§32. Irreversible Data Compression Concepts in Time-order of Particle Trajectory for Visualization of Huge Particle System

Ohtani, H., Hagita, K. (Natl. Defense Academy of Japan), Ito, A.M., Kato, T. (Hiroshima Univ.), Saitoh, T. (Tokyo Inst. of Tech.), Takeda, T. (NAOJ)

The peta-scale computers, e.g., Titan, Sequoia and K computer, have started operation, and ultra-huge-scale simulations can be performed. Though many scientific discoveries are expected, there are big urgent issues, that are, data input and output (data I/O), transmission, saving, and analysis including visualization for the enormous volumes of data. These problems are a major bottleneck in the huge-scale simulation in many fields. In order to resolve the problems of data size, several data compression schemes were proposed. We propose in this paper that a compression of simulation data from a large-scale particle simulation holds favorable prospects for scientific visualization of particle systems[1,2].

Our concepts deal with the data of particle orbits obtained by simulation directly and have following features: (i) By the control over the compression scheme, the difference between the simulation variables and the reconstructed values from the compressed data becomes smaller than an allowed constant for the visualization. (ii) The particles in the simulation are regarded as independent particles and the time-series data for each particle is compressed with independent time-step for the particle. (iii) A particle trajectory is approximated by a polynomial function based on the characteristic motion of the particle. In other words, the parameters of each polynomial are stored in place of the original data. It is reconstructed as a continuous curve by the interpolation from the values of the function for intermediate values of the sample data. In short, the continuous curve which is established by a small number of parameters of the interpolation function approximates the original particle trajectory and the amount of necessary data can be reduced. We name this concept "TOKI (Time-Order Kinetic Irreversible compression)" originated from irreversible compression of time-ordered data sequences, such as a particle trajectory, for visualization.

In this paper, we encode the simulation data as a post process to examine whether our concepts of data compression and the encoder work well or not. That is, we encode the data which has been already obtained by simulation. Moreover, we use a cubic function as an interpolation polynomial for the encoder test.

We first apply our compression scheme to the plasma particle simulation. The simulation code is a three-dimensional electromagnetic particle-in-cell simulation. The boundary conditions are periodic. The data for compression is a part of full simulation data, and the number of particles in the encoder test is 800. The obtained data was calculated during 10,000 time steps.

Figure 1 shows the comparison between the simulation data and the decoded data reconstructed from the encoded data by the decoder for one electron in x direction, and the error between the simulation and the decoded data. It is found that the decoded data reproduces the high-frequency motion of the particle, and that the error between the simulation data and the decoded data is less than the allowed error value (=0.1). The error by the encoder can be controlled due to the allowed error value. Compression rate is defined as the ratio of the encoded data to the raw data. The rate is 5.7% in this case. We apply this concept to the galaxy formation simulation result, and get 32.2% compression rate. These rates are good for practical application.

We proposed the data compression method "TOKI" of particle trajectory data produced by simulation. It was possible to realize a good reproducibility from encoded data and to control an error between the decoded values and simulation data. The next stage is that we install this encoder into the simulation code and encode simultaneously the data with performing simulation.

This work was supported by National Institutes of Natural Sciences (NINS) Program for Cross-Disciplinary Study.

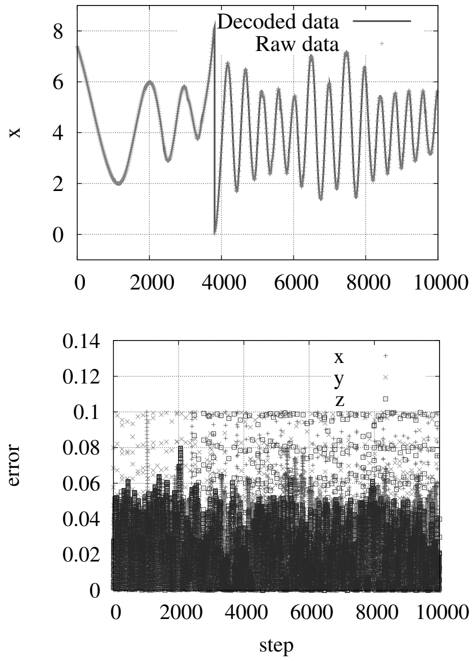

Fig. 1. Reproducibility and error between the simulation and the decoded data. (top) Time evolution of the particle x coordinates is shown. Red line indicates the decoded data, and green plus symbol shows the simulation data. (bottom)Errors (difference) between the simulation and decoded data are shown. Plus, cross and square symbols are the errors for x, y and z coordinates, respectively.

1) Ohtani, H. *et al*: Conference on Computational Physics 2012
2) Ohtani, H. *et al*: J. Physics: Conference Series, accepted.