

# 階層型多変数データ可視化手法「十二単ビュー」 における詳細度制御

山澤舞子<sup>†</sup>

伊藤貴之<sup>†</sup>

山下富義<sup>††</sup>

## Level of Detail Control of a Hierarchical Multi-Variate Data Visualization Technique "JunihitoeView"

MAIKO YAMAZAWA,<sup>†</sup> TAKAYUKI ITOH<sup>†</sup> and FUMIYOSHI YAMASHITA<sup>††</sup>

### 1. はじめに

情報可視化が対象とするデータ構造の分類として Shneiderman は、1次元、2次元、3次元、 $n$ 次元 ( $n > 3$ )、時系列、階層型、グラフ、の7種類を提唱している<sup>1)</sup>。これまで情報可視化技術の研究では、各々のデータ構造に対する可視化手法が多数発表されてきた。最近では階層型多変数データ(今後本論文では、 $n$ 次元データを「多変数データ」と称する)に代表されるように、上述の7種類のデータ構造を複合した情報を可視化する手法が、活発に提案されている。

階層型多変数データの可視化手法として我々は最近、「十二単ビュー」<sup>2)</sup>という手法を提案している。十二単ビューは、大規模階層型データ可視化手法「平安京ビュー」<sup>3)</sup>の拡張手法である。平安京ビューは、階層型データの葉ノードをアイコンで表示し、枝ノードを長方形の枠で表示することで、階層型データの全体を一画面に表現できる手法である。十二単ビューは、階層型データの最下位階層を構成する葉ノードに多変数が割り当てられていると仮定し、個々の葉ノードのアイコンに変数の個数だけの色相を割り当てることにより、階層構造と同時に多変数情報を表現する手法である。十二単ビューは、大規模な階層型多変数データ全体を一画面に全貌表示し、その階層構造と、変数値の

分布を同時に概観できる特徴がある。

本報告では、十二単ビューの機能拡張として、拡大表示時にはデータの葉ノードを単位として変数値を表示し、縮小表示時には下位階層の変数値を統合し、上位階層のみを表示する、というような詳細度制御を実現する手法を提案する。この機能拡張により、画面のサイズやユーザの操作に合った最適な情報量で、階層型多変数データを可視化できると考えられる。

### 2. 関連研究

#### 2.1 多変数データと階層型データの可視化手法

多変数データの可視化手法には、変数の数だけ鉛直線を引き、その各々の上にプロットした変数値を折れ線グラフ状に連結する Parallel Coordinates<sup>4)</sup> や、3次元直交座標系の中にもう1つの小さな3次元直交座標系を表示する、という操作を反復する Worlds within Worlds<sup>5)</sup> 等がある。

また、階層型データの可視化手法として代表的なものには、画面空間を再帰的に長方形領域に分割することで階層型データを表現する Treemaps<sup>7)</sup> や、階層型データを木構造により可視化する Cone Tree<sup>8)</sup>、Hyperbolic Tree<sup>10)</sup> 等があげられる。また、2次元空間で入れ子状に階層構造を構築する手法として、「平安京ビュー」<sup>3),6)</sup>が挙げられる。

#### 2.2 平安京ビュー

1章でも論じたとおり、本論文で用いる可視化手法

<sup>†</sup> お茶の水女子大学大学院人間文化創成科学研究科

Graduate School of Humanities and Sciences, Ochanomizu University

<sup>††</sup> 京都大学大学院薬学研究科

Graduate School of Pharmaceutical Sciences, Kyoto University

本手法は色表現に基づく可視化手法であるため、白黒印刷された予稿集でその効果を示すのは困難である。実行結果等については以下のサイトにカラー形式で公開する予定である。  
<http://itolab.is.ocha.ac.jp/~maiko/>

である「十二単ビュー」は、「平安京ビュー」という階層型データ可視化手法の拡張手法である。そこで本節では、平安京ビューの概要を紹介する。

平安京ビューは、階層型データを構成する葉ノードをアイコンで表現し、枝ノードを入れ子状の長方形の枠で表現し、階層型データ全体を一画面に配置する手法である。配置結果の例を図1に示す。

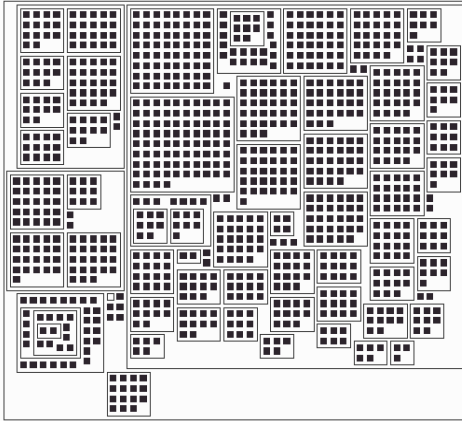


図1 平安京ビューによる階層型データの可視化例。

平安京ビューは、以下の条件を満たすような画面配置アルゴリズムの採用により、図1に示すような可視化結果を実現している。

- 葉ノードや枝ノードを干渉させないように配置する。これにより、画面上の任意のデータ要素を、クリック操作可能な状態で表示できる。
- 画面空間上の占有面積を小さくするように配置する。これにより、限られた画面空間に多くの情報を表現できる。

これ以外にもいくつかの条件が想定されているが、本報告では説明を割愛する。

平安京ビューでは、階層型データの葉ノードが有する変数値を、棒グラフ化されたアイコンの色、高さ、形状などで表現することができる。しかしこの表現方法では、5個、10個といった多変数を同時に表現することは困難である。また例えば、2個の変数値をアイコンの色と高さに割り当てたとき、その2個の変数を平等に認知できるとは限らない。あるときは色のほうが目につき、あるときは高さのほうが目につき、といった変数間の不平等性が生じる。これらの問題を解消し、葉ノードに属する多変数を平等に可視化する手法「十二単ビュー」を、次節にて紹介する。

### 2.3 十二単ビュー

本論文の提案手法では、階層型多変数データ可視化手法「十二単ビュー」の拡張手法である。十二単ビューは、階層構造の最下位階層を構成する葉ノードを、長方形のアイコンで表現する。ここで十二単ビューでは、変数の個数を  $n$  としたときに、長方形のアイコンを、式(1)で算出される縦  $l$  個、横  $m$  個の小領域に分割する。ここで  $[t]$  は、 $t$  を超えない最大の整数を示す。

$$\begin{aligned} l &= \lceil \sqrt{n} \rceil + 1 \\ m &= \lceil \frac{n}{l} \rceil + 1 \end{aligned} \quad (1)$$

この  $lm$  個 ( $lm \geq n$ ) の領域から  $n$  個の領域を用いて、十二単ビューでは以下のような画素値算出手法により、個々の葉ノードに割り当てられた  $n$  個の変数値を算出する。

- $i$  番目 ( $0 \leq i < n$ ) の領域の色相  $H$  ( $0 \leq H < 2\pi$ ) を、式(2)によって算出する。

$$H = \frac{2\pi i}{n} \quad (2)$$

- $i$  番目の領域の変数値  $t_i$  ( $0 \leq t_i < 1$  の値に正規化されている) から、 $i$  番目の領域の明度  $S$  ( $0 \leq S \leq 1$ )・彩度  $I$  ( $0 \leq I \leq 1$ ) を、式(3)によって算出する。

$$S = I = 0.2 + 0.8t_i \quad (3)$$

この結果、各々の変数には固有の色相が割り当てられ、その変数値が大きいほど画面上で明るく表示される。

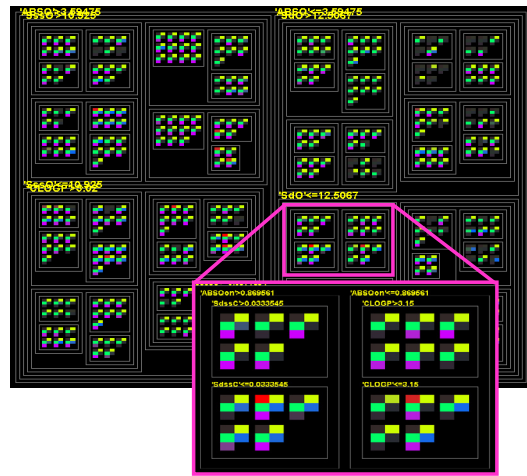


図2 (上) 十二単ビューによる階層型多変数データの可視化結果。(下) (上) の一部を拡大した図。

図2は階層型多変数データ全体を表示した結果、およびその一部を拡大した図である。この例より、個々のノードの変数値を正しく視認するためには、ズーム

イン操作が必要であることがわかる。逆に言えば、階層型データ全体を可視化するためにズームアウトしてしまうと、各々のノードは非常に小さく表示されるので、これらの変数値を視認するのは難しくなる。

### 3. 十二単ビューにおける詳細度制御

前節で論じた問題点の解決のために本報告では、階層型データ中にて同一階層下に属する葉ノードを統合し、この階層を表す代表ノードとして表示することで、画面上に表示されるアイコンの数や大きさの調節することを考える。

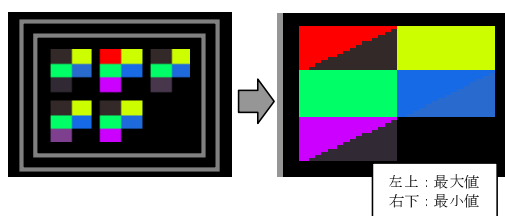


図3 同一階層下にあるノードの統合表示。

図3は、5変数を有する5個の葉ノードを図示した例である。図3(左)では5個の葉ノードをアイコンで表現しているが、提案手法における詳細度制御では図3(右)のように、これら5個の葉ノードを統合して1個のノードとみなし、これを1個のアイコンで表現する。ここで、同一階層に属するノード変数値の傾向を表す数値には、平均値・最大値・最小値・分散などがあげられる。現時点での我々の実装では、アイコンを構成する個々の小領域に対角線を引き、左上部分の三角形で最大値、右下部分の三角形で最小値、をそれぞれ表現している。

以上の表示方法に加えて提案手法では、ユーザによるズーム操作に伴って詳細度を自動制御する機能を実装する。この機能では図4に示すように、ズームアウト操作に伴い、その縮小率に合わせて、下位階層のノードを統合して上位階層全体をアイコン表現する、という操作を再帰的に反復する。またズームイン操作に伴い、自動的に上位階層のアイコン表示を省略して下位階層をアイコン表現する、という操作を再帰的に反復する。以上の機能により、ズームイン操作によって個々の葉ノードの変数値を読み取れるようになり、ズームアウト操作によってデータの全体像と数値的傾向を見渡せるようになる。

### 4. 適用事例

提案手法の適用事例として我々は、以下に示す薬物

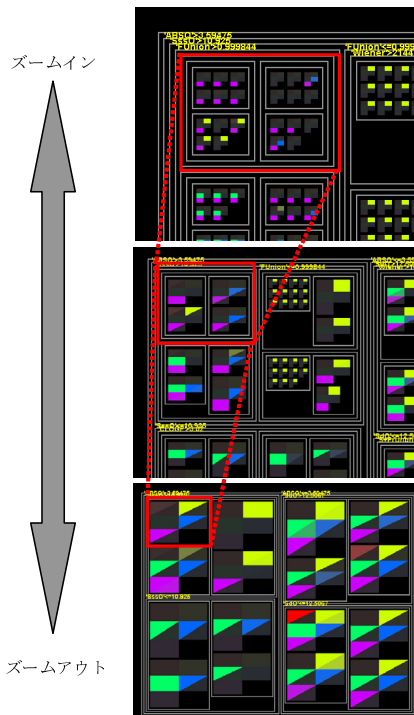


図4 ズーム操作に連動した詳細度制御。

データを用いて可視化を試みている。この成果は製薬企業に実際に導入され、薬物開発の現場にて活用されている。

まず薬物データを構成する多数の薬物を、分子構造上の特徴によって分類する。この分類を再帰的に反復することで、薬物データの階層構造を構築する。また各薬物は、数種類の実験値（本適用事例では数種類の酵素に対する代謝感受性）を有するとする。この数種類の実験値を多変数とみなし、上記の階層構造と組み合わせることで、階層型多変数データが得られる（図5参照）。

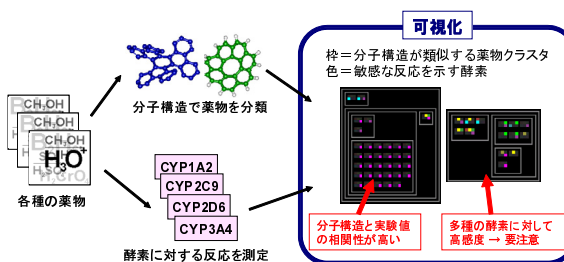


図5 薬物情報の可視化への適用。

階層構造を導く分子構造上の特徴と実験値との相関性が高ければ、同一階層には実験値の類似する薬物が割り当てられる。このような階層構造を構築できたと

きには、これから開発する未実験の新薬について、その分子構造に対応する階層を可視化することで、まだ得られていない薬物の実験値の推測に貢献でき、ひいては薬物開発の期間と費用の縮小に貢献できる。このことから、分子構造に基づく薬物の階層化は、薬物開発において非常に重要である。

分子構造と実験値の相関性が高い階層では、十二単ビューで可視化したときに、階層内の全てのアイコンがほぼ同一の色彩に見える。これをズームアウトすると、各変数の実験値の最大値と最小値の差が小さいために、図 6(右)のように、各小領域の左上部分と右下部分の三角形がほぼ同一の画素値になり、三角形というよりむしろ四角形に見えるはずである。

逆に分子構造と実験値の相関性が低い階層では、この階層に含まれる薬物の実験値の最大値と最小値の差が大きくなり、ズームアウトすると図 7(右)のように、三角形が目だって見えるはずである。このような場合には、どのような分子構造上の特徴を参照して階層構造を構築するか、を再考する必要があると考えられる。

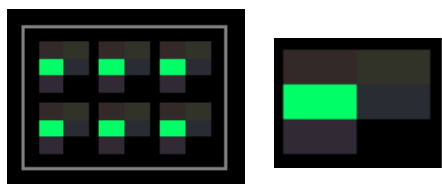


図 6 (左) 分子構造と実験値の関連性が高い階層の例。  
(右) ズームアウトした際に(左)の階層を表す代表ノード。

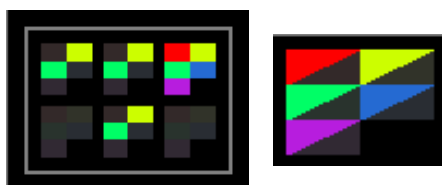


図 7 (左) 分子構造と実験値の関連性が低い階層の例。  
(右) ズームアウトした際に(左)の階層を表す代表ノード。

この適用事例における課題は、各薬物の変数値の傾向という低レベル情報の可視化だけでなく、その上位概念にあたるクラスターを単位とした実験値のばらつき(最大値と最小値の差)という高レベル情報も可視化したい、という点にある。この課題に対して我々は、まず上位階層を単位として実験値のばらつきを確認し、続いて実験値がばらついている階層をズームインすることで個々の薬物の実験値を確認する、というズーム操作に連動した詳細度制御が有効であると考えている。

一方で薬物情報以外の可視化に適用する際には、このままでは提案手法が有効に働かない場合があるかもしれない。あるいは、対角線を引いて最大値と最小値を表現するよりも良好な表現があるかもしれない。今後の課題として、薬物情報以外の適用事例を視野に入れて、提案手法をどのように汎用化するか、という点があげられる。

## 5. ユーザテストと考察

本章では、被験者 11 人により実施した提案手法のユーザテスト結果を示し、提案手法の有効性について考察する。

### 5.1 平安京ビューとの比較

5 変数を有する階層型多変数データを用意し、これを十二単ビューで可視化した図 8 と、変数ごとに 5 個の平安京ビューで可視化して並べた図 9 を用意した。被験者にはこれらを見比べながら、以下の質問に { 十二単ビュー・平安京ビュー・どちらも変わらない } のいずれかを選択させた。

- 問 1 変数値が大きいノードを見つけやすいのはどちらか。  
 問 2 変数値が小さいノードを見つけやすいのはどちらか。  
 問 3 1 つのノードの、すべての変数値を読み取りやすいのはどちらか。  
 問 4 1 つの階層に含まれるすべてのノードの、変数値の傾向(全体的に値が大きい、値にばらつきがある、等)を読み取りやすいのはどちらか。

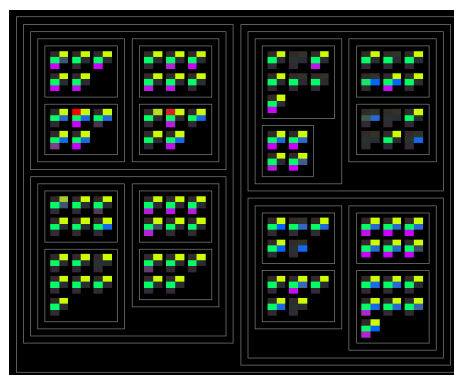


図 8 十二単ビューで可視化した図。

この結果を表 1 にまとめた。問 1 と問 2 の結果から、特定の変数について値の大きい、または小さいアイコンを見つけるためには、平安京ビューが有利だとわかる。本実験で用いた平安京ビューでは変数値の大

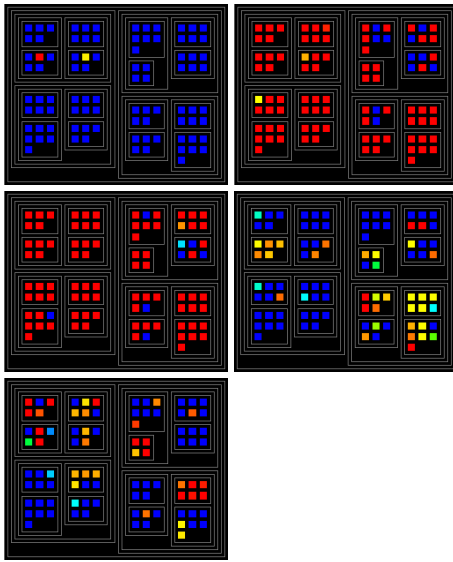


図 9 平安京ビューで可視化した図。

きさを色相で表現する。一方で十二単ビューでは、値の大きさは明度と彩度で表現する。実験結果から、値の大小を見比べるためには、明彩度の違いよりも色相の違いのほうが効果が高いと考えられる。

一方、問3と問4の結果から、1ノードのすべての変数値を読み取るには、あるいは1つの階層に含まれるすべてのノードの変数値の傾向を読み取るには、十二単ビューが有利であるとわかる。十二単ビューでは平安京ビューと違って、1つの可視化結果の1つのアイコンからすべての変数値を読み取ることができる。このため、すべての変数値が一画面上に表現されている十二単ビューのほうが、すべての変数値を読み取るのに適していると考えられる。

前章で論じた薬物情報の可視化の事例では、[問4]に示した「1つの階層に含まれるすべてのノードの変数値の傾向」を視認することが重要であるため、平安京ビューよりも十二単ビューの利用が適切であると考えられる。

表 1 平安京ビューとの比較実験結果。(単位：%)

	平安京ビュー	十二単ビュー	どちらも変わらない
問 1	63.6	9.1	27.3
問 2	63.6	9.1	27.3
問 3	0	90.9	9.1
問 4	18.2	72.7	9.1

## 5.2 より多くの変数への対応

続いて、変数の数が10, 20, 30, 40, 50個の階層

型多変数データを、十二単ビューで可視化した。被験者には各データの可視化結果を自由に操作してもらい、アイコンの小領域を視覚的に区別できるかを回答させた。このとき、詳細度制御を利用せず葉ノードを表示した場合と、詳細度制御を利用して下位階層を代表ノードで表示した場合のそれぞれについて、小領域を判別できる最大の変数の数を回答してもらった。この結果を表2に示す。

この結果は、詳細度制御を利用して代表ノードを表示することが、より変数の数が多い階層型多変数データにおいて有利に働いている、という傾向を示唆しているといえる。

表 2 小領域を判別できる変数の数。(単位：%)

変数の数	10	20	30	40	50
葉ノードを表示	27.3	54.5	0	9.1	9.1
代表ノードで表示	0	27.3	0	36.4	36.4

## 5.3 階層探索時間の計測

階層型多変数データを十二単ビューで可視化し、変数値に固有の特性をもつ最下位階層を文面で指定し、被験者にその最下位階層を探してもらい、その所要時間を測定した。本実験では「固有の特性をもつ階層」として、以下の2種類を指定した。

- 特定のいくつかの変数の値がすべての葉ノードにおいて高く、ほぼ同じ値をもち、その他の変数の値はすべての葉ノードにおいて低く、ほぼ同じ値をもつ。
- 特定のいくつかの変数の値は葉ノード間で大小ばらつきのある値をもち、その他の変数の値はすべての葉ノードにおいてほぼ同じ値をもつ。

本実験では2種類の階層型多変数データ(以下、データ1, データ2, と称する)を用いた。2種類のデータにおける葉ノード数は、9824および729であった。被験者による所要時間の平均時間を、表3に示す。

表 3 最下位階層を見つけるまでにかかった平均時間。(単位：秒)

		データ 1	データ 2
詳細度制御の利用	無	146	91
	有	286	61

この結果からわかるように、データによって詳細度制御の効果に大きな明暗が分かれてしまった。この理由について被験者や我々自身の感想を交えながら分析した結果として、下位階層を統合した代表ノードの大きさや形状が大きな要因であると考えた。図10は、本



実験におけるデータ1の一部を可視化した例である。この例のように、代表ノードの大きさや形状が不均一である場合に、被験者は固有の特性をもつ階層を見つけにくくなることが予想される。特に代表ノードが小さく（または細長く）表示された代表ノードを被験者が見落とす、という現象が容易に推察される。一方でデータ2は、葉ノード数が均一に割り振られた木構造を構成しているために、代表ノードの大きさや形状も均一になる。その結果として、提案手法の詳細度制御を用いて固有の特性をもつ階層を見つけることが容易だった、ということが推察される。

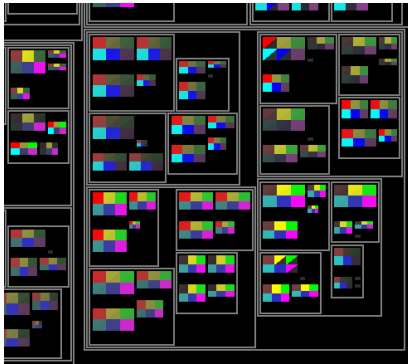


図 10 下位階層を統合した代表ノードが不均一な大きさ・形状で表示された例

提案手法の詳細度制御の目的に限らず、「平安京ビュー」や「十二単ビュー」において、下位階層に属する葉ノードの数にかかわらず、上位階層を均一な大きさの長方形で表現したい、という要求は今後も多くあると考えられる。例えば「平安京ビュー」を応用した大量画像の可視化手法<sup>11)</sup>においても、上位階層を表現する代表画像を均一な大きさで表現したい、という要求は考えられる。今後の課題として、上位階層を表現する代表ノードの大きさや形状の均一化、があげられる。

## 6. ま と め

本報告では、階層型多変数データ可視化手法「十二単ビュー」において詳細度制御を実現する手法を提案した。また、ユーザテストを行い、提案手法の有効性を証明した。

今後の課題として、ユーザテストの考察であげたとおり、上位階層を表現する代表ノードの大きさや形状の均一化があげられる。また、薬物データ以外のデータにも提案手法を適用し、その有効性を検証したい。

## 参 考 文 献

- 1) Shneiderman B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization, *Proceedings of IEEE Symposium on Visual Languages '96*, pp. 336-343, 1996.
- 2) 伊藤, 山下, 階層型多変数データ可視化手法「十二単ビュー」による薬物データベースの可視化, 情報処理学会データベースと Web 情報システムに関するシンポジウム (DBWeb2006), pp. 119-125, 2006.
- 3) 伊藤, 山口, 小山田, 長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善, 可視化情報学会論文集, Vol. 26, No. 6, pp. 51-61, 2006.
- 4) Inselberg A., Dimsdale B., Parallel Coordinates: A Tool For Visualizing Multidimensional Geometry, *IEEE Visualization '90*, pp. 35-38, 1990.
- 5) Feiner S., Beshers C., Worlds within Worlds: Metaphors for Exploring n-Dimensional Virtual Worlds, *ACM SIGGRAPH Symposium on User Interface Software and Technology (UIST'90)*, pp. 76-83, 1990.
- 6) Itoh T., Takakura H., Sawada A., Koyamada K., Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, *IEEE Computer Graphics and Applications*, Vol. 26, No. 2, pp. 40-47, 2006.
- 7) Johnson B., et al., Tree-Maps: A Space Filling Approach to the Visualization of Hierarchical Information Space, *IEEE Visualization '91*, pp. 275-282, 1991.
- 8) Carriere J., et al., Research Paper: Interacting with Huge Hierarchies beyond Cone Trees, *IEEE Information Visualization 95*, pp. 74-81, 1995.
- 9) Koike H., Fractal Views: A Fractal-Based Method for Controlling Information Display, *ACM Transactions on Information Systems*, Vol. 13, No. 3, pp. 305-323, 1995.
- 10) Lamping J., Rao R., The Hyperbolic Browser: A Focus+context Technique for Visualizing Large Hierarchies, *Journal of Visual Languages and Computing*, Vol. 7, No. 1, pp. 33-55, 1996.
- 11) 五味, 伊藤, Li, CAT:大量画像の一覧可視化と詳細度制御の一手法, 画像電子学会 Visual Computing / 情報処理学会グラフィクスと CAD 合同シンポジウム 2007.