AN EXPLORATION OF MULTIVARIATE TIME-VARYING VOLUME DATASETS USING VOLUMETRIC PARALLEL COORDINATES

Shiori Yasuoka, Teppei Tanaka, Naohisa Sakamoto, Koji Koyamada

Kyoto University, Japan

koyamada.koji.3w@kyoto-u.ac.jp (Koji Koyamada)

Abstract

In this paper, we propose “Volumetric Parallel Coordinates” (VPC) as a technique to explore a multivariate time-varying volume dataset. Recently, this type of a large-scale dataset has often required a high-performance computing environment. In addition, such a dataset often includes multiple variables defined on high-resolution grids and thousands of time steps. We can make use of volume visualization, information visualization or animation techniques, which can handle high-resolution grids, multiple variables and time-varying datasets, respectively. However, there is currently no technique that visualizes a multivariate time-varying volume dataset in a single picture, which is often a good starting point for data exploration. The VPC stacks parallel coordinates that represent the relation between multiple variables at a given time step in order to construct a volume dataset. A volume visualization technique can be used to visualize how the relation between variables changes over time. We apply the VPC to results generated from a liquid atomization simulation to confirm its effectiveness, and we find that the relation between the liquid curvature and the velocity maintains a negative correlation over time.

Keywords: Parallel Coordinates, Particle-Based Volume Rendering.

Presenting Author’s biography

Koji Koyamada received a B.S., M.S. and Ph.D degrees in electronic engineering from Kyoto University, Kyoto, Japan in 1983, 1985, and 1994, respectively. He is a professor at Kyoto University. From 1985 to 1998 he worked for IBM Japan. From 1998 to 2001 he was an associate professor at Iwate Prefectual University. From 2001 to 2003 he was an associate professor at Kyoto University. His research interest includes modeling, simulation and visualization.

1 Introduction

A numerical simulation is an effective tool to reproduce a phenomenon that is difficult to analyze experimentally. For example, a simulation is often utilized to investigate a large-scale phenomenon in space or a chemical experiment highly dangerous to the human body. Thus, a numerical simulation can be conducted in various kinds of research areas to understand a complex physics phenomenon or to design a mechanical product. Recently developed high-performance computers enable more complex, more precise and larger simulation models to perform calculations. As the result, a dataset generated by such computers may include multiple variables defined on high-resolution grids and thousands of time steps. We recognize that there are three key characteristics of such a dataset; namely, the numbers of variables, time steps and resolution. To understand a large-scale dataset with respect to all three characteristics, we require an effective visualization technique. Indeed, there are many existing efficient techniques for visualizing a high-resolution time-varying volume dataset.

To analyze a single relation between variables or a temporal change in a single variable, a statistic technique is often utilized [1, 2]. There are few techniques to handle relations among multiple variables over time. In this paper, we propose the Volumetric Parallel Coordinates (VPC) as a technique to explore a multivariate time-varying volume dataset. The VPC stacks a set of parallel coordinates that represent relations between multiple variables at a given time step in order to construct a volume dataset. Note that a volume visualization technique can be used to visualize how a relation between variables changes over time. The rest of the paper is organized as follows. Section 2 briefly introduces related works on visualization techniques for a multivariate time-varying volume dataset. In section 3, we describe our proposed technique (i.e., VPC) in three subsections. First, we provide a definition of the VPC, and then we describe a technique of converting a multivariate time-varying volume dataset into the VPC. Finally, we demonstrate a graphical user interface in which a user can effectively manipulate the VPC. We believe that a graphical user interface has a considerable impact on data exploration. A good GUI facilitates a user in making discoveries. In section 4, we apply the VPC to results generated from a liquid atomization simulation to confirm the effectiveness of the technique.

2 Related Work

2.1 The Time Histogram

The time histogram [2] is a technique used to analyze a time-varying dataset. A histogram is composed of a bar chart that represents a frequency distribution; it is called a time histogram when the distribution is along a time axis. This technique makes it possible to view a temporal change in the distribution of variable values. A time histogram is displayed as a 3D graph, where the axes are value, frequency and time. The time histogram can be displayed as 2D graph, where axes are value and time. Frequency in this case is represented with brightness values, colors, lights or shades. However, if this technique is applied to a multivariate time-varying volume dataset, the temporal change of only one variable can be obtained.

2.2 Parallel Coordinates

Parallel Coordinates [1] is a technique used to analyze a multivariate dataset. This technique makes it possible to analyze "the relation between variables". In Parallel Coordinates, the variable axes are arranged in parallel so that the variable values of observing points are plotted on the axes, and they are connected using line segments. This technique makes it possible to display the relation between variables at the same time by narrowing the distance between the parallel axes, if the number of variables is large. In addition, this technique method makes it possible to analyze the relation between two consecutive variables intuitively. Fig. 1 shows a typical pattern resulting from Parallel Coordinates. The left image represents the negative correlation because the lines intersect between the two axes and are constricted in the middle like the red line in the image. The center image also represents a negative correlation. However, the right image represents a positive correlation because the lines do not intersect between the two axes. If this technique is applied to volume data, the observing point corresponds to a location at which the variable value is defined. However, when using the Parallel Coordinates technique, the relation between variables in multivariate time-varying volume data can only be analyzed during a single time step.

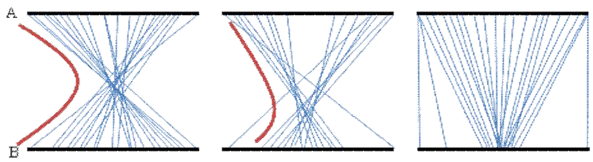

Fig. 1 The typical pattern of Parallel Coordinates (left and center: negative relation, right: positive relation)

2.3 3-D Parallel Coordinates

3-D Parallel Coordinates is an extension of Parallel Coordinates that adds an axis orthogonal to the plane on which the Parallel Coordinates plot is defined. Our VPC can be categorized as a 3-D Parallel Coordinates technique because the additional axis is a time axis. We describe the details of VPC in Section 3. In this section, we introduce three 3-D Parallel Coordinates techniques. These techniques can be distinguished in terms of the use of the additional axis, the dataset used, and the geometry that connects each parallel coordinate along the additional axis.

Honda et al. [3] proposed the 3-D Parallel Coordinates plot, where the additional axis represents the selected variables. This technique aims to clarify a given relation between variables. In this technique, there are two ways to draw polygonal lines. One method involves connecting line segments between adjacent axes at each observing point as a conventional Parallel Coordinates does, and the other method is to connect line segments on the variable axis at each variable. By selecting a standard variable and connecting line segments at each variable, understanding the relation between variables becomes intuitive.

Rubel et al. [4] proposed a 3-D Parallel Coordinates technique, where the additional axis is the arbitrary axis of space. This technique was applied to a volume dataset, where the plot point is equivalent to the coordinates of observed points. This technique makes it easy to understand the difference between the relations among variables according to the position of each variable on the space of observed points.

Wegenkittl et al. [5] have proposed a 3D Parallel Coordinates technique, where the additional axis is the time axis. Like VPC, the purpose of this technique is to analyze temporal changes for the entire dataset. However, the dataset of this technique is limited to cases in which a single observed point is located at each axis. Therefore, a single polygonal line is displayed at each time step. Because consecutive polygonal lines are connected through all time steps, the entire representation becomes a polygonal surface.

2.4 An Exploration of a Multivariate Time-Varying Volume Dataset

Sukharev et al. [6] have proposed a mathematical approach to explore a multivariate time-varying volume dataset using a mathematical approach, where this approach is composed of a self-correlation function, a cross-correlation function and a canonical correlation analysis. The approach makes it possible to integrate variable and temporal information into one variable. This technique enables an exploration of the overall features of the multivariate time-varying volume dataset, since the integrated variable is visualized using volume rendering. However, there are two problems with this technique. First, it is impossible to analyze temporal changes in relations between variables using this technique. For example, a typical case might involve a relation between variables that includes negative and positive correlations at the beginning and end of the time period under investigation, respectively. The other problem is that it is impossible to analyze complex relations that may exist within a single time step. For example, a typical case might involve a relation between variables that includes negative and positive correlations at lower and higher values, respectively.

Akiba et al. [7] have proposed a graphical user interface to explore a multivariate time-varying volume dataset by arranging three display windows that work cooperatively. The contents of the three display windows are Parallel Coordinates representing “the relation between variables”, a time histogram representing “the temporal change of the distribution of variable values” and volume rendering representing “the spatial distribution of data values”. This interface makes it possible to analyze a multivariate time-varying volume dataset by selecting a starting point among ”the relation between variables”, “the temporal change of the distribution of variable values” and “the spatial distribution of data values” for analyzing by cooperating the three displays. In this technique, existing techniques are effectively combined in order to analyze a multivariate time-varying volume dataset. It is thus possible to understand relations with respect to the numbers of variables, time steps and resolution. However, it is difficult to analyze “the temporal change in the relation between variables” since this technique cannot visualize the relations between all variables at all time steps in a single picture, which may nevertheless be a good starting point for data exploration.

3 Proposed Technique

In this section, we introduce our proposed technique, namely, “Volumetric Parallel Coordinates” (VPC). This introduction is composed of three parts in which we present an overview of VPC, a process that converts a multivariate time-varying volume dataset into VPC data and the graphical user interface that facilitates data exploration using VPC.

3.1 Volumetric Parallel Coordinates (VPC)

The basic idea behind VPC involves stacking Parallel Coordinates for each time step along the time axis in 3-D space. The time axis is configured so that it is vertical to the plane in which the Parallel Coordinates is defined (Fig. 2 (left)). We denote the set of all planes as VPC data. VPC data can be viewed as volume data, as shown in Fig. 2 (right), that can be visualized using volume rendering. In other words, the process applied in VPC consists of two steps. The first step is to convert a multivariate time-varying volume dataset into VPC data, and the second step is to visualize VPC data using volume rendering.

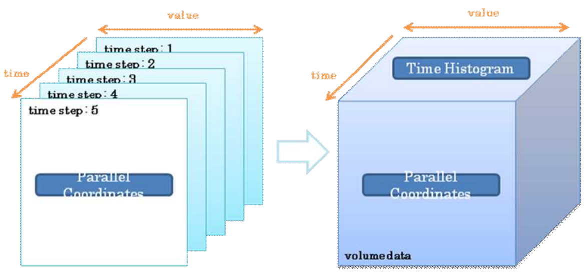

Fig. 2 The basic idea of VPC

VPC data includes information on relations between variables, which Parallel Coordinates is expected to provide. In addition, by adding the time axis perpendicular to the plane of Parallel Coordinates, VPC data includes information on temporal changes in relations between variables. In other words, VPC data maintain variable and temporal information. Therefore, visualizing VPC data using volume rendering makes it possible to explore "the temporal change of relation between variables."

It is also useful that using VPC data makes it possible to analyze the original dataset by specifying either a time step or a variable. First, by using the cutting plane that is perpendicular to the time axis and passes through a time step, Parallel Coordinates can be shown. Therefore, it is possible to analyze a relation between variables at a time step by displaying a cutting plane of VPC data. Second, by using the cutting plane that is perpendicular to the variable axis and passes through a variable, a time histogram can be shown (see Fig. 2). Therefore, it is possible to analyze a temporal change in the distribution of a variable at a given time step. An adequate graphical user interface is required in order for a user to enjoy these features when VPC data is provided.

3.2 The Conversion of a Multivariate Time-Varying Volume Dataset into VPC data

In this section, we describe pre-processes that are required to construct Parallel Coordinates at each time step as well as the structure of VPC data.

3.2.1 The Construction of Parallel Coordinates

As we mentioned in section 3.1, in order to convert a multivariate time-varying volume dataset into VPC data, it is necessary to construct Parallel Coordinates at each time step. In this section, we describe pre-processes for this construction.

Binning

As the first pre-process, binning is employed in order to reduce the data size required to represent Parallel Coordinates. Binning is a method used to express a dataset by the number of observed points that belong to a bin. By definition, a bin is an enclosed space for storing something in bulk, and it is an interval into which the domain of a variable with two or more observations is divided. In order to apply the binning to Parallel Coordinates, our technique refers to the method developed by Novotn\'{y} et al. [8]. The binning process is as follows.

1. Divide the variable axis into b bins and divide the adjacent variable axes into b^2 bins. Then, a bin in the 2-D coordinate system is defined as shown in Fig. 3 (left), and a bin in Parallel Coordinates is defined as shown in Fig. 3 (right).
2. Count the number of line segments that belong to each bin c.
3. Repeat the above treatment for all adjacent variable axes. If the number of variables is k, repeat k-1 times.
4. Set that color and opacity of the parallelogram to correspond to the bin by using the counted number.

In the case of expressing Parallel Coordinates without binning, the number of line segments is equivalent to the number of nodes n between adjacent axes. When the number of variables is k, the total number of line segments on Parallel Coordinates is $(k\text{-}1)n$. If we visualize a large-scale volume dataset using Parallel Coordinates, the resulting image may include enough cluttering to prohibit a user from easily understanding the dataset. However, in the case of expressing Parallel Coordinates with binning, the total number of line segments is the number of bins $(k\text{-}1)b^2$ at maximum. If we adequately determine the number of bins, we can enjoy a Parallel Coordinates image that shows less clutter yet still conveys essential information.

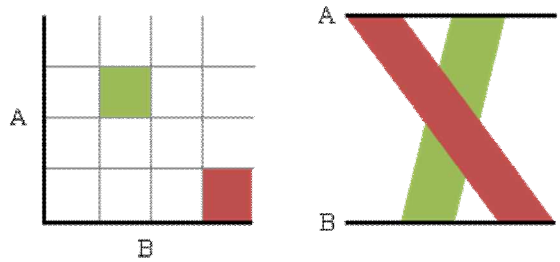

Fig. 3 Binning in Parallel Coordinates
(number of bins on the variable axis: 4)

Biasing

When extremely large values are included in the bins, the differences between other bin values are relatively small. As the second pre-process, biasing is employed in order to avoid a situation in which a change in bin values is not visible. In our proposed technique, biasing refers to the method developed by Perlin et al. [9]. Setting a bias value g, a bin value c, a biased bin value c_0 and a maximum bin value for all time steps M, biasing according to their method is defined as follows.

$$c_o = \left(\frac{c}{M}\right)^{\ln g / \ln 0.5} \qquad (1)$$

We can bias the bin values according to their smallness using the correlation of bin values maintained by setting $g>0.5$. Biasing is applied after binning for all time steps is finished because the maximum bin value for all time steps M is needed to proceed with biasing.

3.2.2 The Structure of VPC data

To construct VPC data, we specify a thickness for the plane in which Parallel Coordinates is defined and arrange them without any gaps (Fig. 4) in order to facilitate the analysis of relations between consecutive time steps. A variable axis appears as a plane due to the thickness given to Parallel Coordinates. We call this plane a variable plane. We described in 3.2.1 that a bin corresponds to a parallelogram under Parallel Coordinates. A parallelogram that corresponds to a bin becomes a parallelepipedon, as shown in Fig. 5, according to the given thickness. An upper plane or lower plane of this parallelepipedon becomes a boundary face between consecutive variable planes. In Fig. 5, the green parallelepipedon corresponds to the third bin on variable A and the second bin on variable B of the third time step.

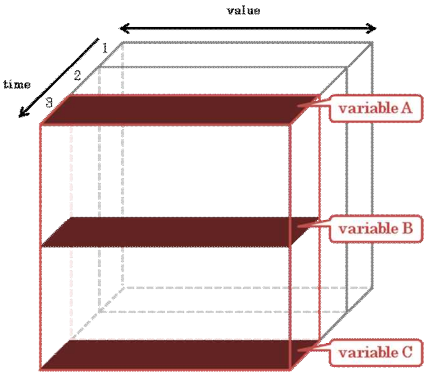

Fig. 4 Parallel Coordinates at a given time step using VPC data (number of time steps: 3)

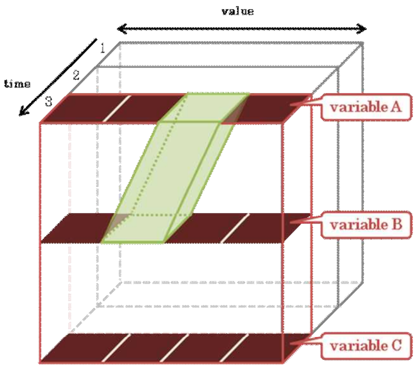

Fig. 5 Bins in VPC data (number of bins on the variable axis: 4; number of variables: 3)

Since a bin becomes a parallelepipedon, and regarding this parallelepipedon as a volume cell, VPC data becomes a type of volume data composed of parallelepipedon volume cells. Generally, a volume dataset can be classified into two types. One type is a regular grid volume dataset, and the other type is an irregular one. VPC data may be categorized into irregular grid volume data. However, unlike a typical irregular volume dataset, VPC data is composed of volume cells that may overlap, since there are overlapping parts on the bins under Parallel Coordinates processed by binning.

3.3 The Interface to Utilize VPC data

In this section, we describe our proposed graphical user interface for VPC data. Our proposed interface is depicted in Fig. 6. This interface has a function to show a volume-rendering image in order to analyze the spatial distribution of a multivariate time-varying volume dataset. In Fig. 6, the left and right sides of the interface are utilized to access VPC data and display the volume-rendering image, respectively.

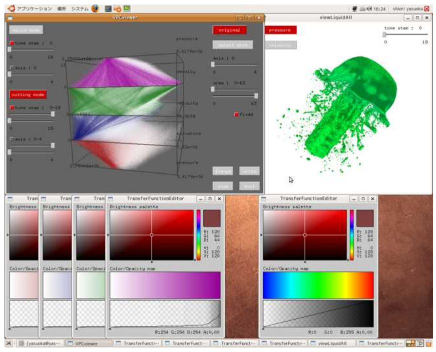

Fig. 6 The interface to utilize VPC data

Recall that we explained in section 3.2.2 that VPC data can be viewed as an irregular grid volume data with possibly overlapping cells. When we visualize this type of volume data, special care should be taken to process overlapping cells. We decided to utilize the Particle-based Volume Rendering (PBVR) technique [10] since it can easily render multiple volumes that may have overlapping regions. PBVR is a rendering technique that generates particles in cells. The generated particle is composed of coordinates and colors. The data value of each color is determined by values in the original volume dataset.

VPC provides a focus and context interface for interactive exploration of a multivariate time-varying volume dataset. By using the VPC, a user can understand an overall feature across all variables and all time steps as well as can focus on some specific regions by displaying a cutting plane of VPC data. In the following section, we show how the VPC interface works interactively through operating buttons and sliders displayed on the screen.

“Display the Cutting Plane” and “Display the Limited Area”

The button labeled “Display the cutting plane” limits the displayed VPC data in order to display Parallel Coordinates for a given step or a time histogram for a given variable. The button labeled “Display the limited area” relaxes these limitations in order to focus on several time steps or variables at a time.

“Setting the Display Area on the Variable Axis”

The button labeled “Setting the Display Area on the Variable” is utilized to select a variable and its value range according to a user’s preferences. In this case, only the bins that are included in the selected range are displayed. By using this function, users can select the nodes on which they wish to focus from the original volume data and analyze the temporal changes in relations between variables from the selected nodes. This function is implemented by creating new VPC data that includes only the cells that correspond to the selected bins and then rendering this VPC data.

"Setting the Transfer Function"

The button labeled "Setting the transfer function" is used to modify a transfer function according to which VPC data is rendered. The transfer function returns a color or opacity based on the number of nodes that are included in the bins. For example, users can set a low opacity value for the bins with very few nodes. This function is implementing by setting the color and opacity on the widget shown in Fig.7.

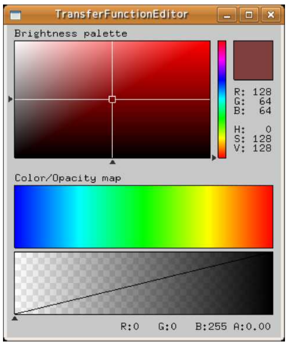

Fig. 7 A widget to set the transfer function

The functions described above are used to analyze the selected time steps or variables using VPC. Users can analyze "the temporal change of the relation between variables", the relation among variables for a given time step or temporal changes in the distribution of a variable.

4 Experimetal Results and Discussion

In this section, we introduce experimental results from applying VPC to a simulation of liquid atomization and discuss these results.

4.1 Liquid Atomization

The dataset used in this experiment is the result of a simulation of liquid atomization and was provided by Dr. Junji Shinjo, a researcher in Japan Aerospace Exploration Agency (JAXA). Liquid atomization is a process by which a liquid is separated into small droplets; in fact, this process is the basic mechanism in the liquid sprayers used in everyday life. Fig. 8 shows that the liquid surface changes at several time steps in the simulation dataset. There are small droplets surrounding the liquid column.

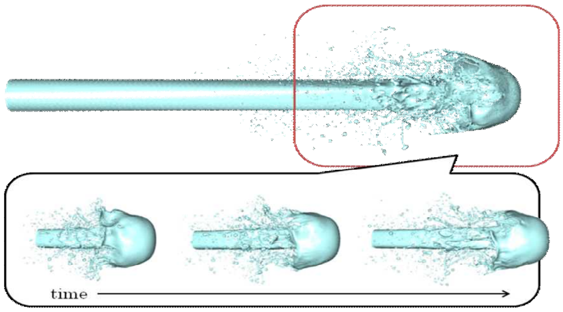

Fig. 8 Liquid atomization

There are four variables contained in the result data, including density Φ, curvature k, pressure p and velocity $\boldsymbol{v}$. Density Φ is a variable that can be used to determine whether the relevant area contains liquid or gas. Its domain is [0, 1], and it is assumed that the area for which the density value exceeds 0.5 and approaches 1 corresponds to liquid. Fig. 9 shows the distribution of curvature, which is determined by volume rendering of droplets. A droplet with high curvature is a small droplet and vice versa, as shown in Fig. 9. The simulation data consists of 19 time steps; the size of each volume dataset is $580 \times 580 \times 1045$.

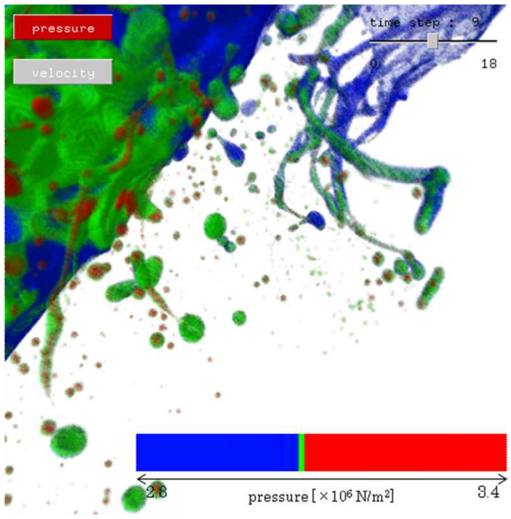

Fig. 9 The distribution of curvature (close to the droplets)

The temporal change in the shape of the liquid surface is shown in Fig. 8. In this figure, liquid is shown as discharging from the left side.

4.2 The Construction of VPC Data

We constructed VPC data that consists of five axes as shown in Fig. 10. The density and curvature variables are used as criteria according which each area is categorized into liquid or gas. In addition, based on these variables, the size of droplets is estimated.

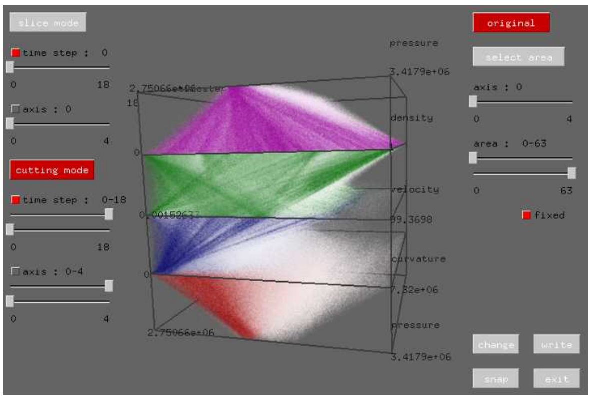

Fig. 10 VPC data that we used, with the variable axes arranged as follows: pressure, density, velocity and pressure, starting from the top.

We arranged the variable axes in the following order: pressure, density, velocity, curvature and pressure, starting from the top. In our experiment, we used three PCs with Intel Core 2 Duo Processor 3.16 GHz CPU, NVIDIA GeForce9800GT 512 MB GPU and 4.0 GB of RAM. It took 96 minutes to construct the VPC data.

We set the bias value and the number of bins on the variable axes at 0.8 and 64, respectively. The transfer function that is used to display VPC data is shown in Fig. 11. The density of line segments on Parallel Coordinates is represented as a color shade. In this transfer function, a white color means that the density of line segments is low, while dense colors indicate that density is high.

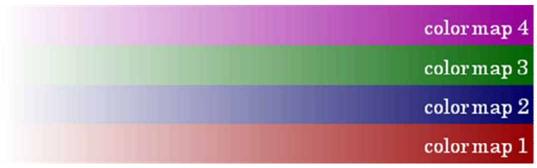

Fig. 11 The transfer function used to display VPC data

4.3 The Relation between Curvature and Velocity

Fig. 12 shows VPC data with respect to the curvature and velocity variable planes. The upper plane corresponds to velocity, and the lower plane corresponds to curvature. This shows that the relation between curvature and velocity is negatively correlated throughout the time period under analysis, because VPC data is constricted in the middle, as indicated by the red line in Fig. 12. From this figure, we can hypothesize that the smaller the size of droplet is, the smaller the speed is.

Fig. 13 shows that volume rendering in which colors and opacity indicate velocity and curvature, respectively. We can easily see that the speed of the liquid column is high, and the speed of droplets is low.

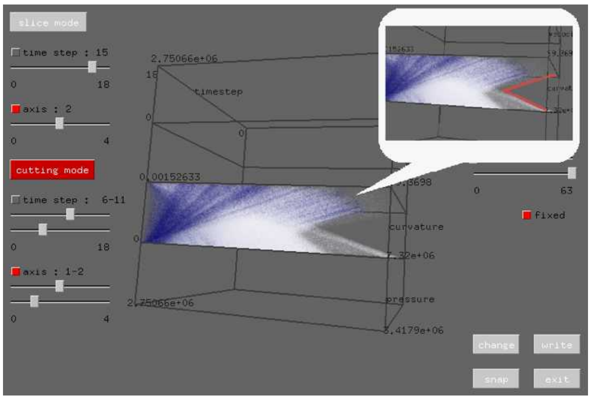

Fig. 12 VPC data that represent the relation between curvature and velocity

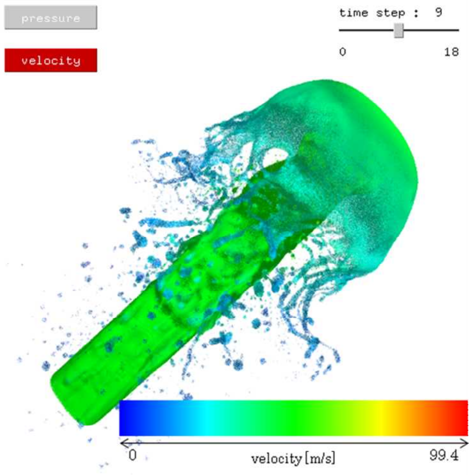

Fig. 13 The distribution of velocity of liquid

4.4 Relation between Density and Velocity

Fig. 14 shows bins with high velocities. Paying attention to the area in which the velocity becomes high, we see that the nodes that have higher velocities are located in gaseous regions. This indicates that velocities are high in a gaseous region that is generated from VPC data.

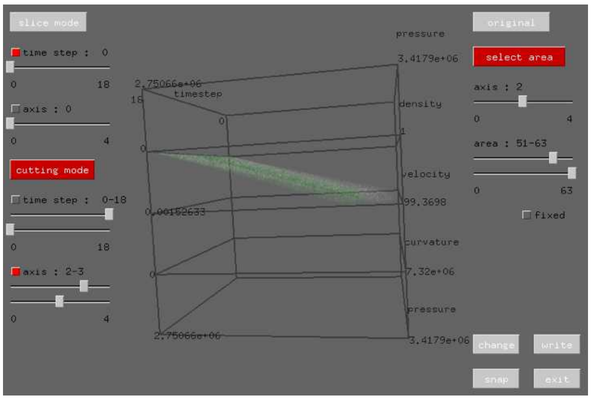

Fig. 14 VPC data that represent the relation between density and velocity (only high-velocity bins are displayed)

To confirm this hypothesis, we apply volume rendering to the volume dataset. Fig. 14 shows the volume-rendering image of the velocity data using the opacity map, which has a peak at the highest velocity. To visualize the boundary between gas and liquid, the volume-rendering image of the curvature data is superimposed. In Fig. 15, yellow and blue colors represent the distributions with higher velocity and curvature, respectively. Based on this image, we can establish that the high-velocity region is surrounded by liquid.

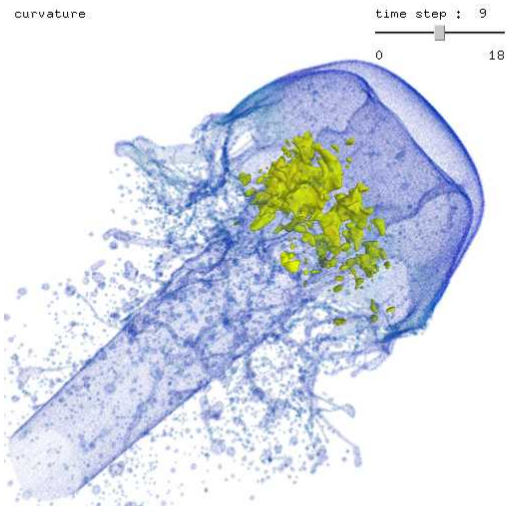

Fig. 15 The distribution of higher velocity in gas (yellow) and the distribution of curvature (blue)

5 Conclusion

In this paper, we propose the use of Volumetric Parallel Coordinates (VPC) to analyze multivariate time-varying volume data. VPC is especially useful to analyze "the temporal change of the relation between variables" when we wish to pay attention to both variables and time steps. We also proposed a graphical user interface to interactively access VPC data in a focused and contextualized manner.

We applied the proposed method to a simulation of liquid atomization in order to evaluate the effectiveness of the proposed technique. As a result, we showed that VPC facilitates the exploration of temporal changes in the relation between variables. Using VPC, we found that the high-velocity region is surrounded by liquid.

In future work, the interactive aspects of the graphical user interface should be improved. Currently, because the order of the variable axes under Parallel Coordinates and the number of bins are fixed, flexibility in these areas requires further attention.

6 Acknowledgements

We wish to thank Mr. Junji Shinjo of JAXA for providing data on the liquid atomization simulation as well as for his comments and suggestions. This research has been partly supported by the National Institute of Information and Communications Technology.

7 References

[1] A. Inselberg and B. Dimsdale. Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry. In Proceedings of the 1st conference on Visualization, 361–378, 1990.

[2] R. Kosara, F. Bendix, and H. Hauser. TimeHistograms for Large, Time-Dependent Data. In Proceedings of VisSym -Joint EUROGRAPICS – IEEE TCVG Symposium on Visualization, 45–54, 2004.

[3] Keisuke Honda and Junji Nakano. 3 dimensional parallel coordinates plot and its use for variable selection. In Proceedings in Compputational Statistics, 187–195, 2006.

[4] O. R¨ubel, G.H. Weber, S.V.E. Ker¨anen, C.C. Fowlkes, C.L.L. Hendriks, L. Simirenko, N.Y. Shah, M.B. Eisen, M.D. Biggin, H. Hagen, D. Sudar, J. Malik, D.W. Knowles, and B. Hamann. PointCloudXplore: Visual Analysis of 3D Gene Expression Data Using Physical Views and Parallel Coordinates. In Eurographics/IEEE-VGTC Symposium on Visualization Proceedings, 203–210, 2006.

[5] R. Wegenkittl, H. L¨offelmann, and Eduard Gr¨oller. Visualizing the Behavior of Higher Dimensional Dynamical Systems. In Proceedings of the 8th conference on Visualization, 119–125, 1997.

[6] J. Sukharev, C. Wang, K.-L Ma, and A.T. Wittenberg. Correlation Study of Time-Varying Multivariate Climate Data Sets. In Proceedings of IEEE Pacific Visualization Symposium, 161–168, 2009.

[7] H. Akiba and K.-L. Ma. A Tri-Space Visualization Interface for Analyzing Time-Varying Multivariate Volume Data. In Proceedings of Eurographics/IEEE VGTC Symposium on Visualization, 115–122, 2007.

[8] M. Novotn´y and H. Hauser. Outlier-preserving Focus+Context Visualization in Parallel Coordinates. IEEE Transactions on Visualization and Computer Graphics, 12(5):893–900, 2006.

[9] K. Perlin and E.M. Hoffert. Hypertexture. Computer Graphics, 23(3):253–262, 1989.

[10]K. Koyamada, N. Sakamoto, and S. Tanaka. A Particle Modeling for Rendering Irregular Volumes. In Proceedings of the 10th International Conference on Computer Modeling and Simulation, 372–377, 2008.