

CAT: 大量画像の一覧可視化と詳細度制御の一手法

五味愛[†] 伊藤貴之[†] Jia Li^{††}

CAT: A Technique for Image Browsing and Its Level-of-Detail Control

AI GOMI,[†] TAKAYUKI ITOH[†] and JIA LI^{††}

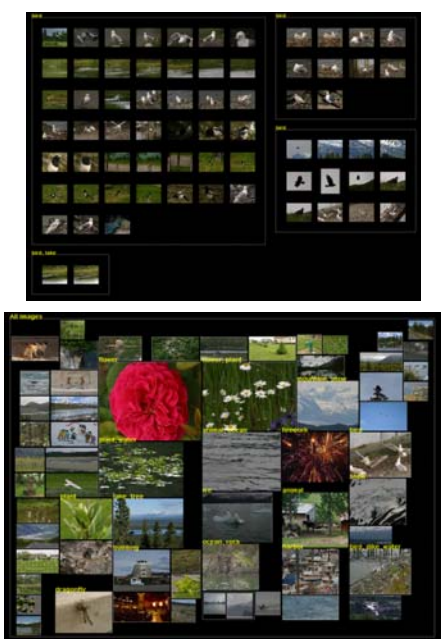


図 1 CAT の概観。(上) CAT では大量画像にクラスタリングを適用し、各々のクラスタを長方形の枠で表示する。(下) CAT ではズームアウト操作と連動した詳細度制御により、クラスタの枠を代表画像に置き換え、全てのクラスタを一覧可視化する。

1. はじめに

近年のデジタル画像撮影技術の急速な発達に伴い、画像の検索や分類に関する研究、および大量画像の一覧表示ブラウザに関する研究が活発である。現時点で

は、両者の研究は独立して進んでいる傾向にある。しかし、これら両者の研究は、本来は共通した目的に基づく研究である。よって両者の融合によって、さらに効果的な画像閲覧技術を構築できると考えられる。

本報告では、画像のクラスタリング手法、およびその一覧表示と詳細度制御のためのインターフェースを持ち合わせた、可視化手法 CAT(Clustered Album Thumbnails) を提案する。

CAT では大量画像の分類のために、注釈キーワードと画素情報を用いた 2 段階のクラスタリングを適用する。この際に CAT では、各クラスタの画像数の均一化を施しながら、ボトムアップ的に画像を分類する。

続いて CAT では、階層型データ可視化手法「平安京ビュー」^{3),4)} をベースにした手法により、階層的にクラスタリングされた画像群を一画面に表示する。さらに CAT では、ズームアウト操作と連動した詳細度制御にともない、クラスタに属する各画像を表示する代わりに、各クラスタの代表画像を表示する。

大量画像の一覧表示ブラウザに関する先行研究^{1),5),12)} と比較して CAT は、対話的操作と連動した詳細度制御によって適応的に画像を表示している、という点が特徴としてあげられる。また、この詳細度制御において重要な点は、各クラスタを表現する領域が、代表画像の縦横比に同等な形状の長方形で表現される、という点である。本報告の 4 章では、長方形領域によるクラスタの表現に「平安京ビュー」の適用が有効に働いていることを実証する。

図 1 は CAT の概観を示したものである。CAT では各クラスタを長方形領域で表現し、その内部に各画像を配置する。ズームアウト操作に伴って CAT は、長方形領域を代表画像で置き換え、大量画像全体のサマリーとなるような表示結果を提供する。

[†] お茶の水女子大学大学院
Graduate School of Humanities and Sciences, Ochanomizu University

^{††} ペンシルバニア州立大学
The Pennsylvania State University

2. 関連研究

2.1 画像検索

有名な画像検索エンジンの多くは、ウェブページ上にあるテキスト記述を参照して画像に注釈をつけ、その注釈に基づいて検索結果を提供する。しかし、これらのテキスト記述から、必ずしも画像の内容を適切に解説する注釈が得られるとは限らない。よってこの仕組みでは、満足な検索結果を得られるとは限らない。

そこでこの問題を解決するために、画像内容検索 (CBIR) の研究が取り組まれている。CBIR 関係技術の多くは、画像の特徴量を算出し、それに基づいて画像内容を検索する。参考文献¹¹⁾では、CBIR において色からの特徴量、続いて周波数特性から得た特徴量が頻繁に使われていると報告している。

一方でここ数年、画像内容に基づく自動画像注釈の研究が活発であり⁹⁾、画像検索エンジンの問題点を解決しうる技術として注目されている。

2.2 大量画像ブラウザ

ウェブサイトの画像検索エンジンなどの検索結果を表示するインタフェースの多くは、検索結果である画像群を、単純に類似性の高い順に並べて表示する。それゆえ、この種のインタフェースでは、ユーザの望む画像を常に効果的に素早く見つけられるとは限らない。この問題を解決する一例として Kang らは、さまざまな質問入力機能や詳細表示機能を搭載した、大量画像の検索手法を提案している⁷⁾。

大量画像ブラウザに関する研究の多くは、画像間の類似性を考慮した画像サムネイルのレイアウトに着目している。初歩的な多くの手法では、多次元尺度構成法 (MDS) や主成分分析 (PCA) を活用して、低次元ユークリッド空間に画像サムネイルを配置する。

木構造やグラフ構造などを対象とした情報可視化技術のいくつかは、大量画像ブラウザの研究にも適用されている。Jankun-Kelly らは、局所拡大表示型のグラフレイアウト技術を大量画像ブラウザに適用している⁵⁾。Bederson らは PhotoMesa¹⁾ という大量画像ブラウザを提案している。PhotoMesa は 2.3 節で紹介する Quantum Treemap²⁾ を採用してディスプレイ空間を分割し、大量画像を階層的に整理して表示する。

画像注釈に基づく大量画像の閲覧手法として、Yang らは SIB (Semantic Image Browser)¹²⁾ を提案している。SIB では画像の木構造化やグラフ化は実現しておらず、Yang ら自身も画像の階層化や詳細度制御表示を今後の課題と位置づけている。この観点において本報告で提案する CAT は、SIB よりも先行している。

2.3 階層型データの可視化

階層型データの情報可視化手法の中でも有名なものに、木を表示する手法と、画面空間充填による手法があげられる。本報告で提案する CAT は、画面空間充填による手法に属する。

CAT は画面空間充填による階層型データ可視化手法「平安京ビュー」^{3),4)} をベースにした手法である。「平安京ビュー」では図 2 のように、入れ子型の長方形の枠で階層を表し、葉ノードをアイコンで表示する。「平安京ビュー」は階層構造の最下位に属する葉ノード群を、クリックしやすく等しいサイズのアイコンとして、しかも互いに重ならずに表示できる、という特徴がある。このアイコンを画像サムネイルに置き換えることで、CAT は大量の画像を、等しいサイズで互いに重ならず、画面上に一覧表示できる。

一方で、2.2 節で紹介した PhotoMesa¹⁾ の中枢技術として適用されている Quantum Treemap²⁾ は、画面空間充填による階層型データ可視化手法 Treemaps⁶⁾ の応用技術であり、「平安京ビュー」に類似した特徴を持つ。文献⁴⁾では Quantum Treemap と「平安京ビュー」の比較評価結果を掲載している。この比較評価結果において「平安京ビュー」は、類似する階層型データのレイアウトの安定性と、長方形領域の縦横比において、よい結果を得ている。また「平安京ビュー」は、長方形空間の配置において柔軟な操作を行えるという特徴もある。文献^{3),4)}で述べるように「平安京ビュー」は、画像配置の理想位置を記述したテンプレートによって、長方形の位置を操作する機能を有する。

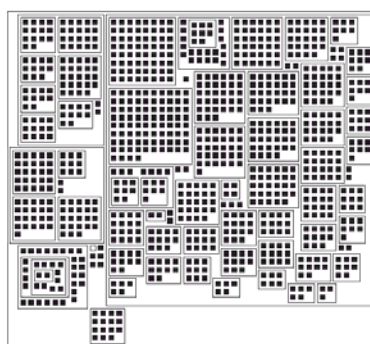


図 2 階層型データ可視化手法「平安京ビュー」の表示例。

3. CAT の処理手順

3.1 概要

本報告が提案する CAT では、各々の画像には 1 つ以上のキーワードが割り当てられていることを前提と

する。この前提の下でCATはまず、以下の2段階処理によって画像をクラスタリングする。

- (1) まずキーワードに基づいて画像をクラスタリングする。この処理結果として得られるクラスタを、高階層クラスタと称する。
- (2) 各々の高階層クラスタに属する画像を、画素情報でクラスタリングする。この処理結果として得られるクラスタを、低階層クラスタと称する。

CATは上記の手順でクラスタリングされた画像群を、以下のように一覧可視化する。まず低階層クラスタに属する画像を並べて、それを長方形の枠で囲むことで、低階層クラスタ自身を表現する。この処理を各々の低階層クラスタにて実行した後、低階層クラスタを表現する長方形領域を並べて、それを長方形の枠で囲むことで、高階層クラスタを表現する。最後に高階層クラスタを表現する長方形領域を並べることで、大量画像群を格納する階層型データ全体を表現する。以上の処理手順の概要を図3に示す。

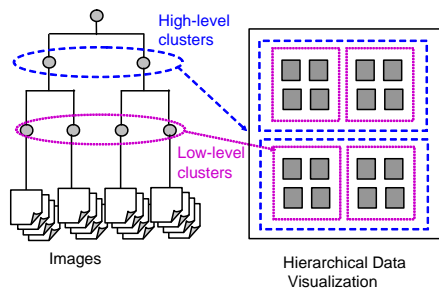


図3 CATの処理手順の概要。

3.2 キーワードからの画像クラスタリング

CATでは文献¹⁰⁾と同様な手法で、キーワードに基づく画像間の距離を定義する。ここで全キーワードを V とし、画像 X_i のキーワードセットを W_i とする。

$$W_i = \{w_{i,1}, \dots, w_{i,m_i}\}, w_{i,j} \in V \quad (1)$$

なお m_i は、画像 X_i に割り当てられたキーワード数である。

ここでCATでは、2画像 X_i と X_j の距離を以下の式¹⁰⁾によって算出する。

$$D_w(W_i, W_j) = \frac{1}{2m_i} \sum_{k=1}^{m_i} \overline{d}_w(w_{i,k}, W_j) + \frac{1}{2m_j} \sum_{l=1}^{m_j} \overline{d}_w(w_{j,l}, W_i) \quad (2)$$

ここで、 D_w は2つのキーワードセット W_i と W_j 間の距離、 $\overline{d}_w(w_{i,k}, W_j)$ はキーワード $w_{i,k}$ とキーワードのセット W_j 間の距離を示している。我々の実装

では、ソフトウェアパッケージ WordNet Similarity によってキーワード間の距離を算出している。

続いてCATでは、画像間の距離に基づいて、画像をクラスタリングする。ここではボトムアップ形式のリンケージクラスタリングアルゴリズムを使用する。ただしCATでは、各々の高階層クラスタに属する画像数の均等化を試みるために、クラスタリング過程において以下の処理も同時に行う。

ここで、 $s_1 < s_2$ となるような閾値 s_1 と s_2 を定義する。あるクラスタに属する画像数が s_1 よりも少なかったら、CATでは積極的に他のクラスタと合併させる。一方、あるクラスタに属する画像数が s_2 よりも多かったら、そのクラスタはこれ以上の合併を避ける。そして、クラスタ間の最小距離値が閾値を超えるまで、距離が最小であるクラスタを再帰的に合併する。

3.3 画素情報による画像クラスタリング

CATでは、キーワードに基づいて生成された各クラスタ内の画像を、画素情報に基づいて再度クラスタリングする。我々の実装では、色情報に基づく特徴量と、周波数情報に基づく特徴量から、各画像の特徴ベクトルを算出し、そのベクトル間の余弦に基づいてクラスタを生成する。

色情報に基づく特徴量として、現時点で我々は、画像空間を格子状に分割し、各々の領域における代表画素値を算出し、この結果を多次元ベクトル化したものを用いている。

周波数情報に基づく特徴量として、現時点で我々は文献⁸⁾と同様な手法を用いている。この手法では画像にウェーブレット変換を施して高周波成分を算出し、その高周波成分が非零である画素を抽出する。続いて抽出された画素について、近隣の画素値を参照して25種類のテンプレートのいずれかに該当することを特定する。この25種類のテンプレートごとに該当画素数を集計し、そのヒストグラムを多次元ベクトルとして用いている。

この2種類の多次元ベクトルを併合したベクトルを用いてCATでは、余弦値が大きい順にボトムアップ的に低階層クラスタを生成する。この処理は、2クラスタ間の最大余弦値が閾値より小さくなるまで反復される。この過程で前節と同じように、各々の低階層クラスタに属する画像数の均等化を試みる。

3.4 代表画像の選択

クラスタリング完了後にCATでは、高階層クラスタ、低階層クラスタ、の両方に対して代表画像を選択

<http://www.d.umn.edu/~tperdese/similarity.html> 参照。

する。現時点で我々は、以下の2種類の代表画像選出方法を採用している。

1つ目の手法は、キーワードに基づく方法である。ここで、キーワード $w_{i,j}$ を持つ画像の数を $k_{i,j}$ とする。このとき $k_{i,j}$ が他よりも大きければ、 $w_{i,j}$ は重要なキーワードと想定できる。よって $k_{i,j}$ は、クラスタ内のキーワードの重要度を表すと考えられる。

ここで以下の式により、画像 X_i のキーワードの重要度合計を定義する：

$$K_i = \sum_{j=1}^{m_i} k_{i,j} \quad (3)$$

多くの場合において、 K_i 値が最大である画像は、クラスタの特徴を表していると言えるので、その画像はクラスタの代表画像として望ましいと判断できる。

2つ目の手法は、画素情報に基づく手法である。特徴ベクトル空間のクラスタの中心に最も近い画像が、クラスタの平均的な画像であるといえるので、その画像はクラスタの代表として望ましいと判断できる。

3.5 階層型画像のクラスタの可視化

続いてCATでは、以下の3ステップにより、大量画像を画面空間に配置する。2.3節で述べた通り、この手法は「平安京ビュー」をベースにした手法である。

- (1) 低階層クラスタに属する画像サムネイルを格子状に配置し、長方形でそれらを囲む。全ての低階層クラスタに対して、同様の過程を繰り返す。
- (2) 低階層クラスタを表現する全ての長方形を隙間なく配置し、高階層クラスタを表現する長方形で囲む。全ての高階層クラスタに対して、同様の過程を繰り返す。
- (3) 高階層クラスタを表現する全ての長方形を隙間なく配置する。

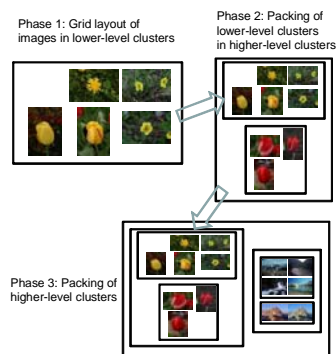


図4 画像配置のための3ステップ

CATは「平安京ビュー」と同様、文献^{3),4)}に示す

以下の値を最小化するように長方形群を配置する。

$$aA + rR + dD \quad (4)$$

ここで a, r, d は、正の定数でありユーザーが指定する。 A は長方形配置における前後間の長方形の縦横比、 R は長方形配置における前後間の長方形の縦横比、 D は長方形の実際の位置とテンプレートに描かれた理想的な位置との間の距離である。CATでは変数 A を、理想的な縦横比 (= 代表画像の縦横比) に対する長方形領域の縦横比の悪化の度合い、として計算する。

CATでは、類似度の高いクラスタを、画面上で隣接配置できる。まず2.2節で説明したMDSやPCAに基づく方法で、クラスタの2次元空間上での理想位置を算出する。続いて2.3節で説明した方法で、理想位置の算出結果をテンプレートとして各クラスタの画面上の位置を操作することで、類似度の高いクラスタを画面上で隣接配置する。

3.6 ユーザインタフェースと詳細度制御

我々のCATの実装では、マウスの左ボタンに平行移動、ホイールに拡大縮小、の各操作を割り当てている。左ボタンのダブルクリックにより、カーソルが指す画像の拡大表示、あるいはウェブブラウザでの表示が可能である。また右ボタンのダブルクリックにより、カーソルが指す部分をウィンドウの中心に移動できる。

また我々の実装では、ズームアウト状態でCATを起動したときには代表画像だけをハードディスクから主記憶に読み込み、ズームイン状態でフォーカス部位の画像サムネイルだけを読み込み、また表示対象から外れた画像は随時主記憶から解放することで、主記憶使用量を節約している。

4. 評価実験

我々はGNU gcc 3.4 およびOpenCVを用いてクラスタリング部を実装し、Java SDK 1.5を用いてブラウザ部を実装した。テスト環境にはWindows XPを搭載したPCを用い、横384画素、縦256画素のJPEG画像2360枚を活用した。紙面の都合により画像表示例は割愛する。

以下、10人の学生被験者による、主観評価のためのユーザテスト結果を示す。

4.1 主観的な外観評価

まず本実験では、以下の5種類の異なる表示条件を課したブラウザを用意した。

<http://www.stat.psu.edu/~jiali/index.download.html> に、本実験に用いた2360枚の画像が公開されている。
<http://itolab.is.ocha.ac.jp/~gomiai/> に、画像表示例を公開する予定である。

- 「クラスタなし」では、全ての画像を、単純に格子状に表示する。
- 「低階層クラスタ・代表画像なし」では、画素情報に基づく1階層のクラスタを生成する。代表画像は表示しない。
- 「低階層クラスタ・代表画像あり」では、画素情報に基づく1階層のクラスタを生成する。代表画像を表示する。
- 「2階層クラスタ・代表画像なし」では、キーワードと画素情報に基づく2階層のクラスタを生成する。代表画像は表示しない。
- 「2階層クラスタ・代表画像あり」では、キーワードと画素情報に基づく2階層のクラスタを生成する。代表画像を表示する。

この5種類のブラウザを被験者に数分間操作してもらい、視覚的印象と使いやすさの観点から、1位が最高評価、5位が最低評価として、主観的に順位をつけてもらった。表1に結果を示す。この結果より、2階層クラスタリングおよび代表画像の有効性がわかる。特に代表画像の表示機能に高い評価が集中している。

この実験で被験者から得た意見として、代表画像の存在は、詳細度制御による画像表示サイズの適正化だけでなく、表示速度の観点からも重要である、という意見があった。代表画像を表示しない場合、初期操作で非常に多くの画像を表示するため、表示速度と主記憶使用量の観点で不利であることが、被験者にも実感されていることがわかる。

表1 主観的概観評価

順位	1	2	3	4	5
クラスタなし	0	0	1	0	9
低階層クラスタ・代表画像なし	0	0	0	10	0
低階層クラスタ・代表画像あり	0	6	3	0	1
2階層クラスタ・代表画像なし	2	2	6	0	0
2階層クラスタ・代表画像あり	8	2	0	0	0

4.2 特定の画像を探し出す所要時間

次に、紙に印刷した特定の画像を被験者に提示し、2360枚の画像の中から探し出す時間を計る、という実験を行った。この実験では「クラスタなし」「低階層クラスタリング・代表画像あり」「2階層クラスタリング・代表画像あり」の3種類のブラウザを使用した。代表画像の有無による所要時間の違いは、ほとんど見受けられなかった。

表2では、時間の測定結果を表示した。最短、平均、最長時間をそれぞれの手法において示している。

この結果から、色分布や周波数特性を用いたクラスタリングだけでは画像を探し出す時間はあまり短縮さ

表2 特定画像を探し出す所要時間(秒)

	クラスタなし	低階層	2階層
最短	22.0	47.3	11.8
平均	129.26	97.52	42.68
最長	279.6	223.6	85.3

れず、キーワードを用いたクラスタリングを併用することで大幅に短縮されていることがわかる。このことから、ユーザが画像を探し出す際にキーワードの存在が非常に重要である、ということが示唆される。

4.3 長方形の縦横比

続いてCATによる画面配置結果から、各クラスタを表現する長方形領域の縦横比を算出し、クラスタの代表画像の縦横比にどれくらい近いかを評価した。本実験に用いた全ての画像の縦横比は4/3なので、本実験における最適な縦横比は、全ての長方形領域において4/3である。

表3に縦横比の算出結果を示す。ここでは、高階層と低階層クラスタの縦横比を個々に表示した。

表3 縦横比率の統計結果

縦横比の区間	高階層	低階層
1.0 ~ 1.1	6	10
1.1 ~ 1.2	3	6
1.2 ~ 1.3	18	38
1.3 ~ 1.4	14	41
1.4 ~ 1.5	7	19
1.5 ~ 1.6	2	16
1.6 ~ 1.7	0	12
1.7 ~ 1.8	1	2
1.8 ~ 1.9	1	3
1.9 ~ 2.0	4	4
2.0 ~	5	0
平均値	1.336	1.383

高階層、低階層ともに、多くのクラスタの縦横比は、理想値に非常に近いといえる。また高階層クラスタにおいて、式4の定数 a を調節することによって、縦横比を改良することが出来る。しかしながら、これによって空領域比が悪化する可能性もある。式4の最も望ましい値の設定については、実験的に、そして主観的に議論する必要があると考えられる。

4.4 長方形領域

最後に、全ての長方形領域について面積を計算し、CATにおけるクラスタリング手法が長方形領域の面積均一化に貢献していることを実証した。

CATのクラスタリング手法は、 s_1 の値を適切に設定することで、小さいクラスタを合併させ、バランスのよいクラスタリング結果を得る。ただし $s_1 \leq 1$ である場合には、クラスタの併合は一切行われない。こ

のクラスタ併合によって CAT では、ズームアウト操作によって代表画像の表示サイズが小さくなったときにも、視認できる十分な大きさで代表画像を表示する。

表 4 では $s_1 = 1$ と $s_1 = 10$ の 2 値について、高階層クラスタを構成する長方形領域面積の最小値、平均値、最大値、最大値と最小値の比、の各々を算出した結果を示している。本研究の目的において、この算出結果は、長方形領域面積の最小値が大きく、最大値と最小値の比が小さいことが望ましい。表 4 からわかるように、我々の実験では $s_1 = 10$ のときのほうが、望ましい長方形領域面積が得られている。この結果から、提案手法によるクラスタの併合が効果的に働いていることがわかる。

	$s_1 = 1$	$s_1 = 10$
最小値	9.28	45.43
平均値	81.46	178.44
最大値	833.17	833.17
最大値と最小値の比	89.78	18.34

5. 結 論

本報告では、クラスタリングされた大量画像の一覧表示と詳細度制御を実現する可視化手法 CAT (Clustered Album Thumbnail) を提案した。今後の課題として、以下について検討したい。

- 何十万もの画像を用いた膨大なデータでの実験。
- 式 4 の a, r, d の最適な値を求めるための主観評価。
- よりよい階層構造の構築手法の考察。「平安京ビュー」は文献³⁾ で述べている通り、深さおよび幅が不均一な階層構造の可視化において関連手法より優位である。この利点を活かして、「不均一であるがゆえに却って可視化に有利な階層構造」の構築方法を追求したい。
- クラスタリング結果の主観評価実験。特に、低階層クラスタリングのアルゴリズムや特徴量ベクトルをキーワードごとに使い分けるような機能を追加することで、低階層クラスタリング結果の主観評価の向上を目指したい。
- ズームインタフェース機能や focus+context インタフェース機能を備えた可視化技術との融合。
- SIB¹²⁾ が目指しているような、画像注釈や画像分類のデバッグツールとしての応用。

なお本研究の一部は、日本学術振興会科学研究費補助金の助成に関するものである。

参 考 文 献

- 1) Bederson B., B., PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps, *User Interface Software and Technology*, pp. 71-80, 2001.
- 2) Bederson B., Schneiderman B., Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, *ACM Trans. on Graphics*, Vol. 21, No. 4, pp. 833-854, 2002.
- 3) Itoh T., et al., Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm, *IEEE Trans. on Visualization and Computer Graphics*, Vol. 10, No. 3, pp. 302-313, 2004.
- 4) 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.
- 5) Jankun-Kelly T. J., Ma K.-L., MoireGraphs: Radial Focus+Context Visualization and Interaction for Graphs with Visual Nodes, *IEEE Information Visualization*, pp. 59-66, 2003.
- 6) Johnson B., Shneiderman B., Tree-Maps: A Space Filling Approach to the Visualization of Hierarchical Information Space, *IEEE Visualization*, pp. 275-282, 1991.
- 7) Kang H., Shneiderman B., Visualization Methods for Personal Photo Collections: Browsing and Searching in the Photofinder, *IEEE Intl. Conf. on Multimedia and Expo 2000*, pp. 1539-1542, 2000.
- 8) Kubo M., Aghbari Z., Makinouchi A., Image Retrieval by Edge Features Using Higher Order Autocorrelation in a SOM Environment, *IEICE, Trans. on Information and Systems*, Vol. E86-D, No. 8, pp.1406-1415, 2003.
- 9) Li J., Wang J. Z., Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach, *IEEE Trans. on Pattern Analysis and Machine Intelligence* Vol. 25, No. 9, pp. 1075-1088, 2003.
- 10) Li J., A Mutual Semantic Endorsement Approach to Image Retrieval and Context Provision, *ACM SIGMM, Multimedia Information Retrieval*, pp. 173-182, 2005.
- 11) Veltkamp R. C., Tanase M., Content-Based Image Retrieval Systems: A Survey, *UU-CS-2000-34, Institute of Information and Computing Science, University Utrecht*, 2000.
- 12) Yang J., et al., Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis, *IEEE Visual Analytics in Science and Technology*, pp. 191-198, 2006.