Development of a Visualization Scheme for Large-scale Data on the Earth Simulator

Group Representative

Yoshio Suzuki R&D Group for Parallel Processing Tools, Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute, Research worker

Authors

Yoshio Suzuki *1 ・ Kadunori Sai *2 ・ Fumiaki Araki *3

Hitoshi Uehara *3 ・ Hirofumi Haginoya *3

*1 R&D Group for Parallel Processing Tools, Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute, Research worker

*2 R&D Group for Parallel Processing Tools, Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute, Operating assistant member

*3 Computer Science R&D Program, Earth Simulator Center, Research Scientist

It is difficult to visualize and analyze efficiently the large-scale data outputted from simulations performed on the Earth Simulator, such as the atmospheric and oceanic simulation using the existing visualization method. Then, research and development of the tool and algorithm for visualizing such large-scale data efficiently have been furthered by the joint research of the Center for Promotion of Computational Science and Engineering at Japan Atomic Energy Research Institute (CCSE/JAERI) and the Earth Simulator Center at Japan Marine Science & Technology Center (ESC/JAMSTEC). Here, the visualization software PATRAS (Parallel Tracking Steering), which CCSE has so far developed in collaboration with NEC, has been transplanted to the Earth Simulator, and the optimization by vectorization and parallelization and the expansion for the improvement in convenience have been performed. In optimization, parallelization of image composition process, JPEG compression process, smoothing process in iso-surface generation and so on has been performed. The parallelization has led about 20 times speed-up for visualization process of atmospheric simulation code AFES (AGCM for Earth Simulator) when 64 CPUs are used. In expansion, the following functions have been developed: high resolution display, geographical feature data display, dynamic title display, axial display, arbitrary section contour line display, color bar detailed setting, the parameter preservation and coupling visualization.

Keywords: Large-scale data visualization, Parallel visualization

Reports of our result

1. Introduction

It is difficult to visualize and analyze efficiently the large-scale data outputted from simulations performed on the Earth Simulator, such as the atmospheric and oceanic simulation using the existing visualization method. Then, research and development of the tool and algorithm for visualizing such large-scale data efficiently have been done by the joint research of the Center for Promotion of Computational Science and Engineering at Japan Atomic Energy Research Institute (CCSE/JAERI) and the Earth Simulator Center at Japan Marine Science & Technology Center (ESC/JAMSTEC). Here, the visualization software PATRAS (Parallel Tracking Steering), which CCSE has so far developed in collaboration with NEC, has been transplanted to the Earth Simulator, and the optimization by vectorization and parallelization and the expansion for the improvement in convenience have been performed. In optimization, parallelization of image composition process, JPEG compression process, smoothing process in iso-surface generation and so on has been performed.

2. PATRAS visualization software

In CCSE, development of the visualization software PATRAS, which aimed at efficient visualization of the large-scale data in the Earth Simulator, has been furthered since 1998. This visualization software consists of a server function and a client one. Since server function is built as a subroutine group using MPI which can be performed on a parallel computer, a user can output a visualization image simultaneously with execution of a simulation by including subroutines in a simulation code as well as create a visualization image after execution of a simulation by including subroutines in a code which reads the existing simulation data. The main subroutines consist of four objects as follows: initialization, address setup of the data arrangement to each lattice type, main processing, and finalization. As a lattice type, the 3-dimensional BFC lattice, the non-structure tetrahedron lattice, and the non-structure sixth page object lattice can be selected. The client function operates on the web browser installed in a user terminal, such as PC and WS. It is possible to display the visualization image created by server function as well as to transmit visualization parameters to a server function.

3. Performance evaluation immediately after a transplant

The transplant to the Earth Simulator and the performance evaluation was performed with operation of the Earth Simulator. As a sample, the atmospheric simulation code AFES (AGCM for Earth Simulator) was used. The evaluation was performed for three cases: T106L20 (320x160x20 meshes), T159L20 (480x240x20 meshes) and T319L20 (960x480x20 meshes). The visualization image outputted for evaluation is shown in Fig. 1. This image does not have a physical meaning, but it is chosen so that processing may dare become heavy for the performance evaluation.

The dependence on the number of CPUs of image generation processing time when 120 images are outputted in case of T106L20 is shown in Fig. 2. Parallelization has done in the direction of latitude. Although the time which the sum of image generation process takes is decreasing with increasing of number of CPUs up to 8 CPUs, it increases after that and has become about 4 times longer when using 64 CPUs.

Figure 2 also shows what process has taken much time in the image generation process. It is found that JPEG compression process, image composition process and smoothing process of iso-surface generation (Phong shading process) come to occupy the time with increasing of the number of CPUs. Although the time of JPEG compression process does not depend on the number of CPUs, the half of the time is occupied when 8 CPUs are used. The similar result is obtained in other cases.

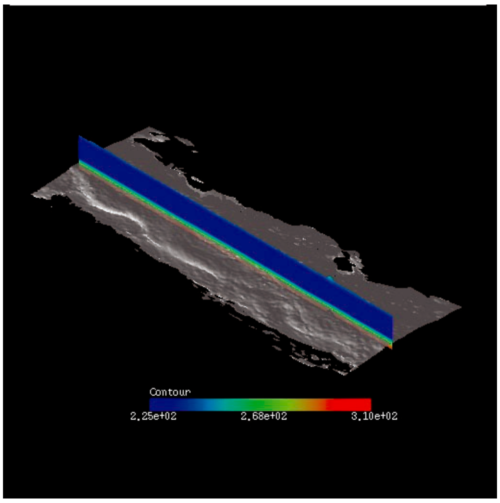

Fig. 1 Visualization image outputted in performance evaluation. The iso-surface (silver along the level surface), color contours (blue to red in the perpendicular section) and the color bar (displayed on the lower part) are displayed for temperature. The resolution of image is set to 512x512.

4. Optimization

Then, the optimization, that is, performance improvement was tried by vectorization and parallelization. Especially this chapter explains in detail the performance improvement made to three processes: image composition process, JPEG compression process and smoothing process in iso-surface generation (Phong shading process).

4.1. Image composition process

Although PATRAS before optimization performed visualization process in parallel by domain decomposition to a simulation domain, image composition was processed only by 1CPU. That is, firstly the visualization process of each

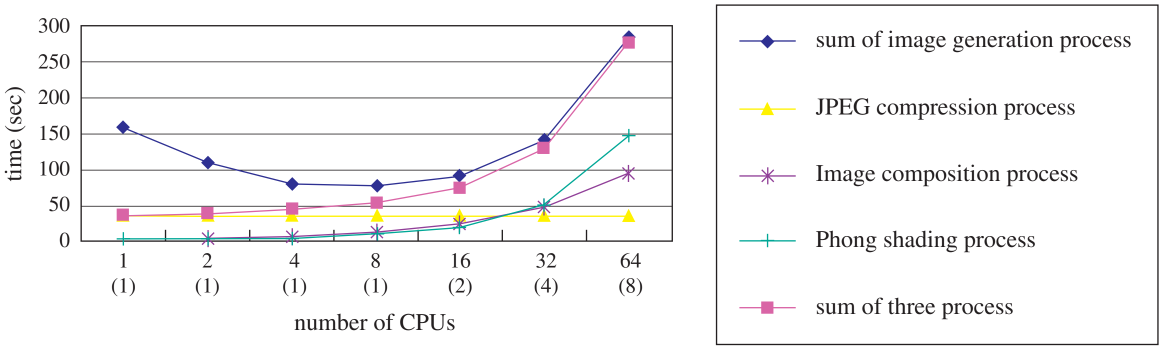

Fig. 2 Performance evaluation using AFES (T106L20) immediately after transplant of PATRAS. A horizontal axis shows the number of CPUs and a vertical axis shows processing time. The inside of the parenthesis of a horizontal axis shows the number of nodes.

domain was performed in each CPU, secondly the image information and Z Buffer information were collected to CPU of rank0, and finally the image composition was performed by rank0. Then, optimization has been carried out so that image composition is processed in parallel. Firstly the image in each domain generated by each CPU is re-divided one dimensionally and transmitted to each rank, secondly each re-divided image is composed in each rank, and finally the composed image in each rank is collected into rank0.

4.2. JPEG compression process

PATRAS before optimization performed JPEG image compression by rank0 after image composition. Then, optimization has been carried out so that JPEG image compression is processed in parallel. Firstly the composed image is divided using MCU (Minimum Coded Unit) and each divided image is transmitted to each rank, secondly each image is compressed in each rank, and finally each compressed image is collected into rank0.

4.3. Smoothing process in iso-surface generation

In PATRAS, smoothing process when generating iso-surface is performed using the Phong shading method. In the Phong shading method, smoothing process is performed as a result that the normal vector of the peak of each polygon is calculated using the normal vector of the center of each generated polygon. Here, on the boundary of each decomposed domain, the information of the normal vector, which different rank has, has to be used. In PATRAS before optimization, the information of a polygon and its normal vector was collected to rank0, and the normal vector on this boundary was calculated by rank0. Then, optimization has been carried out so that the information of a polygon and its normal vector is broadcast to all ranks, and is calculated in parallel.

5. Comparison of performance between before and after optimization

This chapter describes comparison of performance between before and after optimization. The performance to the sum of image generation process as well as each process before and after optimization in case of T106L20 is shown in Fig. 3. When the performance using 64 CPUs before optimization is compared with that after optimization, image composition process, JPEG compression process and Phong shading process required 94.5 seconds, 34.9 seconds and 145.9 seconds, respectively, before optimization, but 2.5 seconds, 1.0 seconds and 4.1 seconds, respectively, after optimization.

The comparison of performance before and after optimization in T106L20, T159L20 and T319L20 when 64 CPUs are used is shown in Fig. 4. This figure shows that the performance is improved about 20 times independently of a simulation scale. The performance of image composition process and JPEG compression process does not depend on the simulation scale although they depend on the image resolution. Since only the Phong shading process depend on the simulation scale, the increase in this processing time causes that in whole processing time.

The amount of the memory averaged by the number of CPUs before and after optimization in case of T106L20, T159L20 and T319L20 is shown in Fig. 5. Moreover, the amount of the memory used for AFES is also shown for comparison. This figure shows that the amount of the memory used in PATRAS after optimization is a little increasing compared with that before optimization. It is found that about 3/4 of the amounts of memory used for AFES is required for using PATRAS. About decreasing of the amount of memories used for PATRAS is a future subject.

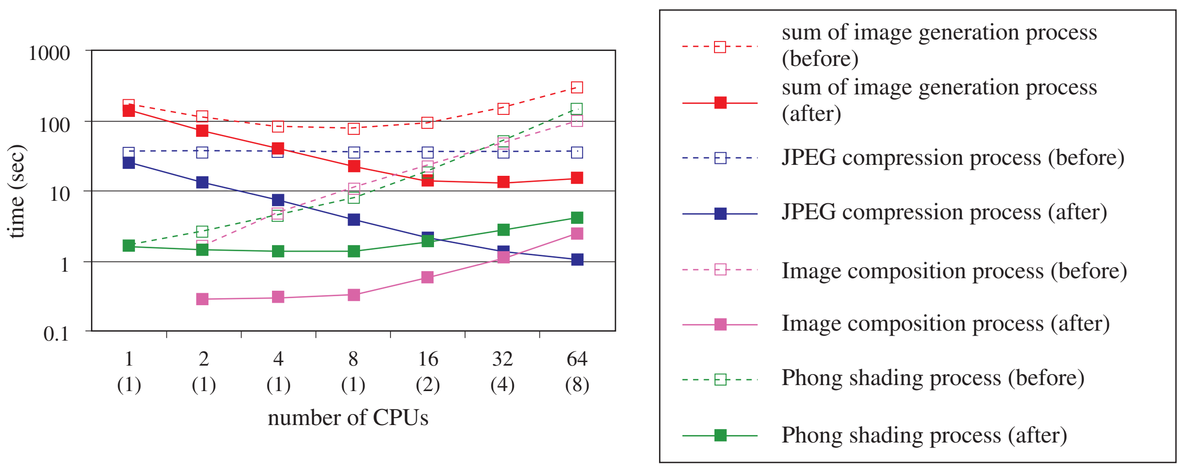

Fig. 3 Comparison of performance before and after optimization in case of T106L20. The dependence on the number of CPUs of the processing time before and after optimization to the sum of image generation process, image composition process, JPEG compression process and Phong shading process is shown. The inside of the parenthesis of a horizontal axis shows the number of nodes.

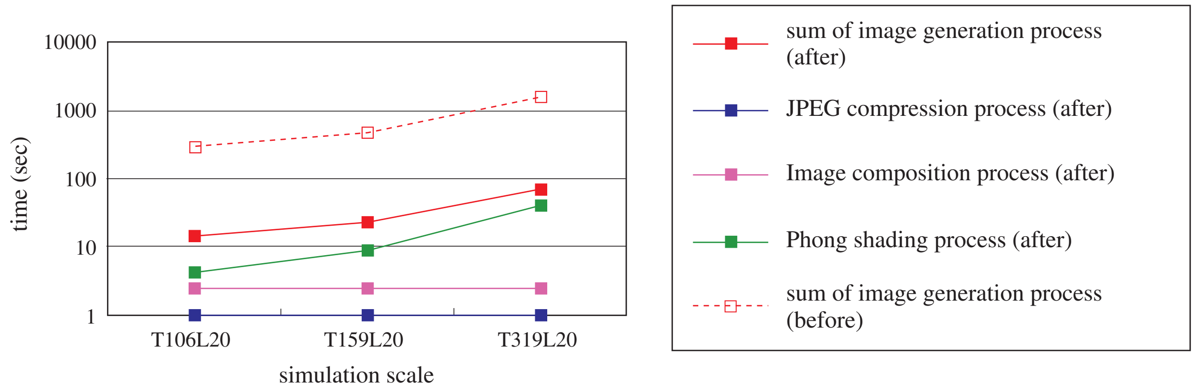

Fig. 4 Comparison of performance before and after optimization in case of T106L20, T159L20 and T319L20. The processing time is shown to the sum of image generation process before and after optimization and each process after optimization.

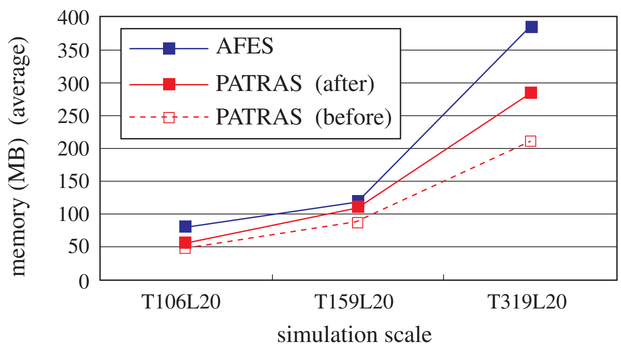

Fig. 5 The amount of the memory averaged by the number of CPUs before and after optimization in case of T106L20, T159L20 and T319L20.

6. Visualization of large-scale data

The performance evaluation in case of T106L20, T159L20 and T319L20 has so far been shown. Furthermore, in order to investigate performance improvement about a large-scale problem, the visualization process in case of T1279L96 (3840x1920x96 meshes) is evaluated. The condition is the same as that in previous three cases. The generated image is the same as that in Fig. 1. The number of outputted images is 120. The comparison of performance before and after optimization is shown in Fig. 6. While processing time is increasing with the increase in the number of CPUs before optimization, processing time is decreasing after optimization. When 64 CPUs are used, processing time before optimization was about 9348 seconds but that after optimization is about 1168 seconds. Thus, the optimization has led about 8 times speed-up for visualization process of such large-scale data.

7. Expansion

The following expansion has been performed for the improvement in convenience. 1) High resolution display function: the display by high resolution has been attained by losing restriction of resolution. 2) Topography data display function: the functional development for displaying the coastline has been performed. 3) Dynamic title display function: functional development of displaying title which is dynamically changed with user's setting has been performed. 4) Axial display function: the display of coordinate value has been attained for both 2-dimensional and 3-dimensional

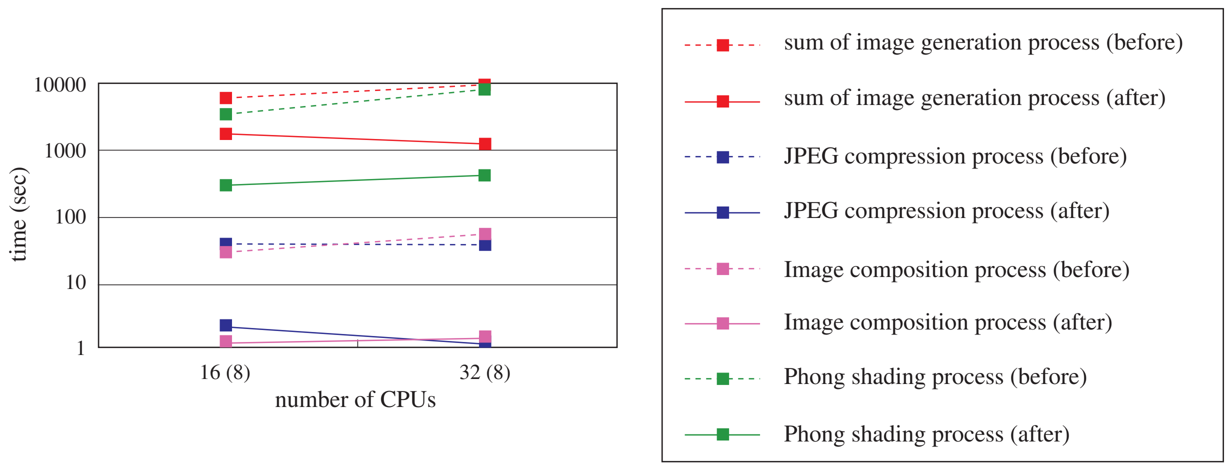

Fig. 6 Comparison of performance before and after optimization in case of T1279L96.

objects. 5) Arbitrary section contour line display function: contour lines have been displayed by identifying arbitrary sections from three points of space. 6) Color bar detailed setting function: functional development which can be set up about a color bar in detail than before has been performed. 7) The parameter preservation function by GUI: function to save the visualization parameter set up by GUI has been developed. 8) Coupling visualization function: function to set up two or more MPI communicators has been developed.

8. Conclusion and future subject

In this research and development, more efficient visualization has been attained by optimization and expansion of PATRAS to the Earth Simulator. However, the amount of the memory used, parallelization efficiency, etc. are considered to be improved much more. The further optimization and expansion is future subject.

Acknowledgement: We are thankful to members of the atmospheric group and oceanic group on Earth Simulator Center in advancing this research. Moreover, we are thankful to user management group at Earth Simulator Center and Inoue GL of use support room at the Japan Atomic Energy Research Institute, which helped about the procedure of use of earth simulator. AFES code used in this work was developed for Earth Simulator by NASDA and JAMSTEC.

地球シミュレータによる大量データに対する可視化処理法の研究

利用責任者
鈴木 喜雄 日本原子力研究所 計算科学技術推進センター 並列処理支援技術開発グループ 研究員

著者
鈴木 喜雄*1 ・ 齋 和憲*2 ・ 荒木 文明*3
上原 均*3 ・ 萩野谷裕文*3

＊1 日本原子力研究所 計算科学技術推進センター 並列処理支援技術開発グループ 研究員
＊2 日本原子力研究所 計算科学技術推進センター 並列処理支援技術開発グループ 業務協力員
＊3 地球シミュレータセンター 計算機技術研究開発領域 研究員

大気・海洋シミュレーションなど、地球シミュレータで実行されているシミュレーションから出力される大規模データを可視化するには、既存の可視化方法では支障をきたすことが予想される。そこで、日本原子力研究所計算科学技術推進センター（CCSE/JAERI）と海洋科学技術センター地球シミュレータセンター（ESC/JAMSTEC）の共同研究により、このような大規模データを効率的に可視化するためのツールやアルゴリズムの研究開発を行った。ここで、これまでCCSEがNEC社と共同で開発してきた可視化ソフトウェアPATRASを地球シミュレータで動作するよう移植を行い、ベクトル化・並列化による最適化、および大気・海洋結合シミュレーションへの適用のための改良などが行われた。最適化では、画像合成処理、JPEG圧縮処理、等値面生成時のスムージング処理に対する並列化などを行うことにより、最適化前後で最大約20倍の高速化が達成された。

キーワード： 大規模データの可視化、並列可視化