

人間の認知特性を利用した画像検索システム

尾 田 政 臣

A T R 視 聴 覚 機 構 研 究 所

従来の画像検索システムでは、検索のための適当な言語キーが思い付かない、検索対象のイメージがはっきりしていない等の点から難しさがあつた。そこでこれらの困難さを取り除こうとする研究が行われている。ところが、それらの方法では検索イメージが検索前にしっかり固まっていることを前提としていた。本研究は、検索前には漠然としたイメージしか無い場合についての画像検索方法を提案する。筆者は、画像概念の形成過程がそれまでの刺激の系列の処理過程に影響されることを心理実験により明らかにし、その結果を利用することにより効率的な検索システムを構成できることを提案した。本稿では、提案するシステムの有効性について述べる。

An Image Retrieval System that uses Human Cognitive Properties

Masaomi ODA

ATR Auditory and Visual Perception Research Laboratories

2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

With many conventional image retrieval systems, it is difficult to select the appropriate verbal key or to have a very clear target image. There have been some attempts to reduce the difficulty of retrieving image data. However, these studies assumed that the images to be retrieved are concrete before retrieval. In the previous paper, context dependency in the successive presentation of image stimuli was studied to examine the influence of prior processes. Based on this result, a new retrieval mechanism for an image database is proposed. In this paper it is shown that the new mechanism is very effective.

1. まえがき

画像データベースの検索方式は言語的な検索キーを用いる方式と画像キーを用いる方式に分類できる。従来の画像データベースでは言語キーを用いていたが、その方式では正確に画像内容が表現できない場合や、適当な言語キーが見つからない場合がある等の欠点があった。そこで近年画像を直接検索キーとして用いる方法が研究されている [1],[2]。しかし画像キーを用いる方法においても種々の問題がある。

即ち、検索すべき画像または類似の画像が既到手元でありそれを読み込ませて画像に付随する情報またはより詳細な画像を得るような使い方では、画像から特徴抽出とパターンマッチングをどのように行うかが課題となる。それが解決できるような使い方では非常に有効なシステムとすることができるでしょう。

しかし、検索すべき画像が手元にな場合いには検索者自身が持っているイメージを絵として表現する必要が生じる。ここでは検索者の描画能力が課題になる。そこでシステムとしてはできるだけ描かれた絵から検索者の意図を汲み取るかが課題になる。この課題に対する取組が文献 [2] 等でなされているが本質的に非常に難しい課題である。

さらに検索者自身のイメージが明確でない場合やそもそも検索当初にはイメージがなくイメージを作るための道具として画像データベースを利用すると言う場合が想定される。このような場合の検索方式については文献 [3] のように思考支援を目指す研究があるものの従来余り議論されていない。

筆者はこのような課題に対し画像概念形成過程が文脈によって影響されることを心理実験により示すと共にその結果を画像データベースの方式として応用することを提案した [4]。本稿では提案した方式が有効かどうかを確認するため、顔画像データベースを想定し実験システムを構築し評価した。その結果良好な結果を得たので報告する。

2. 文脈依存性とその応用

従来の画像データベースの研究では検索者の検索対象が検索前にははっきりしている場合を扱ってきた。例えば地図データベースの中から特定の建物、場所等を特定するとか、部品データベースの中から特定の性能を有する部品を選び出す、といった場合である。このとき検索者の求める検索対象ははっきりしており検索対象の特徴や分類名あるいは性能等を検索キーまたは部分画として入力し検索することになる。現在行われている研究はこれを如何に効率的に行わせるかというものである。

しかし、検索者自身にも検索対象があいまいな状態で行うデータベース検索課題がある。例えば、被服商品のカタログデータベースの中から好みの商品を捜し出す場合を考えて見よう。このような状況では検索者は自分の好みについての漠然とした傾向は認識しているものの明解に言語化して示すことができない場合が多い。そのような状況においても、サンプルを見せられれば好きか嫌いかの判断は可能である。この場合、どのような商品がデータベースの中に存在するかも予め知らない訳であるから、従来のデータベースのようにキーによる検索方式では検索者が適当なキーを考えて検索しなければならず大きな負担となる。結局次々に画像を検索し、好みに合った画像が出てくるとマークしておき最後まで検索が終了すると、マークしたデータを再度呼び出し比較しながら最終的に最も望ましい画像を選択することになるだろう。

これでは非常に多くの画像を呼び出す必要があることと、現在の画像の検索表示速度がそれほど高速には行えないことから、非効率であり何らかの高速化の方法が求められる。

ところで、上記のような画像検索問題を考えると、検索者が種々の画像を見て行くうちに好みのデザインの個々の特徴が形成され、最終的にどのような商品が求めている物かを決定して行く過程と見ることができる。この過程はまさにイメージ概念を形成して行く過程と捉えられるのではないだろうか。

ところで、概念の研究は心理学の主要テーマとして古くから研究されてきてはいる [5] が、概念形成時の動的な特性については調べられていなかった。文

献 [4] では複数の図形を同時に見せられ概念を次々に形成するよう指示された実験で、被験者はそれまでの図形に影響されながら概念を形成することを示した。ここでは概念形成時のこのような特性を画像検索システムに応用することを考える。

これまでの画像検索方の改善を目指す方法として、言語的な連想キーを用いる方法 [6]、例示画に基づく方法 [7]、[8] 等がある。しかし、いずれも検索者の検索対象の明確な意識を必要としている。検索者が意識しないところから検索のための情報が得られればより効率的で、ユーザフレンドリなインタフェースが実現できる。

人間が複数の画像を見た場合、それぞれの画像は独立に処理されるのではなく、以前の処理結果が次の処理に影響するのであるから、検索時においても処理した結果の系列に検索にかかわる情報が含まれることが示唆される。検索系列画像に共通する情報が検索者が注目している情報であり、これをフィードバックし、次の入力情報と共に利用することが効率的な検索を促進させるであろう。

3. システム実現方式

3.1. システム概要

貴方にとって好みの顔はどんな顔と言われたときに明確に言語表現することは難しい。色々な顔を思い浮かべその中から選び出すであろう。具体的な顔を見せられると好みか否かは明確に判断できる。即ち人間はこれまでに見た色々な顔を思い出しながら好みの顔を形成していると言える。本システムは、上記の現象を利用し、このイメージ形成過程を支援すると共にイメージの検索の促進を図る。即ち、画像データベースに格納された顔画像の中から、検索者にとって最も好みの顔画像を選び出す事を支援するシステムである。

本システムは、上記のイメージ検索支援方式を用いた場合と、網羅的にイメージを検索した場合にどの程度の検索時間、検索効率等に差が生じるかを検証する。

実用上のシステムでは検索開始時には、部分的に

も検索すべきイメージが明確な場合にはマウスでキー入力等で直接候補を絞り込めるようにすることが望まれる。これにより検索候補の絞り込みが行われ検索時間の削減が図られる。しかし、本システムでは全く言語を使わず画像の選択だけで実験した。

3.2 ユーザインタフェース

初期メニューで画像検索を選ぶと画面にシステムがランダムに選んだデータが10個表示される。検索者はその中で好みの顔を好きな数だけマウスでクリックする。その後、次の画像の表示を要求するとシステムは選ばれた顔画像から類似の特徴が何かを調べる。次にこの特徴を持つものでこれまで選んだ画像の平均値に近い顔画像をデータベースの中から検索し検索者に提示する。即ちこれまでに選ばれた画像のプロトタイプ的な類似画像を選び出す。この過程を繰り返すことにより検索者は理想のイメージを簡単に固めてゆくことができる。選ばれた画像はシステムが管理し、いつでも検索者が参照可能としている。管理している数は、最大10個までであり、これ以上選ばれた場合には古い方からあふれた分を捨てる。

上記の方法で理想の画像が求められるが、検索途中でイメージの変更の可能性がある。それに対処するメカニズムが必要である。ここでは10個提示する顔画像のうち2個を全くランダムな画像を提示することにした。これまでに選んだ画像の類似画像では満足できない場合や途中で気が変わった場合には、ランダムに選ばれた2個の画像の中から選択すれば良い。また、理想の画像が絞り込まれた段階で初期に選んだ画像では物足りなくなり、求める理想の画像の質が高くなることが予想される。また、画像検索の途中でこれまで固めつつ合ったイメージでは満足できず最初からやり直したくなる場合が想定される。このように、これまでに選んだ画像の再評価メカニズムが必要であろう。この目的のために選択済の画像からイメージからずれた画像を任意の数だけ選んで捨てるメカニズムを入れた。検索者は検索の途中で随時これまでに選んだ画像を呼び出し再評価できる。さらに、本目的のためにいちいち選択済

の画像を表示し評価しなくても良いように、これまでに選ばれた画像の中から常にランダムに1個の画像を選び次候補の画像として提示し再評価させることとした。即ち、10個提示される画像のうち7個はそれまでに選んだ画像と類似性の高い画像を、2個は全くランダムに選ばれたもの、残りの1個はそれまでに選ばれたものの中から再表示する。これらにより検索の集束を速めている。

3.3 画像管理情報

本システムではシステムの有効性を確認するための実験システムである。このため顔画像としてポンチ絵を用いた。顔画像の特徴パラメータを以下のように定めた。

- 1.顔の形状 (面長、標準、三角)
- 2.眉の位置 (下、中、上)
- 3.眉の形状 (下、中、上)
- 4.目の位置 (下、中、上)
- 5.目の形状 (小、中、大)
- 6.鼻の位置 (下、中、上)
- 7.鼻の形状 (小、中、大)
- 8.口の位置 (下、中、上)
- 9.口の形状 (小、中、大)
- 10.耳の位置 (下、中、上)

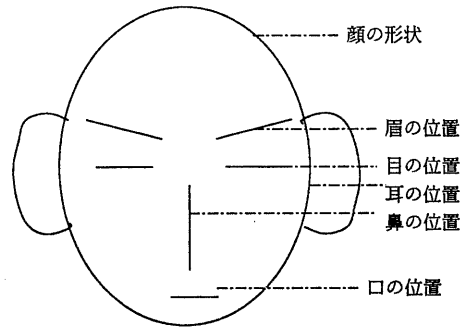
パラメータは下を-1、中を0、上を1と言うように()内左端の値を-1、中央の値を0、右端の値を1とした。実用システムでは目的に合わせて特徴パラメータを定義する必要があり、またパラメータ値も本システムでは3値を用いたが連続値を用いた方が有効な場合もあるだろう。顔画像の例を図1に示す。

これらの画像は画像生成ツールによりパラメータとその値の総ての組み合わせをとり自動生成し、およそ50000個のデータとしてシステムに格納された。

3.4 画像検索方法

本システムでは、特徴パラメータとその値を指定することにより、これらの条件を満たす顔画像をデータベースの中から選択する。

検索者が選んだ画像を用いて次に提示する候補をシステムが選択するが、その方法として画像の類似



特徴パラメータ値

(顔の形状,眉の位置,眉の形状,目の位置,目の形状,
鼻の位置,鼻の形状,口の位置,口の形状,耳の位置)
(0, 0, 1, 0, 0, - 1, 1, - 1, - 1, 0)

図1. 顔画像の例

性を用いた。まず選択された画像の各特徴パラメータの各々についてパラメータ値の平均値 \bar{x}_i ($i = 1, 2, \dots, N$) を求める。次に任意の画像 j を取り出し各特徴値の平均値との差の自乗で定義された類似量 $S_{ij} = (\bar{x}_i - x_{ij})^2$ とその総和である総合類似量 S_j を計算する。

$$S_j = \sum \alpha_i (\bar{x}_i - x_{ij})^2 \quad i = 1, 2, \dots, N$$

α_i : 重み係数

\bar{x}_i : 選択された画像の特徴パラメータ値の平均値

x_{ij} : 任意の画像 j の特徴パラメータ i の値

N : 特徴パラメータの数 (10)。

S_j の値が最も小さい顔画像データが、平均値に最も近いデータと見なす。

ただし、データの検索過程では計算量の低減を図るため次の方法をとる。

特徴の平均値 \bar{x}_i の値を求めその中で小さいものから順に3個の i を求める。それを検索者が最も注目しているパラメータと考える。この特徴を持つデータを検索の第一候補とする。

重み係数 α_i は今回の実験では1に固定した。

4. 検索実験結果と考察

4.1 検索実験結果

文脈依存性のメカニズムを取り入れた場合と、全くランダムに10個づつ顔画像を提示する方法で比較した。文脈依存のメカニズムを使ったシステムでは個人差があるものの平均約200個程度の顔画像を見るだけで満足のゆく顔画像が検索できる。これに比べ文脈のメカニズムを使っていないシステムでは平均約400個程度の検索が必要である。被験者Iの結果を表1及び2に示す。表1は、文脈依存メカニズムがない場合、表2はある場合の最終結果であり対応する顔画像を図2の(b)(C)に示す。ここで F_i は各パラメータのそれまでに選択された顔画像(個数 M ;最大10個)の平均値と各画像の当該パラメータ値との自乗誤差の総和である。

$$F_i = \sum (\bar{x}_i - x_{ij})^2 \quad j = 1, 2, \dots, M$$

さらに総合特徴量 F を計算する。

$$F = \sum F_i \quad i = 1, 2, \dots, N$$

評価値は文脈依存を使用していない場合が $F = 42.7$ 、使用している場合が $F = 10.3$ となり文脈依存のメカニズムがいかに類似度の高い顔画像の検索に有効かが分かる。即ち、理想の画像が固まったことを示している。総てのパラメータとその値の組み合わせをとり画像を生成しているため、格納画像は総て異なり、最も似た画像でもパラメータ値はどこか一か所が1だけ異なる。好みの画像が10個選択された最も評価値の良好な場合の理論値は、 $F = 8.1$ である。表2の結果は10.3であり非常に良い結果であることが分かる。また、図2の(b)と(c)を比較してもその違いが分かる。

表1、2における「検索画像数」は何個の画像を検索し眺めたかを示している。一度に表示する画像数が10であるから、検索画像数を10で割った数が検索回数である。文脈依存性メカニズム無の場合は42回、有りの場合は8回の検索回数となっておりこの点からも検索の容易性が分かる。「選択画像数」は好みの画像として選んだ画像数を示し文脈依存性メカニズムが無の場合は13回、有りの場合は18回とない場合の方が少なくなっている。最終的に10個の画像を選んでいるので選択画像数から10を引いた値が再評価の後捨てた画像数になる。文脈依存メカニズムがある

方が再評価を多く行ったことを示している。これは好みの画像を絞り込むとさらに類似の画像が表示される特性が利いた結果である。表1、2に示した検索者の場合には検索時間は、文脈依存メカニズムがない場合が5分、ある場合が2分半程であり検索結果が良好なばかりではなく検索時間の点でも文脈依存メカニズムが有効に働いた。ただし、他の検索者の例では逆に文脈依存メカニズムがある方が検索時間が長い場合もある。これは検索が進むにつれてイメージが固まって行き、検索結果に対する評価が厳しくなり再評価を繰り返すからであると思われる。

検索例は今のところ少ないものの、検索者の検索途中の心理的变化に対応するために設けた、ランダムに選んだ2個の候補の表示メカニズムは、有効に働いているようである。使用頻度はそれほど多くはないが検索者はランダムの中から選択した後はそれに近い画像を引き続き選ぶ傾向にある。それまでに選ばれた画像の中から1個をランダムに選んで再表示するメカニズムについても、検索開始当初は再選択されないことが多く有効に働いている。しかし、これらについては被験者がどの程度心理的なふらつきがあるかに関わっており、どのような画像を対象としたデータベースかに依存した問題であろう。したがって、システム毎に再表示個数の最適値は異なるであろう。

4.2 考察

本検索方式は文献[2]の分類になじみにくいが、敢えて言えば画像画像型主観的基準に基づく検索方式であろう。文献[2]によれば、利用者毎の主観的類似度空間を作成し、その空間上で提示されたサンプルやスケッチ等のパターンとデータベース中の各パターンとのずれを評価すれば良い。そのためには、比較的少数のキーパターン間の類似度をシステムに示し、システムに物理的画像特徴のパラメータ空間上で、例示学習(統計的学習)をさせれば良く、学習によって、物理的画像特徴パラメータ空間から、主観的類似尺度を反映した主観的画像特徴パラメータ空間に変換する写像を構成できるとしている。今回のように検索者の検索対象イメージがはっきりして

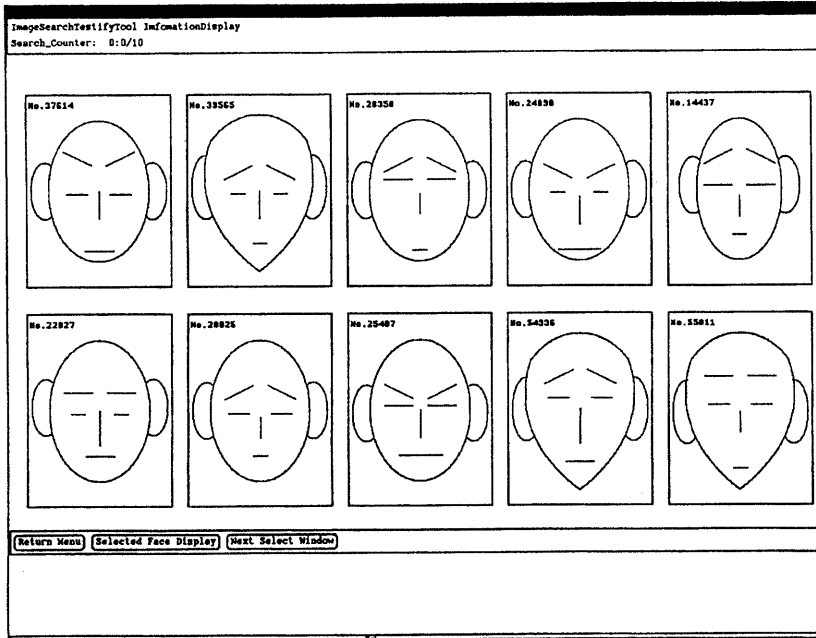


図 2 (a)
初期画面

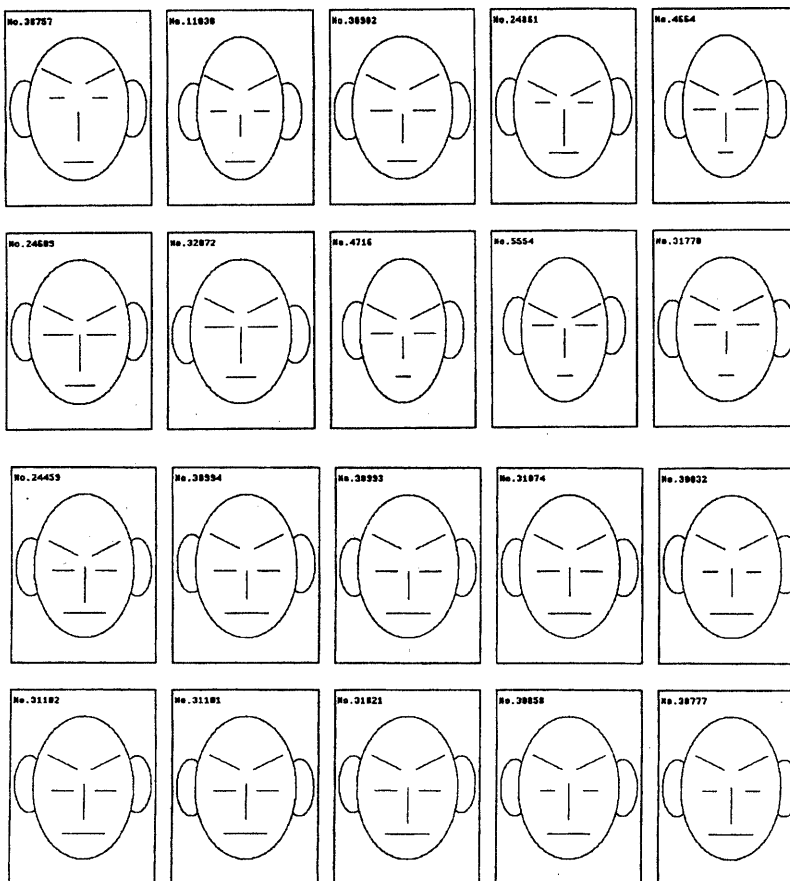


図 2 (b)
文脈依存メカニズム
が無い場合の検索結果

図 2 (c)
文脈依存メカニズム
が有る場合の検索結果

表1. 検索結果 文脈依存メカニズム無

| 顔番号 | 顔形状 | 眉位置 | 眉形状 | 目位置 | 目形状 | 鼻位置 | 鼻形状 | 口位置 | 口形状 | 耳位置 |
|-----------|------|------|-----|------|------|------|-----|-----|------|------|
| 38757 | 0 | 1 | 1 | 1 | -1 | 0 | 0 | 0 | 0 | -1 |
| 11030 | -1 | 0 | 1 | -1 | -1 | 0 | -1 | 0 | 0 | 1 |
| 30902 | 0 | 0 | 1 | -1 | 0 | -1 | 0 | 0 | 0 | 1 |
| 24861 | 0 | -1 | 1 | 0 | -1 | -1 | 1 | 1 | 0 | -1 |
| 4664 | -1 | -1 | 1 | -1 | 0 | -1 | 0 | 1 | -1 | 1 |
| 24609 | 0 | -1 | 1 | -1 | 1 | -1 | 1 | 0 | 0 | -1 |
| 32072 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 4716 | -1 | -1 | 1 | -1 | 0 | 0 | -1 | 1 | -1 | -1 |
| 5554 | -1 | -1 | 1 | 0 | 0 | 1 | 0 | 1 | -1 | 0 |
| 31770 | 0 | 0 | 1 | 0 | 0 | 1 | -1 | 1 | -1 | -1 |
| x_i 平均値 | -0.4 | -0.4 | 1 | -0.4 | -0.1 | -0.1 | 0 | 0.6 | -0.4 | -0.1 |
| F_i | 2.4 | 4.4 | 0.0 | 4.4 | 4.9 | 6.9 | 6.0 | 2.4 | 2.4 | 8.9 |

F = 42.70 検索画像数 420 選択画像数 13

表2. 検索結果 文脈依存メカニズム有

| 顔番号 | 顔形状 | 眉位置 | 眉形状 | 目位置 | 目形状 | 鼻位置 | 鼻形状 | 口位置 | 口形状 | 耳位置 |
|-----------|-----|------|-----|------|------|-----|-----|-----|-----|------|
| 24459 | 0 | -1 | 1 | -1 | 0 | 0 | 1 | 1 | 1 | -1 |
| 30994 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 1 | 1 | 0 |
| 30993 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 1 | 1 | -1 |
| 31074 | 0 | 0 | 1 | -1 | 0 | 1 | 0 | 1 | 1 | -1 |
| 30832 | 0 | 0 | 1 | -1 | -1 | 1 | 0 | 1 | 1 | 0 |
| 31102 | 0 | 0 | 1 | -1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 31101 | 0 | 0 | 1 | -1 | 0 | 1 | 1 | 1 | 1 | -1 |
| 31021 | 0 | 0 | 1 | -1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 30858 | 0 | 0 | 1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 |
| 30777 | 0 | 0 | 1 | -1 | -1 | 0 | 1 | 1 | 1 | -1 |
| x_i 平均値 | 0.0 | -0.1 | 1.0 | -1.0 | -0.3 | 0.5 | 0.6 | 1.0 | 1.0 | -0.6 |
| F_i | 0.0 | 0.9 | 0.0 | 0.0 | 2.1 | 2.5 | 2.4 | 0.0 | 0.0 | 2.4 |

F = 10.30 検索画像数 80 選択画像数 18

いない場合の検索システムでは、この考えは適用出来ないであろう。まず主観的基準は多くの場合しっかりしたものではない。心理実験で示されたように、主観的基準は次々に見た画像の影響を受けるので検索の最中に変動するものと考えなければならない。また、マッピング法則の学習は検索者が限定している場合には有効でも不特定多数の検索者を想定するシステムでは余り有効ではないだろう。実システムの構築にあたってはデータベースの内容と、検索者の検索特徴を十分調べ、配慮する必要がある。本システムでは、検索途中で検索者自身に何度も評価させ、システムがそれに追従するようにすることでシステム構築を容易にすることができた。

人間は情報の処理結果に影響され次の処理するのであるから [4]、情報の提供順序をうまくコントロールすることにより情報検索の支援を行えるものと思われる。従って、あらかじめメニューモードを設定し検索者に各モードで適当な画像を選択させることにより最終目的の画像を絞り込むことができるであろう。ところが、モードのある方式の欠点は人間がモードを意識しなければならず煩わしいことであった。ユーザインターフェースの観点からはモードレスが望ましい。しかし検索対象が割合はっきりしていて対象がはっきり分類できるような場合は、メニューの利用がより効率的になるであろう。検索対象に対する検索時の癖、即ちどのような分類にまず注

目し、どのような分類系列の順序で画像を検索するかを調べておく。被服データベースの場合では、全体の形が先か配色が先かなどの情報を得ておく。または個人毎に検索順序が違うのであれば個人対応に画像データベース利用時にこれらの知識を獲得しておく。人間が概念を形成して行くように、選択される画像の系列から次に選ぶべき項目を推測しユーザーに提示すれば操作性が向上するであろう。

人間の検索する順序には法則性があり、利用する特徴は主要な物から段々末梢的なものになるように思われる。これらの点を調べ、提示順序を人間の思考順序に合わせることにより人間の振舞いに近いものが得られ、さらにユーザインタフェースの向上が図れるであろう。今回はこの方式は採用せず図形を選択だけでどの程度の性能かを確認した。

5. 結論

従来の画像検索方法は、検索者の検索意図が具体的でそのイメージもはっきりしている場合に有効である。しかし、検索意図が検索途中で変化する場合や、設計者の予期せぬ選択戦略をとる検索者にはシステム構成が堅く融通がきかない使いにくいシステムになる。本システムでは文脈に依存させながら次の候補を選ばせると言うように、変更の可能性を取り入れ再評価できるようにした。このため検索者にとって柔軟性のある使い勝手の良いシステムにすることができた。

本方式は従来の検索方式と融合することも可能である。例えば検索意図がある程度しっかりしている場合は候補の絞り込みの段階までは言語的キーまたはメニューにより候補を絞り込ませた方が有効であろう。その後の詳細選択のところでは本方式を用いれば良い。

今回の実験では簡単な類似度計算を用いたがそれでも満足行く結果を得た。実システムでは状況がもっと複雑になるので工夫が必要になるだろう。現実場面での人間の概念形成過程の行動パターンや心理を取り入れればもっと使い勝手の良いシステムが構成できると思われる。

本システムでは好みの顔画像を検索する場合を想定していたが、好みの顔を顔画像を犯人に最も近い顔画像と置き換えることも可能である。ただし、この場合にはシステムを直接目撃者に使用させてはいけない。検索しているうちに検索途中に表示された画像に影響されて犯人のイメージが変わってしまう恐れがあるからである。したがって目撃者の証言を基に第3者が代行検索し結果を示す必要がある。

謝辞

本研究を行う機会を与えて下さったATR視聴覚機構研究所淀川英司社長並びに上野圭一室長に謝意を表します。

参考文献

- [1] 坂内；画像検索技術、信学誌、1988, Vol. 71, No. 9, 911-914.
- [2] 加藤；類似検索機能を持った画像データベースとその設計法、O plus E, 1991, 9月、10月.
- [3] 笠原、岸本；画像データベースナビゲーション、信学技法、1988, IE88-46, 71-78.
- [4] 尾田；図形概念形成過程の文脈依存性とその応用、情報処理学会ヒューマンインタフェース研究会、1991, 35-2, 9-16.
- [5] Medin, D. L., Smith, E. E. Concepts and Concept Formation. Ann. Rev. Psychol. 1984, 35, 113-138.
- [6] 柴田、井上；画像データベースの連想検索方式、信学論D-II, 1990, Vol. J73-D-II, 526-534.
- [7] 栗田、大垣、加藤；主観的類似度に適応した画像検索、情報処理学会論文誌、1990, Vol. 31, No. 2, 227-236.
- [8] 田邊、大谷；形状類似画像における類似尺度の検討、信学技法、1988, PRU88-68, 65-72.