Studies of Large-Scale Data Visualization 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

This paper reports activities of Advanced Visualization and Perception Research Group (AVPRG), Earth Simulator Center (ESC) in the FY2012. We have progressed on three themes; EXTRAWING, virtual reality (VR) visualization and visual data mining techniques for oceanic datasets.

In terms of EXTRAWING, we have been released the source code of VDVGE, the volume visualization software on Google Earth, has been released on the web page of ESC. Cross-disciplinary workshops for VDVGE utilization have also been held regularly.

In terms of VR visualization, a system to select and redraw visualized objects on a mobile device has been implemented into VFIVE as an experiment. When the number of lines of an "operation list" becomes huge, it becomes hard to select particular operation records from this list. To avoid such the difficulty, the function that inserts separators into this list in the middle of user's operation has been added to this system.

Progress of visual data mining techniques for oceanic datasets is summarized as following three topics: 1) Extraction of ocean currents and its evaluation; 2) Optimization of 2D color map using extracted result; and 3) Interactive visual analytics between real and variable space. These works might give our research findings in FY2012 objective assessments and reinforce it further.

Keywords: EXTRAWING, VDVGE, virtual reality, mobile device, visual data mining for oceanic currents

1. VDVGE

We have proceeded EXTRAWING, a project for attractive representation of geophysical fluid simulation results and conveyance of those results to the general public [1] since FY2010. Currently, we utilize Google Earth [2] as a representation platform of our simulation results. To make the Google Earth contents from the simulation results, we have been developing a software program called "Volume Data Visualizer for Google Earth (VDVGE)". Figure 1 shows a screen shot of VDVGE. The source code of VDVGE was released on the web page of the Earth Simulator Center, JAMSTEC [3] on December 14, 2012. VDVGE corresponds to the various operating system such as MS-Windows, Mac OS X and Linux by using Qt, a multi-platform GUI tool kit [4]. The software manual and a sample dataset are also provided on this web page.

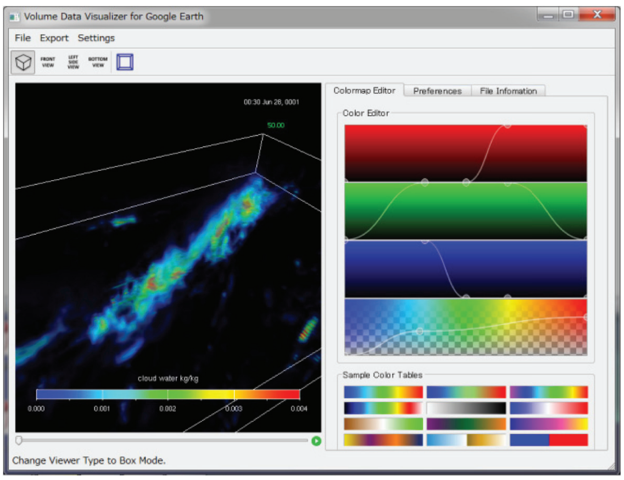

Fig. 1 The graphical user interface of VDVGE.

We have regularly held cross-disciplinary workshops to increase the user of VDVGE. The purpose of these workshops is to practice how to use VDVGE and to discuss role as a communication hub. Main attendees are researchers in various fields, officers of public relations and others. One example of the visualization results obtained on the workshops is shown in Fig. 2. In this figure, a result of the shipboard radar observation (denoted by colored disk) on "Mirai" is put on a satellite

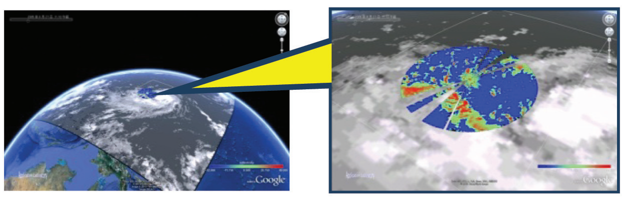

Fig. 2 Overlay of two observational results by the shipboard radar on "Mirai" (color) and the satellite observation (monochrome).

observation image (denoted by monochrome background) on the same day and time. The two kinds of the results which have been obtained from these observations through VDVGE can be overlaid and compared easily on the Google Earth.

We are also planning to compare with observation and simulation of the same period.

2. Development of VR visualization

We has been promoting the virtual reality (VR) visualization project utilizing BRAVE [5], a CAVE-type VR system of JAMSTEC.

In the FY2012, a system to select and redraw visualized objects using a mobile device has been implemented into VFIVE [6] as an experiment. This system sends and receives information between BRAVE and the mobile device via Wi-Fi as shown in Fig. 3.

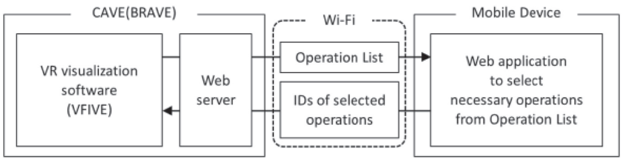

Fig. 3 System configuration.

In this figure, "operation list" is a list-type text file that the visualization methods, data types and several parameters are recorded. The web application on the mobile device picks up the operation list via the web server and displays it as shown in Fig. 4. When the user selects buttons on the diagram, corresponding ID numbers are sent back to VFIVE via the web server. VFIVE then refreshes the visualized scene according to those numbers.

The experiment of the implemented function was conducted in terms of each of isocontouring and streamline drawing using BRAVE. As sample datasets, the wind velocity (vector fields for streamline drawing) and the cloud water (scalar fields for isocontouring) in the near Japan region computed by MSSG

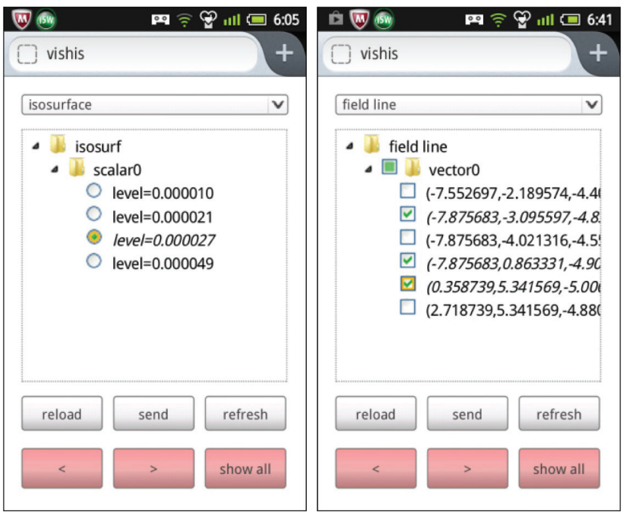

Fig. 4 Screen captures of the web application. Left hand side of the images shows a mode to select a level of isosurface, and right hand side of the images shows a mode to select field lines.

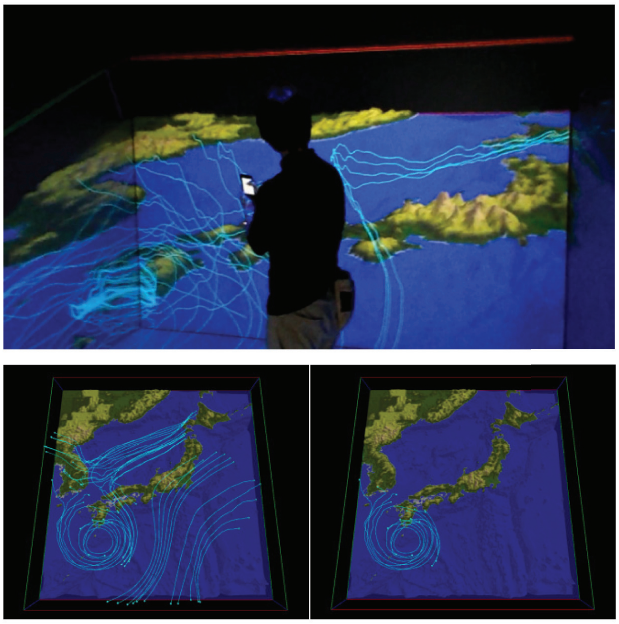

Fig. 5 The results of the streamline selection using the web application; the scene that a user selects streamlines on his mobile device (upper), all of streamlines (lower left) and the selected streamlines eliminated by the proposed function (lower right).

model [7] were used. The results of the streamline selection using the web application on user's mobile device are shown in Fig. 5. The upper side of Fig. 5 shows the scene that a user selects appropriate streamlines on his mobile device in the virtual space of BRAVE. Each of Both images of the lower side of Fig. 5 represents all of streamlines (LHS) and the selected streamlines (RHS) eliminated by the proposed function, respectively. We have been also confirmed that the same way can be applied to the selection of the specific isosurface in isocontouring, as shown in Fig. 6.

Fig. 6 The results of the isosurface selection using the web application.

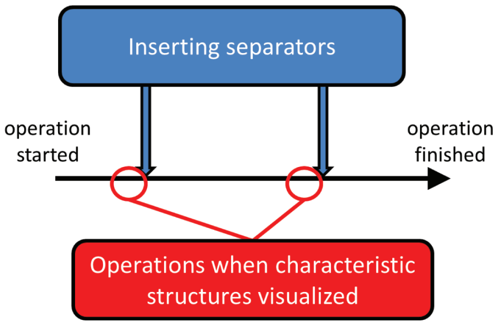

Fig. 7 Conceptual diagram of chapter function.

This system, however, had an essential difficulty. Specific operations stored in the operation list are hard to find out on the small screen of the mobile device, when the number of operations becomes huge. As an experiment, we introduced "separator" to divide the visualization operation sequence, as the conceptual diagram of Fig. 7. If a user inserts the separator at each moment when an interesting feature is found, it will become easy to select each region of the operation list segmented by these separators.

3. Study of visual data mining for oceanic simulation

3.1 Extraction of ocean currents and its evaluation

In the FY2011, we had developed a new 2-dimensional transfer function (color map) which represents two physical values simultaneously and cluster analysis methods which extract emphasis features as shown in Fig. 8 [8]. In the FY2012, we have developed an evaluation method of clustering results and optimization method of 2D color map using clustering results.

Evaluation of clustering results is important to extract ocean currents accurately. In this study, parameter correlation to the validation was examined by means of comparing clustering result and simple correct answer. The simple correct answers are defined as different weight along x and y axes as shown in Fig. 9. The weight given to the boundaries of ocean current and vortex are lower than inside/outside them because of the difficulty of the evaluation of ocean boundaries. Here, the boundary of ocean currents is defined by analyzing the gradient of the velocity magnitude with visual observation.

The accuracy rate of the clustering results with the density-based method and hybrid method was approximately 70-92% and 78-88%, respectively. Analyzing the parameter dependences to clustering results in order to optimize the similarity calculation is future work.

3.2 Optimization of 2D color map using extraction result

Figures 10 (a) and (b) show the 2D histogram and the extraction result of the density-based clustering result mapped to the temperature and the current speed space, respectively. Here, we manually select the boundary of the ocean currents with white closed curves as shown in Fig. 10 (b). Next, 2D color

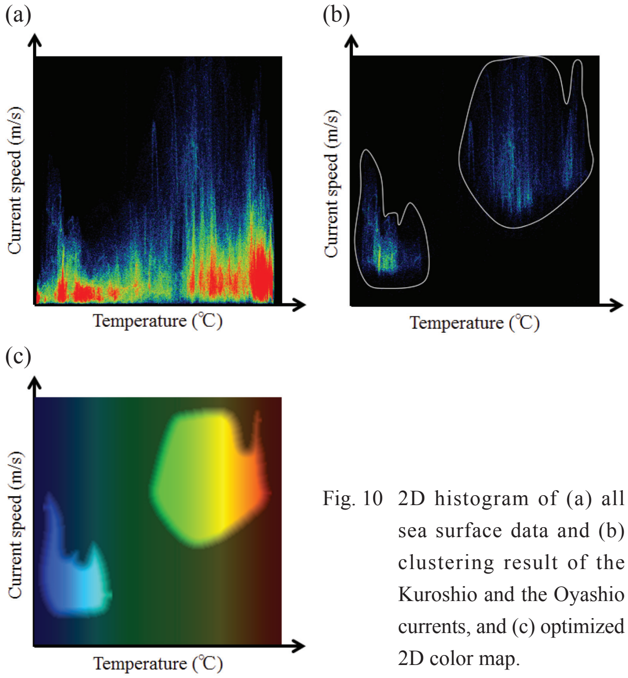

Fig. 10 2D histogram of (a) all sea surface data and (b) clustering result of the Kuroshio and the Oyashio currents, and (c) optimized 2D color map.

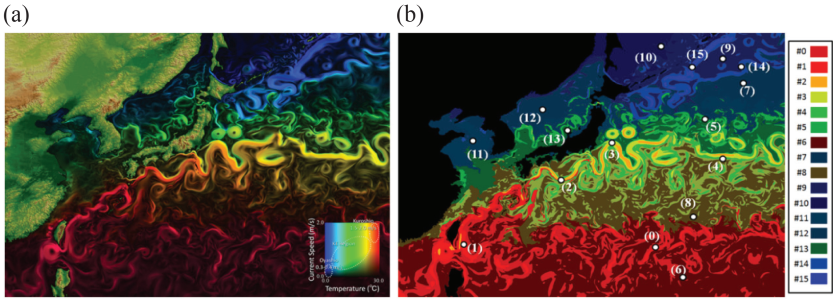

Fig. 8 (a) Multivariate visualization of SST and current speed using 2-dimensional color map and (b) clustering result of ocean currents using k-means.

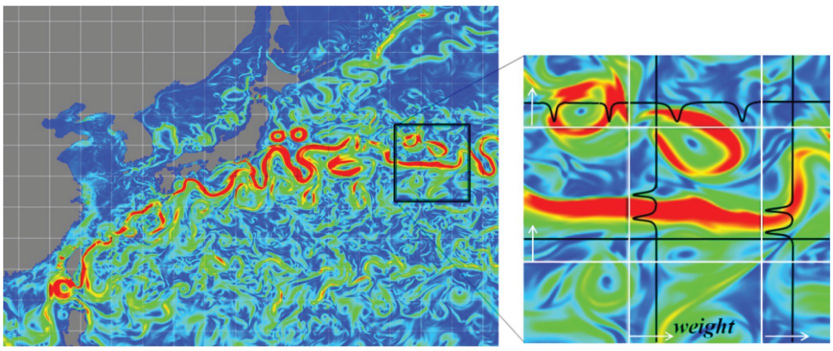

Fig. 9 Evaluation method of clustering results.

map is optimized based on the selected closed curves. Fig. 10 (c) is the 2D color map with refinement of the ocean current region. Finally, visualization result of the ocean currents is obtained by using the optimized 2D color map as shown in Fig. 10 (c). The Kuroshio and the Oyashio currents are clearly emphasized compared with the previous result using non-optimized 2D color map shown in Fig. 11.

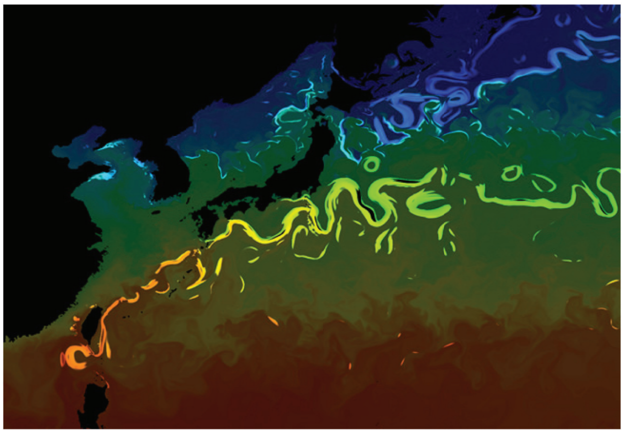

Fig. 11 Visualization result of the Kuroshio and the Oyashio currents using Fig. 10 (c).

3.3 Interactive visual analytics between real and variable space

The optimized 2D color map can also be applied to interactive visual analytics. It enables us to easily analyze the relation between real space and variable space.

Two examples of interactive visual analytics are shown in Fig. 12 and Fig. 13. Characteristic sets of the scatter plot distribution like arch or line are seen in 2D histogram of Fig. 12 (a). Here, we enclose these characteristic plots with closed curve in the same way as specifying the refinement region in Fig. 10 (b). As a result of visualizing enclosed regions, a north boundary of a front of the Kuroshio Extension region and a warm eddy between the Kuroshio and the Oyashio currents are refinement as bright region in Fig. 12 (b).

In the same way with Fig. 12, Fig. 13 is an example in 3D real space and 3D variable space composed of temperature, current speed and salinity. The region with low temperature, low velocity and high salinity is enclosed a white closed surface as shown in Fig. 13 (a). The mapping result to 3D real space is shown in Fig. 13 (b) as red sphere. We can find that the enclosed region is shown as like the Antarctic Circumpolar Current.

(a) (b)

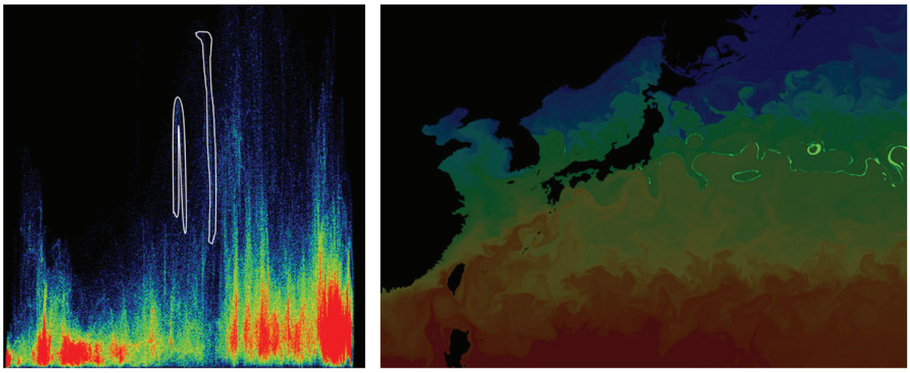

Fig. 12 (a) Manually feature extraction in the 2D variable space (white closed curve) and (b) its mapping to the 2D real space (bright region).

(a) (b)

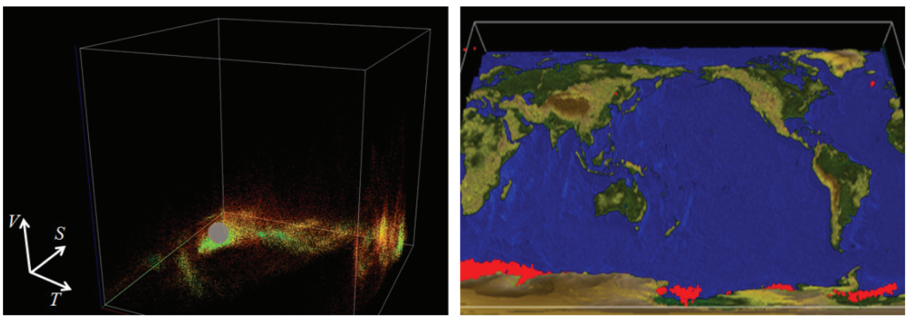

Fig. 13 (a) Manually feature extraction in the 3D variable space (white closed surface) and (b) its mapping to the 3D real space (red scatter plot).

Acknowledgement

This work was partially supported by KAKENHI (24650045), Grant-in-Aid for Challenging Exploratory Research.

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] Google Earth (http://www.google.com/earth/index.html).

[3] VDVGE (http://www.jamstec.go.jp/esc/research/Perception/vdvge.ja.html).

[4] Qt (http://qt-project.org/).

[5] 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.

[6] Akira Kageyama, Yuichi Tamura, and Tetsuya Sato, “Visualization of Vector Field by Virtual Reality”, Progress of Theoretical Physics Supplement, No.138, pp.665-673, 2000.

[7] Keiko Takahashi, Xindong Peng, Kenji Komine, Mitsuru Ohdaira, Koji Goto, Masayuki Yamada, Hiromitsu Fuchigami, and Takeshi Sugimura, “Non-hydrostatic atmospheric GCM development and its computational performance”, *Use of High Performance computing in meteorology*, Walter Zwieflhofer and George Mozdzynski Eds., World Scientific, pp. 50-62 (2005).

[8] Daisuke Matsuoka and Fumiaki Araki, “Visualization for Ocean General Circulation Model via Multi-dimensional Transfer Function and Multivariate Analysis”, Transaction of the Japan Society for Simulation Technology, Vol. 4, No. 4, pp. 168-175, 2012.

大規模データ可視化とビジュアルデータマイニングの研究

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　大祐 [*1]
*1　海洋研究開発機構　地球シミュレータセンター

2012年度に地球シミュレータセンター高度計算表現法研究グループ（AVPRG）が実施した3つのテーマ、EXTRAWING、バーチャルリアリティ可視化、および海洋データに関するビジュアルデータマイニング技術の研究開発、について報告する。

EXTRAWINGは地球シミュレータを用いた地球・環境流体シミュレーションによる結果を魅力的に表現し一般社会へ発信するプロジェクトである。これまでにGoogle Earth上でのボリューム表現を生成するために可視化ソフトウェアVDVGEの開発を進めてきた。またそのソースコードを地球シミュレータセンターのWebページ上に公開した。現在、VDVGEを用いた異分野横断的な利活用ワークショップを定期的に開催し、Google Earth上での3次元的な可視化表現の可能性と発展について議論、探究を進めている。

バーチャルリアリティ可視化については、モバイル装置を用い可視化オブジェクトを選択して再描画するシステムをVFIVEに試験的に実装した。VFIVEを用いて可視化操作を実行すると、その操作履歴がモバイル装置の“操作リスト”に反映される。利用者はこの操作リストからこれまでの可視化オブジェクトの表示／非表示を切り替えることができるようになる。しかし操作履歴の行数が膨大になるとこのリスト中の特定の操作記録の選択がしづらくなるという困難が見出された。この問題を回避するために、操作リスト中に“しおり”を挿入することで選択範囲を限定し目的の操作履歴を探し出し易くする機能を追加した。

海洋データセットのビジュアルデータマイニング研究についての主な内容は以下の3つ、(1) 海流の抽出と評価、(2) 抽出結果を用いた2次元カラーマップの最適化、および (3) 実空間と変数空間の間での対話的可視化解析、にまとめられる。これらの研究は前年度までの研究成果に客観的評価を与えさらに補強するものになるであろう。

キーワード:EXTRAWING, VDVGE, バーチャルリアリティ,モバイル機器, 海流のビジュアルデータマイニング