

XMLによる画像の内容記述と検索

谷口 明正 八村 広三郎
立命館大学 理工学部

XMLによるデータ記述を用いて画像の内容記述を作成し、これにより画像内容に基づく検索を可能にするシステムを構築した。データ入力時に、画像内の対象物の位置、大きさ、および対象物についての名称などのラベルを付与して、XMLデータファイルを生成し、これを検索に利用する。XMLによるデータの記述を、画像データについてのメタデータとして蓄積し、インターネット上でのデータ表現とする。検索時には、画像中の対象物についての概略位置と大きさ、対象物についてのラベルを検索条件とし、これとXMLデータファイル中のデータを比較することで、類似度を算出する。この手法を用いて、静物画などの絵画100枚を対象画像とした検索実験を行った結果、良好な検索結果が得られた。

Image Content Description and Retrieval by Using XML

Akimasa Taniguchi Kozaburo Hachimura

Department of Computer Science, Ritsumeikan University

We have developed a system which enables image content description and retrieval by using XML. In data-input time a user describes contents of target images by interactively specifying location, size and label of each object. These information, together with location of an image file, will be included in XML data format by defining several elements by DTD. In retrieval time the same information will be specified interactively by the user. Image will be retrieved by comparing information contained in XML file with parameters specified.

1 はじめに

デジタルカメラなどマルチメディア技術の進歩により、コンピュータで画像データを簡単に扱うことができるようになってきた。さらに、これらの画像データはインターネットによって公開・流通ができ、だれでも膨大な量のデータを入手できるようになった。また、絵画や写真資料などの文化資産を、デジタル化したデータとして保存し、共用しようというデジタルアーカイブの構想も具体化しつつある。このような背景にあって、コンピュータシステムとネットワーク上の大量の画像データについて、効果的に蓄積・管理・検索を行うための手法が強く望まれるようになってきている。

画像データの検索については、画像の「内容」に基づいて検索を行う「内容検索」や「類似検索」などの必要性が重視され、広く研究が行われてきている。キーワードによっても、ある程度、画像の内容に基づいた記述と検索は可能であるが、構造のない、フラットな複数のキーワードによる表現能力には限界がある。

一方、画像処理、画像認識技術により対象画像を処理し、内容記述を自動的に作成しようという試みも盛んに行われている [1]。これは、対象とする画像の特徴量を自動抽出し、これを用いて画像同士の比較を行うことが基本であり、特徴量としては、色彩分布、領域分布、テキストチャ、対象物の形状などが用いられている。また、最近では、これ

らを組み合わせたシステムも開発されている [2]. 対象とする画像の性質やカテゴリーが限定されている場合には、このような比較的簡単な画像特徴量を用いることで、画像同士の内容の比較を行うことができ、実用に供することも可能である。

しかし、たとえばデジタルアーカイブで対象とするような、広い範囲の、一般的な画像データについて、画像処理によって抽出される簡単な特徴量だけで意味のある内容検索が行えるとは考えられない。また、今後、画像解析やパターン認識の技術が向上しても、一般的な画像データから、その情景や内容について、簡潔に、かつ正確に「記述」が作成できるようになるとも考えにくい。

一方、人間は、画像を与えられたとき、画像中の構成要素(対象物)を解釈して、これを基本として画像を識別・認識・理解する。したがって、画像の内容検索は、画像の中に存在する対象物に基づいて行われることが基本であると考えられる。しかし、この場合でも、画像中から自動的に対象物を識別して、これを付属情報として検索に利用することは、一般に非常に困難である。むしろ、画像の内容の解析は人間の能力にまかせ、これを積極的に利用することの方が現実的である。

このようなことから、本研究では、データ入力時に、対象画像の内容についての説明情報を、自動処理ではなく「人手」で付与し、これを検索時に利用するアプローチをとり、このような画像データの内容記述を、簡潔にかつ効率的に付与するための仕組みについて検討する。この方法により、入力時に手間はかかるが、検索時には、内容についての客観的で正確な情報を利用することができ、必要な画像データを効率的に検索することができる。

内容記述には、画像内のオブジェクトの位置、大きさ、および対象物の名称などのラベル情報を考える。これらの内容記述情報を、画像データと共に保存することになるが、ここでは、これらを画像データの「メタデータ」として、できるだけ画像データそのものと一体として保存するアプローチをとる。こうすることにより、個々の画像データをデータベース管理システムからは切り離して、流通させることができる。

メタデータとしてのデータ記述には、インターネット時代の標準的データ記述システムとして、近年注目を集めている XML を利用する。XML では、独自のタグを定義することができるので、こ

の仕組みによって、画像中の対象物などについての情報を記述することができる。

検索時には、画像データに付随した、メタデータとしての XML 記述を解析して必要なデータを抽出し、これと、検索条件として与えられた画像の内容記述との比較を行うことにより、検索条件に合致した画像を表示出力する。

本報告では、画像中の対象物とその空間的配置状況に着目した内容記述の方法と、それを XML により記述する方法、さらに、これらを利用したデータ入力および画像検索のシステムについて述べる。最後に、実装した実験システムによる検索実験の結果についても述べる。

2 画像の内容記述と XML による内容記述

2.1 オブジェクトに基づく画像の内容記述

人手で内容記述を作成する場合にも、さまざまなアプローチがある。たとえば、最も簡単なものとして、複数個のキーワードにより、画像の内容を簡潔に表現する方法がある。このような方法により作成されたデータの場合、画像データについての書誌的情報による検索と同様の検索手法がとれるので、検索システムの実現には、特別の工夫は必要としない。しかしながら、複数個のキーワードを用いても、それらの間に階層構造などの関連性を持ち込まずに、すべてを同列に扱うのでは、画像の内容を的確に表現することはできない。

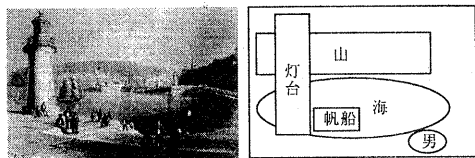
一方、画像の内容を、いわゆるキャプションと呼ばれるような短文や、解説文とでもいうような、ある程度まとまった量の自由記述文により、表現する方法もある [3, 4]。しかし、このような方法をとっても、自然言語には表現の自由度と解釈の曖昧さがあり、画像中の対象物などの細かな様子を適切に表現するのは難しい。

さて、画像の内容を記述するといっても、このためには、その対象分野の専門的知識を必要とする場合も多い。しかし、本研究で考えている内容記述とは、このような専門的な記述と表現を必要とするようなものではなく、むしろ、あくまでも淡々と画像の内容を客観的に、簡潔に、かつ即物的に表現し、検索に利用することを意図している。

専門的な背景や応用を考慮しない場合、人が写

真などの画像を見たとき、これを客観的に記述しようとする際には、まず、画像中の対象物に着目し、これらのそれぞれが何であるか、さらに、これらがどのような属性を持っているかを説明しようとするのが、自然である。また、画像中に同じ対象物が入っている場合でも、それらの、画像内での位置的な相互関係を記述しようとするのも自然である。また、これには余分の専門知識などを必要としない。

したがって、ここでは、対象としている画像について、その中の対象物（オブジェクト）について、それらを指示するラベル（一般的には、オブジェクトの名称）と、オブジェクトの、画像内での位置および概略のサイズをデータ作成時に入力し、これにより画像の内容を記述する。オブジェクトには、画像を記述するのに十分と思われるもののすべてについてとりあげ、この記述データを作成する。位置と大きさについては、それぞれのオブジェクトを取り囲む楕円や長方形を考え、これらの重心位置と面積の値をそれぞれ用いる。



(a) 原画像 (b) 概略図 (内容記述)
図 1: オブジェクトに基づく画像の内容記述

たとえば、図 1(a) の原画像内のいくつかのオブジェクトについて、図 1(b) のような概略図を描くことができる。この概略図で、原画像中のオブジェクトの配置についての特徴を表現できている。このように、それぞれのオブジェクトに対して、位置、大きさ、ラベルといったデータを記述することで、画像の内容に基づいた検索が可能となる。

2.2 XML

XML(eXtensible Markup Language) は、ネットワーク上で文書やデータを交換・管理・蓄積するための標準規格である。構造化データ記述の国際標準である SGML の拡張性と、HTML のインターネット利用性を併せ持った形式として期待されている。

XML では、対象データを幾つかの「要素」で構成されているものと考え、要素は HTML と同様にタグを用いて表現される。

XML で使用するタグは、DTD(Document Type Definition) 中の要素型宣言 (element type declaration) によって定義することができる。

XML データ中では、各要素を DTD で定義された「開始タグ」と「終了タグ」で指定し、この間に要素の内容 (Content) を記述する。図 2 に XML データの例を示す。同図では、XML の先頭 (2 行目から 6 行目) に DTD が付けられているが、この部分は外部ファイルに保存して参照することもできる。

```
<?xml version="1.0"?>
<!ELEMENT report (title,author*)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (name,belong)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT belong (#PCDATA)>
<report>
  <title>画像検索</title>
  <author>
    <name>谷口</name>
    <belong>立命館大学</belong>
  </author>
</report>
```

図 2: XML 文書の例

HTML による記述では、タグに意味を持たないため、値がどのような意味で使われているかを知ることができない。しかし、XML ではタグから値の意味を知ることができる。つまり、タグが対象データのメタデータとしての役割を果たし、検索等のアプリケーションで、これを利用することができる。

2.3 XML による内容記述

2.1 節で述べたような画像の内容記述の仕組みを XML により実現する。ここで用いた XML ファイルのデータ構造を図 3 に示す。また、その構造に従って記述した DTD を図 4 に、XML データの例を図 5 に示す。図 4 の DTD は、img.dtd と名前のついたファイルに保存しておき、図 5 の XML データは、2 行目の文書型宣言 DOCTYPE で、このファイルを指定して DTD を読み込んでいる。

本 XML ファイルは、<img_data>タグをルートとした階層構造を持ち、ルートタグの下に、画像内のオブジェクトについてのデータを記述する <data>タグと、画像ファイルの格納場所 (URL) を記述する <url>タグが定義されている。さらに、<data>の下に、各オブジェクトのラベル、位置、大きさについて記述するために、それぞれ

<label>,<position>,<size>を定義している。さらに、<label>の下には、入力されたオブジェクトのラベルを記述する<word>, 3章で述べる類義語検索を行うために必要な語彙の分類コードを記述した<code>を定義している。また、重心位置を表現するために、<position>の下には、重心の x 座標と y 座標を記述する<gx>,<gy>を定義している。

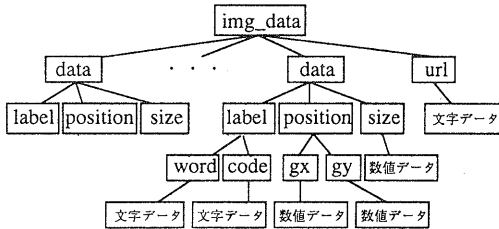


図 3: XML データファイルの構造

```

<!ELEMENT img_data (data* , url)
<!ELEMENT data (label,justin,area)
<!ELEMENT label (word,code)
<!ELEMENT word (#PCDATA)
<!ELEMENT code (#PCDATA)
<!ELEMENT position (gx,gy)
<!ELEMENT gx (#PCDATA)
<!ELEMENT gy (#PCDATA)
<!ELEMENT size (#PCDATA)
<!ELEMENT url (#PCDATA)

```

図 4: DTD(img.dtd)

```

<?xml version="1.0"?>
<!DOCTYPE img_data SYSTEM "img.dtd">
<img_data>
  <data>
    <label>
      <word>花</word>
      <code>1.5530.1.9.12</code>
    </label>
    <position>
      <gx>46</gx>
      <gy>63</gy>
    </position>
    <size>24</size>
  </data>
  </url>flower.tif</url>
</img_data>

```

図 5: XML ファイルによる内容記述例

2.4 XML データファイルからのデータの抽出

XML データファイルからの content データの抽出には、IBM の XML Parser for Java を使用

する。本 Parser によって、XML データファイルの構文解析を行い、要素やテキストといったノードからなる DOM ツリーを生成する。これにより、DOM(Document Object Model) API を利用して、メモリ上での操作が可能になる。メソッドにより、このツリー構造中のノードをたどり、ノードの値から、XML 文書の要素に記述された内容 (content) を抽出することができる。

DOM には、ノードの状態を参照や変更、親子関係で表されたツリー構造へのアクセス、あるいはノードの挿入、置換、削除といった操作をおこなう多くのメソッドが定義されている。したがって、DOM ツリーから様々な XML データを操作を行うことが可能である。

例として、図 6 の要素 A の値を抽出する様子を示す。まず、構文解析を行い、DOM ツリーを生成する。そこで、ルートタグから子ノードにアクセスし、ノードの値を得る。ノードが要素 (Element) ならば、要素名が A であるかを確認する。A であれば、子ノードの値 (Text) を抽出するという流れがある。

本研究では、これと同様の操作で DOM ツリーを走査して、word や gx といったノードにアクセスし、その子ノードであるテキストデータを抽出している。

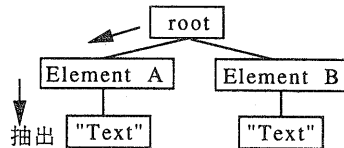


図 6: XML ファイルからのデータ抽出

3 オブジェクト記述に基づく画像検索

3.1 データ作成と検索

データ入力時は、対象画像中のオブジェクトの概略の位置と大きさを、楕円や長方形などの簡単な図形の描画により指定し、また、これらのオブジェクトの名称などをラベルとして入力する。オブジェクトの概形を指定した図形の重心位置と面積をオブジェクトの重心位置と大きさと考え、これらとオブジェクトのラベルを XML により記述してファイルに保存する

検索時には、データ入力システムによるデータ入力時とほぼ同様の操作により、オブジェクトの位置、大きさ、ラベルの指定を行って検索条件を設定する。指定された検索条件と対象のXMLデータファイル中に記述されている画像の内容記述とを比較する。

3.2 オブジェクトに基づくマッチング

検索時の、オブジェクトに基づく内容のマッチングと評価は図7のように行われる。

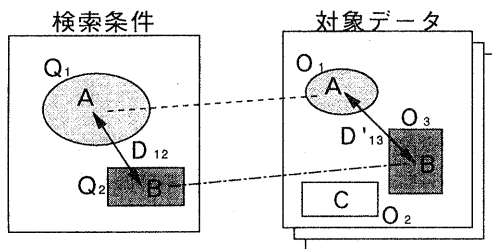


図7: オブジェクトに基づく類似性評価

検索条件に指定されたオブジェクト(図7の場合、 Q_1 と Q_2)のラベル(図7のA、B)および配置状況と、XMLデータファイルに記録されている対象データのオブジェクト(図7の O_1 、 O_2 、 O_3)についての同様のデータ間で、以下に示すようにマッチングをとって評価する。

1. 検索条件中に指定されているすべてのオブジェクトについて、それにつけられたラベルと、同じか、あるいはシソーラスを用いた類似度が、ある閾値以上のものが、対象画像中のオブジェクトのラベルに存在するかどうかを調べる。
2. もし、そのようなものがあれば(図7の場合、 Q_1 と O_1 および Q_2 と O_3)、その対応するオブジェクト間で、オブジェクト同士の重心位置と大きさ(面積)についての評価値を算出する。
3. 1で調べた条件に合致するオブジェクトが複数存在する場合は、検索条件および対象データの中で、それぞれ、これらのオブジェクト間の相対位置関係を表す方向コード(3.3.2で述べる)を求める。さらに、これらの、検索条件中のオブジェクト間の方向コード(図7の D_{12})と、対象データ中でのオブジェクト間の

方向コード(図7の D'_{13})との間の比較を行って評価値を算出する。

4. 2と3で求めた、オブジェクトの重心位置についての評価値、オブジェクトの大きさについての評価値、オブジェクト間の相対位置についての評価値の、3つの評価値それぞれで、1のラベル条件を満たすすべてのオブジェクトの組み合わせについて合計を求める。
5. 最後に、4で求めたこれらの3種類の評価値の加重和をとって、画像データ全体についての評価値とする。

3.3 類似性の評価

3.3.1 ラベル間の類似性

ラベル間の類似性の評価は、文字列比較によって行うことを基本とするが、さらに、類義語や同義語、関連語を考慮に入れた評価も可能にするために、単語を意味的、概念的な類縁関係によって分類・配列した辞書(シソーラス)を用いる。本研究では、シソーラスとして、国立国語研究所で作成された「分類語彙表」を利用した[5]。

分類語彙表では、単語が図8のような、階層構造によって分類されている。階層構造の最上位のレベル(第1レベル)は、品詞により「名詞」、「動詞」、「形容詞」、「その他」と分類される。これを「類」と呼び、順に1, 2, 3, 4と番号が付与されている。さらに、品詞ごとに意味分類され、4桁のコードが付与されている。類の番号と、意味分類したコードを合わせて「分類番号」と呼び、「1.2630」のように表す。さらに、第6レベルの「段落」、第7レベルの「行」、第8レベルの「語」と分類が行われる。

実際の分類語彙表のファイル中では、各単語は以下のような記法で格納されている。

```

スケッチ, すけっち, 1.3220, 2, 1, 2
絵画, かいが, 1.3220, 2, 1, 4
図, ず, 1.3220, 2, 2, 3
画像, がぞう, 1.3220, 2, 2, 4
名画, めいが, 1.3220, 2, 3, 1

```

この分類語彙表に基づいて、2つの単語間の類似度を求める手法は、いくつか提案されている[6]が、本研究では、単語のレベルを利用する、最も基本的な手法を用いた[7]。

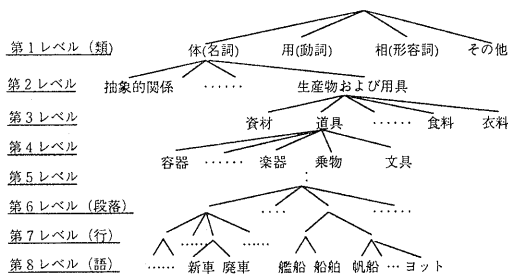


図 8: 分類語彙表の階層構造

2つの単語 w_i, w_j が図 8 の階層構造中で存在するレベルをそれぞれ d_i, d_j とし、それらに共通の上位レベルを d_u としたとき、2つの単語間の類似度 $S_w(w_i, w_j)$ を以下の式で定義する。

$$S_w(w_i, w_j) = \frac{2d_u}{d_i + d_j} \quad (1)$$

2単語の類、すなわち品詞が異なる場合には、 $S_w(w_i, w_j)$ は最小値 0 をなすようにする。一方、第 8 レベルまですべてが一致する場合、すなわち 2つの単語が同一の場合に最大値 1 をとる。

式 (1) による評価は、2つの単語が両方とも分類語彙表に登録されているものでなければならない。一方、あるいは両方が分類語彙表中に存在しない場合は、文字列同士の比較を行って類似度を求める。

3.3.2 オブジェクト間の類似性

画像内のオブジェクトについての検索条件と対象データとの類似性は、オブジェクトの重心位置、オブジェクトの面積、オブジェクト相互の位置関係の 3つのファクターで評価する。

なお、以下のオブジェクトについての類似性の評価は、オブジェクトのラベル間の類似度が設定した閾値以上の場合に実行される。

オブジェクトの重心位置

オブジェクトの重心位置についての評価は、検索条件として与えたオブジェクトの重心位置を (x_q, y_q) 、データファイル内のオブジェクトの重心位置を (x_d, y_d) とするとき、重心間のユークリッド距離 $P_{qd} = \sqrt{(x_q - x_d)^2 + (y_q - y_d)^2}$ を用いる。

ラベルについての条件を満たすすべてのオブジェクト間について求めた P_{qd} の総和をとって、画像全体の評価値 P とする。

なお、この場合、画像のサイズを 100×100 に正規化したときの重心位置の座標値を利用する。

オブジェクトの面積

検索条件に与えたオブジェクトの面積を S_q 、対象データ内のオブジェクトの面積を S_d としたとき、これらの差の絶対値 $S_{qd} = |S_q - S_d|$ を求める。

ラベルについての条件を満たすすべてのオブジェクト間について求めた S_{qd} の総和をとって、画像全体の評価値 S とする。

なお、この場合、オブジェクトの面積には、画像全体の面積に関して正規化した値を用いる。

オブジェクト相互の位置関係

検索条件、対象データそれぞれでの、オブジェクト間の位置関係については、方向コードを用いて、評価をする。方向コードは、あるオブジェクトの重心に原点をとって画像空間を 8 方向に分割したとき、相手のオブジェクトが存在する方向を 0 から 7 の数値で表したものである。図 9 の場合には、オブジェクト 1 とオブジェクト 2 の間の方向コードには、0 が割り振られる。

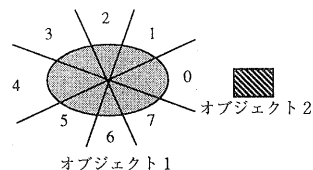


図 9: 領域相互の位置関係と方向コード

オブジェクト相互の位置関係についての類似性の評価値は、以下のようにして求める。すなわち、検索条件におけるオブジェクト相互間の方向コードとデータベース中の対応するオブジェクト間の方向コードの差をとり、この値の 0, 1, 2, 3, 4 に応じて、それぞれ $0, \frac{1}{4}a, \frac{1}{2}a, \frac{3}{4}a, a$ を割り当てる。ただし、 a は、適当な正の値である。

画像全体での評価値 D は、ラベルについての条件を満たすすべてのオブジェクト間について求めた位置関係の評価値の総和として求める。

画像全体での評価値

以上で求めた 3つの評価値に対して重み付けを行って、次式に示す全体的評価値 H を得る。この評価値の低いほど、類似性が高いと考えることができる。

$$H = w_p P + w_s S + w_d D \quad (2)$$

ここで、 w_p, w_s, w_d はそれぞれの評価値に対する重みで、これは、検索時にユーザによって自由に設定することができる。

4 システム概要と検索例

4.1 データ入力システム

データ入力システムのインターフェースを図 10 に示す。表示された画像内のオブジェクトに対し、楕円または長方形で領域を指定し、それぞれのオブジェクトに対応するラベルを、右側のラベル入力部に入力する。

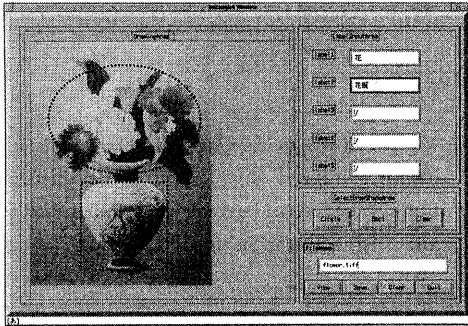
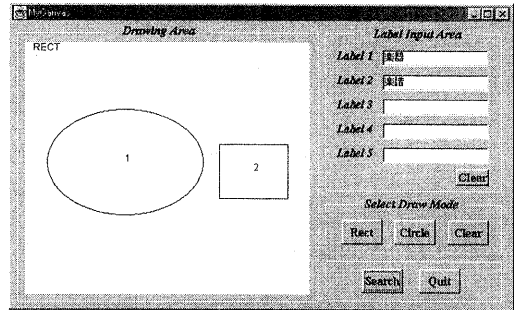
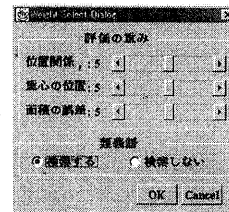


図 10: データ入力システム



(a) 検索画面



(b) 検索条件設定ダイアログ

図 11: 検索システム

4.2 画像検索システム

画像検索システムのインターフェースを図 11(a) に示す。ユーザは、領域とラベルを指定し、Search ボタンをクリックして検索を開始する。

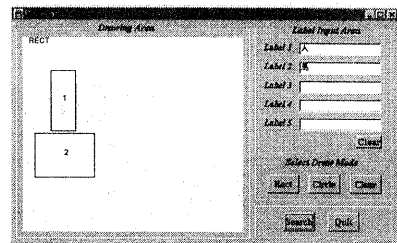
シソーラスを用いた検索を行うかどうかの指定と式 (2) で示した評価値 H の重みの設定を行うためのダイアログ (図 11(b)) が表示されるので、このダイアログで値を決定し検索を実行する。検索条件と合致した画像がディスプレイ上に表示出力される。

4.3 検索例

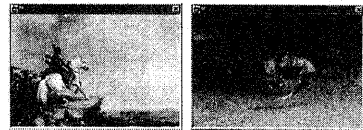
静物画などの絵画 100 枚を対象画像とし、これを利用して検索実験を行った。

検索条件として、オブジェクト「人」、「馬」を、図 12(a) のような大きさと位置関係と与えた。このときの検索結果を図 12(b) から図 12(e) に示す。ここでは、検索結果の最上位の 4 枚の画像を示している。

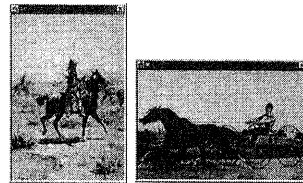
次に、シソーラスを用いた場合の検索例を図 13 に示す。検索条件として「楽器」と「楽譜」を、図



(a) 検索条件の入力画面



(b) $H = 0.93530$ (c) $H = 0.82126$



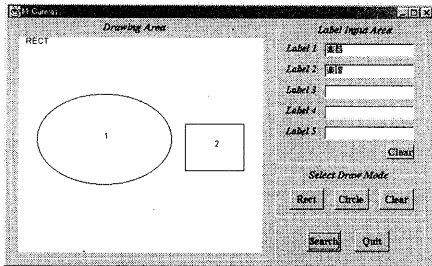
(d) $H = 0.82049$ (e) $H = 0.66194$

$w_s = 2, w_d = 5, w_c = 8$

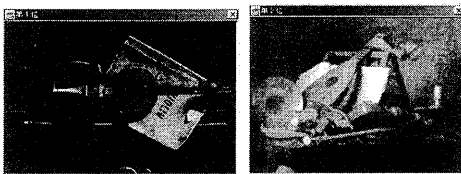
図 12: 検索例 1

13(a)のように指定した。この時の上位4つの検索結果を図13(b)から図13(e)に示す。

図13の検索結果の場合、各画像に対応するXMLデータファイル中には、図13(b)は「バイオリン」、図13(c)は「楽器」、図13(d)は「トランペット」、13(e)は「バイオリン」と記述されていたものであるが、シソーラスを用いているので、それらは検索条件で与えた単語「楽器」の類義語と判断され、検索結果として得られた。



(a) 検索条件の入力画面



(b) $H = 0.60002$

(c) $H = 0.57389$



(e) $H = 0.49812$

(f) $H = 0.35596$

$w_a = 7, w_d = 6, w_c = 3$

図13: 検索例2

5 おわりに

画像内のオブジェクトのラベルと配置に関する情報を用いて画像の内容検索を行う手法と、これらの内容記述情報をXMLを用いて記録・管理する手法について報告した。これにより、画像のメタデータとしての内容記述と、画像データ本体をコンパクトに記録・保管し、これをインターネッ

ト環境で活用する可能性を示した。

なお、内容記述の方法として、本研究の手法だけが優れているということを主張しているのではなく、キーワードによる手法、キャプションや解説文による手法など、それぞれの長所を生かしこれらを相補的に利用することを意図している。さらに、今後、画像処理、画像解析による画像特徴の抽出とそれによる特徴記述データをリンクして検索に利用することも、視野に入れて、検討している。

参考文献

- [1] M. De Marsicoi, L. Cinque, S. Levialdi: Indexing pictorial documents by their contents: a survey of current techniques, Image and Vision Computing, Vol.15, pp.119-141, 1997
- [2] M. Flickner et. al. : Query by Image and Video Content: The QBIC System, IEEE Computer, pp.23-32, September, 1995
- [3] R. K. Srihari : Automatic Indexing and Content-Based Retrieval of Captioned Images, IEEE Computer, pp.49-56, September, 1995
- [4] 伊藤, 八村: キャプションと画像特徴に基づく類似画像検索, 情報処理学会研究報告, 2000-CH-45, pp.33-40, 1999
- [5] 国立国語研究所: 分類語彙表, 秀英出版, 1994
- [6] 小林, 杜, 石井: シソーラスを用いた単語間類似度の判定法に関する研究, 情報処理学会第60回全国大会論文集, pp.169-170, 2000
- [7] 山田, 金, 柴田: インデックス文の類似性に基づく映像検索, 第5回国立国語研究所国際シンポジウム論文集, pp.145-153, 1997