

## 多次元特徴量直方体の相関関係を用いた

### オブジェクト認識手法 FRCM

品田 岳史 安藤 祐樹 渡辺 俊典 菅原 研  
電気通信大学 大学院 情報システム学研究所

#### 概要

新たな画像認識手法として、画像内の認識対象の特徴を多次元直方体で表現し、複数の直方体同士間の相関量によって類似物を認識する手法 FRCM を提案する。単眼画像を用いた簡単なグレースケール 3D シーン内のオブジェクトの認識を試みた。その結果、事前のオブジェクトの切り出し無しにオブジェクトの位置、形状、個数などの変動に対してロバストな認識が行え、かつパターンマッチングのようにいくつものテンプレートを用意することなく非常に少数の教示データのもとで認識を実現できることを実証した。

## FRCM : A New Method of Object Recognition Using Correlations of Multi-dimensional Feature Rectangles

Takeshi Shinada Masaki Ando Toshinori Watanabe Ken Sugawara  
The Graduate School of Information Systems,  
The University of Electro-Communications

#### Abstract

As a new image recognition method, we propose a new method FRCM that expresses the feature of a to-be-recognized object by a multi-dimensional rectangle and recognizes it by using its similarities to pre-instructed objects' feature rectangles. Objects in gray scale monocular images of 3D scenes are used in recognition experiments. Under changes of positions, forms, numbers, etc., robust recognition power could be illustrated. Moreover, the new method needs no pre-segmentation of images, can work under a small number of instruction data without preparing many templates like traditional pattern matching.

## 1 はじめに

人間は、眼で捉えた画像から外界の情報を把握できる。この機能は、まず眼で捉えた画像から物理的な意味での3次元情報を再構成することから始まる。そしてその3次元情報が、人間内部のモデルに沿って解釈され、外界を把握し、外界のコピーが内部に構築されるというものである。

人間のこの外界把握機能の代替を目標とするコンピュータビジョン研究においても、このフレームワークに沿って研究が進められてきた。

一般に、TVカメラあるいは網膜上の画像は、3次元シーンが2次元平面上に写像され、縮退して記録されたものである。したがって、2次元の画像からもとの3次元情報を導出することは不可能である。しかしわれわれ人間は、眼で捉えた画像から外界の情報を把握できる。これは、左右2つの眼による立体視、自分と外界の相対運動による像の動き、射影による像の歪みや明暗変化など様々な手掛かりを用いていると考えられる。このような機能の原理的な解明と効率よいアルゴリズムの開発の為の研究が活発に行われている[1][2]。

この分野は一般に、単眼視に基づくものとステレオ視に基づくものとに大別される[3][4]。特に単一画像からの認識では、対象画像から得られる情報に制限があるため、対象画像の位置や明暗、形などのいくつかの解決しなければならない問題がある。その反面、汎用性や画像を得るためのコストなどを考えると非常に有用なテーマであるといえる。

従来、モノクロ単眼画像を用いた単純なオブジェクトシーンでの陰影や輪郭を抽出する研究が行われてきた[5]。対象画像を認識するための前処理として、画像全体を一定サイズの局所画像に分割し、各局所領域ごとの特徴表現を比較することに焦点を置き、画像全体を処理単位とした場合に比べ、多種多様なシーンへの対応能力を高めるとともに、認識に先立って画像からオブジェクトを切り出す必要をなくするという研究もある[6]。

本報告では、局所画像を最小全域木(MST)を利用してテキスト化を行い、それをデータ圧縮する

ことによって得られた特徴を、あらかじめ登録しておいた教示と相互比較することによって、入力データを認識することを試みる。さらに、ロバスト性の高い特徴表現手法として新たに多次元特徴量直方体という概念を導入し、それらの間の相関関係を利用したオブジェクト認識手法 FRCM (Feature Rectangle Correlation Method)を提案する。そして実験によって提案手法の有効性を確認する。

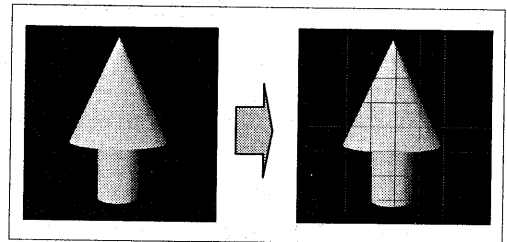


図1 画像の局所領域化(400×400, 64分割)

## 2 多次元特徴量直方体による画像表現

### 2.1 局所領域の特徴抽出方式

#### 2.1.1 画像の局所領域化

対象画像を認識するための前処理として画像全体を一定サイズの局所画像に分割し(図1)、各局所領域ごとに処理を行っていく。複雑なシーンでも局所領域には繰り返し性が高いことから、少数の教示情報で多くの未知情報を処理することができると期待される。

#### 2.1.2 MSTによる局所領域の1次元符号化

画像は通常、離散的な点である画素をグリッド状に整列させた形式で表現される。したがって、各画素を接点 $V$ 、近隣の画素間を枝 $E$ とすればグラフ表現 $G = (V, E)$ が可能である。 $E$ の各枝には隣接接点間のグレースケール値の差( $>0$ )を重みとして与える。最小全域木(MST=Minimum Spanning Tree)[7]とはグラフ $G$ の部分木であり、枝の重さの総和が最小になるようなものを指す。

MSTには、各画素につき近隣の4画素を結ぶ方法と8画素を結ぶ方法が考えられる。ここでは、陰影画像内の面の形状、陰影の変化などの特徴情

報を上手く表現するために8方向 MST を用いる。図 2 から分かるように、MST は重さの大きい枝を避けるように生成されるため、値の類似している画素同士の連結が可能であり、原画像の構成を反映した符号化が行えると考えられる。

対象画像の各局所領域を MST 化し、それを利用して1次元テキストを作成する。MST は上述した規則に従って生成され、MST が連結された順に接点を訪問していき、その都度各接点において符号を割り当てることにより符号化を行う。最終的に、各接点で割り当てられた符号を訪問した順に並べることにより1次元テキストを得る。

各接点での符号の割り当ては図3に示すように、画素の階調値と親接点の方向を考慮して行う。ま

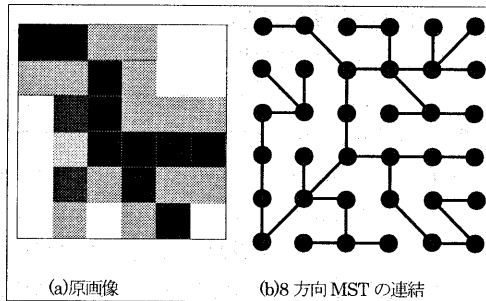


図2 8方向MSTの生成

| 親接点での階調値の範囲 | 親の方向         | 親接点での符号語 | 親接点での階調値の範囲 | 親の方向         | 親接点での階調値の範囲 |
|-------------|--------------|----------|-------------|--------------|-------------|
| 0-31        | 下(down)      | .        | 128-159     | 下(down)      | z           |
|             | 右(rightdown) | ·        |             | 右(rightdown) | [           |
|             | 右(right)     | <        |             | 右(right)     | ∨           |
|             | 右(rightup)   | =        |             | 右(rightup)   | ]           |
|             | 上(up)        | >        |             | 上(up)        | ∩           |
|             | 左(leftup)    | ?        |             | 左(leftup)    | ∪           |
|             | 左(left)      | @        |             | 左(left)      | ∪           |
|             | 左(leftdown)  | A        |             | 左(leftdown)  | a           |
| 32-63       | 下(down)      | B        | 128-159     | 下(down)      | b           |
|             | 右(rightdown) | C        |             | 右(rightdown) | c           |
|             | 右(right)     | D        |             | 右(right)     | d           |
|             | 右(rightup)   | E        |             | 右(rightup)   | e           |
|             | 上(up)        | F        |             | 上(up)        | f           |
|             | 左(leftup)    | G        |             | 左(leftup)    | g           |
|             | 左(left)      | H        |             | 左(left)      | h           |
|             | 左(leftdown)  | I        |             | 左(leftdown)  | i           |
| 64-95       | 下(down)      | J        | 192-223     | 下(down)      | j           |
|             | 右(rightdown) | K        |             | 右(rightdown) | k           |
|             | 右(right)     | L        |             | 右(right)     | l           |
|             | 右(rightup)   | M        |             | 右(rightup)   | m           |
|             | 上(up)        | N        |             | 上(up)        | n           |
|             | 左(leftup)    | O        |             | 左(leftup)    | o           |
|             | 左(left)      | P        |             | 左(left)      | p           |
|             | 左(leftdown)  | Q        |             | 左(leftdown)  | q           |
| 96-127      | 下(down)      | R        | 224-255     | 下(down)      | r           |
|             | 右(rightdown) | S        |             | 右(rightdown) | s           |
|             | 右(right)     | T        |             | 右(right)     | t           |
|             | 右(rightup)   | U        |             | 右(rightup)   | u           |
|             | 上(up)        | V        |             | 上(up)        | v           |
|             | 左(leftup)    | W        |             | 左(leftup)    | w           |
|             | 左(left)      | X        |             | 左(left)      | x           |
|             | 左(leftdown)  | Y        |             | 左(leftdown)  | y           |

図3 使用する符号語

ず画素の階調値(256)を同間隔の8段階に分類し、その値に親接点から次の接点の方向を加える。つまり記号の数は、画素の階調段階(8)に次の接点の方向(8)をかけた64とする。これにより生成した1次元符号語には局所領域の形状的な特徴と色の特徴とを同時に反映させることができる。

## 2.2 局所領域の特徴ベクトル表現法

### 2.2.1 階層文法抽出データ圧縮法

各局所領域の情報を情報特徴空間内での特徴ベクトルとして表現する[8][9]。ここで使用する情報特徴空間は、認識対象となる陰影画像の各局所領域に含まれている形状や陰影などの特徴情報を効率的に反映するために階層文法抽出データ圧縮法[10]を用いる。階層文法抽出データ圧縮法はテキストデータを圧縮する際に対象となるデータに繰り返し現れるあるパターンに対し、階層文法を適用することで圧縮を行うものである。それにより、対象データは短い量のデータで出力され、データの繰り返されるパターンを反映した階層文法が同時に生成される。

図4に階層文法抽出データ圧縮法を示す。まずストリングSに入力データを1文字ずつ加えていく。その過程でSの最後の2文字が前に現れたものであれば、2文字を1文字に置き換える文法を作成し置き換える。このように対象データの最初から読みながら末尾の2文字を処理対象としてい

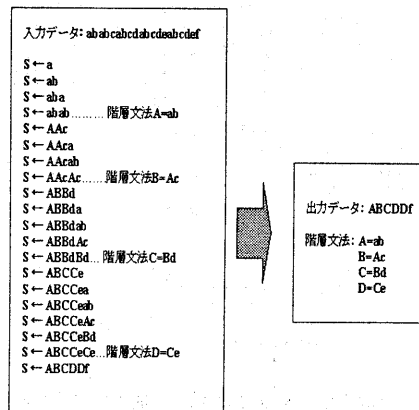


図4 階層文法抽出データ圧縮法

くことから、アルゴリズムは高速でもある。結果として、入力データの構造の特徴を強く反映させた出力ストリングと階層文法が生成される。この階層文法を用いて、入力データの構造を強く反映する1つの指標（生成文法ルール）を生成する。

### 2.2.2 特徴ベクトルの生成

情報特徴空間内で基底を作るための入力テキストを基底データ、基底データから得られた複数の生成文法ルールのことを基底生成文法ルール集合と呼ぶ。情報特徴空間とは、異なるテキストから得られる基底生成文法ルール集合を利用して形成される多次元空間である(図5)。その際に、1つの基底生成文法ルール集合に対し以下の3つの特徴量を算出する。

- 文法の重み付き平均使用率 ( $S$ )
- 文法の重み付き平均使用率の標準偏差 ( $S.D$ )
- 文法を用いた圧縮率 ( $C$ )

以上の特徴量の算出法を図5を用いて説明する。まず基底データとしてテキスト  $T_1 \sim T_n$  を用意する。それぞれに対し、階層文法抽出データ圧縮法

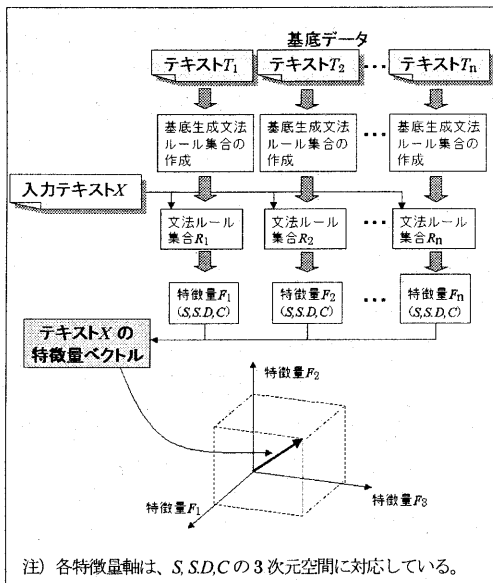


図5 特徴量ベクトルの生成

を用いて基底生成文法ルール集合を作成し、各ルール集合についてその要素文法ルールが使用された回数の多い順に重み付けを行う。ここに未知の入力テキスト  $X$  を入力し、それぞれの基底生成文法ルール集合を用いて、要素文法ルールが使用された回数の多い順に重み付けを行い、それを基底生成要素文法ルールの順に並び替える。各基底生成要素文法ルールごとの重みを  $W$ 、文法使用回数を  $N$  とすると文法の重み付き平均使用率  $S$  は、

$$S = \frac{\sum(W \times N)}{N} \quad (1)$$

と求まり、これを用いて  $S$  の標準偏差  $S.D$  を求める。また圧縮率  $C$  は、圧縮後のテキスト長を圧縮前のテキスト長で割った値とする。

### 2.3 多次元特徴量直方体による表現 (画像内矩形領域の表現)

2.2.2 で得られた特徴量ベクトルと画像平面を用い、多次元のオブジェクト検索空間を生成する(図7)。この空間内であらかじめ登録しておいたオブジェクトの多次元直方体と未知の入力オブジェクトの多次元直方体とを相互比較することにより入力オブジェクトの認識を行う。まず画像内の局所領域の集合 (矩形領域) において、これに含まれる局所領域の特徴量ベクトルの最大値と最

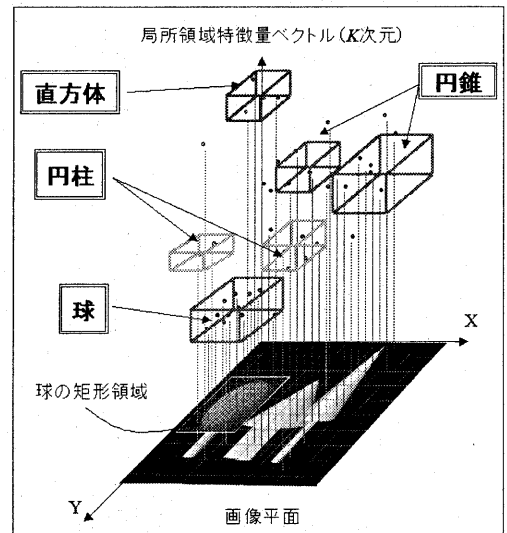


図7 オブジェクト検索空間

小値を求める。この得られた多次元領域の積を求めることで矩形領域の多次元特徴量直方体表現が可能である。以上のことを数式化すると以下の式(2),(3)のようになる。

画像の特徴量ベクトルの次元数を  $K$ 、分類しようとするクラス数を  $n$ 、各分類クラスを  $c_i$  ( $i=1,2,\dots,n$ )とする。ある分類クラス  $c_i$  に属する局所領域は、 $m$  番目の次元  ${}^m x$  において、最小値  ${}^m x_{i \min}$ 、最大値  ${}^m x_{i \max}$  の領域にあると仮定すると、この領域  ${}^m x_i$  は式(2)で表される。

$${}^m x_i = \{ {}^m x \mid {}^m x_{i \min} \leq {}^m x \leq {}^m x_{i \max} \} \quad (2)$$

(ただし、 $m=1, 2, \dots, K$ )

ここで、 $K$  次元空間を考え、式(3)で定義される多次元直方体に囲まれた領域に属する局所領域を分類クラス  $c_i$  に属するものと分類する。

$$x_i = {}^1 x_i \cap {}^2 x_i \cap {}^3 x_i \cap \dots \cap {}^K x_i \quad (3)$$

2.2.2 の方式と式(2),(3)を用い、3つのそれぞれの特徴量ベクトルの  $m$  番目の次元における領域を、

$$\begin{aligned} {}^m R_{in\_S_i} &= \{ {}^m R_{in\_S_i} \mid {}^m R_{in\_S_i \min} \leq {}^m R_{in\_S_i} \leq {}^m R_{in\_S_i \max} \} \\ {}^m R_{in\_S.D_i} &= \{ {}^m R_{in\_S.D_i} \mid {}^m R_{in\_S.D_i \min} \leq {}^m R_{in\_S.D_i} \leq {}^m R_{in\_S.D_i \max} \} \\ {}^m R_{in\_C_i} &= \{ {}^m R_{in\_C_i} \mid {}^m R_{in\_C_i \min} \leq {}^m R_{in\_C_i} \leq {}^m R_{in\_C_i \max} \} \end{aligned} \quad (4)$$

(ただし、 $m=1, 2, \dots, K$ )

と表すことができる。ここでの  $K$  は次元数、すなわち基底数を表している。式(4)より最終的に得られる多次元特徴量直方体  $V_{in}$  は、画像上での局所領域の大きさ (長方形) を  $SIZE_{in}$  とすると、

$$V_{in} = {}^m R_{in\_S_i} \times {}^m R_{in\_S.D_i} \times {}^m R_{in\_C_i} \times SIZE_{in} \quad (5)$$

となる。この  $V_{in}$  が入力局所画像の検索空間内での表現である。

また、あらかじめオブジェクト検索空間内に登録しておく事例に関しても3つのそれぞれの特徴量の  $m$  番目の次元における領域を、 ${}^m R_{case\_S_i}$ 、 ${}^m R_{case\_S.D_i}$ 、 ${}^m R_{case\_C_i}$  とすると多次元特徴量直方体  $V_{case}$  は、画像上での局所領域の大きさを  $SIZE_{case}$  すると、 $V_{case} = {}^m R_{case\_S_i} \times {}^m R_{case\_S.D_i} \times {}^m R_{case\_C_i} \times SIZE_{case}$  (6) となる。この  $V_{case}$  が登録画像の検索空間内での表現である。

### 3 認識方式

#### 3.1 検索空間内での多次元直方体の類似度

登録データの  $V_{case}$  と入力データの  $V_{in}$  との相関関係を表す値を FRC (Feature Rectangle Correlation) を式(5)のように定義する。

$$FRC_{case,in} = -\log \left( \frac{V_{case} \cap V_{in}}{V_{case} \cup V_{in}} \right) \quad (7)$$

FRC 値が小さいほど多次元空間内における登録データの多次元直方体内に、入力データの多次元直方体が含まれている度合いが高い。このことを利用して、未知入力オブジェクトと複数の登録オブジェクトとの間の FRC 値を求めていき、FRC 値が最小の登録オブジェクトが未知入力オブジェクトと似ていると考えることにする (図8)。

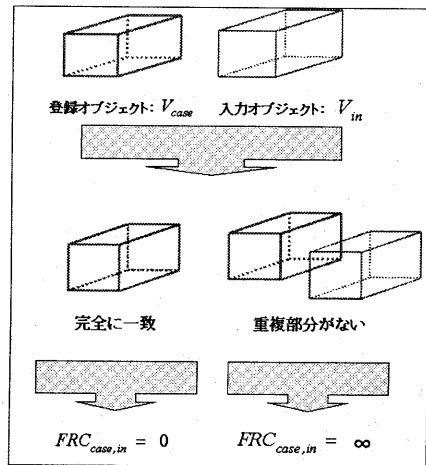


図8 FRC 値の比較

#### 3.2 認識アルゴリズム

登録オブジェクトをその形状から3つの階層に分けて、それぞれ登録しておく。それぞれの階層におけるオブジェクトの例を図9に示す。認識は、階層ごとに登録オブジェクトと入力オブジェクトとを比較し、FRC 値の高い登録オブジェクトに入力オブジェクトを帰属させることで実現する。なお、本報告では3階層処理を例として示すが、必要に応じて階層数を増やすことも容易である。

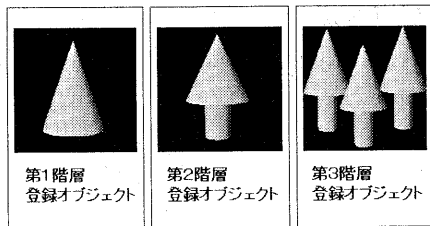


図9 各階層における登録オブジェクトの例

- ① 第1階層 (認識対象は画像内の各局所領域)
- 各局所領域ごとに多次元直方体を生成し、第1階層に登録してあるオブジェクトとの間のFRC値を求める。その結果得られたFRC値を比較し、最小の値のときの登録オブジェクトをその局所領域での第1候補とする。ただし、1つの局所領域に着目したとき、得られる特徴量は単一の多次元

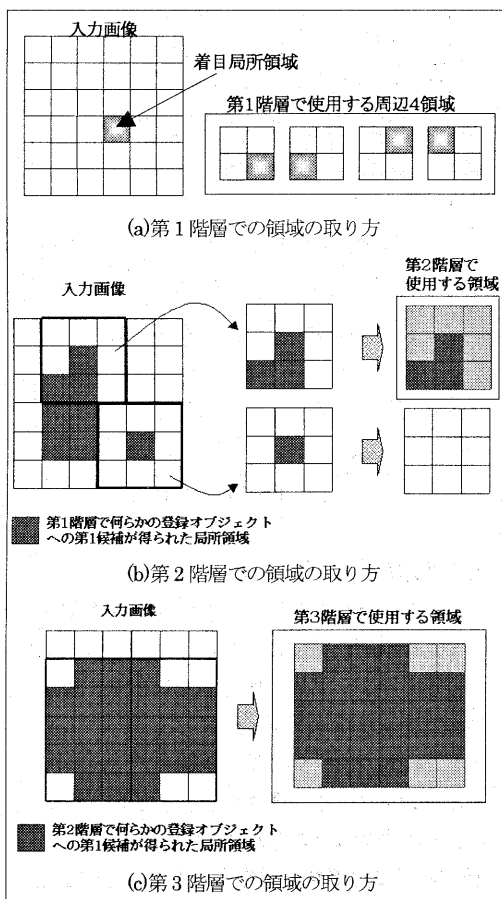


図10 各階層での領域の取り方

ベクトルであるため多次元直方体を形成できない。そこで、対象となる局所領域の近傍で  $2 \times 2$  の領域をとり (1つの局所領域に対して、最大で周囲4領域をとる)、多次元直方体を形成する (図10(a))。この多次元直方体 (4領域~1領域) と第1階層の登録オブジェクトのFRC値を求める。局所領域を中心とする  $3 \times 3$  の領域を新たな認識も着目局所領域と同様の第1候補があれば、着目対象領域とする (図10(b))。逆に近傍に同様の第1候補がない場合には、局所領域が誤認識されているとみなし無視する。得られた  $3 \times 3$  の領域と第2階層の登録オブジェクトとのFRC値を比較し、第1階層と同様に  $3 \times 3$  領域に対する第1候補を取得する。

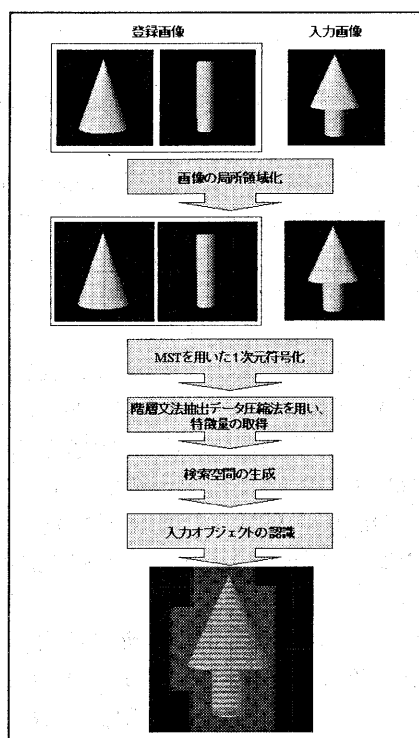


図11 実験システムの概要

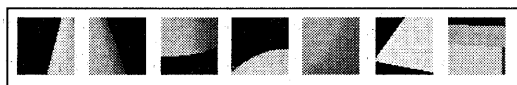


図12 実験で使用する基底画像

② 第3階層（認識対象は画像内の②で得られた領域の外接長方形領域）

同様に第2階層で特定の登録オブジェクトを第1候補とした局所領域のみを対象とする。その際、認識対象領域は第2階層での同一オブジェクトに対する第1候補の領域の外接長方形を用いる（図10(c)）。その外接領域と第3階層の登録オブジェクトとの FRC 値を比較、シーンを認識する。て扱う。つまり、局所領域のサイズは  $50 \times 50$  画素である。図5でも示した基底データとして7個の入力データを用い、図12にその画像を示す。1つの基底に対し2.2.2で述べた3つの特徴量を用いていることから、画像平面内の局所領域のデータ2個を合わせた23次元( $7 \times 3 + 2$ )のデータを扱うことになる。また本実験では、2.3で述べた  $SIZE_{case}$  の大きさを画像平面全体と等しくとる ( $400 \times 400$  画素)。これは非常に少数の教示データでサイズの異なる入力オブジェクトを認識させたいためである。

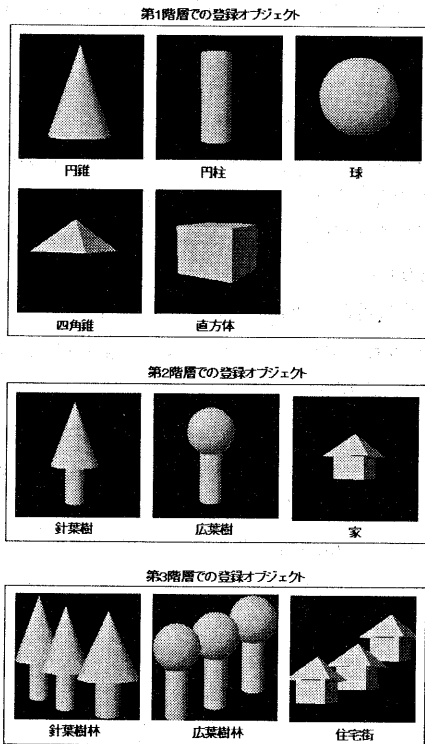


図13 実験で使用する登録画像

#### 4 実験結果及び考察

図11に実験システムの概要を示す。ここで使用する全ての画像は、 $400 \times 400$  画素の画像である。その画像を64分割したものを局所領域とし提案法の有効性を確認するため、複合オブジェクトとしてシーン全体を針葉樹林であると認識すること

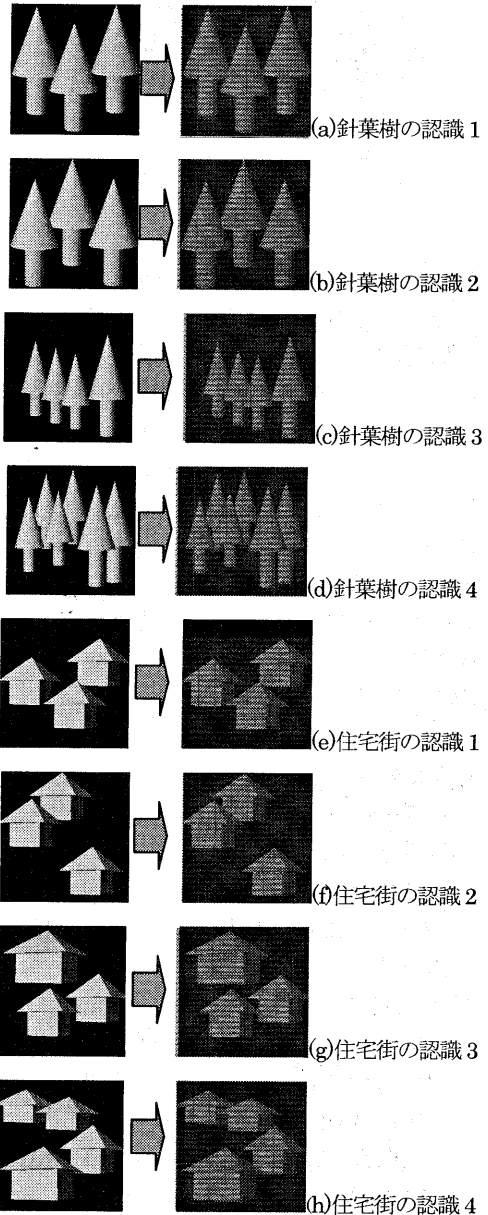


図14 複合オブジェクトで構成されたシーンの認識

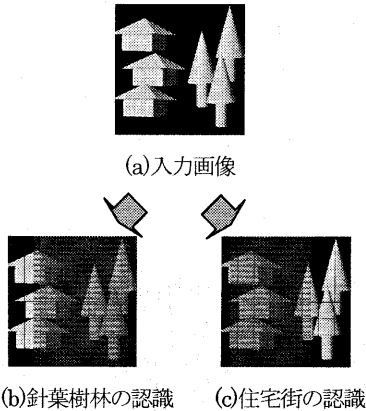


図 15 数の複合オブジェクトで構成されたシーンの認識

ができた。このことは、図 14(e)~(h)の住宅街の認識においても同様のことがいえる。続いて、複数の複合オブジェクトで構成されたシーンの認識を行う。入力画像として図 15(a)のように針葉樹林と住宅街という 2つの複合オブジェクトで構成されたシーンを用いる。図 15(b),(c)より FRCM はシーンをこのように樹林と住宅街とにそれぞれ分割して、双方を別々のオブジェクト集合であると認識していることが分かる。

以上の実験結果より、多次元特徴量直方体とそれらの FRC 値を用いた階層的な認識手法の有効性が確認できた。

## 5 まとめ

画像認識の新技术として FRCM を提案し、以下の有効性を確認した。

- ・ 認識領域の特徴を多次元特徴量直方体で表現
- ・ 多次元特徴量直方体間の相関関係を FRC 値で表現
- ・ 局所領域レベルからシーンレベルに向かって、FRC 値による良好な教示オブジェクト選択
- ・ 事前のオブジェクト切り出しが不要
- ・ ごく少量の教示データでオブジェクトのサイズや位置、構成の異なるシーンに対するロバストな認識

## 参考文献

- [1]L.G.Brown, "A survey of Image Registration Techniques",ACM Computing Surveys,vol.24, no.4, pp.325-376, 1992.
- [2]大田友一, 池内克史, "3次元情報の再構成とモデル化", Oplus E, pp.86-105, 1985.
- [3]B.K.P.Horn, "Hill-Shading and the Reflectance Map", in Proc. of DARPA Image Understanding Workshop, pp.79-120, 1979.
- [4]J.Weng,N.Ahuja,T.S.Huang,"Matching Two Perspective Views",IEEE TPAMI,vol.14,no.8, pp.806-825, 1992.
- [5]K.Ikeuchi, B.K.P.Horn, "Numerical Shape from Shading and Occluding Boundaries", Artificial Intelligence, vol.17, pp.141-181, 1981.
- [6]李容旻, "シーンの局所の特徴を利用した立体認識手法の研究", 電気通信大学大学院 平成 11 年度修士論文, 2000.
- [7]J.R.Leasch,A.E.Iverson,B.N.Webb,K.F.West, "Segmentation of Multiband Imagery using Minimum Spanning Trees", Algorithm for Multispectral and Hyperspectral Imagery II, Proc. SPIE2758, pp.10-18, 1996.
- [8]渡辺俊典, "データ圧縮を用いたパターンの万能特徴表現方式 UPDC", 日本写真測量学会平成 12 年度年次学術講演会発表論文集, pp.253-258, 2000.
- [9]T.C.Bell,J.G.Cleary,L.H.Witten,"Text Compression", Prentice Hall Advanced Reference Series, 1990.
- [10]C.G.Nevill-Manning,I.H.Witten,"Compression and Explanation using Hierarchical Grammars",THE COMPUTER JOURNAL, vol. 40, no.2/3, 1997.