

略画作成による画像検索システムのインタフェース

西山 晴彦 饗場 潔 横山 光男 松下 温

慶應義塾大学理工学部計測工学科

最近、メモリの大容量化・低廉化によって、大量の画像蓄積が可能となり、画像データベースシステムも研究され、開発されるようになった。そこで、非熟練者のために簡単に効果的にこれらの情報にアクセスする方法の実現が必要である。

本稿では、曖昧検索モデルの体系化が非常に求められている画像検索システムについての提案を行う。まず人間の思考に合った画像表現の分析を行い、画像表現における曖昧性をどのように表せば、人間の画像検索に対する曖昧要求処理が我々の感覚通りに行われるかを考察する。そして、人間指向型の画像表現モデルを提案し、これに基づく試作システムを紹介する。

An Interface for Image Retrieval depending on Sketch

Haruhiko NISHIYAMA Kiyoshi AIBA
Teruo YOKOYAMA Yutaka MATSUSHITA

Department of Instrumentation, Faculty of Science and Technology, Keio University

3-14-1 Hiyoshi Kohoku Yokohama 223, JAPAN

Recent technology developments have now made it feasible to store large numbers of pictures or other graphic materials in computer database. So, providing easy and effective access to such information for non-specialists will be important for the success of these new systems.

In this paper, we propose about image database that strongly required systematizing of ambiguity retrieving model. First, we analyze that image expression suited for human consideration. Then, we propose human oriented image expression model and introduce the implementation of an experimental system based on this model.

1 はじめに

最近、メモリの大容量化・低廉化によって、大量の画像蓄積が可能となり、画像データベースシステムも研究され、開発されるようになった。[1-7] そこで、非熟練者のために簡単に効果的にこれらの情報にアクセスする方法の実現が必要である。しかしながら、画像は言葉や数値と違い、そもそも正確な検索が困難なものであるから、画像情報の検索手法としてはこれといったものがなかった。画像そのものが多義的であり、画像に対する解釈が検索者によって異なるからであろう。

このため、曖昧検索が可能なデータモデルを確立する必要がある。従来の画像データベースは、文字・数値型のデータベースと同様に、システムに扱いやすい文字・数値型指向の画像管理情報や質問言語をコミュニケーションのメディアとする、いわば“人間が計算機に歩み寄る関係”であった。次世代の画像データベースでは、画像や主観的表現もコミュニケーションのメディアにする必要がある。いわば、“計算機から人間に歩み寄る関係”が求められるのである。当然のことながら、人間の感覚のうち、画像と最も関連するものは「視覚」であり、視覚を通して歩み寄るということは、画像検索を人間の知覚体系と結び付けることである。[8-10] つまり、問題は人間が見た画像をどのようにとらえ、それらを記憶の中でどのように蓄積しているか、にある。その蓄積している記憶を表現しやすいインタフェースを構築し、人間の主観をある程度システムの側が汲み取ることが可能となれば、有効な画像検索システムとなりうる。

そこで、本稿では、曖昧検索モデルの体系化が非常に求められている画像検索システムについての提案を行う。まず人間の思考に合った画像表現の分析を行い、画像表現における曖昧性をどのように表せば、人間の画像検索に対する曖昧要求処理が我々の感覚通りに行われるかを考察する。そして、人間指向型の画像表現モデルを提案し、これに基づく試作システムを紹介する。

我々のシステムの検索方法は、人間の記憶に残っている曖昧な部分を効果的に使うために、画像中の海・山・空などの領域情報、人・物などのオブジェクト情報、それぞれのオブジェクトに対する属性の3つを用いている。

具体的に示すと、領域情報については、海や空・山などの背景を表す多角形を、その領域を代表するような色で塗る。オブジェクトについては、人・建物など検索したい画像に含まれるオブジェクトに対応するアイコンを配置する。属性はアイコンに対して付加的に

指定することができる。

このように、領域を表す多角形と、オブジェクトごとのアイコンを用いて、ユーザは自分の記憶に合った略画を簡単に作成できるのである。

また、データベースの場合、その操作のしやすさにも注意しなければならない。我々のシステムでは、操作はすべてマウスによって、ディスプレイ上で行うことができ、システムを使う上で特殊な能力を必要とはしていない。アイコン自体の属性の指定もすべてボタンによる選択性であり、自信がなければ選択しなくてもよいようになっている。

2 提案する画像表現モデル

2.1 従来の画像検索モデル

画像検索を人間の感覚に合わせて潤滑に行うためには、実際に我々が手作業で画像を探すときにどこを見て何をキーにするのかを分析するべきである。その分析の中から、我々の頭の中での画像表現モデルを抽出し、それに適応したモデルで検索を行うことができれば、検索要求に曖昧性があったとしても大きく我々の感覚からはずれた検索が行われることはないであろう。

従来の画像検索モデルは、次の3つのアプローチから考えられていた。

1. 各画像データに整理番号やインデックスをつける方式
2. 各画像データに関連のあるキーワードをつける方式
3. 認知科学的に画像を解析してマッチングする方式

1のインデックス法は、明らかにデータベース技術にそのまま画像データを載せたにすぎない。2のキーワード法は、確かに画像の持つ意味的な情報による検索を可能にしているので、それほど人間から計算機に歩み寄ってはいないと思われる。また、キーワード検索に基づくシステムでも、かなり曖昧な感覚で検索が可能なのが提案されている。しかし、そもそもがインデックス法の応用として考えられた方法で、これも人間の思考に適応しているとは言い難い。[11] というのも、人間は画像をキーワードのような言語的な認識もするが、決してそれだけではなく、意味的な認識と感覚的な認識をうまく組み合わせているからである。そして最後の3のアプローチとしては、認知科学の技術の進歩にともなってさまざまなアプリケーションが作られている。これは、人間が見分けるのは厄介な小さな違

いのある、膨大なデータからの検索に効果を発揮している。気象庁の天気図データベースや、医療機関のレントゲン写真データベースなどがその例である。しかし、正確なマッチングに基づいているので、曖昧な記憶を手がかりに所望の画像を探すという観点では、まだ研究段階である。

我々のアプローチにおいては、人間の力で見分けのつかないような小さな違いを認知科学技術で解析しようというアプローチとは正反対に、人間が画像を見分けるときの方法を見習って画像表現モデルを作る。

2.2 人間の思考にあった画像表現

検索モデルを人間の思考に近づける目的を次に挙げる。

1. 検索要求を表現しやすくする。
(人間の思考に似たモデルならば、思った通りに検索要求を表現できる)
2. ユーザのモデルに対する誤解が防げる。
(システム設計者の意図と検索者の間の認識のズレを最小限にできる)
3. 曖昧な検索要求に対して期待通りの曖昧処理を行える。
(ユーザの質問のどこが曖昧なのかということシステムが容易に理解できる)
4. 人間の頭の中で画像がどのように理解されているかを理論的に体系化できれば、あらゆる画像を統一的に処理することができる。
(システムはあらゆる種類の画像に対して同じように対応できる)

つまり、曖昧検索を行うためには、人間の曖昧な思考自体を理解する必要がある、ということである。では、人間の頭の中ではいかにして画像が認識されているのであろうか。

データベースで扱われる各種の画像情報には、さまざまな種類のものがあるが、それらはいずれも表現されている対象物の形や色・濃度や相互位置関係、相互接続関係、分布、さらには美的内容などが情報の中身として意味を持っており、「パターン情報」と総称される。これからは画像情報に、このような「パターン情報」的要素を持ち、もっとも身近なものである絵画を用いて話を進めていく。

さて、図1に示す絵を見たとき、我々が得る情報はこういったものだろうか。また、それがどう表現され、こういったところに曖昧性が現れるのだろうか。

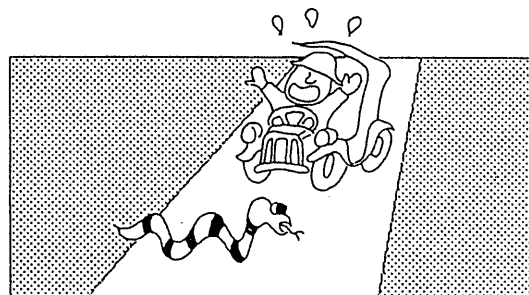


図1: 車と蛇

まず、この絵から得られる情報をピックアップしてみる。

- 道路と、その両わきに芝生がある。
- 道路に、車と蛇が描かれている。
- 車には人が乗っている。
- 車は中央上寄りに描かれている。
- 蛇は車のやや左下に描かれている。
- 車は白い色である。
- 蛇は縞模様である。

このように、我々は1つの絵をながめたときに、さまざまな情報を得る可能性がある。これらの情報のうち、どの情報を覚えていて検索要求を出すかといったことは、人それぞれなので定式化することは難しい。人によっては車と蛇の位置が印象的かもしれないし、車の色や形が印象に残るかもしれない。

絵などの画像をとらえる際に、人間は2つのパターンを持つ。1つは、車や蛇といった、画像の特定のオブジェクトに注目するパターンであり、もう1つは画像全体を「概視」することである。絵を瞬だけ見るような場合には、おそらくほとんどが後者のとらえ方をするはずである。また、この2つの注目のパターンは、我々が画像を記憶するパターンに直接つながるものである。このことは、我々が「見たことのある有名な絵画を1枚、頭の中に思い浮かべて下さい。」といわれたとき、(それは「種をまく人」かもしれないし、「落ち穂拾い」かもしれないが) 空と地面などがどのように塗り分けられていたか、そこに出てくる人物などのオブジェクトはどのような位置にあったか、などを想像することからも明かなことである。

そこで、曖昧検索モデルは、こういったすべての印象に対して検索要求が出せるようなモデルでなければならない。

画像表現モデルを構築するために、図1から得られる情報というものを一般化させると、次のように表現できるだろう。

1. 描かれている場所
2. 画面構成
3. 各オブジェクトの位置関係
4. 目立つオブジェクトの様子

このように一般的な絵は、我々の頭の中では、ほとんど上記の4つに分類される情報で表現されるであろう。

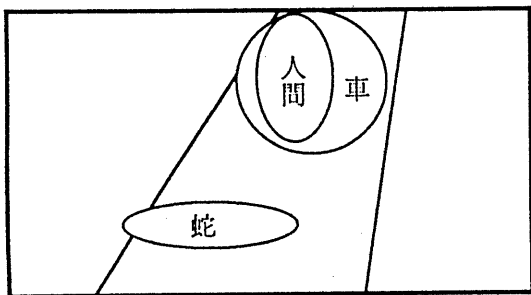


図2: 「車と蛇」の簡略図

先ほどの図1「車と蛇」から得られる情報を簡略化して表すと、図2のようになる。この簡略図が表している情報は、

1. 屋外である
2. 真ん中に道路があり、その両わきに芝生がある
3. 道路に2つの物が置いてあり、その位置関係がわかる

といったことであろう。意外にもこれほど単純な簡略図でも、かなりの情報量を持っていることがわかる。この他に、各オブジェクトが何であるかを示し（たとえば下方のオブジェクトは蛇であるとか）、他の細かい部分（車に乗っている人）を配置すれば、ほぼ我々の頭の中に描かれる画像表現と重なるだろう。

しかし、簡略図にしたことにより失われ、しかも重要な情報がある。それは

4. 蛇は縞模様である

ということである。これは、図2のような簡略図からは得難い情報である。つまり、画面構成・オブジェクトの位置関係（レイアウト）・オブジェクトの様子（オブジェクトの属性）の3つを有機的につないで表現できる画像表現モデルができれば、すべての絵画の表現モデルを体系化できる可能性が生まれるのである。

2.3 領域・オブジェクトレイアウト・属性の3層表現による画像表現モデル

では、前節で行った画像表現の分析に基づき、領域・オブジェクトレイアウト・属性の3層表現による画像表現モデルを提案する。図3の(a)~(d)は、先ほどの「車と蛇」の絵を領域・オブジェクトレイアウト・属性の3層で表現したものである。

まず、絵全体の印象として、道路の部分と芝生の部分に分けて記憶がなされるだろう。これを領域で表すと、図3(a)のようになる。領域2は道路のグレー、領域1と3は芝生の緑色だったということが印象に残る。そしてそれ以上にこの絵で印象に残るのは、道路の上の蛇と車であろう。蛇の右上に車があったということも記憶に残る。(図3(b)) これらのオブジェクトの位置、そしてもし位置を正確に覚えていなかったとしても、それらのオブジェクトの位置関係は覚えているだろう。たとえば、上方の車の位置が大体どのくらいかを忘れていても、蛇よりは上だった、といったように記憶していることはよくある。また、領域とオブジェクトの双方の関係も記憶を助ける1つである。なぜなら、物体1は領域2にあったということは確実に覚えているだろう。つまり、図3(c)のように、領域とオブジェクトの関係を記述しておくことが重要なのである。そこで、どのオブジェクトがどの領域に含まれているかということも、検索のキーにする。

さらに、「縞模様の蛇」といった、オブジェクトの位置よりもそのオブジェクトの様子・特徴がキーになる場合がある。図3(d)の場合では、車が白であるとか、蛇が縞模様であるといったような属性を付与することにより、絵の個別化を助けている。

図4に、領域・オブジェクトレイアウト・属性の3層による画像表現モデルを示す。このモデルは、人間の画像に関する記憶を3層に分けて考えるものである。

1. 実際の画像を大ざっぱに考えて、いくつかの領域に分ける
2. それらのオブジェクトが、どのような位置関係で存在しているかを見る
3. 各オブジェクトに際だった特徴がないかを見る

以上のように、人間の画像を見るとき判断をモデル化してみた。モデル化の方法は、他にもいくつか考えられるかもしれないが、このモデルの特徴は、どの層で切るかによって、その画像の特徴がつかめることである。たとえば、領域層を主に使って構成される絵画は、風景画である。また、中央に大きなオブジェク

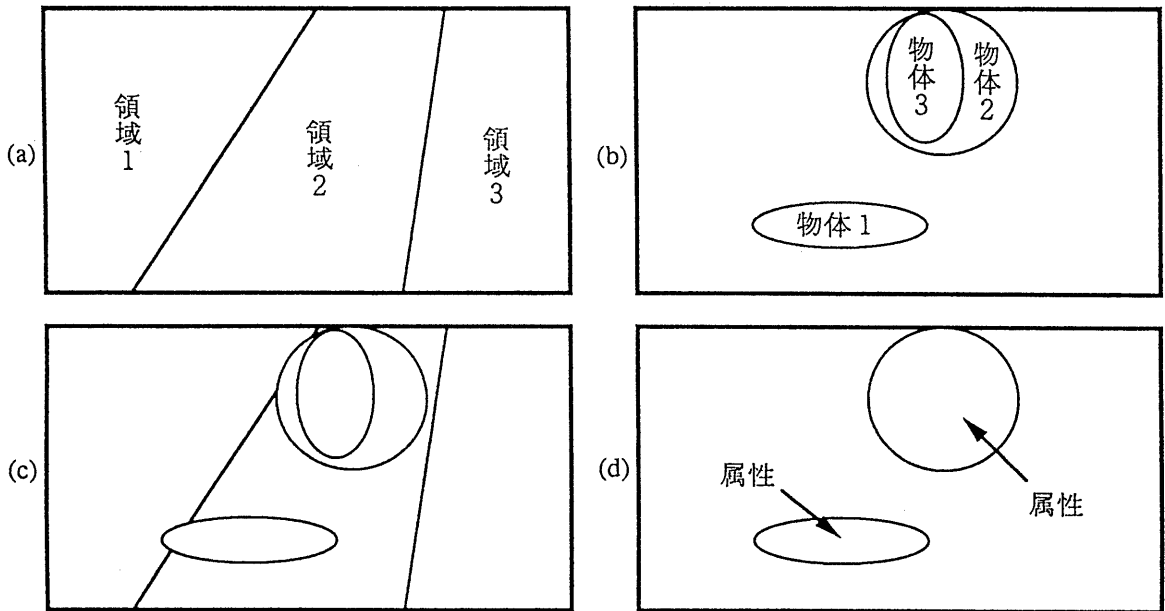


図 3: 「車と蛇」の領域・オブジェクトレイアウト・属性による表現

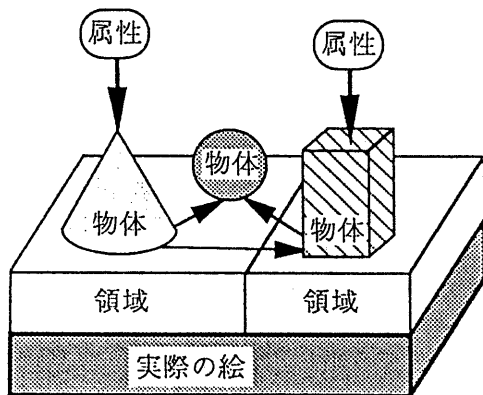


図 4: 領域・オブジェクトレイアウト・属性による画像表現モデル

トを置いて、それに細かい属性をつけていけば、それが人物画か静物画であることがわかる。

用いる層によって画像の特徴がつかめることが、なぜそれほど重要かという、これを用いれば自動インデックス化が図れるからである。オブジェクトの種類、領域の色などを考え合わせれば、かなり綿密に人間が実際に行うのと似通った分類を行うことができるであろう。

3 画像表現モデルの検索システムへの適用

提案する画像表現モデルは、人間が頭で画像を理解する過程に沿ったモデル化である。それゆえ、コンピュータのパターン認識ではこのような理解をすることは不可能である。

パターン認識的に画像を理解し、その抽出された輪郭をキーにして画像検索を行うシステムを「機械指向的理解システム」と呼ぶなら、我々の提案する画像表現モデルに基づく検索システムは「人間指向的理解システム」と呼ぶことができよう。さらに言うなら、前者は人間の網膜のレベルで画像を理解しようとしており、後者の提案モデルは、人間の脳のレベルまで持ち上げて画像を理解するアプローチである。そのため、前者では曖昧性の表現が困難であったのに対し、後者では曖昧性が実際に存在する人間の脳のレベルでモデル化をしているため、容易に曖昧性が表現可能である。

3.1 有効な画像検索形態

検索部分を考えるうえで我々が注目すべき部分は、人間の画像に対する記憶のパターンである。この記憶のパターンを、画像検索に結び付けられれば良いのではないかと考えたわけである。先述したとおり、画像をとらえる際に人間は2つの記憶のパターンを持つ。

まず一方のパターンは特定のオブジェクトに関するものであるが、人間が画像をとらえ、記憶するうえでは、もう一方の、全体像をつかむほうがこれに優先する。なぜならば、画像中の特定のオブジェクトに注目して記憶するということは、その画像内のオブジェクトに関しては既にその存在をはっきりと認識していなければ記憶はできない。つまりそのオブジェクトの存在を前提としているからである。全体像がまったくつかめていない画像について、前者の特定のオブジェクトに注目する記憶のパターンが先行することはまずありえない。また、人間が何度も見たりして、全体像がつかめた画像に関しては、その中の特定のオブジェクトに注目して記憶するはずである。そこで、このどちらのパターンも考慮する必要がある。

さて、我々が見たことのある絵画について頭の中で思い浮かべるとき、一枚のスクリーンの中を、その絵画が塗り分けられていたように塗り分け、その中に絵画の中に存在していた個体（オブジェクト）を、それがあった位置に配置するように思い描く。そして、そのオブジェクトに関して記憶していることがあれば、それを追加して思い浮かべることであろう。人間の記憶を重視した画像データベースを構築するためには、人間の記憶しているものをなるべく変形させることのないままにシステムに伝えられるようにしなければならない。そこで、記憶を表現するためにスクリーンを設け、この上に先述した画像表現モデルに基づく略画を作成する、というインターフェースを実現する。領域情報の入力は、スクリーン内を自由に塗り分けられるようにする。その中に存在するオブジェクトの配置を簡単に表現できるように、個々のオブジェクトを代表する簡略図形（アイコン）を設けて、それをスクリーン内の適当な場所に自由に配置できるようにする。そして更に、オブジェクトに関して細かい属性を設定することにより、検索候補をしぼることができる。

このように、領域、領域内のオブジェクト、オブジェクトの属性、と3段階式の検索ができれば、人間の記憶を柔軟に利用した画像検索システムができるであろう。

3.2 検索手法

領域情報、オブジェクト情報、属性情報のそれぞれに対して以下の方法によりマッチングを行う。一枚ずつ全てのデータ画像に関してマッチングし、似ていれば似ているなりの点数をその画像に与える。最後に、領域・オブジェクト・属性のマッチングで得た得点を合計し、最高点を得た画像から順に表示できるようにする。

3.2.1 領域によるマッチング

人間の画像に関する記憶は、オブジェクトやその位置関係などの意味のあるものに比重が置かれている。よって、領域に関しては、あくまでも“だいたいのどのよう塗り分けられていたか”の表現にとどまる。このため、領域に関するマッチングは、それなりに大まかにとらえることが必要である。ここでは、ユーザが描いた領域の形について細かく解析するのではなく、全体の塗り分けの分布がどのようになっているかを抽出する。

具体的には、ユーザが検索要求を出すスクリーンの縦方向と横方向に等間隔で8点、合計64点の抽出点を設け、その点の色を抽出して、データベースのデータの同じ位置の抽出点とマッチングする。このとき、色の空間をRGBの3次元空間と考え、データと検索要求の対応する点の色の距離に応じた得点を与えるようにする。ただし、領域の塗り残しの部分（白いままの場所）は、比較の対象とはしない。

3.2.2 オブジェクトによるマッチング

オブジェクトを表すアイコンは以下の6種類に分類してあり、同じカテゴリ内であれば検索要求をまちがえてもマッチングの対象とする（‘犬’と‘馬’は“動物”ということマッチングする）。



図 5: アイコンの分類

オブジェクトの位置(x,y座標)、相対的な位置(レイアウト)、色、大きさ、オブジェクトのカテゴリなどによりマッチングを行い、似ている絵には高い得点を与えるようにする。

ここで、レイアウトとはオブジェクトの相対的位置関係のことを指している。相対的位置関係の記述は、画像中に含まれる各オブジェクトを接点として、オブジェクト間の位置関係を枝としたグラフで表現している。

3.2.3 オブジェクトの属性によるマッチング

オブジェクトが“半身-男”・“半身-女”・“全身-男”・“全身-女”のときのみ、属性を指定できるようにした。この理由は、

- 肖像画のような絵画を検索する場合には、画面に登場するのは人物一人というものがほとんどで、オブジェクトやその位置関係による画像の絞り込みが難しいこと
- (肖像画などは) 注目すべきオブジェクトが少ないので、属性に関してよく記憶していること
- 他のアイコンはある程度属性の要素を含んでいること(例えば、‘風車’や‘教会’は“建物”の属性でもある)

などによる。

Man

Close

form: standing sitting lying

direction:

age: young old unknown

hair color: black brown gold unknown

beard: exist not exist unknown

hat: wear not wear unknown

cloth: wear nude unknown

his property: something nothing unknown

図 6: 属性の質問内容

システムは人物のアイコンの属性に関して図 6 に示す 8 個の質問を提示する。この属性設定の特徴は、「unknown」(わからない)という選択を許すことである。基本的には、選択した答えとデータベース中のデータが一致しているとマッチされるが、「unknown」ボタンを選択した場合には、無視する。

3.3 データ画像の登録

データベースを構築する際には、当然データを登録するという作業が必要である。このシステムでは、まずデータベース構築者が後述する検索要求と同様の手順により略画を作成し、システムがこの略画から自動的に領域・オブジェクト・属性の各情報を抽出する。これに図 7 の (g) のように作者名・題名などを加えて、データベースに登録する。よって、前節に述べたマッチングはデータベースの略画と、検索要求の略画を比較することになる。

3.4 システムの操作例

ここでは 2.2 で提案した基本概念を実現するための試作システムを紹介する。このシステムは、サンマイクロシステムのスパークステーション上の Sun View で実現した。画像データはルーブル美術館所蔵の絵 200 枚が登録されている。

1. 色の指定をする

マウスで使用したいパレットの色をクリックすると、その色が指定される。また、所望の色がパレット上にない場合には、パレットの下部にある RGB 3 色の slider bar を適当な位置にマウスで移動して、中間色を作ることもできる。なお、slider bar はそれぞれ 256 段階に設定できる。

2. 領域情報の入力をする

まず、ユーザの記憶における、検索したい絵画がどのように塗り分けされていたか、という「領域」の設定する。1 の要領に従って色を指定した後、マウスにより、略画作成キャンパス上で任意の多角形を描く。

3. オブジェクト情報の入力をする

領域と同じようにアイコンの色を指定し、マウスによりアイコンを図 7 のように略画作成キャンパス上に自由に配置する。複雑なレイアウトもこの方法により簡単に入力できる。

4. 属性情報の入力をする

略画作成キャンパス上のアイコンをマウスで指定する。そのアイコンは黒枠で囲まれて、注目していることが示される。またそのアイコンに対する属性指定パネルが表示されるので、それぞれの属性を決定する。

以上の要領で検索要求を出してボタン群の「Search」ボタンを押すと、3.2 に述べた基準により検索がなされる。検索終了後、最高得点を得たデータ画像の ID 番号がわかるので、これに対応する絵がレーザーディスクから図 7 の検索結果ウインドウに表示される。同時にその絵の作者、題名、得点、次候補、前候補ボタンも表示される。検索された絵が求めるものでないときは、これらのボタンを必要な回数だけ押せばよい。

4 まとめ

本稿では、人間の思考に合った画像表現の分析を行い、画像表現における曖昧性をどのように表せば、人

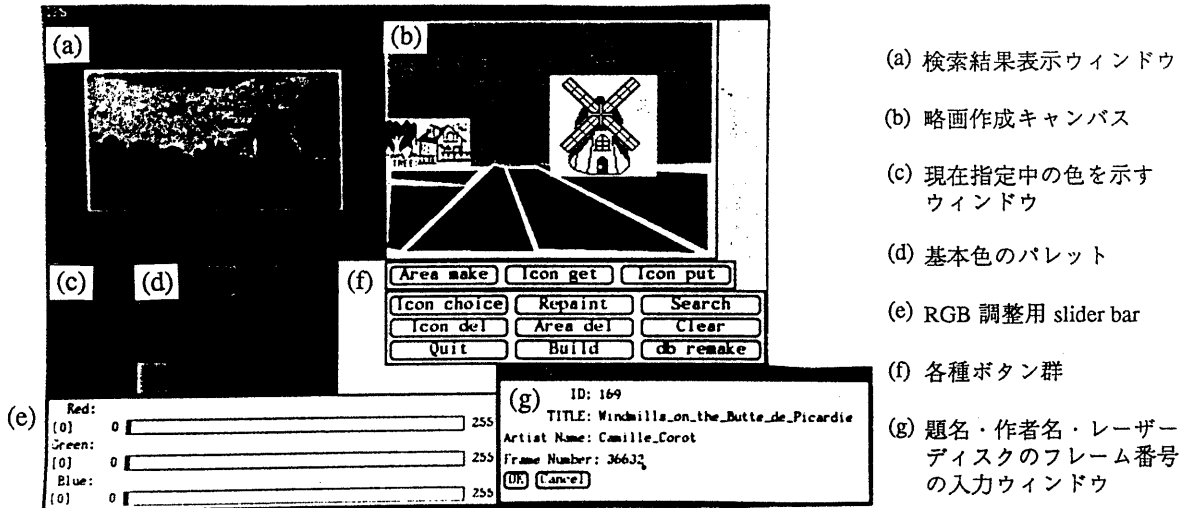


図 7: システムの画面

間の画像検索に対する曖昧要求処理が我々の感覚通りに行われるかを考察した。そして、人間指向型の画像表現モデルとして、領域・オブジェクトレイアウト・属性の3つの情報によるモデルを提案し、これに基づく画像検索システムを紹介した。これにより、従来の方法よりは人間の記憶に沿った検索要求を出せるようになった。

システムに登録してあるアイコンの種類は、絵画に良く登場するものを選択したものであるが、これらが最良のものであるとは言い難い。ごく少数の絵画にしか現れていないアイコンもあった。このようなことから、今後より一般的なアイコンを用意することが課題である。

参考文献

- [1] 坂内 大沢, 画像データベース, 昭晃堂
- [2] 藤田 古郡 畑, 視覚心理空間を利用した画像検索法, 情報処研資, CV-51s-4, pp1-8, 1987
- [3] 長嶋 土方, 人間の主観を重視した類似商標図形の検索の基礎的検討, 電子情報通信学会論文誌 D-2, Vol. J74-D-2, pp311-320, 1991
- [4] 山根 坂内, キーワード自動抽出を考慮した画像データベース, 情報処理学会コンピュータビジョン研究会, 73-1, 1991
- [5] 浦谷 柴田 野口 相沢, 静止画像検索シス

テム FORKS の試作, 情報処理学会論文誌, Vol.28, No.7, 1987

- [6] 柴田 井上, 画像データベースの連想検索方式, 電子情報通信学会論文誌 D-2, Vol. J73-D-2, No.4, pp526-534, 1990
- [7] 加藤 栗田 坂倉, フルカラー絵画データベース ART MUSEUM —色彩感と略画による画像対話—, 信学技報, IE88-118, pp31-38, 1989
- [8] K.Aiba H.Nagata Y.Matsumoto T.Yokoyama Y.Matsushita, An Image Retrieval System depending on Object Layout, 3rd International Conference on Management of Data, pp199-218, Dec. 1991
- [9] 饗場 永田 松本 横山 松下, 画像検索におけるインタフェースの提案, 情報処理学会研究報告, 91-HI-34, 6, 1991
- [10] 饗場 野村 阿部 西山 横山 松下, 画像検索における階層化表現の一考察, 情報処理学会第43回全国大会, 1991
- [11] 小林, 記憶における絵の符号化に関する研究, 風間書房