

In-situ 可視化向け適応的時間サンプリング法 Adaptive time-step sampling for in-situ visualization

- 山岡 義明, 神戸大学, 兵庫県神戸市灘区六甲台町 1-1, E-mail: 180x220x@stu.kobe-u.ac.jp
林 賢悟, 神戸大学, 兵庫県神戸市灘区六甲台町 1-1,
坂本 尚久, 神戸大学, 兵庫県神戸市灘区六甲台町 1-1
野中 丈士, 理化学研究所, 兵庫県神戸市中央区港島南町 7-1-26
吉永 司, 豊橋科学技術大学, 愛知県豊橋市天伯町雲雀ヶ丘 1-1
野崎 一徳, 大阪大学, 大阪府吹田市山田丘 1-1

Recently, In-Situ visualization, which performs simulation and visualization at the same time without data I/O, has attracted attention as simulations become larger. To reduce the required time, many In-Situ visualization techniques and systems that focus on efficient data management between simulation and visualization have been proposed. However, a technique, which can analyze and visualize the dataset efficiently in consideration of interesting phenomena occur in the simulation results, is needed to reduce the time to discover knowledge. We propose an In-Situ visualization method, which can efficiently visualize the simulation results without missing important features. In this method, a metric that is calculated by the statistical information of the physical quantity distribution in the simulation results is used to detect spatio-temporal changes of the distributions and then changes the time steps specified for the visualization process adaptively. As a result, the required time for the visualization is reduced, while capturing important phenomena.

1. 緒言

近年の計算機環境の高性能化及び計算科学技術の高度化によって、スーパーコンピュータなどを使った HPC (High Performance Computing) 環境下での大規模な数値シミュレーションが行われるようになり、物理現象の解明や新素材の開発、ものづくりなど様々な分野で幅広く活用されている。それに伴い出力されるデータもより大規模かつ複雑化し、これらのデータから新たな知見を獲得するための効率的かつ効果的な可視化処理及び分析技術が必要とされている。今後もさらに大規模・複雑化される数値シミュレーションを想定した場合、従来のようなシミュレーション終了後に可視化処理を方法では、HPC 環境下のストレージに保存されたデータを手元の可視化用 PC に転送するコストが著しく増大し、可視化処理に膨大な時間を要してしまう。また、数値シミュレーションによっては、HPC 環境下のストレージに保存することさえ困難なスケールでの計算が行われており、可視化処理が実行不可能な状況も発生している。このような状況に対して、シミュレーション結果データを可視化用 PC に転送することなく、HPC 環境上で可視化処理も行うことで、大規模な数値シミュレーション向けに効率的に可視化することができる In-situ 可視化技術が注目されている。

一般に、In-situ 可視化によって大規模数値シミュレーション結果の可視化を効率良く実行することができるが、スパコン上での実行を想定した場合には従来のような対話的な可視化が困難となり、カメラパラメータや可視化パラメータなどは事前に設定しておく必要がある。そのため、シミュレーション実行前にその結果を想定して、可視化手法およびそのパラメータを複数設定しておき、さらに、視点位置も複数想定して可視化処理を実行する必要がある。従来の可視化処理では、GPU などを活用することによって大規模なデータであっても対話的な可視化が実現可能であり、スパコン上でのシミュレーション時間に比べると極めて高速に処理が可能であった。しかし、In-situ 可視化を想定した場合、データサイズの大規模化だけでなく、処理回数の爆発的な増加によって、可視化処理コストが増大するといった問題が発生している。

2. 関連研究

現象の変化に着目しデータの分布情報から空間方向または時間方向に効率良くデータを間引く技術がいくつか提案されている。Thompson ら(1)や Liu ら(2)は、シミュレーションによって出力されたデータから確率密度分布を推定し、この情報を基に空間方向にデータを間引くことで、データの特徴を捉えつつデータ量を削減して可視化する手法を提案している。しかし、これらの手法はシミュレーションによって出力されたデータから確率密度分布を推定するが、より大規模で複雑な出力データの場合は確率密度分布をうまく推定できないことがある。この場合、分布情報を基に削減したデータを可視化した場合に、重要な特徴を捉えられない可能性がある。一方、Ling ら(3)は、分布情報に基づき特徴重要度を定義して機械学習のアプローチで、重要である時間区間を自動的に識別する方法を提案した。通常、時間方向に一定間隔でデータを間引くと、細かい時間間隔で状態が変化する場合などに重要な現象を見落としてしまう可能性がある。この Ling らの方法では、特徴重要度が高いデータだけをストレージに保存することで、効果的にストレージへのデータ I/O 回数及びストレージコストを削減することに成功した。しかし、Ling らは局所的な領域での変化に注目しており全体としての変化を考慮していない。また In-Situ 可視化とは組み合わせず、In-Situ 可視化にこの技術を適応するためには、適応的に変更された時間間隔で可視化処理を行うためのデータ管理システムが必要になる。

本研究では、これらの問題を解決するために、シミュレーションによって計算される物理量分布の統計情報をもとにして特徴量の時空間変換を考慮した適応的 In-Situ 可視化技術を提案する。まず、統計情報を利用したデータ分析では、物理量の空間分布情報に基づいて時間軸方向の状態変化量を表す指標を計算し、この指標に応じて可視化処理を行う時間間隔を適応的に変更して In-Situ 可視化を行う。この技術を使うことで、計算した指標が高い時間領域に関しては、細かい粒度で詳細に可視化処理を行い、あまり変化が見られない時間領域に関しては、粗い粒度で可視化処理を行う。これによりシミュレーションにおいて興味がある重要な時間空間領域の特徴を捉えつつ、可視化処理回数を削減することで In-Situ 可視化全体の効率化を実現する。

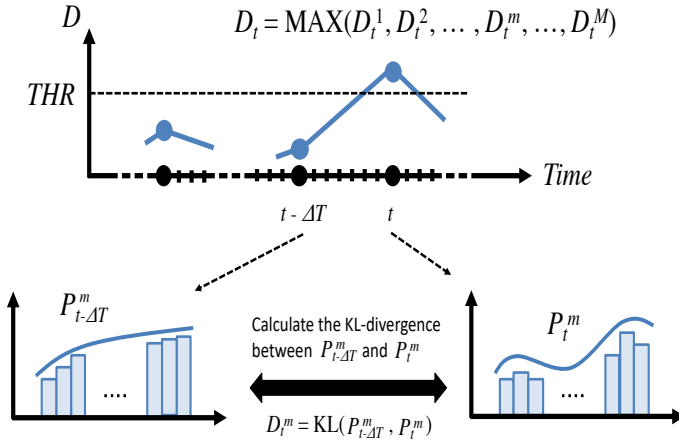


Figure 1 KL divergence between the corresponding volumes.

3. 提案手法

提案する手法は、シミュレーション結果データに対して、時空間変化を評価するための指標計算と、その結果を使って可視化処理を実行する時間間隔を適応的に決定する方法からなる。本章では、各手法の説明とそれらを In-Situ 可視化システムへ組み込む方法について説明する。

3.1. 変化量の計算

本手法では、本手法では、複数の領域に分割された時系列ボリュームデータを対象とし、着目する物理量分布の時空間変化を考慮して適応的に時間間隔を変更して In-Situ 可視化を行う。まず、時空間変化検知するための指標を計算するために、時刻 t における領域 m のボリュームデータ V_{tm} から確率密度分布を推定する。座標 x におけるスカラ値を $V_{tm}(x)$ とすると、時刻 t における m 番目の領域の確率密度分布 P_{tm} は、カーネル密度推定法 (KDE 法) (4) を用いて以下のようにして計算することができる。

$$P_t^m(s) = \text{KDE}(V_{tm}^m, s) = \frac{1}{nh} \sum_x K\left(\frac{s - V_t^m(x)}{h}\right) \quad (1)$$

ただし、 s はスカラ値、 n はデータの格子点数、 K はガウス関数、 h はバンド幅である。

次に、2 つの分布の差を定量的に評価するために、カルバック・ライブラー情報量 (KL 情報量) (6) を計算する。変化量を求める時刻 t とその前の時刻 $t - \Delta T$ の確率密度分布を比較し、以下のように KL 情報量 D_{tm} を計算する。

$$D_t^m = \text{KL}(P_{t-\Delta T}^m, P_t^m) = \sum_{s_{min}}^{s_{max}} P_t^m(s) \log \frac{P_t^m(s)}{P_{t-\Delta T}^m(s)} \quad (2)$$

ただし、 s_{min} , s_{max} は処理対象とする全ボリュームデータのスカラ値の最小値と最大値を表す。本手法では、領域全体を考慮した変化を検知するために、領域ごとに計算された KL 情報量のうち最大値を MPI 通信により全領域で共有し、これを時刻 t における変化量 D_t として以下のように計算する。

$$D_t = \text{MAX}(D_t^1, D_t^2, \dots, D_t^m, \dots, D_t^M) \quad (3)$$

3.2. 時間間隔の適応的決定

前節で計算する変化量 D_t に対して閾値を設定することで可視化するかどうかの判断を行う。このとき、 D_t の時間変化を考え、変化パターンを次の 3 つに分類する。

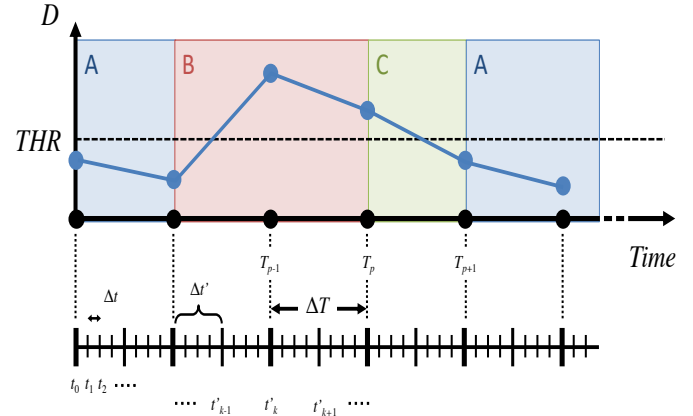


Figure 2 Variation patterns of KL divergence.

- (A) 変化量が閾値よりも低い状態が続く区間 (変化の小さい時間領域)
- (B) 変化量が閾値よりも低い状態から高い状態に変化し、その後高い状態が続く区間 (変化の大きい時間領域)
- (C) 変化量が閾値よりも高い状態から低い状態に変化する区間 (変化が小さくなる時間領域)

本手法では、この 3 つの状態を判断し、それに応じて可視化処理を行う時間間隔を適応的に変更する (図 2)。パターン A では、基本的に細かく可視化処理を行う場合の間隔 $\Delta t'$ よりも粗い間隔で可視化処理を行う。この粗さの度合いを R とすると、パターン A で行われる可視化処理の間隔は $R\Delta t'$ となる。この値が変化量の計算を行う間隔 ΔT よりも大きくなった場合は可視化処理を行わない。パターン B は、変化の激しい領域であるため、 $\Delta t'$ 間隔の細かい粒度で可視化処理を行う。パターン C は、変化の激しい時間領域から変化の小さい領域に移移しているため、パターン A とパターン B の時間間隔を組み合わせた可視化処理を行う。変化量を計算した現在の時刻を T_p 、前回の時刻を T_{p-1} とすると、 T_{p-1} から $T_{p-1} + \Delta T/2$ までは $\Delta T'$ 間隔で可視化し、それ以降は $R\Delta t'$ 間隔で可視化処理を行う。つまり、半分の地点までは細かく可視化し、それ以降は粗く可視化する。

4. 実験結果

本手法の有効性を示すために、歯茎摩擦音の発音シミュレーション (6) に対して提案する In-Situ 可視化手法を適用し評価実験を行った。このシミュレーションは、摩擦音の発生メカニズムを調べることが目的に、口腔内で形成される狭窄流路から発生するジェット流と音の伝播を同時に計算するものである。実験では、CT 画像から生成される口腔形状データから生成される要素数が 3,197,279 個の六面体要素からなる非構造格子データを利用し、時間ステップが 20,000 の計算を実施する。実験は、CPU が Intel Xeon Platinum 8168 (Skylake) (24cores/2.7Ghz/33MB) x 2 (node) の環境上で、48 領域に要素を分割して計算処理及び可視化処理を行った。領域の分割は、それぞれの領域における要素数ができるだけ均等になるように行った。

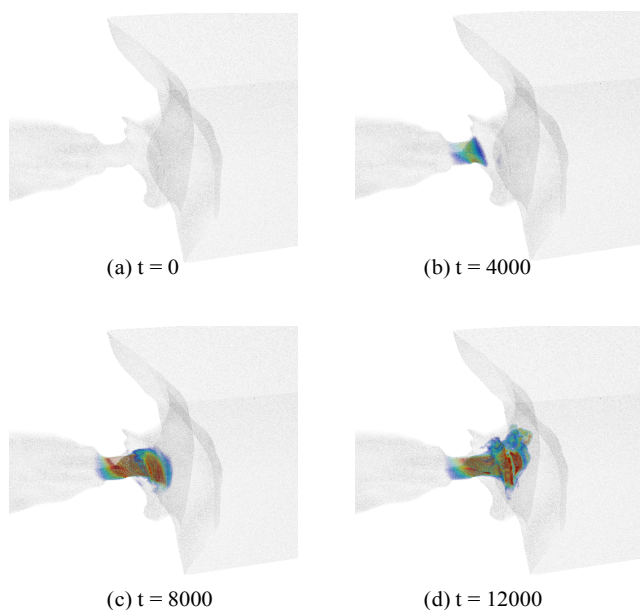


Figure 3 Visualization results.

4. 1. 可視化結果

In-Situ 可視化により出力した可視化結果を図 3 に示す。本実験における可視化には、データの空間分布を把握することを目的として、大規模非構造格子データ向けに効率よくボリュームレンダリングを行うことができる並列粒子ベースレンダリング手法(7)を活用し、計算領域境界を示すポリゴンデータを同時に半透明可視化した。また、本手法を使って適応的に可視化した際に、変化量が指定した閾値を超えた場合と超えなかった場合、つまり変化があると判断された場合と変化なしと判断された場合の可視化結果画像を図 4 に示す。この結果より、変化なしと判断された画像に対して、変化ありとされた画像は赤く可視化されている領域に変化があることが確認できる。しかし、可視化画像は視点や可視化パラメータなどによって結果が変化するため、数値として示される変化量と画像として確認できる変化量が異なる場合があることがわかった。

4. 2. 性能評価

シミュレーションを 20,000 ステップ実行し、本手法を使って In-situ 可視化を行った場合の性能評価を行った。実験では、変化量を計算する間隔 (ΔT) を 100 に、最も細かく可視化する間隔 ($\Delta t'$) を 20 に設定した。つまり、本手法を適用しなかった場合、1,000 回の可視化処理を実行することになる。このとき、可視化処理間隔の粒度 (R) を 5 と 10 に設定して In-situ 可視化を実行した際に、どの程度可視化処理回数が削減されるかを評価した。結果は、 R が 5 の時は 592 回、10 の時は 514 回の可視化処理が実行され、およそ半分の処理回数で大きな変化が期待される時間領域を重点的に可視化できていることを確認した。 R については、5 の場合も 10 の場合も可視化回数に大きな変化はなく、結果画像からも大きな違いは見られなかったが、 ΔT を変更した場合、特にその値を大きくした場合には、特徴的な変化の検知に失敗し違和感のある結果画像となった。今回の提案手法では ΔT を事前に指定しておく必要があるが、この値は結果に対して大きな影響を及ぼすため、適用するシミュレーションに応じて最適な値を推定する技術の開発が必要となる。また、今回の実験では、 R を ΔT よりも小さな値に設定することで、あまり変化が起きていない時空間領域に対してもある程度の間隔で可視化を行うことができ、対象とする現象の特徴をより捉えた In-Situ 可視化処理が可能になることが分かった。

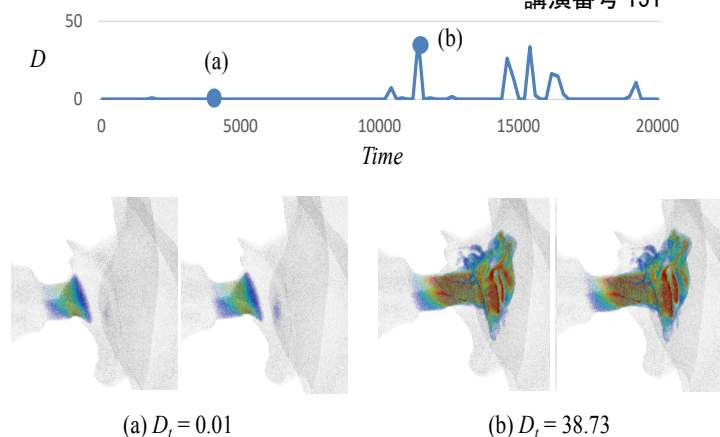


Figure 4 Comparison of images with (a) low and (b) high KL divergences.

5. 結語

本研究では、シミュレーションによって計算される物理量分布から時空間変化を定量的に数値化する手法とその結果を使って適応的に可視化時間間隔を推定する手法を提案した。そして、それらの技術と並列分散環境に適した可視化技術を組み合わせた In-Situ 可視化システムを開発した。時空間変化を検知する分析処理では、シミュレーション結果から得られた物理量分布を利用して変化量を計算した。さらに、計算した指標に応じて適応的に可視化処理を行うことで、シミュレーションで重要となる特徴を捉えながら全体の処理における可視化処理にかかる時間を削減することが可能になった。実験により、分布情報から計算した時間軸方向の状態変化量を表す指標が実行したシミュレーションの変化と対応していることが確認できた。この指標に応じて適応的に可視化処理の時間間隔を変更することで、可視化処理を行う回数を削減し In-Situ 可視化全体の効率化に成功した。今後は、本手法で事前設定していたいくつかのパラメータの自動推定技術に関する研究を進め、さらに、シミュレーション処理と可視化処理を非同期的に実行することによってさらなる In-Situ 可視化処理の効率化を目指す。

謝辞

本研究は、科研費(17K00169, 19H03976)の助成を受けて行った。また、本研究成果の一部は、理化学研究所計算科学研究センターの京クラウドシステムを利用して得られたものである。

参考文献

- (1) David C. Thompson, Joshua A. Levine, Janine Bennett, Peer-Timo Bremer, Attila Gyulassy, Valerio Pascucci, and Philippe P. Ébay, "Analysis of large-scale scalar data using hixels", Proc. of IEEE Symposium on Large Data Analysis and Visualization (LDAV), 1(2011), pp. 23–30.
- (2) Shusen Liu, Joshua A. Levine, Peer-Timo Bremer, and Valerio Pascucci, "Gaussian mixture model based volume visualization", Proc. of IEEE Symposium on Large Data Analysis and Visualization (LDAV), 1 (2012), pp.73–77.
- (3) Julia Ling, W Philip Kegelmeyer, Konduri Aditya, Hemanth Kolla, Kevin A Reed, Timothy M Shead, and Warren L Davis, "Using feature importance metrics to detect events of interest in scientific computing applications", Proc. of IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV), 1 (2017), pp. 55–63.

- (4) Emanuel Parzen, “On estimation of a probability density function and mode”, *The annals of mathematical statistics*, Vol. 33, 3 (1962), pp. 1065–1076.
- (5) Solomon Kullback and Richard A Leibler, “On information and sufficiency”, *The annals of mathematical statistics*, Vol. 22, 1 (1951), pp.79–86.
- (6) Tsukasa Yoshinaga, Kazunori Nozaki, and Shigeo Wada, “Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract model”, *Physics of Fluids*, Vol. 30, 3 (2018), p.035104.
- (7) Naohisa Sakamoto, Jorji Nonaka, Koji Koyamada, and Satoshi Tanaka, “Particle-based volume rendering”, *Proc. of 6th International Asia-Pacific Symposium on Visualization (APVIS)*, 1 (2007), pp.129-132.