301

協調ビジュアルデータマイニングのための テレマージョン環境の構築

小山田 耕二(京都大学)

Teleimmersive environment for collaborative visual data mining

Koji KOYAMADA, Kyoto University

ABSTRACT

To effectively visualize huge volume datasets, we should consider the following view points: 1) A collaborative analysis of the datasets is required. 2) The datasets may not be stored on hand. 3) The datasets may overflow into physical memory. These can cause a problem on interactive visualization. To solve the problem, we will initiate a project in which we plan to develop a teleimmersive environment for collaborative visual data mining.

Keywords: Collaborative Visualization, Visual Data Mining,

1. はじめに

計算機シミュレーションは、実験や観測が困難な現象に対して実験を他のものに変えたり、理論的解析が困難な問題に対して、すでに知られている原理をベースにそれを多数組み合わせ、積み上げることにより解析し、新しい理論を構築する技術である。近年、計算機の性能向上に伴い、より複雑で現実に近いモデルが取り扱う大規模シミュレーションが行われるようになった。

計算機の構成するプロセッサ単体の計算速度は、限界に近づいてきたため、大規模シミュレーションを行うための計算機は必然的に超並列計算機となる。米国の Accelerated Strategic Computing Initiative(以下 ASCI)プロジェクトでは、現在超並列計算機を使って、1億個の格子を使うような大規模シミュレーションが行われている。このシミュレーション結果(以下、ボリュームデータと呼ぶ)を可視化するために、分散メモリに格納されている数値データをひとつの計算機ノードに集める必要が出てくる。

ASCI プロジェクトに参加している Sandia National Laboratory(以下 SNL)では、分散メモリに出力されたボリュームデータ(総格子数1億)をファイルに書き出し、それらを可視化専用システムに転送している。計算ノードの数は、2000 ほどなので、生成されるファイルの数も同数となる。それぞれのファイルに対して、等値面が3角形メッシュとして計算される。同時にデータサイズの

削減も行われ、最終的にはほぼ1000万の3角形から構成される1つのファイルになる。この場合、転送時間を含めて、ボリュームデータの可視化結果表示まで3、4時間かかると報告されているが、研究効率の観点からは望ましいとはいえない。⁴⁾

現在行われている気象シミュレーションでは、地球全球を経度方向360分割、緯度方向150分割、高さ方向に100分割した格子(480万格子)を利用する。将来的には、経度方向3600分割、緯度方向1500分割、高さ方向に150分割した格子(7.2億格子)を利用する予定である。この場合、ASCI プロジェクトの例と同じシステム環境を利用したとすると $(7.2)^{2/3} \approx 3.73$ 倍の時間が必要となる。

これらのボリュームデータは、その結果の解析を必要とする研究者の手元にないことが多い。我々は、この問題を解決するために遠隔地で生成された大規模ボリュームデータに対する、複数人数によるビジュアルデータマイニングを支援するための基盤(テレマージョン環境)を開発する予定である。ビジュアルデータマイニングとは、可視化情報学(そのままでは目に見えない情報を画像化するための学問)の新しい側面で、多量の3次元画像情報の中から意味のあるデータを発掘したり、データ発掘のための新しいアルゴリズムや可視化手法の開発することをさす。データ量が膨大になるにつれて、ひとりだけでのデータ発掘では不十分になってきている

ため、“複数人数による”とした。テレマージョンとは、もともと“バーチャルリアリティとテレビ会議を融合した新しいテレコミュニケーションのメディア”という意味である。本研究では、遠隔地にあるスパコンや高性能計測装置から生成される膨大なデータの生み出すリアリティは、バーチャルリアリティに含まれるものと解釈している。

データ発掘には、まず、データ全体を見渡し、そして、重要だと認識した領域について、詳細に調べる戦略が有効である。最終的な宝の発掘は人間が行うものとし、システムとしては、データ全体を見渡すための基盤を与えることが重要と考える。この基盤を構築するために①遠隔地にある大規模データのための可視化技術、②データ空間を大局的に表示するボリュームレンダリング技術、そして③データの持つ特徴に重きをおく特異点グラフベースの可視化技術について研究開発を行う。本テーマにおける研究では、物理的にはなれた研究者同士が視覚、力覚、触覚等を共有できることが前提となるので多感覚コラボレーション技術についても研究項目④と位置付けた。①についてはプラットフォームに依存しない技術、②については主として CAVE 使用を前提とした技術、そして③については高度 TV 会議システムにおける使用を前提とした技術を開発する。

2. 大規模データに関するリモート可視化技術に関する研究

大規模データの可視化処理の特徴は、ボリュームデータが単一計算ノードの物理メモリに収まりきらないということである。このため可視化処理においては、ディスク入出力に伴うスワッピングが頻発することになる。ディスク入出力においてデータアクセスに要する計算クロック数は、物理メモリにおけるそれに比べてけた違いに大きい。計算時間の増大をもたらす。この問題を解決するために Unix システムコール `mmap()` を使い、ボリュームデータファイル全体を物理メモリにロードするのではなく、必要になった段階で初めて物理メモリにページ単位でロードする機能の開発を行う。また、ストリーミング技術を使い、物理メモリだけを使った可視化機能を開発する。

分散計算ノードに格納されているボリュームデータについて、いったん全計算ノードからボリュームデータを集めてから可視化処理を行うのではなく、各計算ノードで可視化処理を部分的に行う。その結果得られた部分可視化画像を適切な順序で重畳し、最終画像を生成するアプローチはひとつの有効な手法である。この手法は規則正しい順序でデータの定義されている場合について、すでにハードウェア化されている。本研究では、このハードウェアをベースに数値シミュレーションでよく用いられる不規則ボリュームデータについて手法の拡張を行う。

3. 没入型ボリュームレンダリング

ボリュームデータ全体をわかりやすく表示する手法としてボリュームレンダリングがある。このボリュームレンダリングの基本的な考え方は、ボリュームデータを半透明ゲルとみなし、各要素にその物理値に応じて色や不透明度を割当て、ボリューム内部の特徴をとらえることのできる雲状の画像を得ることにある。ボリュームレンダリングには一般的に計算時間が多く必要とされる。対象となるボリュームデータの大規模化、表示装置の高解像度化に伴い、高速なボリュームレンダリングアルゴリズムが求められている。

CAVE に代表される没入型可視化環境においてボリュームレンダリングはそれほど利用されていない。これには、ボリュームレンダリング画像の画質が低い、対話的な計算速度となっていない等の理由が挙げられている。これら問題を解決するために高品位なボリュームレンダリング画像を生成するスケーラブルなアルゴリズムを開発する。

ボリュームレンダリングをレイキャスティングで実装した場合、輝度値方程式の積分計算で使用するサンプリングポイントが視線上で等間隔に配置されていると仮定すると、これらサンプリングポイントは同心球を構成する。視点が変更されるとサンプリングポイント上でボリュームデータを再サンプリングし、各サンプリングポイントで輝度値方程式の積分計算に必要なシェーディングを行い、そして半径の順にテクスチャマップされた同心球を重ねる。^{1,2)}

没入型可視化環境におけるボリュームレンダリングでは、視点がボリュームデータの内部に位置することが多い。ボリュームデータ内での視点移動に対応するために現在点の近傍点に対応する同心球群をあらかじめ生成しておき、移動後の視点については、あらかじめ生成しておいた同心球群から補間計算を用いて、同心球群を生成する。

さらに、没入型可視化環境におけるボリュームレンダリングでは、視覚以外のどのような感覚が有効なのかについて検討を行う。このような環境では、雲に包まれる感じをもつことになるはずであるが、そのような環境において、空間から適切なフォースフィードバックがあった方がボリュームデータの理解に役立つのかを調べる。

4. 特異点グラフを使ったビジュアルデータマイニング

大規模ボリュームデータの場合、データ発掘に備えてボリュームデータの特徴を単純な図形で表現しておくことは有効である。本研究では、特徴を保持した単純図形として、特異点グラフを考える。ボリュームデータの特徴（特異点）を抽出し、その結果を用いて分類や探索を行うための手法の開発を行う。本手法では、ボリュームデータから作成された特異点グラフを用いる事によ

り、ボリュームデータの類似度を判定する。更に、大量のボリュームデータから好みのものを取り出すための検索インデクスとして、この特異点グラフを利用することも可能である。

4-1 特異点グラフ生成アルゴリズム

特異点は、速度ボリュームデータでは、ベクタ 0 の点、そしてスカラボリュームデータでは、勾配ベクタ 0 の点と定義される。これら特異点を速度場、またはスカラ勾配に沿ってつないでできた形状が特異点グラフである。本研究では、与えられたボリュームデータから特異点グラフを効率よく生成する手法の開発を行う。

ボリュームデータには、生成元であるシミュレーション自身が十分に収束していない等の理由によりノイズがふくまれていることが多い。このノイズは、特異点グラフに対して、好ましくない影響を与える。ノイズデータの勾配をとることにより、多くの特異点が発生し、結果として得られるスケルトン形状が複雑なものになってしまう可能性が高くなるからである。この問題を解決するためにボリュームデータの勾配場から得られる特異点に対して適切なラベル付けを行い、重要度の低い特異点を順に除去していくという戦略でノイズ除去アルゴリズムを開発する。

4-2 特異点グラフを使った可視化技術

ベクタボリュームデータにおいて得られる特異点グラフは、与えられたベクタ場の構造を単純明快に表現する。たとえば、翼列周りの空間における数値流体解析結果から得られた特異点グラフを用いることにより、翼列内の 3 次元剥離を伴う複雑な渦流れ構造を的確に解析できるようになった。本研究では、複雑な物理現象の解明における特異点グラフの有効性について調査する。

また、スカラボリュームデータにおいて定義される等値面は必ず特異点グラフに交差する。このことを利用すると、等値面のスカラ値が与えられたとき、ボリュームデータ全体で交差の有無を探索するのではなく、特異点グラフ上で探索し、等値面との交差点から自己増殖的に等値面を生成することができる。この手法では、等値面生成に必要なコストを大幅に削減できることが期待できる。³⁾

商用の可視化システムにおいて利用されている等値面生成ツールのユーザインタフェースは一端が最小値、他端が最大値であるような単純なスライダバーとなっている。このようなユーザインタフェースは、複人数で 3 次元形状を共有する仮想研究環境においては使いやすいたとはいえない。このような環境では、ボリュームデータに直接操作する形で可視化処理できる事が望ましい。直接オペレーションできるものとして、ボリュームデータの特徴を保持した特異点グラフを取り上げ、テレメーション環境における等値面生成のための新しいユーザ

インタフェースについて開発を行う。

ボリュームデータから生成された特異点グラフをテレメーション環境における参加者が共有し、PHANTOM といったフォースフィードバック装置を使って等値面の生成場所を指定できるようにする。例として、同心球状の 2 枚の等値面（どちらのスカラ値も同じ）があり、内側の等値面を可視化したい場合を考える。スライダバー形式のインタフェースでは、スカラ値を指定すると本来ならば出てきてほしくない外側の等値面まで表示されるが、スケルトン形式のインタフェースでは内側の等値面だけを選択的に表示することができる。

4-3 特異点グラフを使ったビジュアルデータマイニング

仮想研究環境において、過去の類似のボリュームデータを参照することが多くなっていくであろう。しかし、現状では、この類似度の判定は、利用者の視覚、経験によることが多い。定量的、客観的に類似ボリュームデータを検索できる実用技術はいまだ開発されていない。

映像コンテンツ等のマルチメディアコンテンツについてはアノテーション（補足情報）を用いた高度利用に関する研究が盛んであり、人間の手によるメタデータ（マルチメディアデータの内容を記述したデータ）を使って、好みのビデオなどを効率よく検索する手法などが提案されている。

過去から蓄積された膨大なボリュームデータから何らかの知見を抽出しようとする場合、それらは、検索しやすい形で整理されていなければならない。この検索のキーのひとつとして特異点グラフを考える。すなわち、特異点グラフ同士を比べることにより、ボリュームデータの類似度を判定する。数値シミュレーションから生成された特異点グラフには、シミュレーション計算を行う際に用いたさまざまな条件（境界条件、初期条件）も属性として付与する。本研究では、特異点グラフをキーとする検索エンジンを試作し、実際のボリュームデータに適用し、その有効性を検証する。

5. 多感覚コラボレーション技術に関する研究

コラボレーションとは、業務上、関連してくる様々な担当者が、互いに協調しながら作業を進めていくことをさす。3D コラボレーションツールといえば、ネットワークを介して共通の 3 次元データを見ながらコミュニケーションを行えるツールを指す場合が多い。多くのツールでは、効果的なコミュニケーションを実現するために、共有化された 3D 形状に印を付けたり、コメントを書き込んだりできる機能が備わっている。ネットワーク環境における 3 次元データの標準は、従来 VRML フォーマットが提案されていたが、最近では、VRML に対して格段の軽量化が実現されている XVL フォーマットが主流になりつつある。

XVLフォーマットは、数値シミュレーションの入出力データに特化していないので、今回のプロジェクトにおいて基盤として利用するには、様々な点で機能追加を行う必要がある。コラボレーションツールを使って、数値シミュレーションの遠隔実行を行おうとしていることを想定し、ある面に境界条件や荷重条件を設定したり、解析結果をカラーマッピングしたりする機能を開発する。また、スパコン側で実行途中のシミュレーション結果やシミュレーションのための制御データを効率よくネットワーク転送するための機能を開発する。

共有化された3D形状に条件設定等の操作を行う場合、GUIとしてはこれまでのものでは不十分で、3D空間を十分に意識したGUI空間GUIを構築するための技術についての研究開発を行う。また、遠隔地にいる共同研究者が力覚や触覚を共有するための技術を開発する。物理的に重要な部分を含むシミュレーション結果を提示する場合には、視点を同期させたりして重要な部分に共同研究者を連れて行くといった操作も重要となる。この場合、研究者同士でアイコンタクトを取れることが望ましい。各サイトで設置された位置センサにより研究者の視線を特定し、誰を注視しているのかをリアルタイムに判断し各サイトにフィードバックするシステムの開発を行う。

6. テレイマージョン環境の検証

これまでに述べた研究開発の結果構築されるテレイマージョン環境について、実応用分野（医療・核融合等）において適用・評価し、その有効性を検証する（図1参照）。検証のための1シナリオを以下に述べる。

1. 物理的に重要なシミュレーション計算が超並列スカラ計算機上で実行され、その経過映像が関連研究機関の表示装置にリアルタイムに流れている。
2. ある研究者がシミュレーション計算の異常に気づき、シミュレーションモデルの変更の必要性を認識する。
3. 研究プロジェクトに所属する研究者を緊急に招集し、コラボレーション作業を開始する。
4. コラボレーションに研究者は、そのときどこにいるか（没入型表示装置、PC、PDA、携帯電話等）に依存するものの結果の可視化映像を共有している。
5. コラボレーションに参加するある研究者が、可視化映像中のある空間部分に着目し、その部分について、視覚、力覚、触覚共有化機能を使って、他の研究者に知らせる。
6. 参加者は、モデルパラメータ値の変更について議論を行い、決定されたパラメータ値で計算を再実行し、良好な結果が得られたことを確認して、コラボレーション作業を終了する。

参考文献

- 1) 小山田耕二、宇野栄、土井章夫、宮沢達夫, “Fast Volume Rendering by Polygonal Approximation,” *Journal of Information Processing* 第15巻第4号, pp. 535-544, 1991
- 2) 小山田耕二、伊藤貴之, “Fast generation of spherical slicing surfaces for irregular volume rendering,” *Visual Computer* 第11巻第3号, pp. 167-175, 1995
- 3) 伊藤貴之、小山田耕二, “Automatic Isosurface Propagation by Using an Extrema Graph and Sorted Boundary Cell Lists,” *IEEE Transactions on Visualization and Computer Graphics* 第1巻第4号, pp. 319-327, 1995
- 4) 小山田耕二, “サイエンティフィックビジュアライゼーション技術の動向,” *情報処理学会誌*, 第42号第1巻, 2001

図1. テレイマージョン環境の検証