

# 網羅的なシミュレーションにおける空間的パターン発見のためのデータマイニング手法の開発

A Data Mining Method for Finding a Spatial Correlation between Two Volume Datasets

石田 愛<sup>♥</sup> 渡辺 知恵美<sup>♦</sup> 城 和貴<sup>▲</sup>

Ai ISHIDA Chiemi WATANABE Kazuki JOE

近年、コンピュータ技術の進歩により様々な要素が複雑に組み合わさっている物理シミュレーションは常識となってきている。技術者らは、膨大なデータの中から複数の要素の分布における空間関係や、複数のパラメータ間における物理現象の相関関係発見の手がかりを発見した後、詳細に分析するが、その手がかりの発見は技術者の経験的知識に基づいた直感によることが多い。そこで、我々は、物理シミュレーション結果の相関関係に関する手がかりの発見を支援するために、考えられる全ての条件を網羅し試行した結果の鳥瞰図を提示するマイニング手法を提案する。これにより、直感的にどの領域でどのような特徴が出ているのか大まかに知ることができ、調査する領域を絞り込むことが可能になる。

In recent years, with the advances in computer technology, computer simulation program can calculate the various factors which are related of the phenomena, and it is important to analyze the correlation of these factors. For efficient analysis of correlations between some simulation data, engineers often look for some clues of correlation feature to analyze in detail. But the way of finding clues depends on an engineer's heuristics and technical knowledge. In this paper we propose a visual mining technique that supports to find some clues of characteristic correlations of two simulation data. We suppose that the simulation data is formatted as 3d-gridded volume data. By using this technique, engineers can find roughly where a correlation feature appears and reduce the time one takes to analysis.

<sup>♥</sup> 正会員 日本アイ・ビー・エム (株) 東京基礎研究所  
[aiishida@jp.ibm.com](mailto:aiishida@jp.ibm.com)

<sup>♦</sup> 正会員 お茶の水女子大学理学部情報科学科  
[chiemi@is.ocha.ac.jp](mailto:chiemi@is.ocha.ac.jp)

<sup>▲</sup> 奈良女子大学人間文化研究科  
[joe@ics.nara-wu.ac.jp](mailto:joe@ics.nara-wu.ac.jp)

## 1. はじめに

近年のコンピュータ技術の進歩により、工業製品の製造過程において物理シミュレーションを取り入れた設計プロセスが常識となっている。一般に物理シミュレーションは温度や圧力、速度場など様々な要素が複雑に絡み合っていることから単一の要素の分布や流れを調べるだけでなく、複数の要素の分布や動きの相関関係を検証する必要がある。

例えば自動車のエンジン内の燃焼において、温度と窒素濃度に何かしらの相関があるかもしれないと考えたとき、技術者は経験的知識に基づいた直感により試行錯誤的に様々な条件で温度と窒素濃度の領域を比較し、そこから何かしらのヒントを得ようとするだろう。その場合、可視化システムなどを用いて対話的に比較検討をする方法もある。しかしながら全てのパラメータによる計算結果を並べてその全容を鳥瞰し、特徴発見する手がかりとする手法も有用であると考えられる。そこで我々は、比較的安価に計算が実行できる物理シミュレーションにおいて、複数のデータの相関関係や、複数のパラメータ間における特徴を発見するための手がかりを得るために、考えられる全ての条件を網羅し試行した結果の鳥瞰図を提示するビジュアルマイニング手法を提案し実装する。

## 2. 提案手法の概要

図 1 では自動車のエンジン内の燃焼における、ある温度領域と窒素濃度領域との相関関係を検証する例に対して本提案を適用した際の処理のフローイメージを示している。

### 2.1 空間的相関関係を表現する特徴ベクトルの生成

本論文で提案するマイニング手法ではまず、相関関係を調べたい2要素を軸とした2次元格子空間を作成する。そしてその各格子(セル)の中に特徴ベクトルを当てはめて行く(図1(a)~(c))。特徴ベクトルとは2種類のパラメータ値によって得られるデータセットの相関関係である。前節でも述べた自動車のエンジン内の燃焼における例では温度と窒素濃度の空間関係を相関関係とすることができる。空間関係として考えられる基準としては、まず Open Geospatial Consortium(OGC)モデルにて定められている a) 距離関係 b) 方向関係 c) 位相関係があげられる[1]。特に位相関係は9-intersectionモデルにより、二つの領域が同じ、一方が他方に含まれている、離れているなどの二つの空間の位置関係や方向関係等を示す8種類の位相関係に分類されており、これらを0~7の値に割り当てることによって1次元の特徴値として表現できる[2]。そしてこれらの空間的関係を複合的に利用し多次元の特徴ベクトルで表現する。

### 2.2 カラーマッピングによるビジュアルマイニング

相関関係を示した特徴ベクトルの格子データは、そのままではどの領域にどのような特徴がでているのか直感的に理解しがたい。そこで特徴ベクトルを図1の(d)で示したように色で表現する。特徴ベクトルを色に割り当てるには、特徴ベクトルを3次元もしくは4次元に縮約し、適切な色空間にマッピングする。色空間とは色を3または4次元の数値ベクトルで表す方法であり、赤(Red)、緑(Green)、青(Blue)のRGB手法や、色相(Hue)、彩度(Saturation)、明度(Brightness)のHSV手法等がある。自動車のエンジン内の燃焼シミュレーションの例では、温度が低く窒素濃度が高い、つまり完全に燃焼が行われていない個所が特定の色で表示され、エンジン内のどこに発生しているかを知ることができると考えられる。このように2次元の格子状にデータを並べてその傾向を見る方法

はHeat Map Viewと呼ばれており [3], 人間の目で直感的にどの領域でどのような特徴が出ているのかを大まかに知ることができ, 調査する領域を絞り込むことが可能になる.

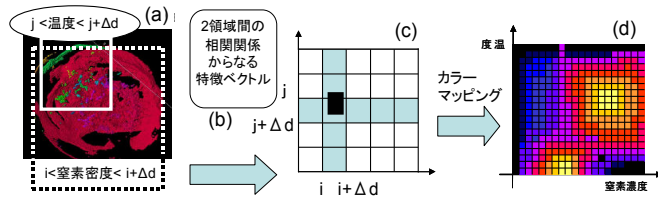


図 1 : 提案するマイニング手法のイメージ  
Fig.1 Image of this Mining Method

### 3. 空間関係による特徴ベクトルの定義

特徴ベクトルとする空間の相関関係には大分して 1)対象となるデータセットやシミュレーションの目的に特化したもの, 2)様々なデータセットに対して汎用的に適用できるもの, の 2 種類がある. 本節ではできる限り汎用的に使用できるよう 2)の多くのシミュレーションデータに対して適用できる特徴ベクトルとして, 位相, 距離, 方向関係からなる特徴ベクトルの求め方を定義する.

以降, 空間関係と特徴ベクトルの求め方を, 前述している自動車のエンジン内の燃焼の例を用いて説明する. 温度に関する条件 A によって得られる空間領域を  $S(A)$  と窒素濃度に関する条件 B によって得られる空間領域を  $S(B)$  とし, これらの空間関係  $R(S(A), S(B))$  を特徴ベクトルとして求めることを考える. データセットに対してある条件の空間領域を抽出した場合, 1つのまとまった領域になるのではなく, 複数のまとまった領域に分散することが多い. そこでまず我々は空間領域  $S(A)$  が複数の領域  $\{O_{a_0}, \dots, O_{a_n}\}$  で構成されているとする. 今回の例では,  $S(A)$  が 2つの領域,  $S(B)$  が 3つの領域を図 2 のような配置で有しているとする.

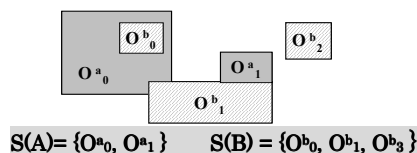


図 2 :  $S(A)$  と  $S(B)$  のオブジェクトの位置

Fig.2 Position of Objects between  $S(A)$  and  $S(B)$

複数の領域を持つ  $S(A_i)$  と  $S(B_j)$  との空間関係に関しては, 我々は文献 [4] において提案されているシーン間の空間関係モデルに基づき, 空間関係  $R(S(A), S(B))$  を, 位相関係  $R_{Top}(S(A), S(B))$ , 距離関係  $R_{Dis}(S(A), S(B))$ , 方向関係  $R_{Dir}(S(A), S(B))$  の三項で表すこととした.

$$R(S(A), S(B)) = (R_{TOP}(S(A), S(B)), R_{DIR}(S(A), S(B)), R_{DIS}(S(A), S(B)))$$

以下, 各々の空間関係について述べる. 今回の例における  $S(A)$ ,  $S(B)$  はそれぞれ複数の領域を有しているため, 各空間関係は  $S(A)$ ,  $S(B)$  の領域の全ての組み合わせについて各空間関係を見ていくことになる. 以下 3.1~3.3 節における定義によって各空間関係は二項関係で表現可能となる.

#### 3.1 位相関係

$S(A)$  の領域  $O_a$  と  $S(B)$  の領域  $O_b$  間の位相関係  $R_{Top}(O_a, O_b)$  は OGC の 9-intersection モデルにより,

1. 離れている (Disjoint( $O_a, O_b$ ))
2. 外接している (Meet( $O_a, O_b$ ))
3. 一部分が重なっている (Overlap( $O_a, O_b$ ))
4.  $O_a$  が  $O_b$  に内接している (Covers( $O_a, O_b$ ))

5.  $O_b$  が  $O_a$  に内接している (CoveredBy( $O_a, O_b$ ))
6.  $O_a$  が  $O_b$  に含まれている (Contain( $O_a, O_b$ ))
7.  $O_b$  が  $O_a$  に含まれている (Inside( $O_a, O_b$ ))
8. 等しい (Equal( $O_a, O_b$ ))

の 8 種類に分類される. これらを利用することによって図 2 の位相関係は図 3 のようになる.

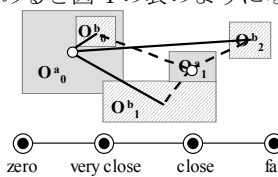
|         |          |         |          |
|---------|----------|---------|----------|
|         | $O^b_0$  | $O^b_1$ | $O^b_2$  |
| $O^a_0$ | Inside   | Overlap | Disjoint |
| $O^a_1$ | Disjoint | Meet    | Disjoint |

図 3 : 位相関係  $R_{Top}(S(A), S(B))$

Fig.3 Phase Relationship between  $S(A)$  and  $S(B)$

#### 3.2 距離関係

2領域間の距離関係  $R_{Dis}(O_a, O_b)$  は  $O_a$  と  $O_b$  がとりうる最大の中距離の値を近い方から very close, close, far に 3 分割し, 距離が 0 の場合は zero とする. 温度と窒素濃度の例では, 距離関係を求めると図 4 の表のようになる.



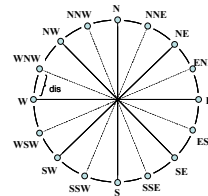
|         |            |            |         |
|---------|------------|------------|---------|
|         | $O^b_0$    | $O^b_1$    | $O^b_2$ |
| $O^a_0$ | Very close | close      | far     |
| $O^a_1$ | close      | Very close | close   |

図 4 : 距離関係  $R_{Dis}(S(A), S(B))$

Fig.4 Distance Relationship between  $S(A)$  and  $S(B)$

#### 3.3 方向関係

$O_a$  を基準とした  $O_b$  との方向関係を二項関係で表現する. 2次元の場合は図 5 の左図に表す 18 の方向関係に分類する図 2 の方向関係  $R_{Dir}(S(A), S(B))$  を図 5 に示す.



|         |         |         |
|---------|---------|---------|
|         | $O^a_0$ | $O^a_1$ |
| $O^b_0$ | NE      | WNW     |
| $O^b_1$ | ESE     | SW      |
| $O^b_2$ | E       | ENE     |

図 5 : 方向関係とその距離  $R_{Dir}(S(A), S(B))$

Fig.5 Direction Relationship between  $S(A)$  and  $S(B)$

### 4. 色空間マッピング

#### 4.1 空間類似度を用いた色空間マッピング

3節にて定義した位相, 距離, 方向関係からなる空間関係  $R(S(A), S(B))$  を特徴ベクトルとし, これを色空間にマッピングすることにより, 図 1(d) に示されるような 2次元図を生成する. 本節では, 特徴ベクトルを色空間にマッピングする方法について述べる.

前節の説明で分かる通り  $R(S(A), S(B))$  は二項関係の行列によって構成されている. 一方色空間は 3次元もしくは 4次元の値によって表される. そのため,  $R(S(A), S(B))$  をそのまま色空間にマッピングするのは困難である. そこで我々はある 1組の条件 a, b による空間領域  $A_0, B_0$  の特徴ベクトル  $R(S(A_0), S(B_0))$  を基準点とし, 基準点との類似度を求めて, 色空間に適用することにする. 図 6 は  $R(S(A_i), S(B_j))$  を色空間にマッピングするための手順を示している 3種類の空間関係の類似度を求め 0~1 までの値になるように正規化した後,

色空間の 1 点に割り当てる. また  $D_{REG}$  とは空間類似度  $D$  を正規化したものである.

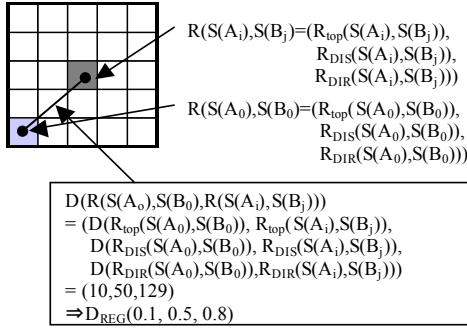


図 6 : 空間類似度を用いた色空間マッピング  
Fig.6 Color Mapping Using Similarity of Space

4.2 空間類似度

$R(S(A), S(B))$  と  $R(S(A'), S(B'))$  との空間類似度  $D(R(S(A), S(B)), R(S(A'), S(B')))$  は以下のように定義する.

$$D_{REL}(R_{REL}(S(A), S(B)), R_{REL}(S(A'), S(B'))) = \sum D_{REL}(R_{REL}(O_a, O_b), R_{REL}(O_a, O_b))$$

$$(O_a \in S(A) \cup S(A'), O_b \in S(B) \cup S(B'))$$

REL は空間関係の意味で TOP(位相関係), DIS(距離関係) もしくは DIR(方向関係)のいずれかとなる. つまり,  $S(A)$  または  $S(A')$  に含まれる任意の領域  $O_a$  と  $S(B)$  または  $S(B')$  に含まれる任意の領域  $O_b$  との空間関係  $R(O_a, O_b) (\in R(S(A), S(B)))$  と  $R'(O_a, O_b) (\in R(S(A'), S(B')))$  との類似度の総和を空間類似度としている.

ここで, 前述している空間関係  $R(S(A), S(B))$  と新たな空間関係  $R(S(A'), S(B'))$  (図 7) との類似度を求める. まず, 同じ領域間の関係において, 二項関係が等しいものは類似度が 0 となるため削除し, 異なる二項関係だけを残す(図 8).

| 位相関係 / 距離関係 / 方向関係 |                          |                          |                          |
|--------------------|--------------------------|--------------------------|--------------------------|
|                    | $O_b_0$                  | $O_b_1$                  | $O_b_2$                  |
| $O_a_0$            | Overlap<br>Close<br>SEE  | Disjoint<br>Close<br>ESE | Disjoint<br>Far<br>E     |
| $O_a_1$            | Disjoint<br>Close<br>WSW | Meet<br>VeryClose<br>S   | Disjoint<br>Close<br>ENE |

図 7 : 空間関係  $R(S(A'), S(B'))$   
Fig.7 Space Relationship  $R(S(A'), S(B'))$

| 異なる空間関係                     |                            |
|-----------------------------|----------------------------|
| Inside( $O_a_0, O_b_0$ )    | Overlap( $O_a_0, O_b_0$ )  |
| Overlap( $O_a_0, O_b_1$ )   | Disjoint( $O_a_0, O_b_1$ ) |
| VeryClose( $O_a_0, O_b_0$ ) | Close( $O_a_0, O_b_0$ )    |
| NE( $O_a_0, O_b_0$ )        | SSE( $O_a_0, O_b_0$ )      |
| WNW( $O_a_1, O_b_0$ )       | WSW( $O_a_1, O_b_0$ )      |
| SW( $O_a_1, O_b_1$ )        | S( $O_a_1, O_b_1$ )        |

図 8 : 空間類似度の評価  
Fig.8 Similarity of Space

次に二項関係同士の類似度を求める. これは文献[5]による空間関係の距離定義を基に算出する. 位相の距離関係は図 9, 距離の距離関係は図 4 の中図, 方向の距離関係は図 5 の左図にそれぞれ記述されており, ある二項関係から別の二項関係にたどるための経路の数を距離とする. 例えば位相 Covers と Inside の距離は equal を通る経路で, 2 となる. 同様に距離 VeryClose と far の距離は 2, 方向 N と WSW の距離は 5 となる.

よって図 8 の距離をそれぞれ測定していくと以下ようになる.

- $Dis(Inside, Overlap) = 2$
- $Dis(Overlap, Disjoint) = 2$
- $Dis(VeryClose, Close) = 1$
- $Dis(NE, SSE) = 6$
- $Dis(WNW, WSW) = 2$
- $Dis(SW, S) = 4$

これらの距離の総和をとることにより, 位相関係の類似度  $D_{TOP}(R(S(A), S(B)), R(S(A'), S(B')))$  = 4, 距離関係の類似度  $D_{DIS}(R(S(A), S(B)), R(S(A'), S(B')))$  = 1, 方向関係の類似度  $D_{DIR}(R(S(A), S(B)), R(S(A'), S(B')))$  = 12 を得ることができる. つまり全体の空間類似度は

$$D(R(S(A), S(B)), R(S(A'), S(B')) = (4, 1, 12)$$

となる.

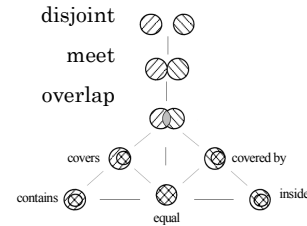


図 9 : 位相関係の距離  
Fig.9 Distance of Phase Relationship

5. 提案モデルの妥当性検証

本稿で提案したモデルは, 2 つのデータの相関関係における特徴をビジュアルマイニングによって発見するものである. そのためには特徴のないデータを本提案モデルに適用した場合は, 各特徴ベクトルおよびカラーマッピング後のビジュアルデータにおいて何も特徴が出ない, つまり偏りなく値が分布する事が求められる. そこで我々は実際のデータに適用する前に, 本提案モデルが実際に妥当であるかを調べるためのシミュレーション実験を行った.

5.1 手順

シミュレーションの大きな手順は以下の通りである.

1.  $1000 \times 1000 \times 1000$  の空間に, 大きさと中心座標の異なる立方体のオブジェクトを 50 個ランダムに発生させたものを 2 つ生成する.
2. ランダムに発生させたオブジェクト同士の空間関係の類似度を求める.
3. 1~2 を 100 回繰り返し類似度の値の分布を見る.

空間関係およびその類似度の求め方については 3, 4 節で説明した通りである.

5.2 結果と考察

空間関係(位相, 距離, 方向)の類似度を求める計算を 100 回試行した結果の分布を図 12~13 に示す.

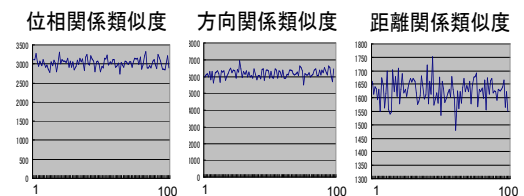


図 10 : 位相, 方向, 距離関係の類似度  
Fig.10 Similarity of Phase, Direction, and Distance Relationship

図 10 から明らかのように位相, 距離, 方向関係の類似度はいずれも特徴となるような値が出ていない. よって本提案モデルを特徴のないデータに適用した場合は, 確かに特徴が出ないことが分かった.

## 6. 実験

本節では特徴が判断し易いテストデータを使用し、3,4 節で述べた提案手法で特徴が正しく可視化されるかどうかを確認する。

### 6.1 テストデータ

特徴を判断し易いデータとして 2 つのテストデータ (data1,data2) を作成した。data1 は  $x = 50, y = 50, z = 50$  の格子状空間の  $(x,y,z) = (0,0,0)$  から  $(x,y,z) = (50,50,50)$  に引いた対角線にそって球をランダムに発生させたものである。ランダムに発生した球は 1 以上 50 未満の値を持っており、その値は  $(x,y,z) = (0,0,0)$  に近いほど大きく、 $(x,y,z) = (50,50,50)$  に近づくにつれて徐々に小さくなっていく。data2 も同様の格子状空間で  $(x,y,z) = (50,0,0)$  から  $(x,y,z) = (0,50,50)$  に向かって値が小さくなっていく物体をイメージして作成したデータである (図 11)。

### 6.2 実装

今回カラーマッピングを行うのは data1,data2 を軸とする  $50 \times 50$  の 2 次元格子空間である。それらの格子 1 つ 1 つに入る値 (類似度) を求めるためまずは空間関係を求める。空間関係を求めるために必要な各シーン内のオブジェクトは、あるシーンのデータをデータベースから取得した後クラスタリングを行う事によって定義される。クラスタリングをする事によってデータはまとまった領域ごとに分割され、その領域を全て含むような直方体 MBB (Maximum Bounding Box) をオブジェクトとする。data1, data2 の各シーンを  $S(A)$ ,  $S(B)$  とすると  $S(A_n)$  は  $n \leq \text{値} < n+1$  のデータの集合、 $S(B_n)$  は  $n \leq \text{値} < n+1$  のデータの集合になる。例えば  $S(A_{45})$  と  $S(B_{50})$  のオブジェクトは、data1 における  $45 \leq \text{値} < 46$  の等値面と data2 における  $48 \leq \text{値} < 49$  の等値面をそれぞれ覆う直方体 (MBB) になる。上記のようにシーン内のオブジェクト定義した後、3,4 節で述べたように空間関係、空間類似度を求めていく。そして位相関係の類似度を R、距離関係の類似度を G、方向関係の類似度を B とする RGB 値 ( $0 \leq R, G, B \leq 255$ ) をマイニング用の格子状空間にカラーマッピングしていく。

### 6.3 結果と考察

テストデータ data1, data2 に本提案手法を適用した結果、図 12 のようなビジュアルマイニング用の可視化データが得られた。

図 12 の結果では、data1,data2 の値が 20~30 の間で色が極端に明るくなっているところから、この値において何かしら空間関係に特徴があるのではないかと推測できる。そこで実際のデータの等値面画像を見てみると、この 20~30 の値において 2 つのデータが交差していることが分かった。さらに方向関係を表す青色が  $y = -x + 50$  の線を境に明るさが反転している事から、オブジェクトの位置関係が逆になっている事が推測できる。等値面画像を見ると実際にオブジェクトの方向関係が逆になっていることが分かる。

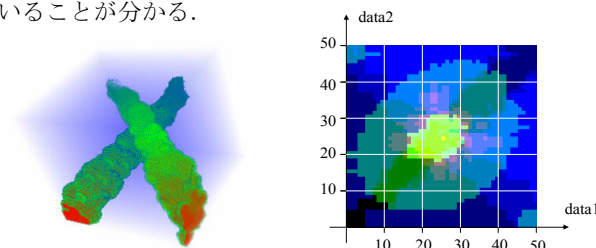


図 11 : 作成したテストデータ Fig.11 Test Data 図 12 : 適用結果 Fig.12 results of the experiments

以上の事からテストデータ data1,data2 の空間関係の主

った特徴がどの範囲の値に出ているかは、図 12 の可視化データからビジュアルマイニングによって大まかに推測することができることが分かった。

## 7. まとめと今後の課題

本稿では複数の要素の分布における空間的關係や、複数のパラメタ間による物理現象の相関關係を発見する手がかりを得るためのマイニング手法を提案した。本提案は空間的相関關係を表現する特徴ベクトルを生成し、それを格子状空間に並ベカラーマッピングする事によって、空間的相関關係のオーバービューを作成しビジュアルマイニングを行うものである。特徴ベクトルは対象データや目的によって大きく異なってくるため、我々は今回比較的汎用性のある空間關係を特徴ベクトルとするものを提案した。これにより、人間の目で直感的にどの領域でどのような特徴が出ているのかを大まかに知ることができ、調査する領域を絞り込むことが可能になると考えられる。

今後の課題としては、さらに複雑な特徴をもったテストデータを作成し、本提案手法を適用する事によって正しく特徴がでるかどうかなを確認する。その後実際のシミュレーションデータに本手法を適用して、さらに汎用的に使用できるよう本手法を改良していく予定である。

### 【文献】

- [1] OCG モデル : <http://www.opengeospatial.org/>
- [2] M. Egenhofer and R. Golledge "Spatial and Temporal Reasoning in Geographic Information Systems", Oxford University Press, New York, p.304, 1998.
- [3] Usama Fayyad, Georges G. Grinstein and Andreas Wierse (editors), "Information visualization in Data Mining and Knowledge discovery", Morgan Kaufmann Publishers, p.407, 2001.
- [4] Tom Bruns and Max Egenhofer, "Similarity of Spatial Scenes", Seventh International Symposium on Spatial Data Handling (SDH '96), Delft, The Netherlands, M.-J. Kraak and M. Molenaar (eds.), pp. 4A.31-42, August 1996.
- [5] M. Egenhofer and D. Mark, "Modeling Conceptual Neighborhoods of Topological Line-Region Relations", International Journal of Geographical Information Systems 9 (5): 555-565, 1995.
- [6] Max J.Egenhofer, "Spatial Relations: Models, Inferences, and their Future Applications", Proceedings of Advanced Database Symposium '96, p.9, 1996, December.

### 石田 愛 Ai ISHIDA

日本アイ・ビー・エム (株) 東京基礎研究所在中。2006 奈良女子大学人間文化研究科修士課程修了。ビジュアルデータマイニングシステムの研究・開発に従事。日本データベース学会正会員。

### 渡辺 知恵美 Chiemi WATANABE

お茶の水女子大学大学院人間文化研究科博士後期課程在学中。2000 お茶の水女子大学大学院人間文化研究科博士前期課程修了。仮想世界データベースシステムの研究・開発に従事。情報処理学会学生会員。日本データベース学会正会員。

### 城 和貴 Kazuki JOE

奈良女子大学人間文化研究科教授。1996 奈良先端科学技術大学院大学情報科学研究科博士課程修了、工学博士。自動並列コンパイラ、ニューラルネットワーク、画像処理等の研究・開発に従事。情報処理学会 (数理モデル化と問題解決研究会主査)。