

キーワード自動抽出を考慮した画像データベース

山根 淳 坂内正夫
東京大学生産技術研究所

近年、大量に蓄積された画像情報に対して効率よく必要な情報を取り出すことのできる画像データベースに対する要求が高まってきた。本研究においては、キーワードを自動獲得する画像検索において大部分を占めていた認識側への比重を軽くし、検索側での工夫によってこの画像認識技術の未成熟さを補うひとつのアプローチとして、新しい枠組を持った画像データベースを提案する。これは画像の自動認識を前提とし、画像認識・理解のモデルを検索側でも共有する枠組を持つ画像データベースである。また数種類のスポーツ画像を対象とした画像データベースをこの枠組を用いて構成し検索を行うことにより、このシステムの有効性を評価するものとする。

Image Database System with Capability of Automated Image Keywords Extraction

Jun Yamane Masao Sakauchi
Institute of Industrial Science, University of Tokyo
7-22-1, Roppongi, Minato-ku, Tokyo 106, Japan

Recently, flexible image database retrieval system is strongly required, where image keywords need to be captured automatically to manage a practical number of image data successfully.

A new image database system framework which has a devised method to realize image retrieval with keywords which represent image contents is proposed. In this framework, a state-transition-type image recognition model which is used for capture image keywords is shared by an image retrieval part of the system.

And implemented experimental system using a various type of sports scene images reveals the effectiveness of this framework.

1 まえがき

大量に蓄積された画像情報に対して効率よく必要な情報を取り出すことのできる画像情報を含んだデータベース、すなわち画像データベースに対する要求が高まってきた[3, 4, 5]。しかし、画像は量的および質的に従来のデータベースに適さない性質のため、対象画像や利用方法に依存したものが多く、汎用性のある画像データベースの実現は困難な状況にある。

従来の画像データベースの研究は、画像情報個々についてその表現内容をいかに効果的にインデックスとして付与するかに技術的焦点が当たられ、そのため画像認識技術に対する依存性が大きくなっている。画像認識技術はまだ未成熟な状態であり、画像から必要な情報を自動的に抽出して検索に用いることは非常に困難な状況である。

そこで本研究においては、キーワードを自動獲得する画像検索において大部分を占めていた認識側への比重を軽くし、検索側での工夫によってこの画像認識技術の未成熟さを補うひとつのアプローチとして、新しい枠組を持った画像データベース[1, 2]を提案する。これは画像の自動認識を前提とし、画像認識・理解のモデルを検索側でも共有する枠組を持つ画像データベースである。

そして十数種類のスポーツ画像を対象とした画像データベースをこの枠組を用いて構成し検索を行なうことにより、このシステムの有効性を評価するものとする。

2 画像データベースの実現

近年、光ディスク等の画像情報に適した情報蓄積媒体や入出力媒体の普及により、さまざまな応用分野において画像情報のファイル化とそのコンピュータ利用が行なわれている。

そこで、大量の画像を蓄積し、必要な時に必要な画像情報を取り出すことのできる画像データベースに対する要求が高まってきた。

これまでに文字、数値を扱うデータベース技術はめざましい発展をとげてきた。しかし画像情報には次のように従来のデータベースが扱ってきた文字や数値情報と異なる二つの特徴があり、従来のデータベース技術の拡張としての画像データベースの実現を困難なものとしている。

量的な特徴—画像データの冗長性 利用する情報量に対して画像の持つ情報は非常に大きく、利用者にとって必要な情報のみを取り出すことは非常に困難である。

質的な特徴—画像データの抽象度の低さ 計算機上の画像データは信号レベルのものであり、一方人間がその画像から読みとる情報は一般には非常に抽象度の高い概念レベルのものであり、両者には大きな隔たりがある。

そのため、画像情報をそのままデータベース化するのは難しく、データベース構成時に従来のデータベースの枠組に適したなんらかの情報を抽出しておき(抽出された情報をデータモデルと呼ぶ。)利用の際にはその抽出された情報をもとに検索を行なうという方法が現実的である。データモデルにはさまざまな抽象レベルのものが考えられるが、その中でも概念レベルで画像の内容を表現する意味的な情報(検索キー情報、あるいはキーワードと呼ばれる)が抽出されれば画像データベースの高度な利用が可能になる。画像情報から概念レベルのキーワードを獲得するには、人間がすべての画像を見ながらキーワードを手で抽出するか、あるいは画像認識・画像理解技術を用いて画像情報を概念レベルまで抽象化必要がある。すべての画像に対して人間が手でキーワードを付与するのは効率的ではなく、画像認識・理解技術を用いたキーワードの自動抽出に期待がかかるが、現状の画像認識・理解技術が未成熟であるためキーワードを完全に自動獲得することのできる画像データベースは実現されておらず、現状では、個々の対象ごとにさまざまな工夫がなされているのみである。

それら対象による工夫点を大きく分けると次の三つの形態があげられる。

1. 人間による意味的検索キーの付加

これは、意味的な検索キーを画像の入力過程で人間が直接付加するものである[7]。現状の画像認識・理解システムにおいては、これら検索キーとして用いるべき情報の完全自動抽出は非常に困難であり、ほとんどの場合人間による手動入力に頼らざるを得ない。この方法には、キーの付加にコストがかかるという欠点があるため、この方法は一度入力したデータの価値の高い画

像データ(地図や配管図など)の維持管理などには有効である。

2. 画像から物理的特徴量を自動抽出

画像から比較的自動抽出の容易な物理的特徴量を抽出して検索に用いる方法[8]である。この方法を用いたものとしては、画像から抽出した特徴量から類似度を計算して例示画検索を行うものがほとんどである。この方法は入力が容易であり実現しやすいが、画像の内容に踏み込んだ処理がなされていないので、ユーザに対する柔軟な検索が実現できているとはいえない。

3. 意味的検索キー自動抽出の試み

意味的なキーワードを画像処理などによって自動的に抽出しようという試みもいくつか行われている。風景画像について領域分割を用いて対象物を認識し、ユーザの検索要求に応えるシステム[9]や位置関係についてのみ自動抽出を行ない、検索要求に備えるシステム[6]等が研究されている。これらは、非常に限られた対象について認識を行なおうとするものであり、一般的な対象に拡張するのは容易ではない。

今後は画像認識技術の発展に期待することはもちろん、現在の画像認識技術で得られた情報を有効に利用することのできる検索の枠組の作成が望まれる。

3 提案システムの構成

3.1 システムの基本構造

図1に、提案システムの基本構成を示す。

キーワード抽出部は、データ画像からトークン抽出を行ない、状態遷移モデルを用いた汎用の画像認識システムによってそれぞれのトークンの抽象度を高め、最終状態をキーワード管理部に渡す。

キーワード管理部は、抽出部から渡されたトークンを、多次元データ構造を用いて管理する。

キーワード照合部はユーザの要求と管理されているキーワードとを照合させる。以下に述べるように、ここにおいてもキーワード抽出で用いた認識・理解モデル(状態遷移モデル)を利用する。

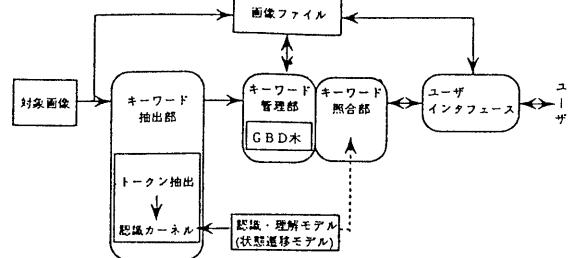


図1：提案システムの基本構成

3.2 画像認識・理解モデルの検索側における利用

提案システムにおいて最も大きな特徴を述べる。前章でも述べたように、現状の画像認識・理解技術では概念レベルのキーワードを自動抽出することは非常に困難である。そこで、現状の画像認識・理解技術を利用し、検索側の工夫により認識結果が不十分な抽象レベルのものであっても内容依存の検索を可能とすることを考える。そのひとつのアプローチ法として、画像の認識・理解の枠組を検索側でも利用する、すなわち画像認識・理解モデルを認識側と検索側とで共有する画像データベースを提案する。

具体的には、不完全な抽象度のキーワードを許した検索キーデータベースを形成し、内容依存の検索を可能とするものである。

3.2.1 状態遷移モデルを用いた画像認識システム

ここで提案する手法は、状態遷移型の認識モデルを利用した画像認識・理解システム[11, 14]を用いた検索キー抽出に適用されるものである。このシステムは、認識手法を認識システムカーネルから独立した抽象的なルールの形で与えることにより、多目的で柔軟な画像理解システムの枠組を目指すものである。この枠組の概略を図2に示す。

状態遷移型の認識モデルを用いた認識システムの特徴は、各トークンはそれの持つ属性によって意図した抽象レベルに達しなくとも必ずなんらかの状態を持つという点である。ここで述べる手法はこの特徴を生かしたものとなる。

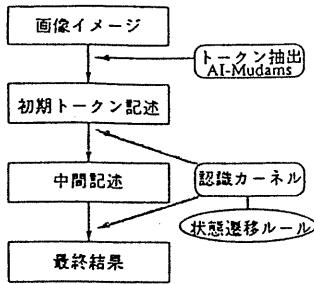


図 2: 状態遷移モデルを用いた画像認識システムの概略 [11]

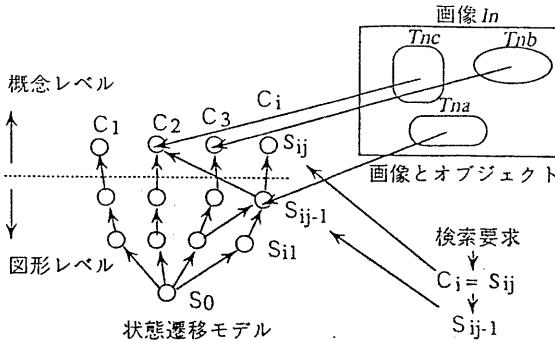


図 3: 状態遷移モデルの検索側での利用

3.2.2 検索側での画像認識・理解モデルの利用

対象画像からトークン抽出を行ない、それぞれのトークンを図 3のような状態遷移モデルによって状態遷移をさせることにより画像認識を行うとする。本来は状態遷移モデルにおいて最上位の抽象度を持つ概念レベルまですべてのトークンが達することが望ましいが、それは困難であり多くの場合途中の抽象レベルで止まったまま認識を終了してしまう。従来は概念レベルに達したトークンのみキーワードとして記録されて検索に利用されてきた。しかし、上に述べてきたことからもこれは十分に画像を認識した、すなわち十分に検索キー抽出がなされたとはいがたい。

そこで、次のようなことを考える。図 3を用いて説明すると、画像中から抽出された各トークンは、初期状態 S_0 を出発点として、トークン自身の持つ属性や、周囲、あるいは画像全体の状況などを条件に状態遷移を行ない、状態遷移モデル中を移動する。画

像 I_n のあるトークン T_{na} が图形レベルの状態 S_{ij-1} に達し、ここからどの遷移条件も満足できず、この状態 S_{ij-1} を最終状態として得たとする。

システムは非概念レベルであるこの状態を記録し、検索時にこの最終状態の上位の概念レベルの状態である C_i が検索キーとして与えられた場合、この状態を持つトークンを含む画像、すなわち I_n も候補として挙げるのである。

この方法は、状態遷移を起こす度に徐々にトークンの抽象度が変化するという状態遷移型の特徴を利用したものであり、これまで画像認識技術の未成熟さによりなかなか実現できなかった検索キーの自動獲得を前提として内容依存の高度な画像検索を実現しようとするものである。

4 キーワード獲得・検索手法

4.1 キーワード獲得手法

4.1.1 トークン抽出

状態遷移モデルを用いた画像認識システムを用いて画像中のオブジェクトを認識することによりキーワードを獲得する。状態遷移モデルを用いた画像認識システムにおいては、まず状態遷移をするあるまとまった処理単位であるトークンをビットマップの画像から抽出する必要がある。

地図や機械設計図面等を対象とした図面理解システムにおけるトークン抽出手法(トークンエクストラクタ)としては AI-Mudams [12, 13] を用いたものが提案されている [11, 14]。カラー画像の場合はまずカラーセグメンテーションを行なう必要がある。カラーセグメンテーションはよく似た画素値を持つ画素を併合していくことによって画像をある大きさをもった領域(セグメント)に分割するものである。セグメンテーション手法にはさまざまなもののが提案されている。セグメンテーションの結果、ラベルづけされた領域は認識カーネルに渡されるトークンとなる。

4.1.2 認識カーネルと認識結果の記述

抽出されたトークンは認識カーネルに渡される。ここではトークンの持つ图形的性質やトークンの周囲の状況、および状態遷移モデルを用いてトークン

の抽象度を高めていく。認識終了後、その最終状態を記録する。最終状態は概念レベルの状態でなくとも記録する。

属性について ここでいう属性とは、オブジェクトを理解する際に用いられる、すなわちトークンの状態遷移に用いられる色、形等の图形的性質のことである。

これらの属性は、検索要求に含まれる場合もあり得るので、なんらかの方式で記録する必要がある。状態遷移モデルにおいて、概念の上位概念として属性を含んだ概念の状態を置くこともできるが、これは状態数が非常に多くなり、非現実的である。このため属性は認識結果とは独立に記録するものとする。

位置の記述 抽象度を下げた範囲からも検索を行なった場合検索された画像数が多くなり過ぎることが予想され、それへの対策として画像の中央付近にあるものを重視するというオブジェクトの画像中での位置を基準にした検索方針が考えられる。ここではオブジェクトの外接矩形の中心位置を記録することにより位置関係を利用した絞り込みに対応するものとする。

图形的性質の多次元データ構造による管理 画像中の主なトークンそれぞれに固有の情報(属性、外接矩形の中心位置)は、認識結果を除き多次元空間上の領域として記録される。これらの領域は、多次元データ構造 GBD 木によって管理・照合される。

4.2 検索アルゴリズム

4.2.1 オブジェクト検索

検索アルゴリズムは以下の通りである。この検索要求は一般に主体 S_i が属性 a_1, a_2, \dots, a_m を持つ形とする。

1. 最初の単語は概念名概念 C_n であるので、(概念 \leftrightarrow 探すべき状態) 対応表から検索すべき状態を得、検索を行なう。
2. 以下の各単語は属性属性 a_m を表す単語なので、(属性名 \leftrightarrow 属性値) 対応表から値を調べ、その属性を表す座標軸の検索範囲を決定する。次の単語についてまたこの項目を調べる。

3. 概念の検索によって得られた各オブジェクトについて、属性によって決定した検索範囲について画像空間における範囲検索を行なう。検索の結果数が多過ぎる場合は、画像中の中心付近に位置する画像のみに絞り込む。これは検索範囲をオブジェクトの中心位置を記述した軸について範囲を狭くすることによる。

4.2.2 提案システムの利用

提案システムはこれまで述べてきたように概念レベルに達しないオブジェクトも検索に用いるものである。よって検索要求の概念を含まない画像が検索される可能性もあるし、あまり抽象レベルの低いオブジェクトまで検索すると検索結果の画像数が膨大なものになるといった質・量的な問題が当然生じてくる。

しかし複数のオブジェクトを同時に要求するような検索の場合は、質的な問題は解決されないが、量的な問題はユーザのブラウジングによる判別が可能な画像数にまで軽減することが可能になると考えられる。

そこで、複数指定の検索の一例としてスポーツシーン画像におけるシーン検索を挙げる。スポーツ画像の検索にはさまざまなものが考えられるが、中でもスポーツシーン、特にスポーツ名による検索が可能であることが望ましい。ある画像のスポーツ名を認識するには、画像中のオブジェクト、オブジェクトの内容、動き、位置関係、画像の縮尺、立体的なシーンを映した画像については撮影角度等を総合して判断する必要がある。しかしそれを計算機により自動的に行なうことは非常に難しく、オブジェクトが何であるかの認識すらまだ不十分な状態であることは第2章等で述べた通りである。

そこで、今回はスポーツ名を構成する要素のうち、画像内のオブジェクトのみを考慮した検索法を提案する。そのスポーツを構成するために必要であると考えられる器具、競技場、オブジェクトの概念名を列挙し、記録しておく。

検索要求としてあるスポーツ名が与えられると、それらの構成要素であるオブジェクトをひとつひとつ検証し、すべて、あるいは一部の条件を満たした画像を提示していく。

4.3 ユーザインターフェース

4.3.1 「らしさ」の評価

提案システムにおいては画像認識によって概念レベルに達しなかった場合も検索に用いるため、ある要求がなされた場合かなり多くの画像が提示される可能性がある。これらを同一に扱い、あとはユーザの判断に任せるものとすると、結局絞り込みが完全でユーザに多くの負担を負わせるという結果に陥ることが予想される。

そこで、これらの提示画像に対して要求された画像「らしさ」を評価するためのなんらかの評価値を算出し、その大きさによって絞り込みを行なうこと必要になる。

この評価値の算出にはさまざまな方法が考えられるが、ここでは画像からのキーワード獲得によって得られたものを用いる方法を考える。そこで、検索要求を構成する要素それがどの程度概念レベルに近いかを用いて要求された画像「らしさ」を表現し、評価するものとする。

画像 i に対して、検索要求 R の構成要素のすべてが概念レベルに達している場合に 1 の値をとり、構成要素がどれも存在しない場合に 0 の値をとる評価関数 $E(R, i)$ を定義する。

4.3.2 評価関数算出

上で述べたように状況は構成要素であるオブジェクトの集合で表される。例えば、「サッカー」という状況を表す画像には、「サッカー選手」「サッカーボール」「サッカーゴール」「サッカー場」という構成要素がある。「サッカー」という検索要求に対して、システムは構成要素それについての検索を行なう。それぞれの構成要素が概念レベルまで認識が高められた画像が存在することが望ましいが、それは稀、あるいはほとんどない場合もあり得る。

そこで、要求 r に適合する可能性のある多数の画像が得られるのであるが、まず要求 r のそれぞれの構成要素 $c_j(r)$ に重み $w(c_j(r)) (\sum w(c_j(r)) = 1)$ を与えておく。対象画像 i において、それぞれの構成要素が状態遷移モデルにおいて抽象度がどの程度であるかを表す $A_j(i) (0 \leq A_j(i) \leq 1)$ を求める。重み w に抽象度 A を重畠し、すべて加えたものを評価値 $E(r, i)$ とする。(図 4)

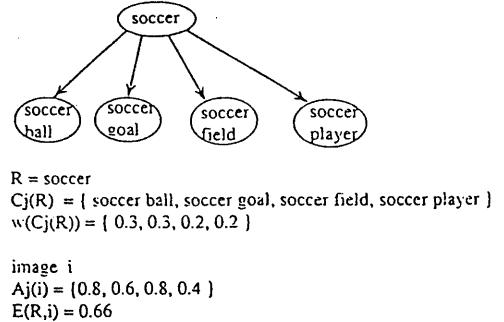


図 4: 評価値の算出

この評価値は、対象画像 i の検索要求 r に対する各構成要素の抽象度を表しており、「らしさ」を表しているといえよう。次に状況における各オブジェクトの持つ重み、抽象度の数値化について考察する。

オブジェクトの持つ重み ここで提案する評価値の算出には各状況におけるオブジェクトの持つ重みが大きな意味を持ち、効果的な重みの与え方が要求される。一方、この評価値は、画像認識の立場にから考えるとオブジェクトのみから状況認識を行なう場合の評価値ということになる。この値は画像認識技術においても問題となり、なかなか一般的な定式化は難しい。

ここで、今回提案するシステムにおいては、この重みはデフォルト値として構成要素である各オブジェクトに均一に与えられるものとし、ユーザがインタラクティブにその重みを変換していくという機能をユーザインターフェースに持たせるものとする。

5 提案システムの実装

提案システムを計算機上に実装した。今回は自然画像データベース構成の第一段階としてスポーツ画像を対象とした。8種目(サッカー、ラグビー、プロレス、体操、新体操、ダイビング、ボクシング、テニス)のスポーツを選び、それぞれ3~4枚ずつの写真を選んだ。それらの写真はカラースキャナ GT-6000 を用いてデジタル入力した。

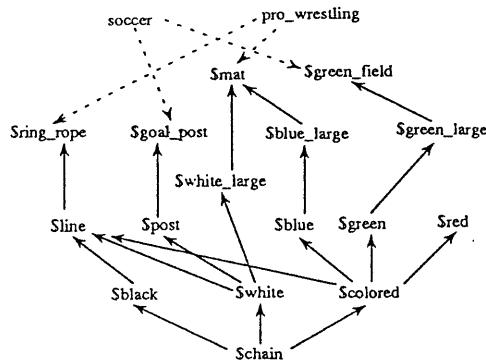


図 5: オブジェクト認識用の状態遷移モデル (抜粋)

5.1 キーワード抽出部

5.1.1 トークン抽出部

カラー画像を認識するために、画像から認識対象となるトークンを切り出す。ここではトークンとして、同じような色によって構成される画像中の領域(セグメント)を用いる。すなわちカラーセグメンテーションによってトークン抽出を行なう。カラーセグメンテーションの手法として、ここでは限定色表示法を用いる。限定色表示法は、画素値の似たものをまとめて同一色とし、色数を 16 とか 8 とかの限られたもので表示する手法であり、本来は同時に表示可能な色数の少ないディスプレイ装置のために開発されたものである。

限定色表示の後、各代表色によってラベリングされた領域について、チェインコードを抽出し、面積 100 未満の微小面積のオブジェクトを除去し、折れ線近似を行ない、このチェインコードをそれぞれ認識カーネルに渡すトークンとする。

5.1.2 オブジェクト用状態遷移モデル

十数種類のスポーツ画像を対象とした画像内オブジェクト認識用の状態遷移モデルを作成した。(図 5)

5.1.3 属性について

第4章で述べたように、各トークンについて属性を用いた検索要求への対応、および認識モデルの変更への柔軟な対応のために、認識の結果得られたキーワードとは切り離した状態で管理することを考える。

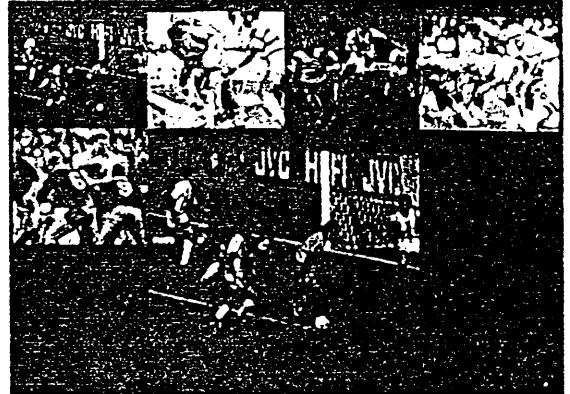


図 6: 「サッカー」という要求に対する検索結果

ここでは、カラーのスポーツ画像を対象にしているため、(1) 色(セグメンテーションされた領域の代表色の HSI 座標系表示)、(2) 形(複雑度)、(3) 面積(画素数)、(4) 外接長方形の中心位置をそれぞれ記述するものとする。

5.2 キーワード管理

第4章で述べたように得られたキーワードは多次元データ構造 GBD 木によって管理を行なう。実装システムにおいては、画像内オブジェクトの属性である(1) 代表色の HSI、(2) 複雑度(3) 外接長方形の中心座標 (x, y) の 6 次元の画像空間を構成した。

5.3 検索

「サッカーの画像」という検索要求をシステムに与えて検索を行なった。システムは「サッカー」という状況に必要なオブジェクトの検索を行ない、それぞれ検索されたオブジェクトの抽象度から算出される評価値の高い順に画面に表示した(図 6)。

この結果によるとサッカー画像として入力された画像はすべて提示されている。提示結果の中にラグビーの画像も含まれている。サッカーの要素オブジェクトをすべて含んでしまっているため、これを分離させるためにはプレイヤーの動き、ユニフォームを利用したより高度な認識ルールが必要であり、今回の認識ルールからしてやむをえない結果である。

6 むすび

本研究においては、画像認識・理解に用いるモデルを検索時にも利用することにより、画像認識・理解

技術の未成熟な点を補い高度な内容依存の検索を実現する画像検索の枠組を提案した。具体的には、状態遷移型の画像認識モデルを用いた認識・検索キー抽出において、概念レベルに達しない不完全な図形レベルの抽象度で認識を終了した画像内のオブジェクトについてもその最終認識状態を記録しておき、検索時にその上位レベルである概念が検索キーとして与えられた場合に候補として挙げるというものである。

この枠組においては各オブジェクトを検索する際に質的・量的な問題が生じるが、量的な問題に関しては複数のオブジェクトを要求する検索に用いることにより、質的な問題は参考のために検索されたオブジェクトの抽象度をもとに算出した評価値を用いることにより対処した。

そしてこの枠組の有効性を確かめるために、十数種類のスポーツ画像を対象とした画像データベースを、この枠組を用いて構成した。画像認識には限定色表示をもとにしたトークン抽出部を持つ状態遷移モデルを用いた汎用の画像認識システムを用いた。その結果、この枠組の有効性が確かめられた。

今後は、このシステムをあらゆる対象の画像に適用し、このシステムが適した対象群を浮き彫りにするとともに、ユーザインタフェースをインテリジェントなものにして画像の多義性に対応することによりこのシステムの欠点を補っていく必要があるであろう。

参考文献

- [1] 山根, 坂内:キーワード獲得の自動化を考慮した画像データベースの基本構成, 信学会秋全大 D-59(1990).
- [2] 山根, 坂内:キーワード獲得の自動化を考慮した画像データベースにおける検索アルゴリズム, 信学会春全大(1991).
- [3] 坂内:画像検索技術, 信学誌, Vol.71, No.9, pp.911-914(1988).
- [4] 坂内:これからの画像・図形情報のデータベース化, 生産研究, Vol.42, No.4, pp.227-234(1990).
- [5] 坂内, 大沢:画像データベース, 昭晃堂(1987).
- [6] 位置情報を手がかりとする画像検索法, 情処学論, Vol.31, No.11, pp.1636-1643(1990).
- [7] 浦谷他:静止画検索システム FORKS の試作, 情処学論, Vol.28, No.7, pp.758-767(1987).
- [8] 加藤, 下垣, 藤村:画像対話型商標・意匠データベース TRADEMARK, 信学論 D-II Vol.J72-D-II No.4 pp.535-544(April 1989).
- [9] 岡崎, 美濃, 坂井:風景画像の認識とその認識対象属性による画像検索—風景画像中の山を例として, 信学技報 IE86-117, pp.57-64.
- [10] Y.Ohsawa, M.Sakauchi:A New Tree Type Data Structure with Homogeneous Nodes Suitable for a very Large Spatial Database, *Proc of IEEE Sixth International Conference on Data Engineering*, Los Angeles, 13-1(1990).
- [11] Shin'ichi Satoh, Yutaka Ohsawa, and Masao Sakauchi:Drawing image understanding framework using state transition models, *10th ICPR*, pp.491-495, IEEE(1990).
- [12] Wei Lu, Yutaka Ohsawa, and Masao Sakauchi: A database capture system for mechanical drawings using an efficient multi-dimensional graphical data structure, *9th ICPR*, pp.266-269, IEEE, 1988.
- [13] M.Sakauchi:Two interfaces in image database systems, *Proc. of the International Workshop on Industrial Applications of Machine Intelligence and Vision*, pp.22-27, IEEE(1989).
- [14] 佐藤, 大沢, 坂内: 状態遷移モデルを用いた多目的図面理解システム, グラフィック・イメージディアの高度利用に向けて機能図形情報シンポジウム講演論文集, pp.117-122(1990).
- [15] 龍, 鳥海, 大沢, 坂内: 重視領域の指定によりデザイナーの好みを反映し得る限定色表示法, テレビ学誌, Vol.45, No.1, pp.86-93(1991).
- [16] 鈴木, 大沢, 坂内: 色選択に柔軟性を持たせた限定色表示手法, テレビ学誌, Vol.43, No.3, pp.268-275(1989).