コンピュータビジョン　32－4
（1984. 9. 21）

類似画検索のためのプロダクションシステムとルール変更ツール

A production system for similarity retrieval of image database and a rule modification tool

友田大輔，竹田裕幸*，カンシット マイットリー，田中稔，市川忠男
Daisuke Tomoda, Hiroyuki Takeda, Cunshit Maitree, Minoru Tanaka, Tadao Ichikawa
広島大学 工学部 第二類
Faculty of Engineering, Hiroshima University

This paper concerns a production system which works on image databases, in which the similarity between images is estimated for retrieval. Generally, production systems serve to attaining a flexible management of image processing rules. However, the retrieval efficiency goes down as the volume of image data increases. The fact requires re-organization of production rules, and hence the authors are proposing a tool for modifying production rules. The modification is based on the interaction between the system and the designer. The validity of the proposed tool is demonstrated through experiments.

1. はじめに

今日，画像データベースシステムの重要性が高まってきているが，扱うデータが画像であるという特性にあわせた検索方法が必要であり，検索効率も検討する必要がある．本研究では，例画像を入力してそれと類似した画像を検索するという方法を前提として，この方法の基礎となる類似性評価のためのプロダクションシステムを紹介する．さらに，標本画像の増加に伴う検索効率の低下を防ぐためのツールとして，プロダクションルールの変更を行うプロダクションルール変更ツールについて述べる．

ここで述べる類似画像判定は次のように行う．[1)]

i) 標本画像の特徴を調べ，繰り返し2分割法を適用して，あらかじめ標本画像をいくつかのカテゴリに分類する．

ii) 入力された例画像がどのカテゴリに属するかをi)と同じ手法を用いて判定する．

iii) 判定されたカテゴリに属する標本画像を類似画像として出力する．

プロダクションシステムを使用することによって次のような利点がある．

i) それまでの抽出処理をもとにして，次に行うべき特徴抽出処理を指定できる．したがって不要な特徴を抽出しないですむ．

ii) 画像データベースの成長・発達に伴うルールの追加・変更が容易に行える．

2. 類似性評価システム

類似性評価システムは，多種類の画像からなる画像データベースに対して例画像を入力し，それと同一の，あるいは類似した画像を検索するシステムである．

2.1 類似性評価の方法

各画像から特徴を抽出し，特徴の有無に基づいて画像の類似性の評価を行う．類似性評価は，次のような繰り返

*）現在，三菱重工

し2分割法を用いて行われる．

1) 1つの特徴を用いて標本集合を2分割し，他の特徴を用いて各分割集合をさらに2分割する．これを繰り返すことによって標本画像を類似したもののグループに分類する．

2) 1)の分割結果で得られる標本の部分集合の中で意味的に等価なものを1つのグループにまとめる．これを画像のカテゴリとして定義する．

3) 例画像から逐次的に特徴を抽出して，その画像の属するカテゴリを判定する．この特徴の抽出過程では，その時点の分割状況に応じて次に抽出すべき特徴が指定される．

4) 3)で判定されたカテゴリに属する標本画像を類似画像として出力する．

2.2 システム構成

図1にシステム構成を示す．

1) 画像データベース
カテゴリに分類された画像データ，および画像の特徴に関する2次データが蓄積されている．

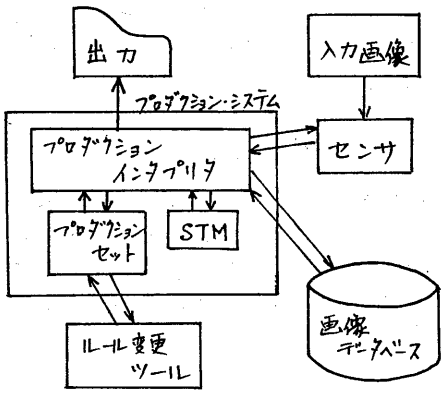

図1. 類似性評価システムの構成

2) センサ
入力された例画像からプロダクションシステムの指定する特徴を抽出し，特徴の有無，あるいは特徴値を返す．

3) プロダクションシステム
例画像の特徴に基づいて画像を分類し，そのカテゴリを判定する．判定された カテゴリ内にある画像を画像データベースから取り出し，類似画像として出力する．3.で詳細を示す．

4) ルール変更ツール
画像データベースの変化に対応するために，プロダクションルールの追加，およびプロダクションセットの変更を行う．4.で詳細を示す．

2.3 プロダクションシステムの導入

例画像の属するカテゴリを判定する過程では全ての特徴を抽出する必要はなく，カテゴリを判定するのに必要な特徴だけを抽出すればよい．つまり，その時点までの画像の分割状況に応じて次に抽出すべき特徴を指定する．また，画像をカテゴリに分類する過程では繰り返し2分割法を用いる[1]．ここでプロダクションシステムの考え方を導入する．プロダクションシステムを用いることで，

① 分割の状況に応じて次に行うべき特徴抽出処理を指定できる．

② ルールを追加・変更することで画像データベースの変化に対応できる．

といった利点がある．特に②のルールを修正する機能は，常に増大しつづける画像データベースに対し必要不可欠なものである．

画像をカテゴリに分類する方法を図2を用いて説明する．特徴f_1によって，画像集合$\{P_1, P_2, P_3, P_4, P_5\}$が$\{P_1, P_3, P_4\}$

と{P_2, P_5}の2つの集合に分割できる．特徴f_2を集合{P_1, P_3, P_4}に適用すると{P_3}，{P_1, P_4}に，特徴f_4を集合{P_2, P_5}に適用すると{P_2}，{P_5}にそれぞれ分割される．さらに特徴f_3を{P_1, P_4}に適用すれば{P_1, P_4}は{P_1}，{P_4}に分割される．このように分割を繰り返すことによって画像の分類を行う．分類の様子をグラフ表現した木を分類木という．分類木を用いることによってプロダクションシステムの動作が容易に把握できる[3]．

カテゴリに分類された標本画像のデータベースに新しいカテゴリPに属するような画像データが追加された場合，図3のように新しい特徴抽出処理fを追加してその新しい画像データを分類する．

2.4 分類木のコスト

分類木に対するコストの算定は次に従う．各特徴抽出処理に対してそのコスト（$0 \leqq cost \leqq 1$）を与える．ある画像に対する処理コストは，その画像がある1つのカテゴリと判定されるまでに行われる特徴抽出処理の和で求まる．分類木のコストは，すべての画像に対する処理コストの平均によって求まる．分類木を図4のように考えると，コストは次式によって再帰的に求まる[2]．

$$Cost = C_x + \frac{m}{m+n} C_L + \frac{n}{m+n} C_R \qquad (1)$$

ただし，

C_x：ノードxの処理コスト
C_L：部分木Lのコスト
C_R：部分木Rのコスト
m：Lに含まれる画像の数
n：Rに含まれる画像の数

画像データの追加に伴って分類木は成長してゆくが，木の成長に伴って画像の分類効率が悪くなることがある．効率の改善については4.で述べる．

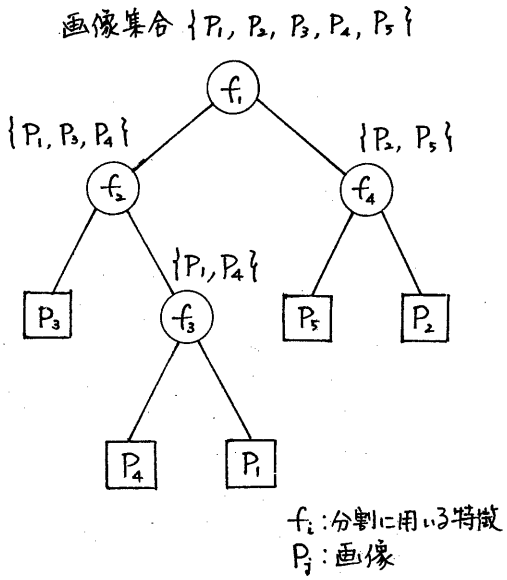

図2．繰り返し2分割法の適用例

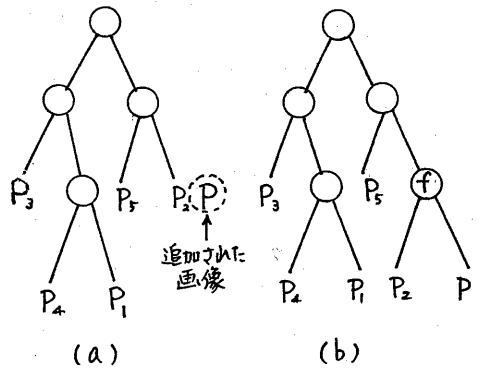

図3．画像データの追加に対する分類木の変化

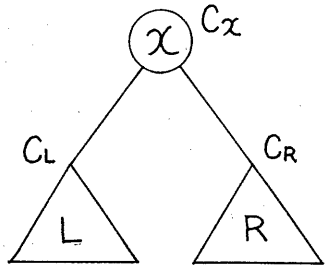

図4．分類木のコスト

3. 類似性評価のためのプロダクションシステム

3.1 システム構成

プロダクションシステムは次の3つの基本的な構成要素からなる。（図1参照）

(1) プロダクションセット
プロダクションルールの集合。その時点までの特徴に応じて次に行うべき特徴抽出処理を指定するルール，および，抽出された特徴からその画像の属するカテゴリを判定するルールからなる。

(2) STM（Short Term Memory）
例画像のその時点までに抽出された特徴が格納されている。

(3) プロダクションインタプリタ
STMの内容とマッチする条件をもつプロダクションルールを選び出し，その行動部分を実行する。

3.2 メカニズム

例画像が入力されたとき，初期記号がSTMに書き込まれる。プロダクションセットの中のプロダクションルールは，そのルールが起動される条件と，条件が満たされたときに実行される行動の対からできている。プロダクションインタプリタによって現在のSTMの状況とマッチする（すなわちプロダクションルールの条件の全ての要素がSTMの中の記号とマッチする）ルールが選ばれ，そのルールの行動部分が実行される。この行動の実行の結果，STMの状況が変化し，新たなSTMの状況とマッチするルールが逐次起動される。

あるSTMの状況とマッチするプロダクションルールは一般には複数個あると考えられる。ルールの選択もプロダクションインタプリタが行う。本システムでは，条件がマッチした複数個のルールのうちのどれを起動するかを決定する選択基準を次のように定めている。

(1) 条件の先頭がSTMに最後に入れられた記号であるようなルールを選び出す。

(2) (1)で選ばれたルールのうちで条件の全ての要素がSTMとマッチするものをコンフリクトセットに登録する。

(3) コンフリクトセットの中で条件の要素の数が最大のものを適用すべきルールとして選ぶ。

(4) (3)において条件の要素の数が最大のものが複数個ある場合は，最初のルールを選択する。

プロダクションセットは分類木で表される。プロダクションセットの各ルールは分類木の各ノードと1対1に対応している。特徴を抽出するルールが分類木のルート，あるいは中間のノードに対応し，カテゴリを判定するルールが分類木のリーフノードに対応する。

4. ルール変更ツール

画像データベースは常に変化するため，画像の分類を行うためのプロダクションセットに対するルールの追加，および変更が不可欠である。また，画像データの追加に対してルールの追加を繰り返すと，画像の分類効率が低下するという問題が生じる。

この問題の解決のために，本システムではプロダクションルール変更ツールを用意している。プロダクションセットは分類木としてディスプレイ上に視覚的に表現される。ユーザは対話的に分類木を変更することができる。改善された分類木はプロダクションルールに変換され，プロダクションセットが再構成される。プロダクションルール変更ツールには次の4つのコマンドが用意されている。

特徴	画像 P_1	P_2	P_3	P_4	P_5	P_6
f_1	1	1	0	0	1	0
f_2	1	0	1	1	1	0
f_3	1	2	1	0	1	2
f_4	1	0	1	1	0	0
f_5	0	1	1	0	2	1

表1. 真理値表の例

0, 1, 2の値からなっている. 画像がその特徴を持てば1, 持たなければ0, その特徴に対して情報がない, つまり1か0かわからなければ2を与える

	処理コスト
f_1	0.8
f_2	0.4
f_3	0.3
f_4	0.1
f_5	0.7

表 2. 特徴抽出処理のコストの例

$$H(f_i) = P(f_i) * Separate(f_i) * Cost(f_i)$$

ただし

- $P(f_i) = (m/n)^3$
- $Separate(f_i) = -4\left(\frac{m_Y}{m} - \frac{1}{2}\right)^2 + 1$
- $Cost(f_i) = 1 - (\text{処理コスト})$

ここで

n: 画像の数
m: 真理値表の値が2以外の画像の数
m_Y: 真理値表の値が1である画像の数

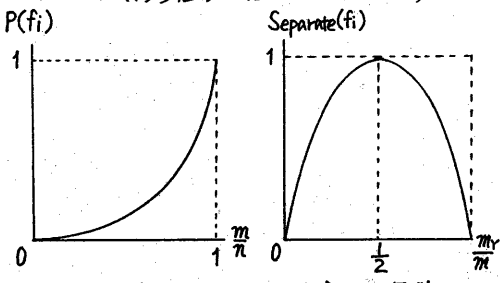

図5. 特徴f_iに対する評価関数

(1) REMAKE

各画像の各特徴に対する真理値表(表1)と特徴抽出のためのコスト(表2)を与え, 評価関数(図5)の値が最もよい特徴で画像集合を2分割する. 分割された各画像集合も同様な操作で分割する. 画像集合が同じカテゴリに属するものだけになるまでこの操作を繰り返して新しい分類木を作る.

(2) MERGE

ある1つのノードの2つの子に対応する特徴抽出処理が同じである場合, その2つのノードを1つにまとめる. 分類木のルートからリーフノードへ至るパスの同じものがある場合, MERGEを適用すれば効率を改善できる.

(3) INSERT

分類木の途中にノードを挿入する. このコマンドはこれ自体で分類効率を改善するものではないが, REMAKE, MERGEの前処理として使用することによって効果的な効率改善を行うことができる.

(4) DELETE

不要なノードを取り除く. インサートなどによって画像を1つもたないカテゴリがリーフノードに現れた場合に限り, そのノードを消去する.

以上のコマンドを用いて効率の改善された分類木が求まると, 分類木のリーフノードをカテゴリの判定ルールに, またその他のノードを次に抽出すべき特徴を指定するルールにそれぞれ対応させて, プロダクションセットを再構成する. 再構成のためのフローチャートを図6に示す.

5. 実験システム

本研究では, 類似画検索システムのうち, 画像の類似性を評価するための

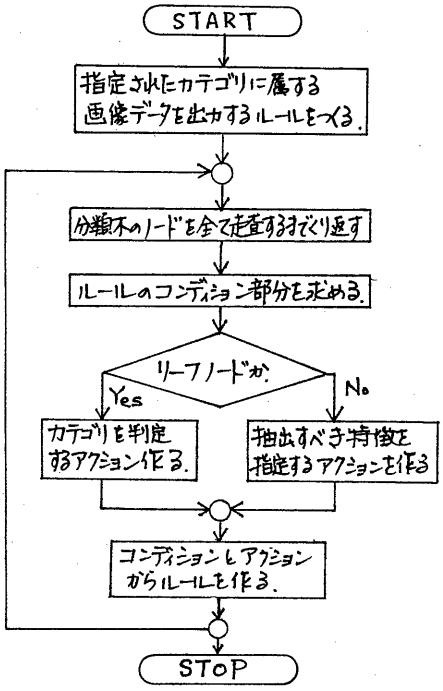

図6. 分類木をプロダクションセットに変換するフローチャート

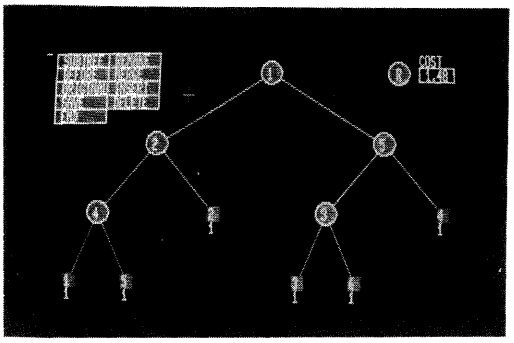

写真 1

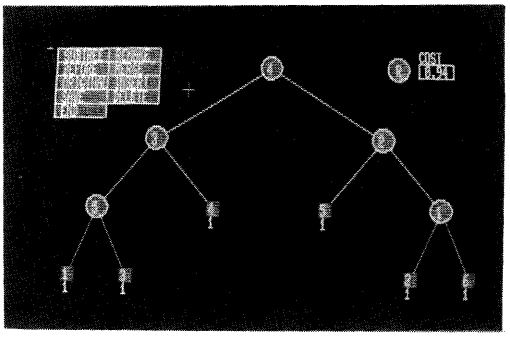

写真 2

プロダクションシステム，および画像データベースの変化に対応するためのプロダクションルール変更ツールを開発した．プロダクションシステムは，HITAC M-200H上で稼動している．画像データベースおよびセンサは本研究では開発していない．現在の実験システムでは，模擬的な画像データベースを用い，センサの役割を人間が代行している．

ルール変更ツールは，HITAC M-200HとFM-11上にインプリメントされている．FM-11とHITAC M-200Hの間は，1200ボーの専用回線によって接続されている．FM-11上にグラフィックを用いて分類木を視覚的に表示する．ユーザは対話的に分類木を効率よいものに変更することができる．実験結果の例を示す．

(1) REMAKE

写真1の分類木にREMAKEを適用した結果を写真2に示す．ここで，○の中の数字は特徴の番号，□の中の数字はカテゴリの番号，□の下の数字はそのカテゴリに属する画像の数を表している．真理値表，処理コスト，評価関数を用いて分類木を作りかえると，コストは1.48から0.94へと改善された例である．

(2) MERGE

写真3において，処理1の2つの子

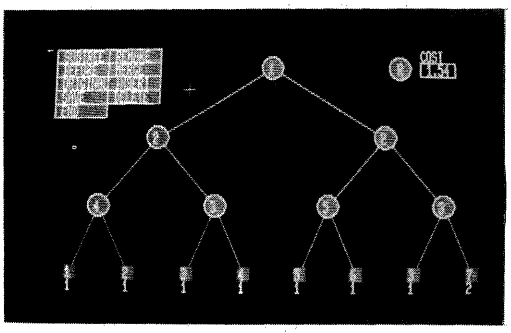

写真 3

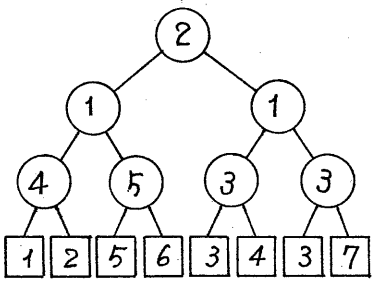

図7. MERGEの途中経過

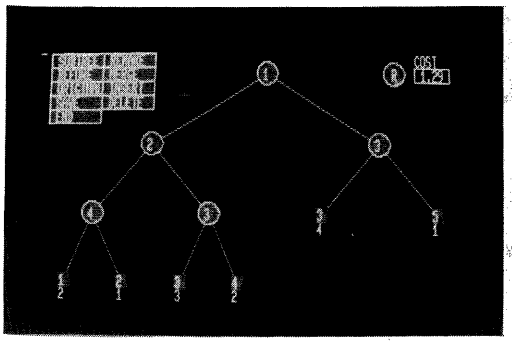

写真 5

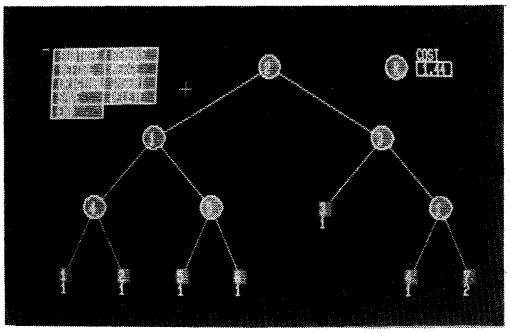

写真 4

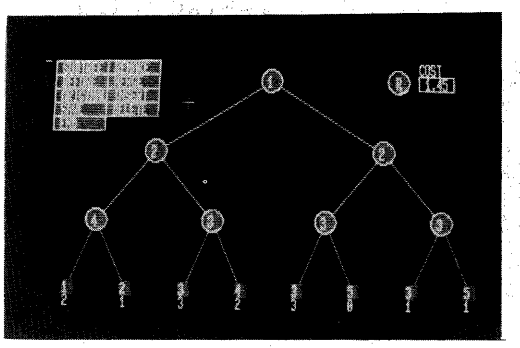

写真 6

の特徴抽出処理はどちらも処理2である．これにMERGEを適用すると図7のようになる．効率は何ら改善されていない．図7の右の部分木に対して，さらにMERGEが自動的に適用され，その結果写真4になる．処理1が1つ削除された分だけ効率が改善されている．

(3) INSERT

写真5の場合，右の部分木に処理2のノードをINSERTすると写真6になる．これにMERGEを適用すると写真7になる．

(4) DELETE

写真7のように，画像を1つもたないカテゴリが現れた場合に適用する．写真8のようになる．

最後に，実際に使用した模擬データに対する分類木を写真9に示す．効率改善後の分類木が写真10である．これらの写真中の△は部分木を示しており，

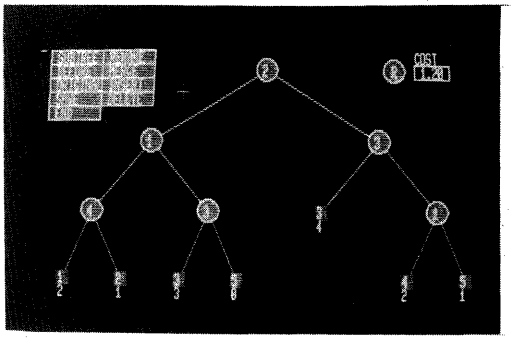

写真 7

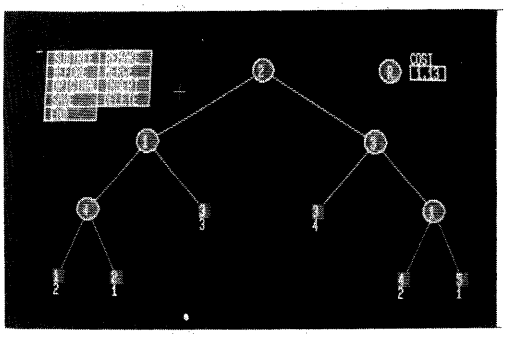

写真 8

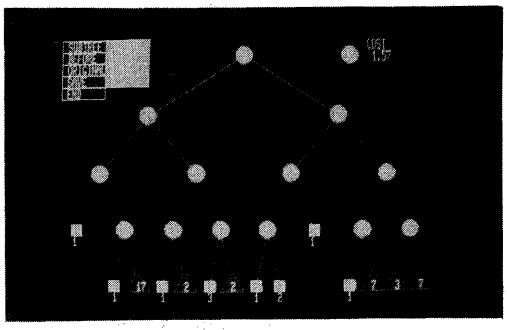

写真 9

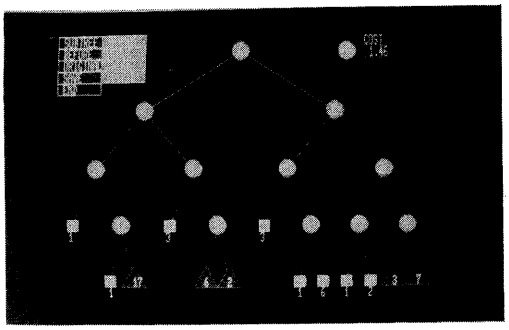

写真 10

その中の数字は部分木のルートノードにある特徴抽出処理の番号である．これらの部分木はスクロールによって任意の箇所から見ることができる．

6. おわりに

本研究では，画像データベースシステムの検索において，例画像を入力してその類似画像を検索するという方法を前提とした類似性評価のためのプロダクションシステム，および画像データベースの変化に対応するためのプロダクションルール変更ツールの開発を行った．

プロダクションシステムを用いることによって，その時点に応じた特徴抽出処理が可能となり，不要な特徴を抽出するといった無駄を省くことができた．ルールの変更については，ユーザが現在のプロダクションセットの状態を分類木として視覚的にとらえることができ，対話的に分類木を効率よいものに変更できるルール変更ツールを用意した．

なお，本報のルール変更ツールは，知識ベースのコンシステンシィを視覚的にユーザに伝える手段であるとともに，視覚情報による操作で知識ベースの保全を行う手段と見なすことができる．

今後の研究の方向として，プロダクションルールを変更することで検索効率の改善を図るとともに，検索メカニズム自体に知識を持たせる方法を用い，学習するシステムへの発展を考えている．

謝辞

本研究を行うにあたり，有益な御討論，御助言をいただいた情報システム研究室の諸氏に感謝する．また，情報処理センター計算機の使用，およびUTILISPの使用にあたり御協力いただいた関係者の方々に感謝する．

文献

1) 田中，田中；"例画像入力による画像検索について"，昭和56年情報処理学会第23回全国大学予稿集 5C-8．

柳原；"木構造による図形・画像の認識と分類"，学位論文，大阪大学，昭和59年．

3) 森島，原島，宮川，斉藤；"医療診断システムにおける診断ルール自動抽出に関する一検討"，信学技報PRL83-33（1983-10）