社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－CVIM－153 (9)
2006／3／16

選択注視点における特徴ベクトルの階層化クラスタリング

斎藤 純† 山名 早人††,†††

†早稲田大学大学院理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1
††早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1
†††国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: †john@yama.info.waseda.ac.jp, ††yamana@waseda.jp

あらまし コンテントベース画像検索において，(i)検索の高速化及び(ii)検索における意味性を向上させるためには，画像の分類が必要不可欠となる．本研究では，画像の中から選択注視点を特定したのち，注視点とその周辺情報を特徴ベクトルとし，階層化クラスタリングにより画像を分類する．選択注視点とは，視覚内の，すなわち空間的なはずれ値である．本稿では，選択注視モデルを拡張し，時間的なはずれ値を検出するために，残差情報を用いた階層化クラスタリング手法を提案する．また，本稿では，獲得したカテゴリの重心ベクトルを用いた重み付けによる選択注視点の移動方法を述べる．結果，選択注視手法のはずれ値検出の考え方は，階層的なカテゴリの獲得に応用可能であることがわかった．また，獲得したカテゴリに属する注視点を画像の中から探索することが可能となった．

キーワード 選択注視，コンテントベース画像検索，残差情報を用いた階層化クラスタリング

Hierarchical Clustering Of Feature Vectors at Visual Attentional Points

Jun SAITO† Hayato YAMANA††,†††

††Science and Engineering Waseda, University Okubo3-4-1, Shinjyuku-ku, Tokyo, 169-8555 Japan
†††National Institute of Informatics Hitotsubashi2-1-2, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: †john@yama.info.waseda.ac.jp, ††yamana@ yama.info.waseda.ac.jp

Abstract In Content-based image retrieval, the classifications is needed for better performances of (i)speeds of retrieval and (ii)semanticity of retrieval. Our system extracts the most attentional points by using a selective visual attention model which extracts feature vectors of attentional points in images. And our system classifies feature vectors by hierarchical clustering with residuals. An attentional point in an image is outlier in an image, or special outlier. We propose extension of selective attention model to extract temporal outlier with residual vectors, and the method of moving attentional points weighted by a cancroid of a category extracted. This paper shows that the outlier extraction idea of selective visual attention is extensible to extract hierarchical categories. And also, this paper shows that our method can select an image point which belongs to an extracted category.

Keyword Selective Visual Attention, Content-based Image Retrieval, Hierarchical Clustering With Residual

1. はじめに

コンテントベース画像検索を対象としたデータベース作成は，画像セット内の画像の特徴ベクトルを求め，それを照合用のデータベースに格納する．しかし，特徴ベクトルの全要素を対象とした検索は，膨大な時間を必要とする．そこで，データベース中の画像から抽出した特徴ベクトルに対し事前にクラスタリングを行うことにより，検索空間を小さくすることが必要となる．

また，画像の適確な分類は，高速化だけではなく，セマンティックな検索にも必要である[1][2]．現在の画像検索システムの検索結果は，ユーザの求める結果と異なるという問題がある．これは，画像の領域分割や低次特徴抽出から得られた情報のみによる検索と，ユーザの求める意味的な検索との間にギャップが存在するためである．こうしたギャップを埋めるためには，検索に先立ち，画像のカテゴライズ・ラベリングが必要となる．

本研究で用いる特徴ベクトルは，選択注視点[4]における低次特徴量とその周囲のヒストグラムである[3]．選択注意モデルとは，人間や霊長類の被検体が，画像を提示された際「眼球運動の起こる前に，どこに注意が向けるか」のモデルである．選択注視点における特徴ベクトルのクラスタリングの研究として，最小全域木の密度を用いたも

の[9]がある．この手法は，ノンパラメトリックであるので，入力特徴ベクトルに依存しないクラスタリングが可能だが，サンプルが無数にある場合に計算量が膨大になる．また，選択注意モデルでは，視覚内の（空間的な）はずれ値(outlier)を求める．平易な言葉で言い換えると，視覚内のはずれ値とは，「珍しいもの」ということである．しかし，空間的かつ時間軸的なはずれ値を重要と考えるクラスタリング手法はない．そこで本研究では，学習後の学習器にとって誤差の大きい入力は，時間軸的にはずれ値であると考え，情報量が高い入力であると考えた．これより，学習器の下層ユニットの処理段階において誤差の大きい入力を，上層ユニットで処理するモデルを考えた．このようなモデルで，時間軸的はずれ値の検出・学習が可能になると考えられる．以上より，学習器は，階層型であり，残差情報を入力する上位層を備えているものを用いた．

以下，2 節では，特徴ベクトル抽出手法である，選択注視モデルを説明する．3 節では，残差情報を用いた階層化クラスタリング，4 節では，評価方法とその結果，5 節では考察を述べる．

2. 選択注視モデル

画像は高次元な情報であるため，次元を落とした特徴ベクトルを画像から抽出し，画像の「指紋」として検索に用いるのが一般的である．本研究では，画像の特徴ベクトルとして，画像中の選択注視点とその周辺の情報（ヒストグラム）を用いる[3]．特徴ベクトルに選択注視点のベクトルを用いた理由は，画像中の局所的に情報量の高い点を検出し，集中的に処理することにより，システムの識別能力を高めることが可能と考えたからである．選択注視点を求める手法は，[4]の選択注視モデルをベースにした．

また，画像内の重要な点は，1 点のみであることは稀である．トップダウンの選択注視手法では，事前知識を用いて複数ある選択注視点のそれぞれの重要度を変更することができる．

以下では，低次特徴のみから選択注視点を決める手法と，画像検索に用いる特徴ベクトルについてと，選択注視点の変更方法について述べる．

2.1. ボトムアップ選択注視

選択注視モデルとは，人間や霊長類が「画像などの視覚情報を提示された際に，眼球運動が起こる前に，画像中のどの部分に注意が高くなるか」をモデル化したものである．選択注視モデルで抽出された点は，脳にとって情報量が高いことを示している．本研究で用いた選択注視モデル[4]では，次の過程を経て，注視点が決定される（図 1 参照）．

(i) 入力画像から，輝度変化・色・方位に関する低次特徴情報を値として抽出し，入力画像と同じ大きさのマップを作成する．

(ii) それぞれ得られた低次特徴マップを各点で加算した「サリエンシーマップ」を作成し，その最大値をとる箇所を選択注視点とする．

本研究では，

- 12 段階の周波数の輝度変化情報
- 10 段階の周波数の色情報
- 6 方位の方位情報

から，選択注視点を抽出した．サリエンシーマップは低次特徴量から作成されるため，サリエンシーマップの最大値を示す点を注視点とする選択注視手法は，ボトムアップ選択注視と呼ばれる．

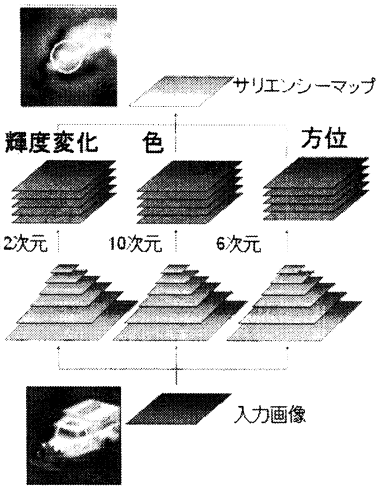

図 1. ボトムアップ選択注視モデル．入力画像に対して，サリエンシーマップが作成され，注視点が決定される．

2.2. 注視点とその周辺情報の特徴ベクトル

本研究で用いる画像検索用の特徴ベクトルは，注視点とその周囲の情報である．注視点の情報とは，2.1 で述べた各マップの注視点における値である．しかし，注視点のみの情報では，十分な特徴抽出が難しい．そこで，注視点の周囲の情報として，ヒストグラムを用いた．本研究では，次のヒストグラムを用いた．

- 色ヒストグラム：HSV 空間の H を 5 分割，S を 5 分割の 25 ビン
- ・方位ヒストグラム：6 方位×4 周期の 24 ビン

ヒストグラムを得た範囲は，画像の幅の 20%を一

辺とする正方形である．以上より，注視点ベクトルは，輝度変化，色，方位それぞれ 12，10，6 次元，ヒストグラムは色，方位それぞれ 25，24 次元である．よって，注視点とその周辺の情報を持つベクトルは，合計 77 次元となった．この特徴ベクトルをデータベースに格納し，検索に用いる．

2.3. トップダウン選択注視

2.1 において，低次特徴マップを加算するが，ここで重み付けをすることにより，選択注視点を変更させることが考えられる．例えば，画像中にオブジェクトが 2 つ存在しているとき，どちらか一方のオブジェクト上に選択注視点を配置させたい場合が考えられる．このとき，重み付けパターン 1 によってはオブジェクト 1 上に注視点が選定され，重み付けパターン 2 によってはオブジェクト 2 上に注視点が選定されるような手法が必要になる．

ここで，重み付けのパターンの抽出が問題となる．そこで，画像集合 1 にはオブジェクト 1（たとえば黄色い）のみが存在していて，画像集合 2 にはオブジェクト 2（たとえば白黒の縞模様）のみが存在している場合を考える． 2.1 より，選択注視点は，各低次特徴量の値が高い部分が選択される．オブジェクト 1 の画像集合から得られた注視点特徴ベクトルは，オブジェクト 1 に共通して，ある「特徴的な」変数（元）の値が高いと考えられる．例では，オブジェクト 2 の特徴ベクトルにおいて，色の変数の値は，0 となっている確率が高く，少なくともオブジェクト 1 の色の変数の値よりは低い値を示すと考えられる．

以上より，本研究では，この重み付け方法は，次の手順で得た．

(i) 重み付けを獲得したいオブジェクトの画像集合 n 枚から，それぞれ注視点ベクトル（2.2 参照）を n 個抽出する．
(ii) n 個の注視点ベクトルの重心（各変数の平均値）を求める．
(iii) 重心ベクトルが重みそのものとする．

重心ベクトルは，オブジェクトから得た「知識」であり，このように事前知識を利用した選択注視手法はトップダウン選択注視と呼ばれる．

この方法の有効性を示すため，画像中に 2 つのオブジェクトが存在する画像を用いて，片方のオブジェクトに対する検索精度の検証を行った（4 節参照）．

しかし，実際の画像においては，単独のオブジェクトが存在している場合は少ないため，カテゴリの平均値を直接抽出することはできない．そこで，続く節では，注視点ベクトルの教師（ラベル）なしの分類を行う手法について述べる．

3. 残差情報を用いた階層化クラスタリング

本節では，選択注視モデルで得られた注視点の特徴ベクトルを教師なしで分類する手法について述べる．本研究では最終的に，検索対象の画像は Web 画像を想定しており，対象とする Web 画像の量は膨大となる．例えば，Web クローリングの結果，収集期間 2004 年 1 月－2005 年 7 月で，30 億ページ中 78,598,283 の jpeg リンクが存在することがわかった．このため，クラスタリングにおいて，バッチ処理は妥当ではないと考えられる．また，事前にカテゴリをマニュアルで設定することには，以下の 2 つの問題がある．

(1) あらかじめ用意したカテゴリが妥当とは限らない．
(2) カテゴリが多量になったとき，ユーザは求めるカテゴリの選択が困難になる．かつ，検索に時間がかかるようになる．

(1)より，教師なし学習が必要となる．(2)の問題は，階層型にすることにより解決が可能となる．以上より，本研究で用いるクラスタリング手法は次の 3 つの特性を備えているものを用いた．

(i) オンライン型
(ii) 教師なし学習
(iii) 階層型

そこで，本研究で用いるクラスタリングには，(i)(ii)を備えた SOM(自己組織化マップ)を用いた．ただし，階層化については，時間的はずれ値を検出する拡張を加えた．

3.1. SOM（自己組織化マップ）

特徴ベクトルのクラスタリングには，SOM(Self-Organization Map，自己組織化マップ)[6]を用いた． SOM は，近傍どうしのユニットが，類似した入力に反応するように学習が進む．このため，全ユニットが入力ベクトルの重心に近づきやすくなる．

SOM は，K 個のユニットを備える．それぞれのユニットは，入力する特徴ベクトルと同次元(q 次元)の選択性ベクトル $\mathbf{w}_k$ を持つ．この q 次元ベクトルは，ランダム値で初期化されている．以下が，学習アルゴリズムである．

(i) K 個のユニットについて，入力特徴ベクトルとの誤差を計測する．
(ii) 誤差の一番小さいユニットが勝者となる．
(iii) 勝者ユニットとその周辺のユニットが，誤差に従って学習する．

ユニット k の更新（学習）式は次のものである．

$$\vec{m}_k(t+1) = \vec{m}_k(t) + \alpha(\vec{x}(t) - \vec{m}_k(t)) \qquad (1)$$

ここで，$\vec{m}_k(t+1)$は，$\vec{m}$が更新されていることを示す．すなわち，時刻 t の入力ベクトル$\vec{x}$に対して，k 番目のユニットが勝者となり，学習係数 αで，選択性ベクトル$\vec{m}$の学習が進むということである．学習の結果，K 個のクラスの重心が求められる．学習係数は α=0.01 で実験を行った．

3.2. 残差情報を用いた階層化

本節の冒頭で挙げたように，クラスタリングには階層化が必要となる．ここで階層化にあたり，「時間的なはずれ値」を検出する方法を考えた．これは，選択注視手法の拡張である．Itti の選択注視手法[4]では，「選択注視点は空間的はずれ値である」として，注視点の選定の有効性を向上させる方法を用いている．空間的はずれ値であるとは，「同時に存在するものの中で珍しい」ということである．さらに Itti は，Surprise Theory で，時間的はずれ値の考え方も取り入れ，入力と予測の誤差の情報の重要性について論じている[5]．

そこで，本提案手法では，SOM の下層ユニットの選択性ベクトルを予測ベクトルととらえ，入力ベクトルと予測ベクトルの差である残差ベクトルを上層への入力ベクトルとした．これにより，下層において誤差の大きな入力，すなわち，「学習器にとって珍しい入力」の検知が可能となる．かつ，下層で誤差の小さい入力は，下層ユニットの持つ選択性ベクトルによるカテゴライズが，すでに十分であるような入力であることを示す．このような入力はすなわち，「学習器にとってありふれた入力」であるが，「過去には珍しかったが，学習が十分に終わった入力」であるともいえる．

今回は，下層のユニット k についてそれぞれ上層ユニット L 個を用意した．すなわち，上層は合計 $K \times L$ 個のユニットが存在することになる．ただし，k ごと付属する L 個のユニット群の間は独立に動作をすることになる．今回は，K=9,L=9 で実験を行った（図 2 参照）．

3.3. 本提案手法と他の学習機械との関連

階層化において，残差情報を扱うモデルとしては，Predictive Coding がある[8]．残差情報を用いた SOM とは，次のように関係している．

SOM は，学習において，近傍のユニットも同時に学習するという制約をはずすと k-means クラスタリングと等価になる．k-means クラスタリングで，勝者ユニットの値が 1，敗者ユニットの値が 0 という離散値をとるのではなく，ユニットの値がスパースに分布するような拡張を加えると Sparse Coding[7]となる．さらに，Sparse Coding を拡張して，上層への残差ベクトル伝播を取り入れたモデルに Predictive Coding がある．SOM, Sparse Coding, Predictive Coding はニューラルネットワークの 1 種であり，脳のトップダウンの信号の役割のモデル化をしている．よって，本提案手法もニューラルネットベースの脳のモデルと考えることも可能である．

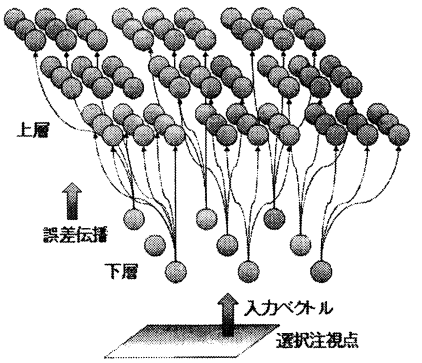

図 2. 本研究で用いた階層化 SOM. 下層のユニット 1 個について，上層の 9 つのユニットが付属する．上層のユニット群について，群の中では競合学習が行われるが，群間の関係性はない．

4. 評価方法と結果

本提案手法に関して評価するべきポイントは，以下の 2 つである．

(1) 階層化 SOM が，選択注視点とその周辺情報の特徴ベクトルを，適確に分離できる選択性ベクトルを獲得できるか．

(2) 階層化 SOM で得られたクラスによるトップダウンの信号（クラスの平均ベクトル，SOM の選択性ベクトル）が，目的オブジェクトへ選択注視点の移動をさせることができるか．

以下，上記 2 点について，評価方法と結果を述べる．

4.1. 検索結果の情報エントロピー

画像検索結果のばらつきが少ないほど，分類手法が有用であると考えられる．本提案システムの有用性の評価方法としては，ばらつきの尺度である，情報エントロピーを用いた．情報エントロピーHは，

$$H = -\sum_{i}^{N} p_i(x) \log p_i(x) \qquad (1)$$

であり，ここで，$p(x)$はオブジェクト $i(i=1...N)$の出現頻度を表す．N はオブジェクト（カテゴリ）のラベ

ルの数である．今回，画像セットは coil-100[10]を用いた．coil-100 には，100 オブジェクト×72 枚の画像がある．

画像を分類（学習）した結果，階層化 SOM の各ユニットの持つ選択性ベクトルがカテゴリの平均値として得られる．この選択性ベクトルをクエリとして，画像検索を行う．類似度の近かった上位 72 枚を用いて，エントロピーを計算した．

表 1 に，分類方法とその検索結果の情報エントロピーを示す．ここで，オブジェクト平均とは，オブジェクトごとの平均値を検索クエリとしたものである．オブジェクト平均の検索結果のエントロピーは，教師信号をじかに用いていることから，理想結果としての比較対象である．また，階層化の有無を比較するために，ユニットが 81 個の単層 SOM による検索結果のエントロピーを調べた．結果，階層化 SOM との違いはあまり見られなかった（片側 t 検定の結果，p 値は 15.4%）．しかし，識別能力が同性能の場合，階層化されているほうが，高速な検索やセマンティックな検索が可能となる．

表 1．分類方法によるエントロピーの違い

	オブジェクト平均	SOM(ユニット 81 個)	階層化 SOM（上層）
H	1.249	1.773	1.692

4.2. トップダウン選択注視の有用性

選択注視に事前知識を用いる手法（トップダウン選択注視）の有用性を調べるために，次のような実験を行った．

- 画像中にオブジェクトが複数（2 つ）存在している画像を作成し，クエリとする．
- 選択目的オブジェクトの特徴ベクトルの平均値を重みとして，サリエンシーマップを作成し，注視点を選定する．
- 注視点から得られた特徴ベクトルで検索し，距離で順位を付ける．
- 検索結果上位 40 個中で，目的オブジェクト数を精度とする

オブジェクト（カテゴリ）は，coil-100 から任意に選択した 10 個を用いた．以上のようにして，トップダウン選択注視によって，検索精度が向上するかどうかを調べた．トップダウン選択注視による，検索精度の違いを表 2 に示す．表の番号は，coil-100 におけるオブジェクトの番号を示す．また，トップダウン信号の有無によって，サリエンシーマップと検索結果が異なる様子を図 3 に示す．

表 2 より，オブジェクトによっては大きく検索精度が向上することがわかる．ただし，検索精度が上がらないオブジェクトも存在する．これは，選択注視点が目的オブジェクト上に選定されたとしても，注視点の特徴ベクトル自体の検索精度が悪いためと考えられる．

表 2．トップダウン選択注視による検索精度

番号	TD 無し	TD 有り	精度の差
1	31.9	38.9	+7.03
21	18.2	40.1	+21.9
27	16.9	30.1	+13.2
31	31.2	31.5	+0.30
43	8.40	16.6	+8.15
51	6.45	9.50	+3.05
57	4.30	15.7	+11.4
71	0.40	34.0	+33.6
81	13.4	21.0	+7.55
100	12.9	53.5	+40.7
平均	14.4	29.1	+14.7

単位：%

5. 考察

本稿では，選択注視点における特徴ベクトルを，階層化 SOM でクラスタリングをして，その有効性を示した．階層化には，入力と下層勝者ユニットの選択性ベクトルとの残差を伝播する手法を用いた．これは，「珍しいものを検知する」という考え方である選択注視手法に，時間軸的な検知能力を加える拡張である．近年，はずれ値を検出する手法に注目が集まっている[11][12]．本研究では，学習器の階層化にあたって，学習済みの学習器にとってのはずれ値に着目した．

また本稿では，階層化によって得られる平均値ベクトルを重みとして扱い，トップダウンの選択注視の有効性を示した．これは，はずれ値検出とは逆に，「既知の部分」に着目する手法である．結果，検索精度の向上が示された．動画検索では，動き情報という際立った特徴量があるため，重要部位の選定がしやすい．しかし，静止画検索では，検索タスクにとって重要な特徴量が変わるため，重要部位の選定には，事前知識が必要になる．

今後の課題としては，ボトムアップ選択注視の改良が大きい．これは，適確なカテゴライズを得るためにも，入力ベクトルには独立な変数が必要なためである．また，現在の本提案手法において，画像中の方位情報は用いているが，より高次な特徴である，形状の情報を用いていない．ボトムアップ選択注視に形状の情報を含めることで，検索精度の向上が期待できる．

文　献

[1] C. Tsai, K. McGarry and J. Tait "Image Classification Using Hybrid Neural Networks," SIGIR'03 p.p.431-432,2003

[2] Y. Chen and J.Z. Wang, "Image Categorization by Learning and Reasoning with Regions," Journal of Machine Learning Research, vol. 5, p.p. 913-939, 2004

[3] 斎藤純，山名早人「選択注視を用いた画像検索システムの提案」信学技法 2006 年 2 月 23,24 日開催 PRMU(2006)

[4] L. Itti.; C. Koch;.and E. Niebur "A model of saliency-based visual attention for rapid scene analysis," Pattern Analysis and Machine Intelligence, IEEE Trans. on vol. 20, Issue 11, p.p.1254 - 1259 ,1998.

[5] L. Itti and P. Baldi "A Surprising Theory of Attention," In: Proc. IEEE Workshop on Applied Imagery and Pattern Recognition (AIPR) , 2004

[6] T．コホネン著，中谷和夫監訳，"自己組織化と連想記憶"，シュプリンガーフェアラーク東京,1993

[7] B.A. Olshausen and D.J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?," Vision Res, vol.37, no.23, pp.3311-3325 ,1997

[8] R.P.N. Rao and D.H. Ballard, " Predictive Coding in the Visual Cortex : a functional interpretation of some extra-classical receptive-field effects" Nature Neuroscience, vol.2, no.1, pp.79-87, 1999

[9] N.T. Mundhenk, V. Navalpakkam, H. Makaliwe, S. Vasudevan and L.Itti, "Biologically inspired feature-based categorization of objects," Proceedings of the SPIE,Human Vision and Electronic Imaging IX (HVEI04),vol 5292, pp. 330-341, 2004

[10] S.A. Nene and S.K. Nayar and H. Murase "Columbia Object Image Library (COIL-100)" Technical Report CUCS-006-96, 1996

[11] 井出剛，鹿島久嗣「Web 系システムからの特徴抽出とオンライン障害検知手法」Proceedings the seventh workshop on information-based induction Science(IBIS), p.p.7-14

[12] 丸山祐子，山西健司「動的モデル選択とそのセキュリティ，障害検知手法」Proceedings the seventh workshop on information-based induction Science(IBIS), p.p.15-22

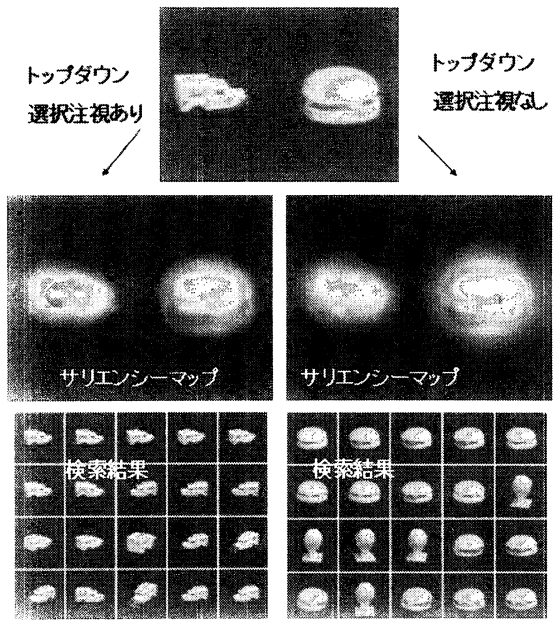

図 3. トップダウン選択注視．上がクエリ画像である．左の列は，ボトムアップ選択注視において，クエリの右のオブジェクト（トラック）を取り出すバイアスをかけ，注視点を選定し，その周辺情報を用いてデータベースから画像を検索した結果である．検索結果は，左上から右へ，類似度の順位が高くなっている．右の列がバイアスをかけずに同様の検索を行った結果である．