Studies of Large-Scale Data Visualization: EXTRAWING and Visual Data Mining

Project Representative

Fumiaki Araki Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

Authors

Fumiaki Araki[*1], Shintaro Kawahara[*1] and Daisuke Matsuoka[*1]

*1 Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

Progress of EXTRAWING, the research and development project for attractive representations and transmissions of results of geophysical and environmental fluid simulations is reported. The software tool to make KML contents of simulation results has been improved as following three points. First, the GUI of the software is modified for improving usability. Second, color-contour plots become available on horizontally- and/or vertically-sliced planes. Third, the functions to output animation data is implemented for analysis of time-developing simulations. An effective method to render for a set of laminated layer surfaces is newly designed.

Progress in terms of visual data mining techniques for semi-automatic feature extraction of oceanic currents has been also reported. It contains two themes which techniques of cluster analysis are applied. One is about a density-based method and another is a hybrid method combined between non-hierarchical and hierarchical methods. The density-based method has successfully extracted both of Kuroshio and Oyashio currents from OFES simulation data, but its computational cost is too high. The hybrid method has been also tested to solve such the difficulty from the density-based method.

Keywords: EXTRAWING, tool development, feature extraction, cluster analysis, oceanic currents

1. EXTRAWING

EXTRAWING [1] is a project for novel and attractive representations of results of geophysical and environmental fluid simulations and effective transmission of those results to the general public. In the FY2011, we describe about following two progresses.

1.1 Volume visualization software tool

We have developed a program, called "Volume Data Visualizer for Google Earth (VDVGE)", to make EXTRAWING contents since FY2010. Fig. 1 shows the GUI of this program. VDVGE can read a volumetric dataset, visualize it using our original volume visualization method [1], and output a content file written in KML format for Google Earth. Such the specification means that VDVGE can be utilized for not only contents-making but also three-dimensional visual analysis of earth-scientific data. To strengthen it further, we implemented additional visual analysis functions, especially two-dimensional contour-plot in the extracted surface from the volume dataset, that is the cross-section, and file-saving function of time-sequential rendering results.

Fig. 2 shows an example of the visualization result by the contour-plot function. This function extracts cross-sections from the target volume data via interpolation process and plots two-dimensional color-contour on the cross-section. The normal direction of the cross-section is selected from the axes of polar coordinates. In terms of file-saving, the function not only exports a KML file but also saves rendering results on the main window in Fig. 1 as sequential images or a movie file of MPEG format.

With implementation of these additional functions, the GUI of VDVGE had been complicated. So we also reconstructed the program design of VDVGE and improved the GUI for better usability.

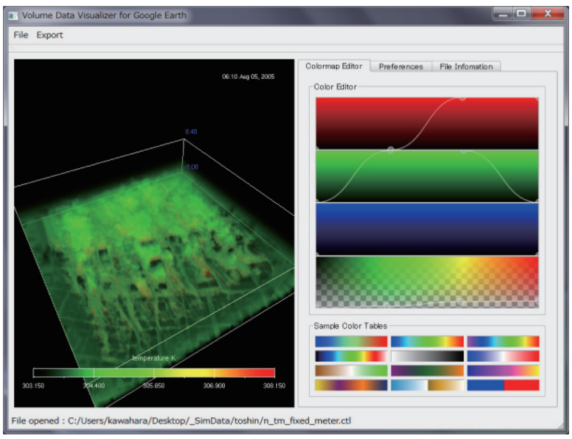

Fig. 1 Main window of VDVGE.

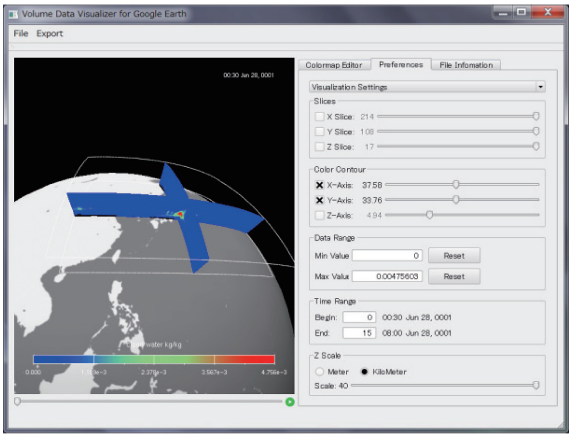

Fig. 2 Color slices on VDVGE.

1.2 Novel method for laminated surfaces

We devised a novel surface-ordering method to render effectively a set of laminated surfaces with color-contour images mapped on each ones. Usually, when sliced planes with transparencies are rendered, an accurate rendering result can be got by using a combination of a couple of techniques, z-sorting and alpha blending. The z-sorting process must perform at every change of viewing position. On the other hand, our algorithm enables us to get a fixed rendering result by devising the rendering order of each sliced plane without re-sorting the order, even when viewing position changed. Fig. 3 shows examples of the rendering order in this method. The number of 1 to 5 in Fig. 3 means those orders. The signs A and B mean viewing directions of users. In these examples, a slice of #1 is set as the reference plane and the number of other slices become larger with increasing of the distance from the reference plane. Fig. 4 shows a rendering diagram based on proposed method,

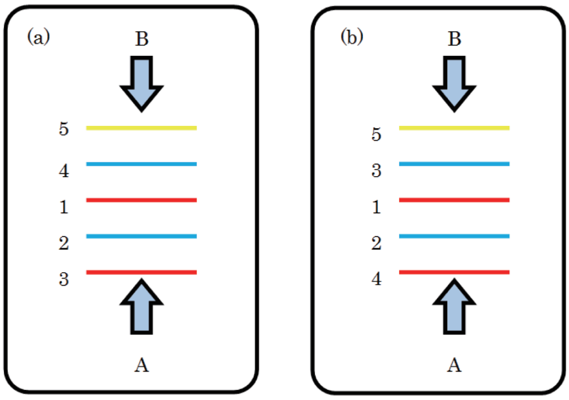

Fig. 3 Rendering order of slice planes.

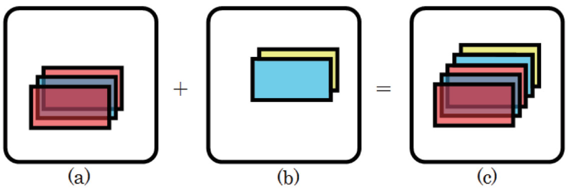

Fig. 4 Diagram of rendering based on proposed method; (a) Planes which are nearer than the reference plane, (b) Planes which are farther than the reference plane and (c) visualization result.

when rendering orders are set up as in Fig. 3. Alpha-blending is carried out nearer than the reference plane (Fig. 4 (a)) and z-buffering is carried out farther than the reference plane (Fig. 4 (b)). Fig. 5 shows differences of rendering results in the case with (right) / without (left) this method. These figures mean that we can get an accurate result about planes close to the viewing position from the reference plane.

We applied for a patent about this algorithm (Application Number (2012) 005541). Use in the visual representation on low-resource devices such as mobile phones can be expected.

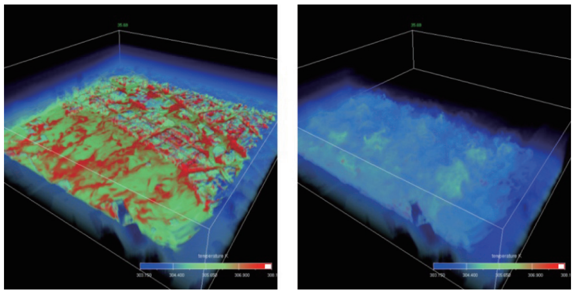

Fig. 5 Differences of rendering result with/without use of the proposed method. Right hand side of the images was the case which our method applied and left hand side was not.

2. Visual data mining for oceanic simulation

Visualization of numerical simulation results is effective way to understand complex phenomena and/or configuration intuitively. Nevertheless, there is still a challenge to extract information from huge amount of data obtained from large-scale and high-resolution simulation on the massively parallel super computer. In the fiscal year 2010, we developed a novel 2-dimensional transfer function (color map) which can simultaneously represent two physical values by using H-V space of HSV color [5]. We visualized ocean currents effectively by applying this method to OFES data. While this approach extracts the oceanic currents successively in the case of two variables, it is still difficult to expand to higher dimensions. In order to establish feature extraction of oceanic currents in the further multi-variable cases, in the fiscal year 2011, we have developed two different approaches applying techniques of cluster analysis.

2.1 Extraction of ocean currents via density-based cluster analysis

In our method, it is necessary to set the following parameters, density in the variable space, range for calculating density, range for exploration and range for clustering. For instance, regarding to set parameters of density and range, we visualize "similarity density" which represents density of elements in arbitrary range within the multi-dimensional variable space as shown in Fig. 6. In this figure, the red (or blue) region indicates that many elements of the surrounding area have similar (or different) characteristics. In addition to these parameters, initial seeds for extraction of the Kuroshio and Oyashio are shown in

Fig. 7. Clustering results are also visualized in Fig. 7. The red and blue area indicates the Kuroshio and Oyashio, respectively. Both results can be seen that the clustering is roughly correct.

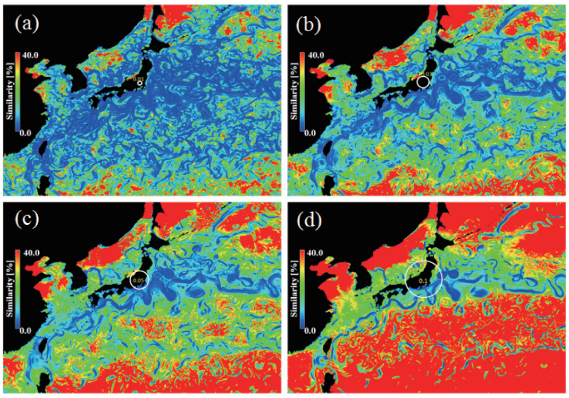

Fig. 6 Similarity density map. Eps = (a) 0.1, (b) 0.3, (c) 0.5 and (d) 1.0.

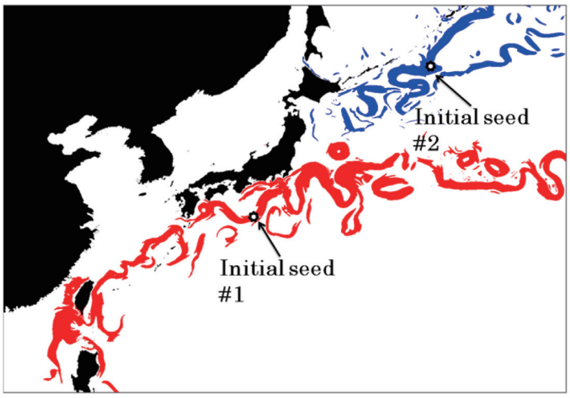

Fig. 7 Extraction result of Kuroshio and Oyashio by density-based method.

2.2 Extraction of ocean currents via non-hierarchical and hierarchical cluster analysis

The density-based method, which is one of the cluster analysis techniques, can extract ocean currents but enlarges its computational cost too high. We study also the case adopting the non-hierarchical cluster method, which is another cluster analysis technique.

One of the important steps of cluster analysis is to select a distance measure, which can determine how to calculate the similarity between two elements. The distance between two elements (i and j) in 4-dimensional Euclidean space is defined as:

$$D^2_{ij} = C_T(T_i - T_j)^2 + C_V(V_i - V_j)^2 + C_X(X_i - X_j)^2 + C_Y(Y_i - Y_j)^2$$

where, T, V, X and Y is temperature, flow speed and x- and y-positions, respectively. The coefficients, C_T, C_V, C_X and C_Y, are constant values to make the similarity dimensionless.

The resulting image of non-hierarchical cluster analysis and 16 centroids, initial points of non-hierarchical clustering, are shown in Fig. 8. In this figure, each cluster is indicated in different colors. Not only the Kuroshio (cluster #1 and #2), the Kuroshio Extension region (cluster #3 and #4) and the Oyashio (cluster #14 and #15) but also the mixed layer in the Kuroshio/Oyashio Extension region and the vortices in the Sea of Japan seem to be clustered almost correctly.

Furthermore, let us also adopt hierarchical method to extract the hierarchical structure of these small clusters obtained by the non-hierarchical method. We can recognize from the resulting diagram by the hirerarchical method as shown in Fig. 9 that Kuroshio is composed from cluster #1 - #4 and Oyashio is done from cluster #14 and #15, respectively.

How the validity of clustering results and optimization of the parameters should be evaluated is a key issue in the future works.

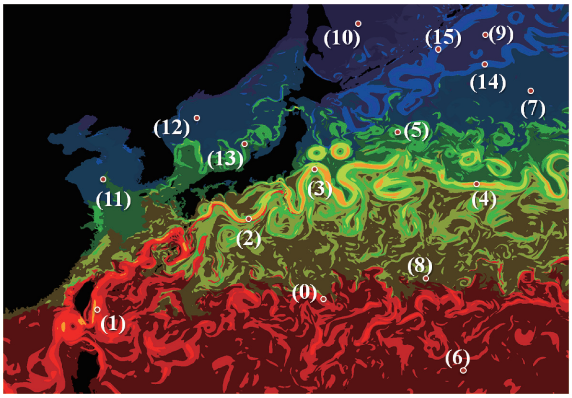

Fig. 8 Extraction result of Kuroshio and Oyashio by the non-hierarchical method.

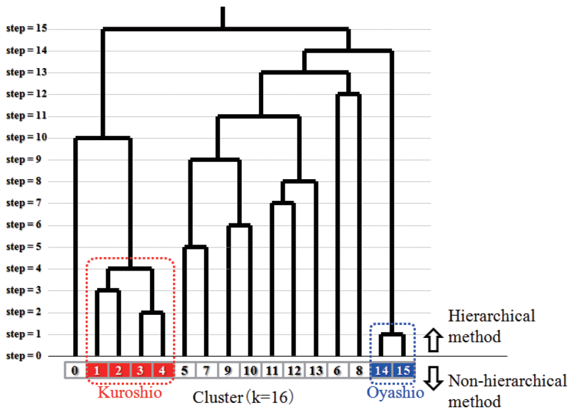

Fig. 9 A diagram in case of non-hierarchical and hierarchical hybrid method.

References

[1] Fumiaki Araki, Shintaro Kawahara, Daisuke Matsuoka, Takeshi Sugimura, Yuya Baba and Keiko Takahashi, "Studies of Large-Scale Data Visualization: EXTRAWING and Visual Data Mining", Annual Report of the Earth Simulator Center, April 2010-March 2011, pp.195-199, 2011.

[2] GrADS, http://http://www.iges.org/grads/

[3] Fumiaki Araki, Hitoshi Uehara, Nobuaki Ohno, Shintaro Kawahara, Mikito Furuichi and Akira Kageyama, "Visualizations of Large-scale Data Generated by the Earth Simulator", Journal of the Earth Simulator, vol.6, pp.25-34, 2006.

[4] Conduit, http://www.mechdyne.com/conduit.aspx

[5] N. Ohno and D. Matsuoka, Visualization of Global Geophysical Fluid Simulation, NAGARE, Vol.30, No.5, pp. 409-414, 2011.

大規模データ可視化研究：
EXTRAWING とビジュアルデータマイニング

プロジェクト責任者
荒木　文明　　海洋研究開発機構　地球シミュレータセンター

著者
荒木　文明 *1，川原慎太郎 *1，松岡　大輔 *1
*1　海洋研究開発機構　地球シミュレータセンター

地球環境流体シミュレーションの結果を魅力的に表現し一般社会へ発信する研究開発プロジェクト、EXTRAWING の進捗を報告する。2011 年度は、シミュレーション結果の KML コンテンツを作成するソフトウェアの改良を以下の 3 点：(1) 利便性の向上のための GUI の変更、(2) 水平 / 垂直断面上でのカラーコンタ表示、(3) 時間発展シミュレーションを解析するための連番画像および動画ファイルの出力機能の実装、を中心に実施した。また積層面の全体をレンダリングする上で効果的な新しい方法を考案した。

海流の特徴抽出を半自動的に実行するためのビジュアルデータマイニング技術についての進捗も報告する。本年度はクラスタ分析法を応用した二つのテーマに取り組んだ。一つは要素間の類似度密度に基づいた方法、もうひとつは非階層的クラスタ分析法と階層的クラスタ分析法を組み合わせたハイブリッドな方法である。一つ目の類似度密度に基づいた方法では、黒潮と親潮の両方の海流を正しく抽出できることが確かめられた。しかしながら計算コストは極めて高い。そこでこの困難を解決するために二つ目のハイブリッド手法も試みられた。

キーワード : EXTRAWING, ツール開発 , 特徴抽出 , クラスタ分析 , 海流