Editor: Theresa-Marie Rhyne

Linking Multidimensional Feature Space Cluster Visualization to Multifield Surface Extraction

Lars Linsen, Tran Van Long, and Paul Rosenthal, *Jacobs University*

Data sets resulting from physical simulations typically contain a multitude of physical variables. So, visualization methods should take into account the entire multifield volume data rather than concentrate on one variable. We have developed a visualization approach based on surface extraction from multifield volume data. The extracted surfaces segment the data with respect to an underlying multivariate function. Decisions on segmentation properties are based on the analysis of a multidimensional feature space. We perform feature space exploration using automated multidimensional hierarchical clustering. The hierarchical clusters appear as a cluster tree in a 2D radial layout. In this layout, the user can select clusters of interest. A selected cluster in feature space corresponds to a segmenting surface in object space. On the basis of the segmentation property induced by the cluster membership, we extract surfaces from the volume data.

We applied this approach to the 2008 IEEE Visualization Design Contest data set (see the sidebar, next page). This time-varying data set simulates the propagation of an ionization front instability. It includes a variety of attributes, including density, temperature, mass abundances of eight chemical species, and velocity. Each of the 200 time steps is 37 × 106 points, resulting in 1.7 Gbytes of data per time step. For performance purposes, we spatially sampled the data. Because no numerical details were available on how the simulation was produced, the best way to produce an unbiased sampling was to sample randomly. As a result, we obtained a set of uniform randomly distributed points in the volumetric space, where each point carries multiple properties. One million points per time step were sufficient to robustly reproduce clustering results. We directly extracted surfaces from the data by computing individual points on the surface, supported by an efficient neighborhood computation. We rendered the extracted surface points using point-based rendering. Our approach combines methods in scientific (spatial) visualization for object-space operations with methods in information (nonspatial) visualization for feature-space operations.

Hierarchical Density-Based Clustering

For large data sets with many records, clustering has proven extremely useful. Clustering partitions a data set into subsets of similar observations. Each subset is a cluster, which consists of observations that are similar among themselves and dissimilar to observations of other clusters. Cluster analysis for multidimensional data includes finding areas where individual records group together to form a cluster.

We analyzed the data by looking into the density of the distribution of the multidimensional feature space. Areas of locally high density form clusters. Density computations must be performed in the original high-dimensional space, because projection to lower-dimensional space imposes assumptions and thus distorts the data. We based our density computations on subdividing the domain into hypercubes and evaluating density functions over the spatial subdivision. The clustering is done hierarchically by iteratively splitting each density cluster into subclusters (if any). We record any splitting of a cluster into multiple higher-density subclusters.[1]

Cluster Tree Visualization

To display the density clusters' hierarchical structure, we generate a hierarchical density cluster tree and show it using a 2D radial layout. We place the root in the origin of a unit circle and evenly distribute the leaves on the circle. We place the internal nodes on concentric circles corresponding to their depth. We visually represent the clusters as colored disks whose radii encode the size of the clusters and whose colors encode the position in the radial layout using the hue, saturation, value (HSV) color space (see Figure 1a, page 87).

Linked Views

The cluster tree visualization supports comprehension of the clusters' structure within the multidimensional feature space. To also support the

IEEE Visualization 2008 Design Contest

Amit Chourasia, ***University of California, San Diego***

Russell M. Taylor II, ***University of North Carolina at Chapel Hill***

The IEEE Visualization Design Contests aim to foster development of new visualization tools, techniques, and solutions to real-world problems. They provide curated data sets and scientific questions for visualization researchers and students. The scientific questions drive the development of effective visualizations. The curation goes beyond data storage and includes sample programs to read and display the data and sample images to visually validate the processing. All past contest questions, data, and submissions are archived and available at the Web site (http://viscontest.sdsc.edu).

The theme for the 2008 contest was *Multifield 3D Scalar Data*, using an instability simulation of an ionization front data set provided by Michael L. Norman of San Diego Supercomputing Center and Daniel Whalen of Los Alamos National Laboratory. The goal was to understand the formation of galaxies—particularly the effect of shadow instabilities, where radiation ionization fronts scatter around primordial gas.[1,2]

Stars form when clumps in molecular clouds collapse owing to gravity; but if only gravity acted, stars would form so quickly that all the gas in the galaxy would have been converted into stars eons ago, with none being formed today. One agent thought to slow star formation to the rates observed in the Milky Way is turbulence, which supports cloud cores and opposes collapse. The origin of turbulence in molecular clouds is therefore important. Scientists posed the following questions:

- The shadow instability forms in the ionization front when it encounters a spherical bump in the gas that centers on the *x*-axis. Is the instability symmetric around this axis? If not, how is the symmetry broken?
- H_2 enables primeval gas clouds to collapse and form the first stars before galaxies later coalesce. Where is H_2 most prevalent in the simulation?
- How thick are the first fingers of radiation that break through cracks in the shock front?
- We can think of turbulence as a cascade of fluid flow from large to small scales that tends to dissipate the bulk motion of flow. What usually triggers the cascade are shear motions in which fluid parcels slide past one another. Researchers have recently suggested that dynamical instabilities in the ionization fronts of massive stars in the clouds stir up turbulence. Is there any evidence of this in the shadow instability?
- Turbulent flows by themselves don't form H_2; only the presence of free electrons can. If free electrons are present (as signified by $H+$ fractions), however, can turbulence enhance H_2 formation? If so, is it because turbulent eddies create overdensities in which reactions occur more rapidly? If not, is it that even though free electrons are present, the turbulence disrupts $H-$ and H_2+ formation (the key precursors of H_2)?
- General question: what is causing the turbulence?

Submissions included a two-page document describing the solution (how the design addresses the scientific questions, and which software systems and algorithms were used), up to five color images, and at most one movie shorter than 15 minutes showing the visualization. The contest posed no restrictions on hardware or software requirements for the contestants. Providing source code with the submission was optional. The judges received three entries for the contest.

Judging

The judges based their evaluation on two metrics: the visualization's effectiveness and completeness. Effectiveness counted for 80 percent of the total points and was judged by an astrophysicist. The scientist evaluated on a five-point scale, with 5 meaning "I could see the answer immediately and clearly" and 1 meaning "I know the answer already, but I still can't see it in the visualization." The two contest cochairs judged completeness on the basis of whether you could recreate the visualizations (including parameter settings) with the submitted materials.

visualization of the value ranges for the individual dimensions, we employ parallel coordinates. We use linked views to select clusters in the cluster tree representation and display them in a parallel-coordinates representation (see Figure 1b). The applied coloring scheme assures unique color coding. In Figure 1, several clusters have been selected simultaneously. In the radial layout, they're represented by disks colored with higher saturation and exhibiting a small red dot in their centers. In parallel coordinates, we visualize the selected clusters' values with respect to the multivariate field.

In addition, we support another linked view that renders the spatial distribution of clusters in the volumetric object space. Figure 1c shows the volumetric distribution of the points belonging to the selected clusters. The user interacts with the cluster visualization and gets multiple linked views on the data.

Surface Extraction and Rendering

The multidimensional cluster membership induces a segmentation in object space. On the basis of this property, we can extract a surface from the volume

Winner

The winner was "Shadow Clustering: Surface Extraction from Nonequidistantly Sampled Multifield 3D Scalar Data Using Multidimensional Cluster Visualization." It scored 4 or 5 on four of the science questions and was linked to a publication completely describing the implementation. The main section of this article describes this entry's details.

Acknowledgments

Daniel Whalen at Los Alamos National Laboratory and Michael L. Norman at the San Diego Supercomputer Center (SDSC) provided the simulation data. Data Central at SDSC provided data storage and archiving for the visualization contest.

References

1. D. Whalen and M.L. Norman, "Ionization Front Instabilities in Primordial H II Regions," *Astrophysical J.*, vol. 673, issue 2, 2008, pp. 664–675.
2. D. Whalen et al., "Photoionization of Clustered Halos by the First Stars," *First Stars III: First Stars II Conf., AIP Conf. Proc.*, vol. 990, American Inst. of Physics, 2008, pp. 381-385.

Amit Chourasia *is a visualization scientist at the San Diego Supercomputer Center at the University of California, San Diego. Contact him at amit@sdsc.edu.*

Russell M. Taylor II *is a research professor at the University of North Carolina at Chapel Hill. Contact him at taylorr@cs.unc.edu.*

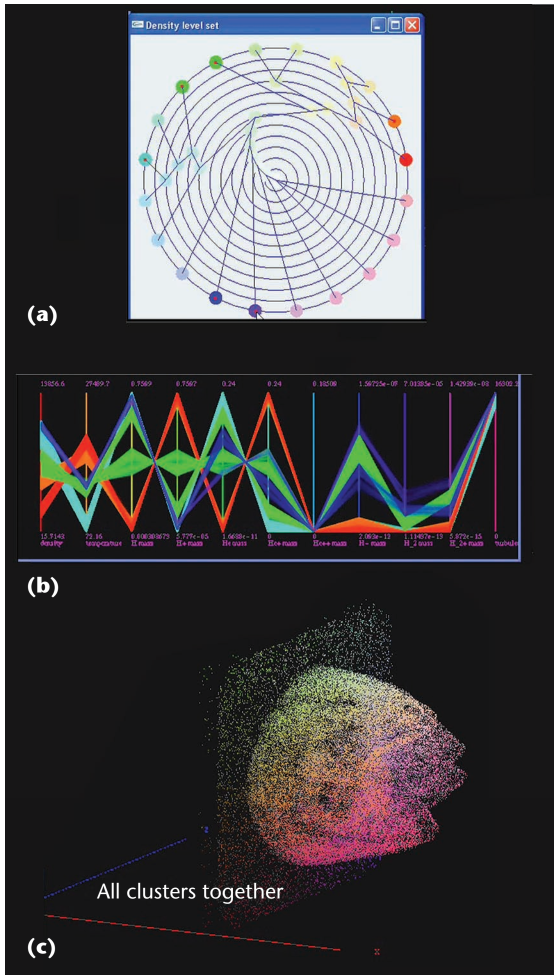

Figure 1. Linked views of (a) the hierarchical cluster tree visualization to (b) parallel coordinates and (c) object-space rendering. The automatically detected hierarchical clusters are interactively selected in the cluster tree visualization and simultaneously displayed in the respective linked views.

data using direct surface extraction.[2] More precisely, we extract the isosurface with isovalue 0.5 to the characteristic function of the cluster. We directly extract our surfaces without prior resampling or grid generation. The surface extraction computes individual points on the surface, supported by a neighborhood computation using *k*-dimensional trees (*k*d-trees) and an efficient indexing scheme.

We use point-based rendering operations to render the extracted surface points. An interactive approach is provided using image-space point cloud rendering.[3] The illuminated surface points are directly rendered on the screen. Image-based filters produce hole-free surface renderings (see Figure 2c, next page).

A second approach to rendering the surface points is splat-based ray tracing,[4] producing renderings with global illumination but not at interactive rates. For each surface point, a circular splat is fit to the surface using a least-squares approach, which also provides a normal map of the splat. The resulting splats are ray traced using the normal map (see Figure 2a).

Scientific Questions

The scientists had specified several questions that they had about the data set, and we designed the system to display the answers to these questions. We describe here how the system can be used to answer each.

Symmetry of the Shadow Instability

First we want to clarify the 3D shape of the observed shadow instability. So, we randomly resampled the data set of time step 90. Then, we clustered it in feature space according to all attributes. We extracted the surface segmenting the

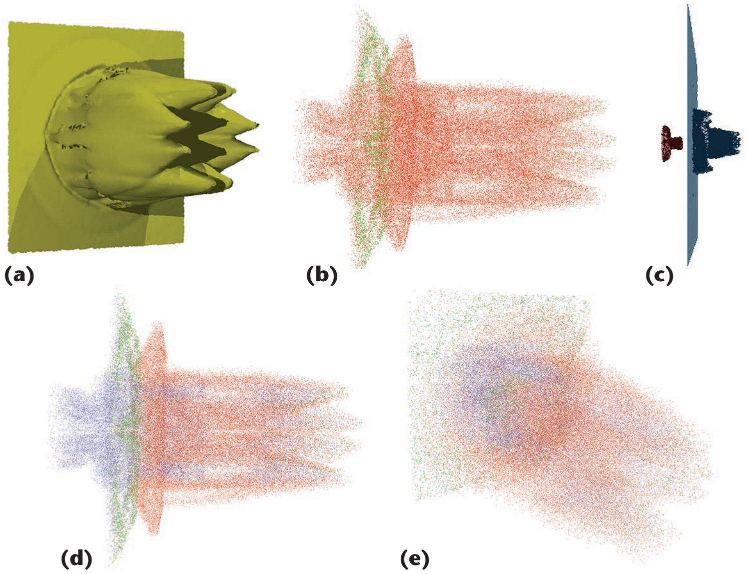

Figure 2. Results of the interactive visual exploration of the ionization front instability propagation: (a) Extracted surface (using splat-based ray tracing for point cloud rendering) exhibiting symmetry of shadow instability, (b) points with highest H_2 prevalence (green) and highest turbulence (red), (c) development of radiation fingers, (d) and (e) multidimensional clustering in feature space, which provides three clusters.

cluster of the ambient gas with a temperature of approximately 72 K from the shocked and ionized gas. This surface represents the front of the ionization process. Figure 2a shows a ray-traced picture of the surface.

Using the same view as in Figure 1, we can clearly see that the shadow instability isn't symmetric around the *x*-axis. Instead, it's symmetric with respect to the $y = 124$ and the $z = 124$ planes. Furthermore, it's symmetric with respect to the two planes that are perpendicular to the *x*-axis and lie diagonal in the *yz*-plane. The symmetry around the *x*-axis is broken mainly by the eight "fingers" representing the instability front.

Prevalence of H_2

We chose time step 99 to show the regions where H_2 is most prevalent. We again clustered the randomly resampled data set in feature space. Figure 2b shows a point rendering of two important clusters. We rendered the points of the cluster with the most H_2 in green. To have a context for this cluster, we also rendered the points of the cluster with highest turbulence in red. These points indicate the zone of the shocked gas in the ionization process.

The picture shows that most of the H_2 is generated at the very beginning of ionization. It's most prevalent directly at the beginning of the zone with shocked gas and high turbulence. The extracted clusters' properties—more precisely, the average attribute values—also show that in the regions where H_2 is most prevalent, most of the He was already ionized to He+ but only some H+ was ionized from H.

Radiation Fingers

To see the thickness of the first fingers of radiation that break through the front, we must study data early in time. We chose time step 10, where this phenomenon first clearly appears. Again, we applied a nonequidistant sampling followed by a multidimensional clustering. Then, we extracted and rendered two segmenting surfaces using image-space point cloud rendering (see Figure 2c).

The bluish surface segments the ambient gas from the rest and gives the context for a better understanding of the picture. The red surface segments the first finger of radiation breaking through. We rendered the picture with a viewport parallel to the *xy*-plane to allow a better judgment of the size. The diameter of the radiation finger is approximately 0.017 parsecs.

The Cause of Turbulence

To understand whether turbulence is stirred up in the front of the shadow instability, we again refer to Figure 2b. As mentioned earlier, we resampled and clustered time step 99 in feature space. To have an indication for turbulence, we introduced the magnitude of the curl for the provided velocity field as an additional data dimension. We rendered the points of the cluster with high turbulence in red.

Clearly, most turbulence is present in the region of shocked gas—that is, directly behind the ionization front. So, we have a strong indication that the shadow instability stirs up the turbulence in the gas.

H_2 Formation and Turbulence

To answer the last two key questions on the cause of turbulence and whether turbulence enhances H_2 formation, we refer to Figures 2d and 2e. The data set for time step 99 was resampled nonequidistantly and clustered in feature space. Then, we extracted three clusters. We rendered these clusters' points with different colors to provide a 3D view of the distribution.

The points of the cluster with the highest H_2 are green. Red and blue points represent high turbulence (see Figure 2b). We split these points into two clusters representing points with low H+ density in red and points with high H+ density in blue. Most of the H_2 is generated at the very beginning of the turbulence cluster, where H+ density is low—that is, where it is lower than 0.2. This indicates that turbulence is essential for the formation of H_2 but that the needed electrons mostly don't come from H^+. Instead, a high density of He^+—that is, greater than 0.15—is observable in the regions of H_2 formation. This also heavily correlates with the presence of H^- and H_2^+.

More precisely, the procedures causing the turbulence give the following impression: When the ionization front reaches the ambient gas with $\rho H \approx 0.76$ and $\rho He \approx 0.24$, both chemical species are ionized at approximately the same speed. However, a high amount of H^+ is converted into H_2 with the help of the free electrons from He^+. This process, and probably also the breakup of H_2, afterward lead to the observable turbulences behind the H_2 front. We can observe this behavior at the front of the instability as well as at the front of the "normal" ionization front.

Multidimensional clustering in feature space can be a powerful tool when appropriate interactive mechanisms are provided to explore the clusters. Our cluster tree visualization supports an intuitive understanding of the data distribution in feature space. When the cluster tree visualization is linked to parallel coordinates, the user can investigate the range of values of each cluster in the individual dimensions. This linked view proved extremely useful for answering questions from the application domain and for validating hypotheses. When linked to object space visualization, the user can observe and analyze the spatial location of the detected clusters of interest.

We provide a visualization of cluster points in object space as well as direct surface extraction from the randomly distributed sample points. The surfaces represent the boundary of an object-space segmentation induced by the feature-space clustering. We extract them in a point cloud surface representation and use point-based rendering techniques for their display. With this linked-view setup, we could answer questions involving both object- and feature-space investigations.

References

1. L. Linsen et al., "Surface Extraction from Multi-field Particle Volume Data Using Multi-dimensional Cluster Visualization," *IEEE Trans. Visualization and Computer Graphics*, vol. 14, no. 6, 2008, pp. 1483–1490.
2. P. Rosenthal and L. Linsen, "Direct Isosurface Extraction from Scattered Volume Data," *Proc. Eurographics/IEEE-VGTC Symp. Visualization*, Eurographics Assoc., 2006, pp. 99–106.
3. P. Rosenthal and L. Linsen, "Image-Space Point Cloud Rendering," *Proc. Computer Graphics Int'l*, Computer Graphics Soc., 2008, pp. 136–143.
4. L. Linsen, K. Müller, and P. Rosenthal, "Splat-Based Ray Tracing of Point Clouds," *J. WSCG* (Int'l Conf. Central Europe on Computer Graphics, Visualization and Computer Vision), vol. 15, nos. 1–3, 2007, pp. 51–58.

***Lars Linsen** is an associate professor at Jacobs University. Contact him at l.linsen@jacobs-university.de.*

***Tran Van Long** is a graduate student at Jacobs University. Contact him at v.tran@jacobs-university.de.*

***Paul Rosenthal** is a graduate student at Jacobs University. Contact him at p.rosenthal@jacobs-university.de.*

Contact editor Theresa-Marie Rhyne at tmrhyne@nscu.edu.