

## 多様体学習を用いた近似 contour tree プロット

小林 潤<sup>1</sup> 高橋 成雄<sup>2</sup> 藤代 一成<sup>3</sup> 岡田 真人<sup>2</sup>

<sup>1</sup> 東京大学 大学院情報理工学系研究科

<sup>2</sup> 東京大学 大学院新領域創成科学研究科

<sup>3</sup> 東北大学 流体科学研究所

contour tree は、一価関数により表現されるフィールド内の等値面の位相変化を効果的に表現できるため、ボリュームやその時系列データの特徴を抽出する手段としてよく用いられてきている。しかしながら、既存手法ではサンプルデータを補間するために、データの定義域を単体で分割する必要があり、結果として時系列ボリュームデータのような高次元データにおいては、補間処理の実装が複雑化してしまう問題点があった。本報告では、多様体学習を用いて、このような高次元データから、簡単かつ効果的な contour tree の近似表現を、補間処理を用いずに構築する手法を提案する。本手法の基本原理は、高次元データの局所的な接続性を頼りにデータの分布を追跡し、それを低次元空間に埋め込む次元圧縮をする点にある。本手法の有効性を示すために、構築した contour tree を介したボリュームデータやその時系列データの特徴解析の事例についても示していく。

### Plotting Approximate Contour Trees Using Manifold Learning

Jun Kobayashi<sup>1</sup>, Shigeo Takahashi<sup>2</sup>, Issei Fujishiro<sup>3</sup>, and Masato Okada<sup>2</sup>

<sup>1</sup>Graduate School of Information Science and Technology, The University of Tokyo

<sup>2</sup>Graduate School of Frontier Sciences, The University of Tokyo

<sup>3</sup>Institute of Fluid Science, Tohoku University

A contour tree is a powerful tool for delineating the topological evolution of isosurfaces of a single-valued function, and thus frequently used as a means of extracting features from volumes and their time-varying behaviors. However, existing algorithms for constructing such contour trees need simplicial decomposition of the data domain, which result in intricate software implementation to introduce appropriate interpolation over sampled values especially for higher dimensional cases such as time-varying volumes. This report presents a simple yet effective approach to plotting approximate contour trees of high-dimensional data in 3D space without any interpolation of its sampled values. Our main idea is to take advantage of a manifold learning technique so that we can elongate the distribution of high-dimensional data samples to embed it into a low-dimensional space while respecting its local connectivity. Examples are provided to demonstrate that our dimensionality reduction scheme is powerful enough to analyze the feature of volumes and their temporal behaviors through constructed trees.

## 1 はじめに

ボリュームやその時系列データから重要な特徴を抽出する作業は、入力データを可視化するための適切なパラメータの値を提供してくれるがゆえに、サイエンティフィックビジュアライゼーションにおける主要な問題のひとつとなっている。そのように抽出された特徴を表現する手法の中で、contour tree は、スカラ値に関する等値面の位相変化を追跡し、入力データの骨格を表現するために、いままで重要な役割を果たしてきた。そのような contour tree を構築するための秀逸なアルゴリズムも既にいくつか提案されており、あわせて物理学や生物学等を含めた具体的な研究テーマへの応用事例も示されてきた。

これらのアルゴリズムは、原理的には、一価関数  $s = s(p) = s(x_1, \dots, x_n)$  で表されるデータの

サンプル点であれば、任意次元のデータを処理することができる。(ただし、ここでは  $s$  がスカラ値を表し、 $p = (x_1, \dots, x_n)$  は定義域の座標を表すとする。) この表現は、いくつか典型的なデータ形式を表現することができ、 $n = 2$  のときは地形データ、 $n = 3$  のときはボリュームデータ、さらに  $n = 4$  のときはその時系列データを表すことができる。しかしながら、従来手法は一般的にデータ定義域におけるスカラ値の補間処理が必要となり、それゆえサンプル点の単体分割を用いて単純な線形補間を導入することがしばしば行われる。これは、特に時系列ボリュームのような高次元データに対しては、3次元空間に埋め込むことのできないプリミティブを用いた定義域の分割が必要になり、従来手法の実装が難しくなることを意味している。

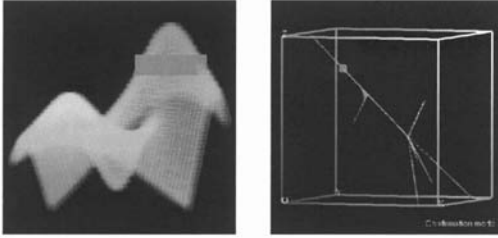


図 1: 提案手法による 2 次元スカラ (高さ) 場の解析: (a) 入力サンプル点分布 (縦が高さ方向), (b) 計算された近似 contour tree. (b) において,  $x, y, z$  座標に RGB を対応させることで, それぞれのサンプル点に異なる色を割り当てている.

本報告は, 多様体学習 (manifold learning) を用いて簡単かつ効果的に, 高次元データから contour tree 抽出する手法を示す. 多様体学習は, 非線形の次元圧縮法であり, 固有値問題のための数値計算ツールが導入できれば, 先に述べたような複雑な補間処理の実装を必要としない. 提案手法により抽出された contour tree は, サンプル点の 3 次元空間への投影であり, そのため近似表現ではあるが, 与えられた高次元データにおける等値面探索を効果的に行うことができる. 図 1 は, 我々が実装したシステムの出力のスナップショットを示しており, 2D スカラ (高さ) 場 (図 1 (a)) のように 2 つの頂上と 1 つの谷底をもつ) の位相骨格との対応関係を, 図 1 (b) のように求められた contour tree 上の点を選択することで, 追跡することができる.

## 2 関連研究

contour tree は, スカラ関数の位相骨格を表現する Reeb graph [1] の特殊な事例と考えることができる. Reeb graph は, 形状表現の道具として Shinagawa ら [2] により初めてコンピュータグラフィックスの分野に導入された. スカラ関数  $s = s(x)$  の Reeb graph は, 数学的にはあるスカラ値  $C$  の逆写像において, 同じ連結成分に属するサンプル点を同一視することにより得られる商空間として定義される. contour tree は,  $s(x)$  が一価関数であるとき, 閉路をもたない Reeb graph として定義される. contour tree は, Bajaj ら [3] によって初めて可視化分野に導入され, 等値面の位相変化を探索するのに利用された. アルゴリズムの観点からは, Carr ら [4] が提案した手法が, 任意次元のデータから contour tree を構築できる標準的な手法とされている.

先に述べたとおり, 既存のアルゴリズムでは, 通常三角形や四面体などのプリミティブによる単体分割を用いたサンプル点の線形補間が必要となる. しかしこの単体分割表現は, 特に高次元データにおいては, プリミティブ自身が 4 次元以上の空間をもち, 実装に必要なデータ構造や

アルゴリズムの実装を複雑にする. このため一般的には, 入力として与えられた時系列ボリュームデータから, 直接 contour tree 抽出を行うのではなく, 各時間ステップにおける contour tree の集合を取り出し, その時間変化を追跡する手法が用いられている. 例えば, Mascarenhas ら [5], Fujishiro ら [6], Sohn ら [7] は, 時系列ボリュームデータを解析するために, 各時間ステップにおけるボリュームデータの contour tree を抽出して利用する手法を示している.

この問題への我々のアプローチは, まず最初にサンプル点の分布から多様体構造を解析したのち, データ定義域において湾曲している多様体を引き伸ばし, 低次元空間への写像を求めることである. 本手法の特筆すべき点は, 前述の単体分割を必要とせず, 特に時系列ボリュームデータなどの高次元データの可視化においては, 簡単かつ効果的にデータを解析できる点である.

## 3 近似 contour tree プロット

本手法は, 多様体学習の一手法である Isomap [8] の簡単な改良と位置付けることができる. そのため, Isomap と同様, 近似 contour tree プロットの処理は以下のような 3 つの手順から成る.

1. 入力サンプル点からの多様体の局所的接続性の構築 (図 2 (a))
2. 各サンプル点对間の, 多様体上の最短経路に沿った距離の計算 (図 2 (b))
3. 多次元尺度構成法 (MDS) を用いた, 各サンプル点の 3 次元空間への埋め込み (図 2 (c))

以降, 各手順について順を追って説明していく. ここで, 注意したいのは, 本手法と一般的な Isomap との違いが, サンプル点对間の距離の定義にあることである. これは, 3.2 項で詳述する.

### 3.1 サンプル点の多様体接続性構築

本手法の最初の処理は, 与えられたサンプル点分布から誘導される低次元多様体の局所的接続性を抽出することである. ここでは, 多様体は局所的に平坦な位相空間と考えられる. しかし, 与えられたサンプル点から誘導される多様体の全体形状は, データ定義域に複雑に湾曲して埋め込まれることが多く, 低次元空間へ写像するためには, まずその形状を局所的接続性を保持しながら引き伸ばす必要がある.  $m$  個の  $n$  次元サンプル点の集合を  $D = \{p_1, \dots, p_m\}, p_i \in R^n (i = 1, \dots, m)$  とする. 局所的接続性を求めるため, 各サンプル点を中心とする座標近傍を導入し, サンプル点上に定義されるグラフ構造を構築する. ここで, グラフの頂点がサンプル点を, 辺が抽出されたサン

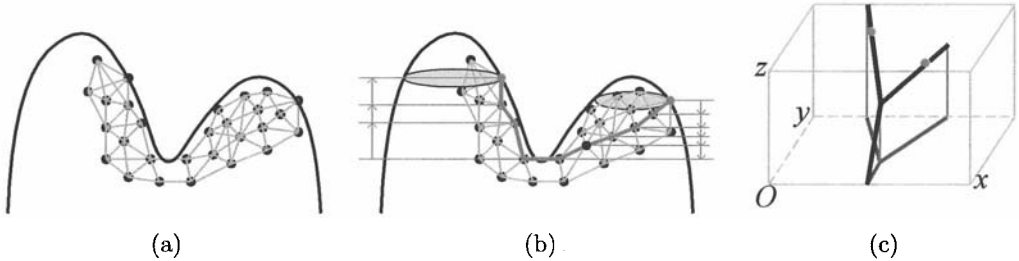


図 2: 提案手法の計算手順: (a) サンプル点分布から多様体接続性を構築, (b) 各サンプル点対間のスカラ値の差の絶対値を最短経路に沿って積算することで相対的距離を評価, (c) 接続性を最大限保持しながらサンプル点を 3 次元空間に埋め込む。

プル点間の接続性を表すものとする。この場合、構築するサンプル点間の接続性は、単体分割のような厳密なものに限定する必要はない。むしろ、図 2 (a) のように比較的自由的な接続性を利用することに注意したい。

任意のサンプル点  $p_i, p_j$  間における距離を  $|s(p_i) - s(p_j)| \times |p_i - p_j|$  と定義する。(ただし  $s$  はサンプル点に対応するスカラ値への一価関数を表すとする。) この上で、各サンプル点の隣接点を決める指標として、本手法では 2 つの基準を導入する。1 つは、幾何的基準で、各サンプル点間の距離が事前に定めたある閾値  $\delta$  より小さければ、接続関係があると判定する。もう 1 つは、位相的基準であり、各サンプル点において最も近い  $k$  個サンプル点を隣接点とする。ただし、 $k$  はあらかじめ与えられた定数とする。幾何的基準では、データ定義域のいずれの場所でも統一された基準で接続性を評価するので、最終的に得られる contour tree に、各サンプル点間の幾何的距離が如実に反映される。一方、位相的基準は、各局所領域におけるサンプル点の密集度に応じて、適応的に接続関係が定まるため、不均一に分布したサンプル点に対しても頑健に接続性が抽出できる。

接続性を表すグラフにおいて、辺の重み値の割り当ても重要な問題である。本手法では接続性の有無に応じて 1 か 0 を割り当てる離散的な重み値と、 $\exp(-|p_i - p_j|^2/t)$  を重みとする連続的重み値のどちらかを利用することとする。ただし、 $p_i$  と  $p_j$  は対応する辺の端点となるサンプル点を表し、 $t$  は定数である。これら 2 つの手法は、最終的に出力される contour tree の幾何学的な違いはもたらずものの、我々の予備実験からその位相構造は概ね同じになることがわかった。本手法では、特段の精度が必要な場合を除いて、離散的な重み値を利用している。

### 3.2 サンプル点間の相対距離の計算

次に、サンプル点上に定義される多様体接続性を利用して、各サンプル点対間の距離を評価する。ここでは、局所的接続性から解析された多様体を低次元空間に適切に引き伸ばすため、データ定義域における直線距離ではなく、誘導された多様体

上の最短経路に沿って定義される距離を計算する必要がある。このため、本手法では Isomap を利用し、最短経路上に存在する中間サンプル点を辿って局所的距離を足し合わせることで、最終的にサンプル点間の測地線距離を計算していく。

しかしながら、一般的な Isomap をそのまま適用すると、局所的距離として隣接サンプル点間のユークリッド距離を用いることになり、contour tree そのものを抽出することにはならない。本手法の特筆すべき点は、この距離定義を変更することで、次元圧縮処理の結果として contour tree を近似的に構築する点にある。本手法では、新しい距離として図 2 (b) に示すように、最短経路に沿った、各辺の両端点のスカラ値の差の絶対値を合計することで計算していく。例えば、サンプル点  $p_i, p_j$  間における最短経路が一連のサンプル点  $p_i, p_{\sigma(1)}, \dots, p_{\sigma(k-1)}, p_j$  で構成されるとする。ただし、 $\sigma$  はサンプル点のインデックスへの写像を表すものとする。結果として  $p_i$  と  $p_j$  間の距離は、 $|s(p_i) - s(p_{\sigma(1)})| + \dots + |s(p_{\sigma(k-1)}) - s(p_j)|$  となる。ただし、 $s$  は対応するサンプル点のスカラ値への一価関数を表すものとする。contour tree は、等値面の同じ連結成分におけるサンプル点を同一視して得られる商空間として定義される。そのため、サンプル点対間の相対的配置は対応する最短経路におけるスカラ値の差によってのみ決り、結果としてこの新しい距離の定義は十分正当化されることになる。各サンプル点対間の最短距離は、前の手順で得られた接続性を表すグラフへ、ダイクストラ法を適用することで容易に求めることができる。

### 3.3 サンプル点の 3 次元空間への埋め込み

先に定義した距離により、構築されるべき contour tree における、任意のサンプル点間の相対距離を評価することができた。そこで各サンプル点対の相対的位置関係を最大限保持しながら、サンプル点集合から誘導される多様体の接続性を、まず 2 次元平面へ埋め込むことを考える。これが可能となるのは、本来 contour tree は平面グラフであり、辺の交差なく 2 次元平面に埋め込むことができるためである。ここでは Isomap で用い

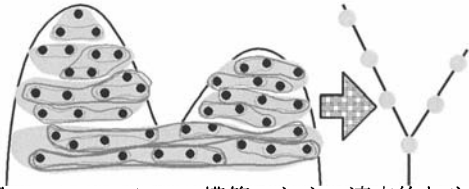


図 3: contour tree 構築のための適応的なサンプル点のクラスタリング

られる、**多次元尺度構成法 (MDS)** [9] を用いることで、各サンプル点対間の距離の歪みが最小2乗になるよう、高次元空間におけるサンプル点の分布を低次元空間に埋め込んでいく。

MDS を利用した次元圧縮を行うため、最初に距離行列を作る。ただし、第  $(i, j)$  要素は、先に求めた  $p_i, p_j$  間の距離の2乗を表すものとする。その後、2重中心化公式により、距離行列の正規化を行う。最終的なサンプル点の写像は、正規化された行列の固有値問題を解くことにより得られ、一番大きな固有値  $\lambda_1$ 、と二番目に大きな固有値  $\lambda_2$ 、それぞれ対応する固有ベクトル  $e_1$  と  $e_2$  とすると、サンプル点  $p_i$  は、 $(\sqrt{\lambda_1}e_{i1}, \sqrt{\lambda_2}e_{i2})$  にプロットすることで2次元平面上の点に変換する。ただし、 $e_{ij}$  は  $j$  番目に大きな固有ベクトル  $e_j$  の第  $i$  要素を表すものとする。これで contour tree の大局的な構造を2次元平面上に求めることができるが、スカラ値による等値面の位相変化を視覚的にとらえるため、最終的には図 2 (c) のように写像された空間の第3座標に、各サンプル点のスカラ値も付加することにする。

## 4 種々の拡張

本節では、膨大なサンプル点数を処理するための種々の拡張について示す。

### 4.1 サンプル点の階層的クラスタリング

前節で示したように、提案手法では、最終的に固有値問題を解くことで contour tree の写像を得ていた。このような固有値問題は、利用可能な数値計算ツールを導入すれば解くことができるが、それでも計算量の観点からサンプル点数がある程度までに制限する方が実用的である。本手法ではこの目的のため、サンプル点の逐次的な単純化によるクラスタリングを導入し、図 3 に示すような contour tree の点の分布に対して適応的なデータのサンプリングを実現していく。

ここではまず、接続性を表すグラフにおける辺の配列をその重み値に従って昇順にソートする。ここで各辺の重み値は、端点となるサンプル点におけるスカラ値の差の絶対値として計算される。次に、配列から重み値の最も低い辺を取り出し、その端点を新しいクラスタとして併合し、その端点集合の重心をそのクラスタの位置とする。この

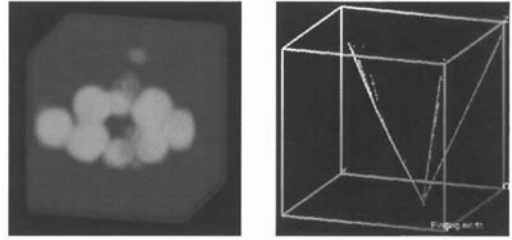
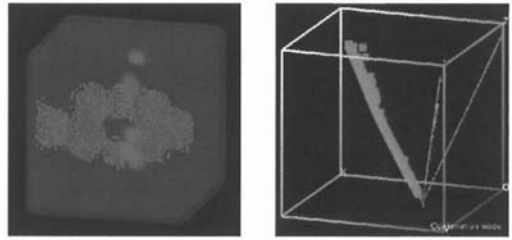
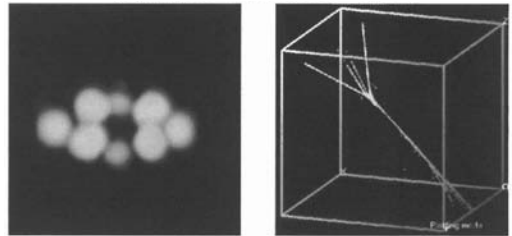


図 4: スプラッティング法による“neghip”データのダイレクトボリュームレンダリング。



(a) 抽出する部分領域の指定。



(b) 抽出された部分領域の拡大表示と解析結果。

図 5: データの部分特徴を解析するための、近似 contour tree を介した対話的セグメンテーション。

ような併合操作を繰り返し適用することで、結果として最小全域木に基づくサンプル点の階層的なクラスタを得る。このような階層表現は、膨大な個数のサンプル点を比較的少ないクラスタの代表点に置き換え、適応的に contour tree 上のサンプル点をプロットすることができるようになる。このような階層的サンプリングを用いた contour tree の抽出法は、contour tree の構造が本質的に1次元であり、その形状を適度な個数のサンプル点でプロットできる事実にも照らしても、合理的な手法である。

### 4.2 ダイレクトボリュームレンダリング

与えられたサンプル点集合は、本手法によって適切に3次元空間に contour tree 上の点として写像されるため、対応する3次元空間の位置に応じて、色や不透明度のような光学的特性を容易に割り当てることができるようになる。その光学的特性を利用してレンダリングを行うと、そのまま

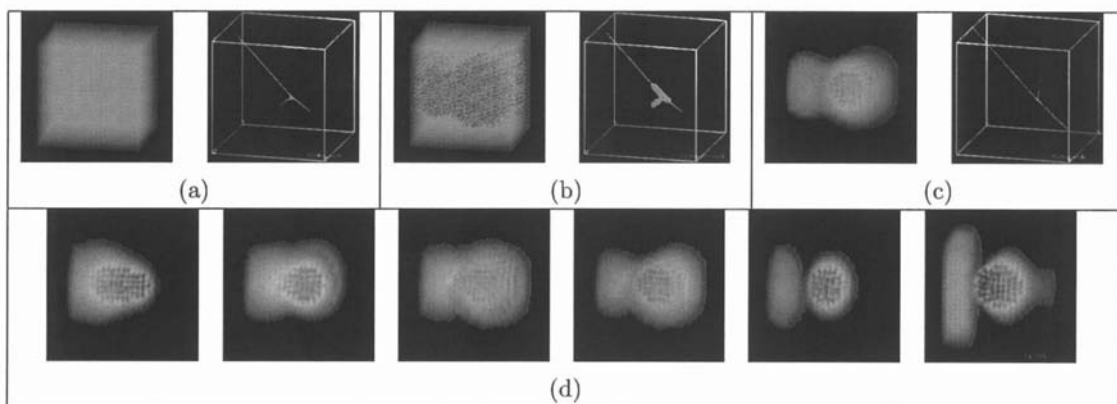


図 6: 陽子・水素原子衝突データ ( $21^4$ ): (a)contour tree による大域的な解析結果, (b)contour tree から部分領域の指定, (c) 部分領域の拡大表示と対応する contour tree による局所的な解析結果, (d) データの時間変化.

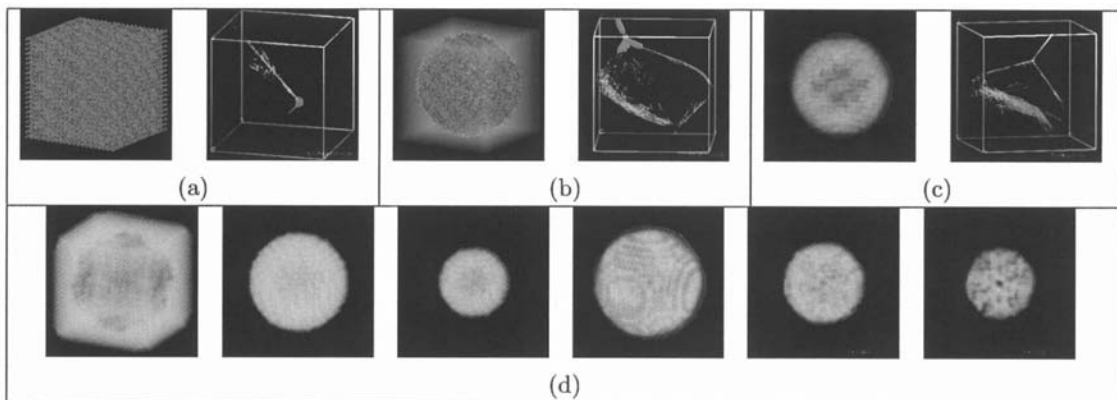


図 7: レーザー核融合による爆縮データ ( $21^3 \times 20$ ): (a)contour tree による大域的な解析結果と 1 回目の部分領域の指定, (b) 局所的な解析結果と contour tree から 2 回目の部分領域の指定, (c) 最終的な部分領域の拡大表示と対応する contour tree による解析結果, (d) データの時間変化.

与えられたデータの位相構造を強調することができ、そのデータ内容を分かりやすく解析することができる。この点において、我々の手法は、従来の代表的なレンダリング手法 [10] と比較しても、簡単かつ効果的にダイレクトボリュームレンダリングにおける伝達関数を設計することができる。我々の実装では、デフォルトで各サンプル点の  $x, y, z$  座標に応じてそれぞれ RGB の色を割り当てている。さらに、スカラ値が大きくなるに従い不透明度を増加させることで、与えられたボリュームデータの内部の方を強調して閲覧できるようにしている。図 4 は、“neghip” データの可視化結果を示している。このデータは、高電位タンパク質分子における電子の空間確率分布をシミュレーションしたものである。この図において、右の立方体状の枠における色の変化が、プロットされたサンプル点の位置と割り当てられている色の対応を示していることに注意されたい。

### 4.3 対話的セグメンテーション

元来多次元尺度構成法は、与えられたデータの大きな構造を特定するものであるため、contour tree の全体像を閲覧しているときに、その局所的な特徴を抽出することが難しい場合がある。この問題は、contour tree を部分的に拡大する機能を補い、局所的な分岐構造を検出できるようにすることで、解決することができる。そのため、我々のシステムでは、contour tree の部分領域を対話的に指定すると、対応するサンプル点の部分集合を抽出できるようにしている。図 5 は、我々のシステムにおいて、どのようにボリュームデータの部分特徴領域を抽出できるかを示したものである。まず、図 5 (a) のように contour tree の解析したい部分領域をマウスで指定した後、その部分領域に含まれるサンプル点を抽出する。さらに、図 5 (b) では、その抽出部分を拡大表示してさらに詳細な接続性を解析し、対応する contour tree

の局所特徴を検出する。この拡大表示機能は、近似 contour tree 表現において、潜在化する局所特徴を効果的に探索する機能を提供し、さらに大規模データに対する詳細度制御機能を実現することができる。

## 5 結果

本節では、提案手法の有効性を示すため2つの時系列ボリュームデータの可視化結果を示す。実験環境として、メモリ 3.2GB, Intel Core2Duo E6750(2.66GHz, 2MB cache)を搭載したデスクトップPC上で我々のプロトタイプシステムを実装した。また、固有値問題を解く数値計算ツールとしては SLEPc [11] を利用している。

図6は、陽子-水素原子衝突をシミュレーションしたデータであり、データのサイズは  $21^4$  である。図6(a)は、近似 contour tree により、その時系列ボリュームを大域的に解析した結果である。この contour tree を介して、図6(b)のように分岐部分を指定することで、陽子と水素原子の相互作用を表現する部分領域を特定し、さらに図6(c)のようにその部分領域を拡大表示し、局所の特徴を解析している。最終的には、この部分領域の時間変化を追跡することにより、時系列ボリュームの重要な特徴を得ることができる。図6(d)はその時間変化を示したもので、陽子の正電荷が水素原子周辺における電子の動きに著しい影響を与えていることを示している。

図7はレーザ核融合による爆縮シミュレーションの時系列ボリュームの例である。データのサイズは  $21^3 \times 20$  である。この例では、停留における内側の燃料層と外側のプッシャー層の接触面の時間変化が示されている。図7(a)のような大域的なデータの解析において、2つの層の接触面周辺部分を contour tree の分岐部分として同定することができる。さらにその特徴部分を図7(b)のように抽出することで、さらに詳細な解析を行うことができる。この2回の拡大表示による解析の結果、図7(c)が示すように、接触面の時間に関する特徴的な振る舞いを検出できる。図7(d)は、接触面の時間変化を表し、ここでは3枚目と4枚目の画像の間に爆縮の瞬間を特定することができる。

## 6 まとめ

本手法による多様体学習を用いた contour tree 構築は、他の従来手法と比べて実装が簡単であり、ボリュームやその時系列データといった複雑な構造を解析するのに有効である。本枠組みにより得られる contour tree は近似表現ではあるが、ユーザーは局所的な部分領域を拡大して解析することができる。

**謝辞** 本研究の一部は、文部科学省科学研究費補助金若手研究(B)(課題番号17700092)、基盤研究(B)(課題番号18300026)、基盤研究(B)(課題番号20300033)、およびカシオ科学振興財団の助成による。

## 参考文献

- [1] Reeb, G.: Sur les points singuliers d'une forme de Pfaff complètement integrable ou d'une fonction numérique [On the Singular Points of a Completely Integrable Pfaff Form or of a Numerical Function], *Comptes Rendus Acad. Sciences Paris*, Vol. 222, pp. 847-849 (1946).
- [2] Shinagawa, Y., Kergosien, Y. L., and Kunii, T. L.: Surface Coding Based on Morse Theory, *IEEE Computer Graphics & Applications*, Vol. 11, No. 5, pp. 66-78 (1991).
- [3] Bajaj, C. L., Pascucci, V., and Schikore, D. R.: The Contour Spectrum, *Proc. of IEEE Visualization '97*, pp. 167-173 (1997).
- [4] Carr, H., Snoeyink, J., and Axen, U.: Computing Contour Trees in All Dimensions, *Computational Geometry: Theory and Applications*, Vol. 24, No. 2, pp. 75-94 (2003).
- [5] Mascarenhas, A. and Snoeyink, J.: Implementing time-varying contour trees, *SCG '05: Proceedings of the twenty-first annual symposium on Computational geometry*, New York, NY, USA, ACM, pp. 370-371 (2005).
- [6] Fujishiro, I., Otsuka, R., Takahashi, S., and Takeshima, Y.: T-Map: A Topological Approach to Visual Exploration of Time-Varying Volume Data, *Proc. of 6th International Symposium on High Performance Computing*, Lecture Notes in Computer Science, Vol. 4759, pp. 176-190 (2008).
- [7] Sohn, B.-S. and Bajaj, C.: Time-Varying Contour Topology, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 12, No. 1, pp. 14-25 (2006).
- [8] Tenenbaum, J. B., de Silva, V., and Langford, J. C.: A global geometric framework for nonlinear dimensionality reduction., *Science*, Vol. 290, No. 5500, pp. 2319-2323 (2000).
- [9] Cox, T. and Cox, M.: *Multidimensional Scaling*, Chapman & Hall, London (1994).
- [10] Takeshima, Y., Takahashi, S., Fujishiro, I., and Nielson, G. M.: Introducing Topological Attributes for Objective-Based Visualization of Simulated Datasets, *Proc. of the 4th International Workshop on Volume Graphics (VG2005)*, pp. 137-145, 236 (2005).
- [11] Hernandez, V., Roman, J. E., Tomas, A., and Vidal, V.: SLEPc Users Manual, Technical Report DSIC-II/24/02 - Revision 2.3.1, Dept. of Information Systems and Computation, Technical University of Valencia, Spain (2004).