variable Y has decreased when value x is observed. On the other hand, a high negative $I_2(x;Y)$ value indicates that the uncertainty of target variable Y has actually increased. According to information theory, data values that are less probable or unpredictable contain more information representing salient regions in the data with diverse characteristics that are worth deeper exploration. Therefore, the surprise and predictability measures provide different statistically meaningful results, an important consideration in the workflow.

SMI-driven Opacity Mapping Functions

These two SMI measures can now be applied to the image data to identify and highlight statistically salient regions. Since each pixel in the data has a scalar value, SMI measures can be estimated at every spatial pixel location. Note that high surprise regions and high / low predictable regions indicate salient variable relationships. We

want to emphasize such regions where statistically significant multivariate properties exist between the selected variable pair. One of the ways to highlight the regions is by modulating the opacity channel of the image. This suppresses unimportant pixel values while directing focus to important regions. In the following, we show how different types of opacity mapping functions for SMI values can be used to automatically highlight important regions in the images. The design goal of such opacity functions is to make the regions containing high SMI values more opaque so that they are clearly visible and suppress regions with low SMI values by making them transparent. The choice of opacity mapping functions is quite broad and we consider linear and nonlinear mapping functions.

Linear Mapping Strategy of SMI Values

A linear mapping function can be trivially designed. We normalize the values of I_1 and I_2 in the range of $[0,1]$ using the following linear function.

$$f(x) = Constant \tag{4}$$

As shown in Figure 3, for the surprise measure, I_1, 3a(i) shows a linear relationship representing the I_1 values between $[0,1]$ for a pair of variables. Since predictability, I_2, produces both positive and negative values, we model them separately. We normalize positive values between $[0,1]$ and negative values between $[-1,0]$. Combining both at 0, we get a 'V-shaped' plot, as shown in Figure 3b(i). By designing linear mapping functions such as these, lower SMI valued or unimportant regions will be transparent and higher SMI valued or informative regions will become opaque.

Nonlinear Mapping Strategy of SMI Values

The linear mapping strategy computes opacity value as a linear function of SMI values. However, this may not provide sufficient differentiation in the opacity to highlight the most salient regions. In order to design a mapping strategy where the higher SMI valued regions are clearly visible by further suppressing the low valued regions, we introduce nonlinear mapping functions, where the transparency value mapping can be modulated exponentially, giving us more control during analysis. We define the following nonlinear exponential function:

$$f(x) = e^{1-\frac{1}{x^a}}; a >= 1 \tag{5}$$

where a is the exponential control parameter. As a increases, higher SMI values are assigned higher exponential weight. a provide a control parameter that a user can use to set a threshold on the measures that are improtant for a specific analysis. Figure 3, columns(ii), (iii) and (iv), illustrates how the function changes with increased values of a from 1 to 3. In the case of I_1, as a increases, the plot gets steeper by assigning less weight to lower values and more weight to higher values. For example, in the case of Figure 3a(iv), the regions with highest I_1 values will be most opaque making anything below threshold transparent, thus highlighting the significant regions in the images.

This approach is extended for the I_2 analysis by using the function separately for positive and negative values. As seen in Figures 3b(ii), b(iii) and b(iv), with higher orders of a, the V-shape from the linear mapping becomes more 'U-shaped' with steepening curves emphasizing the most significant positive and negative I_2 values.

With the parameter a, the user can set the opacity threshold for results useful to their specific analysis and achieve control over the images they want to visualize for further exploration.

RESULTS

The results of our work are presented using an interactive visual analytics tool, CinemaView, to study salient regions in image datasets. CinemaView is a browser-based viewer that allows interactive exploration of image databases stored as a Cinema database. Figures 4 and 5 show the user interface of the CinemaView tool. Figure 4(a) shows the color mapped ground truth images of two selected variables, pressure and cloud, followed by the images representing the analysis of the variables using surprise (I_1) and predictability (I_2) as opacity mapping functions. Images containing both linear and nonlinear mapping can be visualized simultaneously using this tool as shown in Figure 4(a). In this study, we present results by using order up to 3 for the nonlinear opacity mapping functions. The right panel of the CinemaView interface provides interactive widgets that can be used to adjust image size and

to explore the results over time. There is a drop-down menu where the user can select the dataset to view. CinemaView is intuitive and user-friendly and it allows interactive exploration of multiple image databases simultaneously in a side-by-side fashion. Users can easily compare and contrast the relationships among multiple variables and study their evolution over time (supplementary video).

Hurricane Isabel Dataset

Hurricane Isabel data was produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF). This dataset consists of 13 variables and 48 timesteps with a spatial resolution of $250 \times 250 \times 50$ for a single timestep. In this work, we show analysis results obtained using the pressure and cloud variables.

Figure 4(a) presents analysis results for timestep 7. The pressure is the reference variable and cloud is the target variable. Thus the specific mutual information measures are calculated for values of pressure. After computing I_1 measures, the results are stored as images for visual analysis. Since each pixel in the raw data has a pressure value and each pressure value has an associated surprise (I_1) value, we create a new image where each pixel contains the I_1 value and the opacity at each location is also controlled by a linear/nonlinear mapping function using the associated surprise values. This is then repeated for each timestep. The corresponding opacity mapping functions used to modulate the opacity for timestep 7 are shown in Figure 4(b), where the goal is to highlight regions that have high surprise value. As shown in Figure 4(b), we modulate the order of the opacity function so that we can emphasize regions with high magnitude of I_1 values.

In Figure 4(a), the high I_1 valued regions are presented with different shades of blue where the different shades indicate the opacity modulated regions with darker blue depicting higher surprise values. From the I_1 linear mapping results, we can observe that the areas around the hurricane eye are highlighted as having high I_1 values and indicate that such regions have become more probable after the cloud variable is observed. These regions coincide with the hurricane eyewall – a salient region in the pressure data. It is also observed that, by increasing the ordering of the nonlinear mapping, we can refine the most significant and surprising regions around the hurricane eyewall.

The second row of Figure 4(a) (except the first image) presents I_2 analysis results. As the I_2 values can be both positive and negative, for visualization purposes, those regions are highlighted using shades of blue and red. Blue and red indicate negative and positive I_2 values, respectively. From the I_2 analysis results, we see that the hurricane eye region is red (positive I_2) which means it is a highly predictable region when pressure and cloud variables are analyzed. It is known that in the hurricane eye region, pressure values are typically low and cloud values are mostly homogeneous and thus such region is detected as a predictable region. If we focus at the region around the hurricane eye's boundary, we find that a region is identified as uncertain and has negative predictability values and so has blue color. This is also a consistent observation since this region is known as the eyewall and the target/observed variable cloud has high variability and so is less predictable. Finally, moving away from the hurricane eyewall, the cloud values again become less varying and such regions are detected as more predictable regions (red color) away from the hurricane eye. The white regions in these images indicate regions where both the positive and negative I_2 values are relatively low and so they are transparent. From the predictability plots in Figure 4 b(i), b(ii), b(iii) and b(iv), the white areas represent the parts where the 'V-shape' flattens into 'U-shape' as we increase the order of the nonlinear mapping. As the order is increased, stronger predictable and uncertain regions become highlighted as significant regions.

Turbulent Combustion Dataset

The Turbulent combustion simulation data is made available by Dr. Jacqueline Chen at Sandia Laboratories through the US Department of Energy's SciDAC Institute for Ultrascale Visualization. This dataset has 5 scalar variables and 122 timesteps with a spatial resolution of $240 \times 360 \times 60$ for a single timestep. During combustion process, fuel and oxidizer react and the flame exists where fuel and oxidizer are in stoichiometric proportions [12]. The mixture

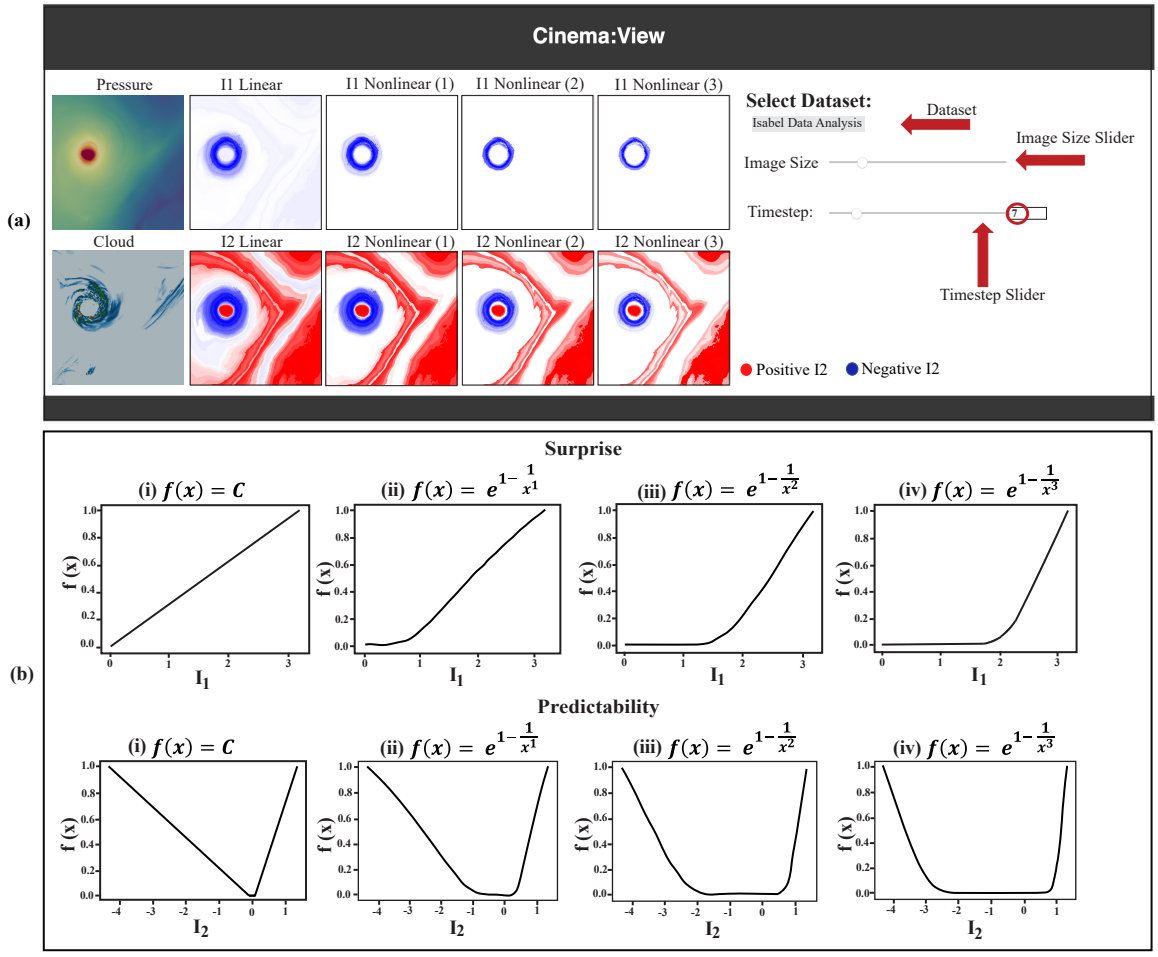

Figure 4. (a) presents salient regions between pressure and cloud variable analysis from the Hurricane Isabel dataset at timestep 7 using CinemaView. The first images of each row are the color mapped images of the reference variable pressure and target variable cloud. The first row shows the combined analysis using surprise (I_1) as the opacity mapping function. The blue regions represent detected salient areas. The second row shows combined analysis using predictability (I_2) for the opacity mapping function. The red regions represent positive predictability and the blue regions represent negative predictability. The elements annotated with red arrows and circles show the interactive tools of CinemaView. (b) presents function plots of the opacity mapping for modulating transparency in the corresponding images. Upper row shows surprise (I_1) plots and lower row shows predictability (I_2) plots. Column (i) represents linear mapping and columns (ii), (iii) and (iv) represent increasing order of nonlinear mapping. x-axis of the plots shows the values from the SMI measure and y-axis shows the mapped values from the corresponding functions.

fraction is an important variable in this dataset that indicates the fraction of mass at fuel stream origin. So, we have used the mixture fraction (mixfrac) as the reference variable and hydroxyl radical (Y_OH) as the target variable since both of these can be used to study the flame regions of the simulation [12]. By analyzing the interacting relationship of these two variables, important features can be studied and detailed information about the combustion process can be gleaned.

In Figure 5, we show results from timesteps 5, 41, and 80 as three different representative timesteps, highlighting three stages of the time-varying simulation. Timestep 5 in Figure 5(a) shows the initial state of the combustion variables interacting when the flames just started burning.

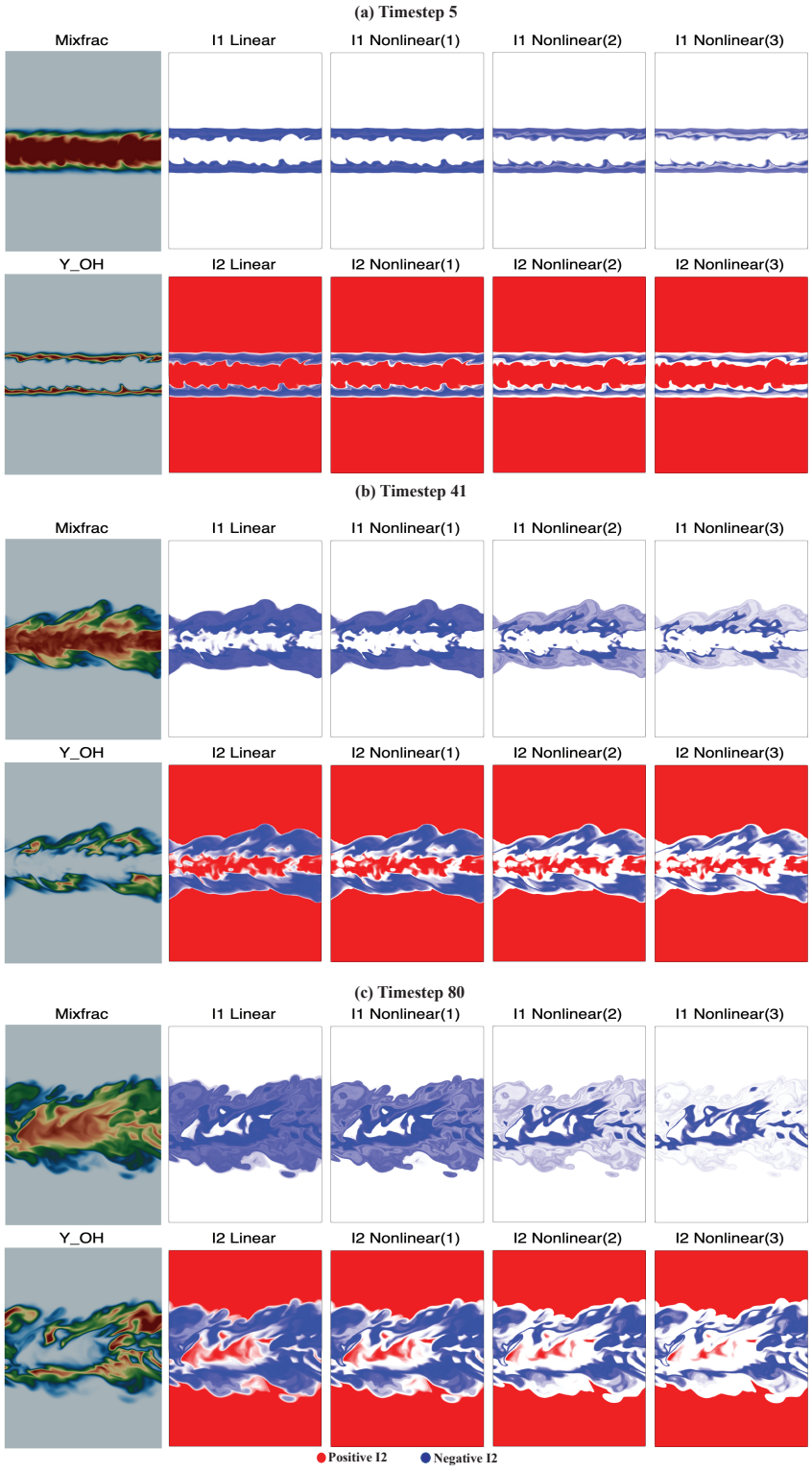

Figure 5. Salient regions between reference mixfrac and target Y_OH variable analysis from Turbulent combustion dataset at (a) timestep 5, (b) timestep 41 and (c) timestep 80. The first images of each row are the color mapped images of the reference variable mixfrac and target variable Y_OH. After the color mapped image, the top row from every timestep shows combined analysis of the variables using surprise (I_1). The blue regions represent the salient areas (flames). Similarly, the bottom row shows analysis using predictability (I_2). Red and blue regions represent positive and negative predictability respectively.

Timestep 41 in Figure 5(b) represents an intermediate time when the combustion process is active and and finally, Figure 5(c) presents the result from a later timestep 80 when the flame has expanded. From these three figures, the salient regions clearly change their shape and position over time, indicating how this method is able to capture temporal changes.

The salient regions detected from the I_1 analysis signifies the areas where the combustion process is happening around the flames. I_1 analysis shows blue regions identifying the areas with combustion flames. As we proceed to nonlinear mapping with increased order, higher I_1 valued regions get highlighted with dark blue and lower I_1 valued regions become transparent with lighter shades of blue, displaying the flame regions in a more refined manner.

From the I_2 analysis results, we see two types of regions, blue and red. As before, the blue regions show the locations where the values of the target/observed variable (Y_OH) are not homogeneous when observing the reference variable mixfrac. From all of the three timesteps, we find that the blue regions coincide well with the regions detected by the I_1 analysis, i.e., the regions where the flame is. In this region, the complex chemical reactions take place and so is hard to predict. From our I_2 analysis, such regions are detected as having negative I_2 values which means such regions have higher uncertainty, therefore, less predictable. On the other hand, the red regions in these results show predictable regions of Y_OH when mixfrac is observed. The two outer red regions (the top and the bottom part) are the background regions where the combustion is not happening and hence the data values are mostly homogeneous. As a result, such regions are correctly identified as the highly predictable regions. The red regions in between two blue uncertain regions indicate that at the center of the simulation, there are some places where the variable Y_OH is more predictable and hence has positive I_2 values. It is also observed that as we increase the order of our opacity mapping function for both linear and nonlinear approaches, we can obtain further refined views of these predictable and uncertain regions where the darker (more opaque) regions indicate locations with higher magnitude of I_2 values. From these analysis results, we observe that both I_1 and I_2 analysis on the Turbulent combustion dataset bring out salient regions that the user can further study in more detail for exploring important characteristics of these variables over space and time.

Conclusions and Future Work

Our work successfully enables scientists to explore and extract salient regions in time-varying multivariate data sets. This technique is generalizable and is not limited to the data sets analyzed in this work. In future work, we plan to accelerate the computation of the information measures by using GPU-based parallel computing. The computation for each timestep can be further parallelized since the computation at each timestep is independent. We also plan to design more sophisticated optimization functions for opacity mapping. Instead of generating different orders for opacity modulation, an optimization-based approach could generate regions that are most useful to the domain scientists.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). This research is released under LA-UR-21-21824 version 2.

REFERENCES

1. S. Li, N. Marsaglia, C. Garth, J. Woodring, J. Clyne, and H. Childs, "Data reduction techniques for simulation, visualization and data analysis," *Computer Graphics Forum*, vol. 37, no. 6, pp. 422–447, Sep. 2018. [Online]. Available: https://doi.org/10.1111/cgf.13336
2. J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D. H. Rogers, and M. Petersen, "An Image-Based Approach to Extreme Scale in Situ Visualization and Analysis," *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, vol. 2015, no. January, pp. 424–434, 2014. [Online]. Available: https://doi.org/10.1109/SC.2014.40
3. D. Banesh, J. A. Schoonover, J. P. Ahrens, and B. Hamann, "Extracting, Visualizing and

Tracking Mesoscale Ocean Eddies in Two-dimensional Image Sequences Using Contours and Moments," in *Workshop on Visualisation in Environmental Sciences (EnvirVis)*, K. Rink, A. Middel, D. Zeckzer, and R. Bujack, Eds. The Eurographics Association, 2017. [Online]. Available: https://doi.org/10.2312/envirvis.20171103

4. T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*. Wiley-Interscience, 2006. [Online]. Available: https://doi.org/10.1002/047174882X
5. M. Chen, M. Feixas, I. Viola, A. Bardera, H.-W. Shen, and M. Sbert, *Information Theory Tools for Visualization*. USA: A. K. Peters, Ltd., 2016. [Online]. Available: https://doi.org/10.1201/9781315369228
6. H. Tasnim, G. Fricke, J. Byrum, J. Sotiris, J. Cannon, and M. Moses, "Quantitative measurement of naïve T cell association with dendritic cells, FRCs, and blood vessels in lymph nodes," *Frontiers in Immunology*, vol. 9, no. JUL, 2018. [Online]. Available: https://doi.org/10.3389/fimmu.2018.01571
7. R. Bramon, I. Boada, A. Bardera, J. Rodríguez, M. Feixas, J. Puig, and M. Sbert, "Multimodal data fusion based on mutual information," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1574–1587, 2012. [Online]. Available: https://doi.org/10.1109/TVCG.2011.280
8. R. Bramon, M. Ruiz, A. Bardera, I. Boada, M. Feixas, and M. Sbert, "Information theory-based automatic multimodal transfer function design," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 870–880, 2013. [Online]. Available: https://doi.org/10.1109/JBHI.2013.2263227
9. A. Biswas, S. Dutta, H. W. Shen, and J. Woodring, "An information-aware framework for exploring multivariate data sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2683–2692, 2013. [Online]. Available: https://doi.org/10.1109/TVCG.2013.133
10. S. Dutta, X. Liu, A. Biswas, H.-W. Shen, and J.-P. Chen, "Pointwise information guided visual analysis of time-varying multi-fields," in *SIGGRAPH Asia 2017 Symposium on Visualization*. ACM, 2017. [Online]. Available: https://doi.org/10.1145/3139295.3139298
11. M. R. Deweese and M. Meister, "How to Measure the Information Gained From One Symbol," *Network*, vol. 10, no. 2, pp. 123–32, 1999. [Online]. Available: https://doi.org/10.1088/0954-898X_10_4_303
12. H. Akiba, K. Ma, J. H. Chen, and E. R. Hawkes, "Visualizing multivariate volume data from turbulent combustion simulations," *Comput. Sci. Eng.*, vol. 9, no. 2, pp. 76–83, 2007. [Online]. Available: https://doi.org/10.1109/MCSE.2007.42

Humayra Tasnim, is pursuing her Ph.D. in Computer Science at the University of New Mexico. She works in Professor Moses's Biological Computation Lab. She received her M.S. in Computer Science and Engineering from University of Dhaka, Bangladesh in 2015. Her research interests include computational biology, modeling complex systems; digital image processing; statistical data analysis, exploration and visualization; machine learning; artificial intelligence. Contact her at htasnim30@unm.edu.

Soumya Dutta, is a staff scientist at Los Alamos National Laboratory. He received his Ph.D. degree in Computer Graphics and Visualization from the Ohio State University in 2018. His research interests include big data science & visualization, statistical techniques for big data, in situ analysis, machine learning for visual computing, high performance visualization, and time-varying multivariate data analysis. Contact Soumya at sdutta@lanl.gov.

Terece L. Turton, is a staff scientist at Los Alamos National Laboratory. Her current research interests include in situ workflows for exascale simulations, scientific visualization, and user evaluation. She received her Ph.D. in Physics from the University of Michigan in 1993. She is a member of the IEEE and the IEEE Computer Society. Contact her at tlturton@lanl.gov.

David H. Rogers, is the team lead of the Data Science at Scale Team at Los Alamos National Lab. He joined LANL in 2013, after a decade of leading the Scalable Analysis and Visualization Team at Sandia National Labs, where he was instrumental in bringing in situ analysis and visualization into production. He now focuses on interactive analysis tools that integrate design, scalable analytics and principles of cognitive science to promote scientific discovery. Contact him at dhr@lanl.gov.

Melanie E. Moses, is a Professor of Computer Science at the University of New Mexico and External Professor at the Santa Fe Institute. She received her PhD in Biology from the University of New Mexico in 2005. She studies the scaling properties of complex biological and computational systems, particularly how the processing of energy and information changes with scale in natural and engineered computational systems. Her current research focuses on: i) computational immunology and understanding

how viruses and immune cell movement through the complex spatial structure of the lung determines viral dynamics; ii) bio-inspired swarm robotics and how collective algorithms can guide effective autonomous collective search in real-world environments; and iii) ethical artificial intelligence. Contact her at melaniem@unm.edu.