Efficient data compression of time series of particles' positions for high-throughput animated visualization

Katsumi Hagita
National Defense Academy
1-10-20, Hashirimizu,
Yokosuka, 239-8686, JAPAN
+81-46-841-3810
hagita@nda.ac.jp

Takaaki Takeda
VASA Entertainment, Inc
4-14-17, Sennin-cho,
Hachiouji, 193-0835, JAPAN
+81-42-673-7867
takedatkgm@gmail.com

Tsunehiko Kato
Hiroshima University
1-3-1, kagamiyama,
Higashi-hiroshima, 739-8526, JAPAN
+81-82-424-7362
kato@hiroshima-u.ac.jp

Hiroaki Ohtani
National Institute for Fusion Science
322-6, Oroshi-cho,
Toki, 509-5292, JAPAN
+81-572-58-2556
ohtani.hiroaki@nifs.ac.jp

Seiji Ishiguro
National Institute for Fusion Science
322-6, Oroshi-cho,
Toki, 509-5292, JAPAN
+81-572-58-2540
ishiguro.seiji@nifs.ac.jp

ABSTRACT

We tried to improve data-reading throughput by data compression on animated visualization of time series data generated from simulations of huge particle systems. It is considered that data compression is expected to reduce problems related to relatively slow reading from storage device. In the present paper, we examined efficient data compression schemes for huge time series data of plasma particle simulation. For example, to read positions of 300,000 particles with double precision text, required Read Throughput is almost equal to that of recent SSD. Thus, we considered data compression enables visualization of larger system. For smooth animations, compression of trajectories in time order is effective, because correlation of positions of a certain particle in time orders is higher than spatial correlation. Especially, lossy compression with polynomial functions in time order shows good performances. We called this scheme as TOKI (Time-Order, Kinetic, and Irreversible) compression.

Categories and Subject Descriptors

I.3.6 [**COMPUTER GRAPHICS**]: Methodology and Techniques – *Graphics data structures and data types.*

General Terms

Experimentation

Keywords

Animated visualization, Particle system, Data compression, CAVE based virtual reality system

1. INTRODUCTION

Data supply is one of important elements for smooth virtual reality (VR) of huge particle systems, such as N-body problems of galaxy formation, plasma behaviors on fusion science, and particle formation and dynamics on statistical physics and material sciences. Especially, data-reading from storage device is rate controlling as well as data-filtering and rendering. Recently, there are many studies about massively parallelization for data-filtering and rendering with distributed storage system[1 ,2] with MPI-IO and/or parallelized HDF-5 [3]. For usual virtual reality system, such as CAVE VR system, single storage device is implemented, such as raid system of HDD and/or SSD. Data compression is effective for data supply problem for time series of huge number of particles' positions generated from massively parallel particle simulations. We studied effectiveness of the data compression in time orders by using HDF-5 (Hierarchical Data Format ver. 5) library [3], XTC library [4], and our TOKI (Time-Order, Kinetic, and Irreversible) compression scheme [5]. TOKI compression scheme is lossy compression scheme of trajectory (sequence of positions) of each particle with polynomial functions. Here, we defined allowable error under encoding to find coefficients of polynomial and length of encoded data. For fast computation, we used 3rd polynomial functions. It is considered that correlation among the sequential data is much because time interval of the sequential data is set to be enough small to produce smooth animation.

In the present analysis, we focus on sequential through-put of reading files based on file size of compressed files. It should be noted that total performance including with parallelized I/O system is important for implementation to software.

2. METHOD

2.1 TOKI compression

In TOKI compression algorithm, we find longest interval to express 3rd polynomial function within (given) allowed error ε.

$$\max_{i=s_j}^{s_j+l_j} \left| x_i - \sum_{n=0}^{3} a_{n,j} t_i^n \right| < \varepsilon \qquad (1)$$

Here, x_i denote time series of positions of a certain particle at time i. s_j and l_j is starting time and interval of jth block, respectively. $a_{n,j}$ are coefficients of 3rd polynomial function. We record these $a_{n,j}$ instead of x_t. It achieves reduction of file size. When eq. (1) does not satisfy, we record x_t directly. Then, error of the compress data is always within the allowed error ε by definition. Figure 1 shows cartoon of TOKI compression.

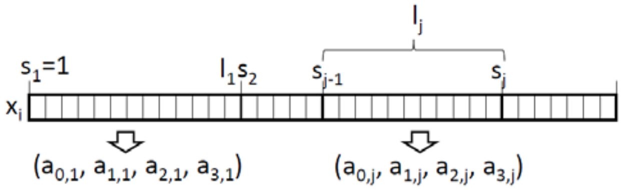

Figure 1. Cartoon of TOKI compression

2.2 Test data (Plasma particle simulation)

We performed a plasma particle simulation to obtain 24,000 time steps (about 800sec with 30frames/sec) of 3d positions of 150,000 electron particles and 150,000 ion particles. In this simulation, we solved the dynamics of particles and the electromagnetic field self consistently by explicit electromagnetic particle-in-cell (PIC) method [7]. The initial condition is given by a one-dimensional Harris type equilibrium [8], in which magnetic field is parallel to x-axis and a function of y-coordinate. The initial velocity distribution of particles is a Maxwellian with a uniform temperature, in which the thermal velocities of electron and ion are respectively 0.256c and 0.0256c, where c denotes the speed of light. The ratio of ion mass to electron mass is 100. The boundary condition is a periodic boundary, and the simulation box size is $3.85c/\omega_{ce}$ x $1.92\ c/\omega_{ce}$ x $3.08\ c/\omega_{ce}$, where ω_{ce} is the electron cyclotron frequency. Here, we set the numbers of grids for field variables (electromagnetic field, charge and current densities) 16 x 15 x 16 as a test case. In plasma physics, $\Delta t = (0.5/c)\ \min(dx, dy, dz)$ is empirically used to avoid discretization errors. In order to obtain smooth trajectories of particles, we set time step $0.016/\omega_{ce}$, which is 1/4 times of $\Delta t = (0.5/\omega_{ce})\ \min(3.85/16, 1.92/15, 3.08/16)$.

In plasma physics, smooth animation is considered as one of tool to analyze complex plasma phenomena, for example, particle acceleration in magnetic reconnection and shock wave. Thus, it is needed to record positions of all plasma particles in short time interval. Therefore, the sequence of data of a position has much correlation.

3. RESULTS AND DISCUSSIONS

Data size of 24,000 time steps of 3d positions of 300,000 particles is about 525GB in FORTRAN double precision text format. For 30 frames per sec, it corresponds to 657MB/sec, which is faster than SSD. Here, we set chunk size of HDF-5 to (32, 3, 100) under chunk dimensions for particle id, space dimension, and time steps. In the present paper, file sizes are examined for precision determined by decimals 0.001 and 0.01. Results of sizes of compressed files are shown in Table. 1. Required speed for 30 frames/sec animated visualization is also presented in Table. 1.

Table 1. File size and required speed of compressed files

Filesize (GB) / Speed (MB/sec)	24000frames Δt=.016/ω_{ce}	6000frames Δt=.048/ω_{ce}	1000frames Δt=.384/ω_{ce}
Text (double)	525.6 / 657	131.4 / 657	21.9 / 657
Binary (double)	172.9 / 216	43.2 / 216	7.20 / 216
HDF5(double)	163.2 / 204	41.4 / 207	6.92 / 208
HDF5(single)	78.4 / 98.0	19.7 / 98.5	3.30 / 98.9
HDF5(0.001)	60.7 / 75.8	16.1 / 80.5	2.73 / 82.0
XTC(0.001)	43.7 / 54.7	11.0 / 54.8	1.82 / 54.5
TOKI(0.001)	17.1 / 21.4	5.13 / 25.6	2.77 / 83.1
HDF5(0.01)	36.2 / 45.3	11.3 / 56.5	2.14 / 64.3
XTC(0.01)	34.8 / 43.7	8.69 / 43.5	1.45 / 43.5
TOKI(0.01)	10.0 / 12.5	3.55 / 17.7	2.31 / 69.2

For all formats except for our TOKI format, data size is almost proportional to number of frames. It is found that increase of data size of the TOKI format is weak for increasing of number of frames. We considered TOKI is good for smooth animation in precision required for visualization.

It is noted that size of TOKI compressed trajectory of ions with Δt=.016/ω_{ce} is 1/3.5 to 1/6.1 times of that of electrons because mass of an ion is smaller than that of an electron. For other cases, compressed file size for ions is almost same as that for electrons.

4. CONCLUSIONS

We implemented data compression by HDF-5, XTC, and our TOKI format to OpenGL based visualization programs in order to improve performance of data reading from storage device and/or file system. We confirmed TOKI is good for smooth animation in precision required for visualization. It is confirmed that data compression is good solution to resolve the problem of slow reading of files.

5. ACKNOWLEDGMENTS

This work is partially supported by "Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures" in Japan, NIFS/NINS under the project of Formation of International Scientific Base and Network, and HPC project of Nagoya University. The authors thank to Dr. A. M. Itoh, and Dr. T. Saitoh for useful discussions at early stage of this work.

6. REFERENCES

[1] http://www.paraview.org/

[2] K. Ono. Private communication. http://labs.aics.riken.jp/ono.html

[3] http://www.hdfgroup.org/

[4] http://www.gromacs.org/

[5] H. Ohtani, K. Hagita, A. M. Ito, T. Kato, T. Saitoh and T. Takeda. J. Phys: Conf. Ser., 454, 012078, 2013.

[6] H. Ohtani, K. Hagita, A. M. Ito, T. Kato, T. Saitoh and T. Takeda. IEEE VIS2013 Poster presentation, 2013.

[7] C.K.Birdsall and A.B.Langdon: Plasma Physics Via Computer Simulation. McGraw-Hill, New York (1985).

[8] E.G.Harris: Nuovo Cimento, Vol.23, p.115, 1962.