多重解像度と固有空間表現による3次元物体のイメージスポッティング

村瀬 洋[†] シュリー K. ナイヤー[††]

本論文では，2次元照合により任意の方向を向いた3次元物体を複雑背景から切り出し，同時にその方向と大きさを検出する手法について述べる．3次元物体は見る方向により見かけ画像が大きく変化するため単純な2次元照合により物体の位置を検出することは困難である．一方，すべての見え方を辞書に登録し，それと画像のすべての部分領域とを照合する手法も考えられるが，この手法では記憶容量，計算量の点で実現が容易でなく，従来あまり試みられてこなかった．ここでは，変形する画像系列の集合を固有空間上での多様体で記述するパラメトリック固有空間法を提案した．これにより3次元物体を2次元画像の集合体として少ない記憶容量で記憶できる．この表現法により，幾何学的な特徴抽出をすることなく，2次元照合により多様な見え方を持つ3次元物体を複雑な背景画像から切り出し，同時に物体の方向を推定をすることが可能となる．本論文では，パラメトリック固有空間表現を多重解像度表現した画像に階層的に適用し物体領域を切り出す手法を提案し，実験結果によりその効果を示した．

Image Spotting of 3D Objects Using Multi-Resolution and Eigenspace Representation

HIROSHI MURASE[†] and SHREE K. NAYAR[††]

This paper proposed a novel method to find three-dimensional object regions from a complex image and simultaneously to measure their direction and size. We refer to this process as image spotting. The method consists of two stages: object learning and image spotting. In the learning stage, for a sample object to be learned, a large set of images is obtained by varying pose and size. This large image set is compactly represented by a manifold in compressed subspace spanned by eigenvectors of the image set. This representation is called the parametric eigenspace representation. In the image spotting stage, a partial region in an input image is projected into the eigenspace, and the location of the projection relative to the manifold determines whether this region belongs to the object, and what its pose is in the scene. This process is sequentially applied for whole places with different resolutions of the input image. Experimental results show this method accurately extracts the target objects in arbitrary pose and size.

1. はじめに

3次元物体を2次元画像から認識する3次元物体認識は産業用ロボットの要素技術や一般環境内での物体の監視技術など幅広い応用がある．画像中の物体を認識するためにはまず背景から物体を切り出す処理が必要となる．この切り出し処理は画像処理においては一般的な処理であるため，従来より後段にくる処理内容に対応して多数の研究がなされてきた．ここでは特に物体認識のための切り出し処理に着目する．これを実現するための単純な手法としては，背景画像との差分を取る手法，あるいは物体が背景に比較して濃淡値が異なる場合にはしきい値などを利用する手法が考えられるものの，これらの単純な手法では一般環境のように複雑に変動する背景の中から物体を切り出すことは困難である．この処理を精度よく行うためには，物体を認識することにより物体を切り出す手法が必要となる．つまり，物体の切り出しは物体の認識と同時に行われるべきである．

一方，物体認識に関しては，これまで多数の研究がなされてきた[1),2)]．従来の3次元物体認識は大別すると，物体のモデルと入力画像との照合に3次元モデルを利用する手法と2次元モデル[3),4)]を利用する手法に分類される．上記の認識手法の分類にもとづき物体の切り出し手法を分類してみると以下のようになる．

† NTT 基礎研究所
NTT Basic Research Laboratories
†† Columbia University

3次元モデルを利用する方法は，2次元画像からまずエッジやコーナなどの幾何学的特徴や，表面の3次元的特徴を抽出し，これとあらかじめ用意してある3次元モデルとを照合するものである．このアプローチでは，物体の部分特徴を基本的に使用するために，物体の切り出しや物体の隠れにも対処しやすいという長所は持っているものの，2次元画像から3次元特徴を精度よく抽出する処理は容易でなく，現在も研究レベルにとどまっている．また3次元モデルの作成も同様の理由から自動的に行うことは容易ではない．

他方，3次元物体の見かけの2次元画像をあらかじめ2次元モデルとして記憶しておき入力画像とこれとを比較する手法が考えられる．2次元の画像に着目して物体の領域を検出する手法をここでは特に「イメージスポッティング」と呼ぶことにする．イメージスポッティングを行うためには，テンプレートマッチングや，マッチトフィルタなどを利用して物体領域を切り出す手法が考えられる．しかし，3次元物体はその向きにより物体の見かけが大きく変化するため，もし仮に見かけ画像のあらゆる場合を登録すれば膨大な画像データ量となる．そのうえ入力画像中での物体の大きさが不明の場合には，その大きさの変化までも登録して照合しなければならず，更に膨大な画像間の照合が必要となる．そのため，記憶量，計算量の観点から，これらの手法はあまり実際的ではない．本論文で提案する手法は，見かけ画像の膨大なデータを圧縮表現して，この表現を用いて入力画像と2次元画像モデルとの2次元照合（画像間相関）を行うことにより3次元物体をイメージスポッティングする手法である．

従来，2次元照合の観点から効率的な領域抽出を目指した研究としては，まずピラミッドを使ったテンプレートマッチング[5]による手法があるが，従来の研究ではエッジやコーナなどのサイズ不変な特徴の抽出に限られ，ここで扱うようなサイズの変化によって見かけ画像が変化するような物体の抽出ではなかった．また，マッチトフィルタの計算時間を高速化するために2段階処理を利用した領域抽出[6]や，Coarse-fineによるテンプレートマッチング[7]が報告されているがこれらの手法は物体の任意の方向に対処するものではなかった．また，顔画像を対象としてニューラルネットを使った手法[8]なども報告されているが，この手法はヒューリスティックな特徴記述を基本とした手法で，本手法のように2次元の画像間相関の観点から物体を抽出する手法ではなかった．

本手法では，物体の方向や物体の大きさに対して多様に変化する2次元の見かけ画像を，固有空間（固有ベクトルにより構成される空間）上での多様体でコンパクトに表現するパラメトリック固有空間表現を導入する．このコンパクトに表現された空間上での演算により，高速に2次元照合が可能となる．ここではこの考え方に基づき，任意の方向を向いた3次元物体を複雑な背景から切り出す手法を提案した．

本手法は部分空間法によるパターン認識[9),10)]と関係が深い．従来，画素値の固有ベクトルを認識へ応用した例としては，投影法や部分空間法による文字認識手法[11)]，あるいはEigenface法による顔画像認識手法[15),16)]などが知られている．しかし，これらはいずれもパターンの分類に主眼をおいたものであり，本手法のように3次元物体の領域を切り出したり，物体の向きなどのパラメータを検出したりするものではなかった．

本論文では，まず3次元物体を2次元画像の集合体で効率よく表現するパラメトリック固有空間法について述べる（2，3章）．次にこのアイデアを，複雑背景からの物体領域の切り出し（スポッティング）に適用した例（4章）について述べ，その実験結果（5章）について述べる．

2. パラメトリック固有空間による物体の表現

3次元物体の見かけの画像は，その物体の方向により大きく変動する．例えば，ある物体を一回転させただけで図1に示すような多様な画像が得られる．これをいかに記憶するかが，ここでの学習の問題となる．ここでは符号化を基本とする画像の表現法としてパラメトリック固有空間法を提案する．これは膨大な2次元の画像集合からその画像の情報の本質を抽出し，少ない記憶容量で記憶するという画像符号化と同様の発想に基づくものである．パラメトリック固有空間法は近年，物体の認識[13),14)]のために開発された画像表現

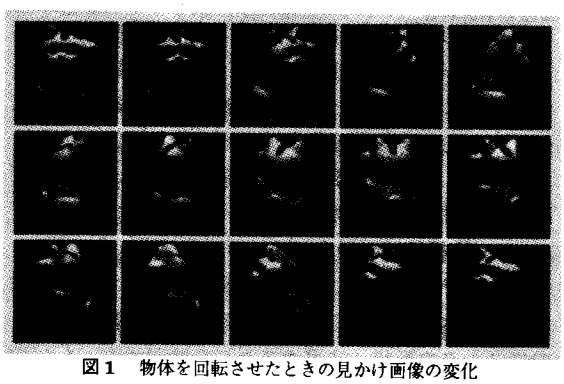


図1 物体を回転させたときの見かけ画像の変化
Fig. 1 A variety of the appearance when varying the pose.

法であるが，イメージスポッティングに対しても有効な表現法と考えられる．イメージスポッティングでは物体の大きさ（カメラから物体までの距離に対応）も不明である．そのため，認識の場合と異なり，ここでは物体の方向と物体の大きさをパラメータとするパラメトリック固有空間を構成し，イメージスポッティングの為の表現として活用する．

処理は大きく分けて，物体を学習する段階と，学習した情報を利用して実際にイメージスポッティングする段階から構成される．学習段階では，入力画像の集合からパラメトリック固有空間を構成する．この処理は2段階から構成される．1段階目は学習画像集合から固有ベクトルによる部分空間（固有空間）を構成する段階，2段階目は連続的に変化する学習画像の系列を固有空間上に投影し，この空間上の多様体（曲線や曲面等）によりもとの画像系列を表現する段階である．イメージスポッティングの段階では，まず入力画像中の部分画像に着目してこの部分画像を固有空間に投影し，次にこの点が固有空間上の多様体にのっているかどうかで，その領域が物体かどうか判別される．この注視領域を多重解像度表現された入力画面中でラスタスキャンし同様の処理を繰り返すことによりイメージスポッティングは終了する．

3. 物体の学習

3.1 学習データの収集

まず物体のサンプルから学習に用いる画像データを収集する．ここでは物体をターンテーブルの上に乗せそれを一回転させ黒い背景のもとでデータを取り込む．今回は台に置いたときに安定な状態を持つ物体の場合を考えた．もし安定状態を持たず連続的な傾きを許すような物体に対しては傾いた状態の画像も収集する必要があるが，これは今後の検討事項とする．更に大きさの変動に対処するために，ここではその物体に対して大きさの異なる画像の集合も計算機上で生成する．ここでは一例として6段階のサイズ（例えばもとの画像に対して1倍，1.1倍，...，1.5倍の拡大率）で，すべての方向の物体の画像を拡大することにより生成した．そのようにして準備した画像データセットの例を**図2**に示す．これを学習データセットとする．

3.2 注視領域の設定

本手法では基本的にはフィルタを利用して背景画像中から物体を抽出する手法である．ここでは，まずそのフィルタの形状（注視マスク）を検討する．もし注視マスクを大きくしすぎると注視領域に背景画像が混入し物体との相関値が正しく評価できない．また注視

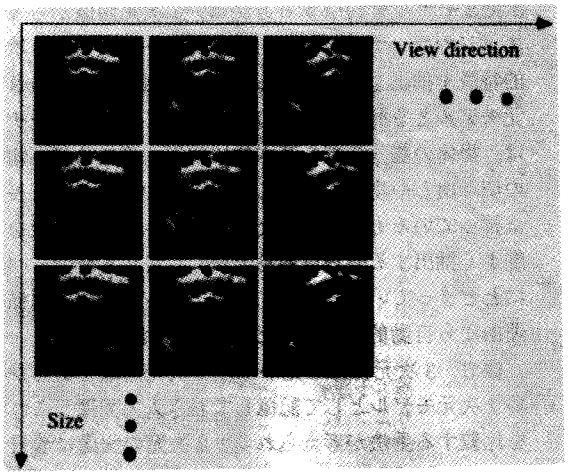
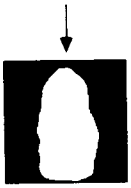


図2 学習データセットの例
Fig. 2 A learning image set.

(a) Object region of the learning images



(b) Search window

図3 注視マスクの例
Fig. 3 An attention window.

マスクを小さくすると有効な特徴を持った領域が注視マスク内に入らない場合が発生し，やはり相関値が正しく評価できない．単純に円形の注視マスクを設定してこの内側を注視領域とする手法も考えられるが，ここではなるべく領域面積を大きく且つ背景の影響を除去するために以下のように対象に適応した注視マスクを導入する．注視マスクの作成方法の例を以下に示す．まず個々の物体の画像 $f(x.y)$ にガウスフィルタをかけ，それをしきい値処理し $h(x,y)$ を得る．

$$h(x,y)=\begin{cases}1 \text{ if } \int \frac{1}{2\pi\sigma^2}\exp(-\frac{(x-u)^2+(y-v)^2}{2\pi\sigma^2}) \\ \qquad\qquad \times f(u,v)dudv \geq Th \\ 0 \text{ else}\end{cases}$$

図3(a)に学習サンプルにガウス分布を重畳しそれをしきい値処理（しきい値 Th）した画像の例を示す．次に，画像集合全体でこの2値パターンのAND領域をとることにより注視マスクを作成した．図3(b)に注視マスクの例を示す．物体の形状によっては1回転

する時にその領域形状が変化し，ANDにより作成される注視マスクの大きさが非常に小さくなる場合が考えられるが，その場合には複数のクラスタに分けて注視マスクを生成することにより解決される．例えば0度から100度まではある形状，100度から360度までは別の形状とし，後の処理をそれぞれに対して行う．

3.3 画像の正規化

学習データの画像はまず注視マスクにより取り出される．この領域の画素値をラスタスキャンし，画素値を要素とするベクトル $\hat{\mathbf{x}}$，

$$\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N]^T$$

でもとの画像を表現する．ここでベクトルの次元 N は注視マスク内の画素数であり，マスクの面積に依存する．

次に，センサ感度の影響を除去するために，明るさの正規化を行う．正規化後の画像を $\mathbf{x}$ とすると，ここではベクトル $\mathbf{x}$ の大きさ，つまり画像のエネルギーが1 ($\|\mathbf{x}\| = 1$) となるように式，

$$\mathbf{x} = \frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|}$$

により正規化する．

3次元物体の見かけ画像は，物体の向きによって変化する．ここでは物体の向きが1軸の周りで回転する場合を考え，これを第1パラメータとする．物体の任意のポーズを扱うためには，回転の軸数つまりパラメータ数を増やすことにより拡張できる．ここで物体の大きさがカメラからの距離に比較して小さいような弱いパースペクティブを仮定をすれば，物体の見かけはカメラから物体までの距離により大きさだけが変化する．各方向の画像に対して物体の見かけの大きさの変化に対応した画像を生成し，これを第2パラメータとする．ここで物体を1回転し，さらに大きさを変化させた画像集合を，

$$\{\mathbf{x}_{1,1}^{(p)}, \cdots, \mathbf{x}_{R,1}^{(p)}, \mathbf{x}_{1,2}^{(p)}, \cdots, \mathbf{x}_{R,L}^{(p)}\}$$

で表現する．R は回転方向の刻み数を，L は大きさの刻み数を表す．また画像の見かけは物体の種類によっても当然異なる．これを第 p 物体の画像集合と呼ぶ．また各画像ベクトルを学習サンプルと呼ぶ．我々の実験ではこの学習サンプルの収集に，計算機制御で回転可能なターンテーブルを用いた．つまり学習サンプルは，対象となる物体をターンテーブル上に乗せることにより，すべて自動的に収集できる．

3.4 固有ベクトルの計算

図1の画像系列の例からもわかるように隣会った2つの画像は極めて相関が高い．まず第1段階としてこの相関の性質を利用して，画像を圧縮する．ここでは画像集合に対して，2乗誤差の観点から最適に圧縮することが可能なKarhunen-Loeve展開を採用する．これは，画像集合の共分散行列の固有ベクトルが張る部分空間（固有空間）により，もと画像を表現しようとする手法である．ここでは物体 p の固有空間を計算する．

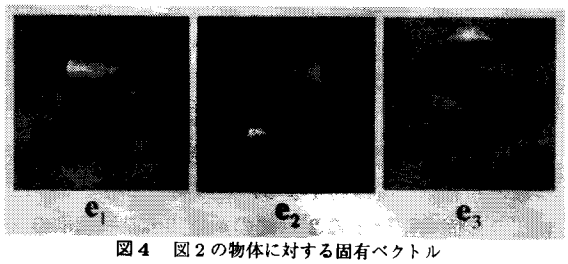


図4 図2の物体に対する固有ベクトル
Fig. 4 Eigenvectors for a learning set shown in Fig. 2.

まず，物体の画像集合の平均 $\mathbf{c}^{(p)}$，

$$\mathbf{c}^{(p)} = \frac{1}{RLP} \sum_{r=1}^{R} \sum_{l=1}^{L} \mathbf{x}_{r,l}^{(p)}$$

を計算し，つぎに各学習サンプルから平均画像を差し引き，行列 $\mathbf{X}$ を作る．

$$\mathbf{X} \equiv [\mathbf{x}_{1,1}^{(p)} - \mathbf{c}^{(p)}, \cdots, \mathbf{x}_{R,1}^{(p)} - \mathbf{c}^{(p)}, \cdots, \mathbf{x}_{R,L}^{(p)} - \mathbf{c}^{(p)}]$$

画像集合の共分散行列 $\mathbf{Q}$ は，

$$\mathbf{Q} \equiv \mathbf{X}\mathbf{X}^T$$

により計算される．固有空間（例えば k 次元）は次の固有方程式

$$\lambda_i \mathbf{e}_i = \mathbf{Q}\mathbf{e}_i$$

を解き，k 個の大きい固有値 ($\lambda_1 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_N$) に対応する固有ベクトル ($\mathbf{e}_1 \cdots \mathbf{e}_k$) を基底ベクトルとすることにより得られる．一般的に画像の共分散のように次元数（今回は16,384次元）の大きな行列の固有ベクトルの計算は困難である．しかし，画像数が少ない場合には，特異値分解やSTA法[12),17)]などを利用することにより解くことが可能である．ある物体に対する固有空間はその物体の集合を表現するのに適した空間である．図4に図1で示した画像から作成した固有ベクトルの例を示す．

3.5 固有空間の性質

固有空間は，画像間の相関値が空間上の距離値に対応するため，画像間相関の計算に対して優れた性質を持った空間である．これは以下の式により容易に示すことが可能である．ある画像は固有空間上の点に投影される．ここで2枚の画像 $\mathbf{x}_m$，$\mathbf{x}_n$ とそれに対応する固有空間上の点

$\mathbf{g}_m([g_{m1}, g_{m2}, \cdots, g_{mk}]^T)$, $\mathbf{g}_n([g_{n1}, g_{n2}, \cdots, g_{nk}]^T)$ を考えてみる．画像 $\mathbf{x}_m$ は固有ベクトルを用いて $\mathbf{x}_m = \sum_{i=1}^{K} g_{mi}\mathbf{e}_m \approx \sum_{i=1}^{k} g_{mi}\mathbf{e}_m$ と近似的に表現される．部分空間の次元 k の値を全空間の次元 K に近づける程近似精度は高くなる．この関係を用いることにより，固有空間上の2点間の距離 $\|\mathbf{g}_m - \mathbf{g}_n\|$ の2乗は

$$\|\mathbf{g}_m - \mathbf{g}_n\|^2 = \|\sum_{i=1}^{k} g_{mi}\mathbf{e}_m - \sum_{i=1}^{k} g_{ni}\mathbf{e}_n\|^2 \approx \|\mathbf{x}_m - \mathbf{x}_n\|^2$$

と近似される．

また，各画像のエネルギは3.3節で示したように大きさ1に正規化されているため以下の式が成り立つ．

$$\|\mathbf{x}_m - \mathbf{x}_n\|^2 = (\mathbf{x}_m - \mathbf{x}_n)^T(\mathbf{x}_m - \mathbf{x}_n) = 2 - 2\mathbf{x}_m^T\mathbf{x}_n$$

上記の2式により，

$$\|\mathbf{g}_m - \mathbf{g}_n\|^2 \approx 2 - 2\mathbf{x}_m^T\mathbf{x}_n$$

の関係が導かれる．つまり，画像間の相関 $\mathbf{x}_m^T\mathbf{x}_n$ が高ければ，それに対応する固有空間上の点間距離 $\|\mathbf{g}_m - \mathbf{g}_n\|$ は近くなるような順序関係が保たれていることが分かる．

これは具体的には，連続的な画像の変化，例えば連続的な向きの変化や連続的な大きさの変化を考えた場合には，連続する画像間の相関は高くなるため，このような場合には，その画像に対応した固有空間上の点はスムースな系列を描くことになる．

3.6 見かけ画像のパラメトリック固有空間表現

次に物体の方向と大きさの変化により連続的に変化する3次元物体の見かけ画像を固有空間上の多様体により表現する．学習サンプルから平均画像を引いたベクトルを式

$$\mathbf{g}_{r,l}^{(p)} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k]^T(\mathbf{x}_{r,l}^{(p)} - \mathbf{c})$$

により固有空間に投影すると，1枚の画像は固有空間上の点に対応する．更に1回転分の学習サンプルを固有空間に投影するとそれは一次元の点の系列になる．これらの点系列は補間により連続的な変化として表現する．ここでは補間にはキュービックスプラインを用いた．更に大きさを変化させた画像も同様に固有空間上に投影すると，物体の方向と大きさの2パラメータにより表現される多様体（曲面）が固有空間上に構成される．この曲面を $\mathbf{g}^{(p)}(\theta_1, \theta_2)$ で表現する．θ_1, θ_2 はそれぞれ回転と大きさのパラメータに対応する．学習サンプルに存在しない方向や大きさ（中間の方向や大きさ）に対する画像も，この曲面は補間により表現していることになる．図2に示した画像に対する固有空間上での曲面を図5に示す．実際には多次元部分空間での曲面であるが，表示の都合上3次元で表示した．この表現をパラメトリック固有空間表現と呼ぶ．このパラメトリック固有空間表現をつくることにより学習は終了する．

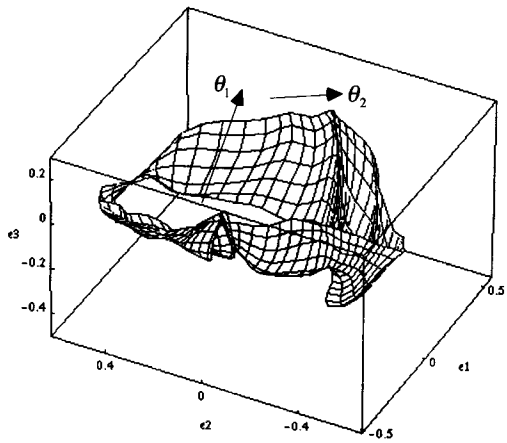


図5 図2の物体に対するパラメトリック固有空間
Fig. 5 A parametric eigenspace representation for the object shown in Fig. 2.

4. イメージスポッティング

まず入力画像から注視マスクによりある特定の場所の部分画像を抽出する．その部分画像に対して，学習段階と同様の正規化を行い，次にパラメトリック固有空間上の点に投影する．次に，この点と多様体との距離を計算する．もし，領域が物体上の点であるならば，この点は多様体に乗っている．この多様体は方向と大きさによってパラメータ化されているために，この点の多様体上での位置によってこの領域の物体の大きさと方向が同時に検出されることになる．注視マスクは順次入力画面上を走査し，上記の処理を繰り返す．

4.1 多重解像度による大きさ変動への対処

物体の大きさの変化はパラメトリック固有空間上での多様体によって表現されているが，実際にはこの多様体によって大きさの大きな変動を吸収することは困難である．その理由は，多様体による大きさの大きな変動を許すと照合回数は減少できるものの，それに対応する注視マスクのサイズが小さいものになってしまう為に照合に有効なパターン領域のサイズが小さくなり，精度の低下につながるためである．そこで，ここでは変動する範囲を適当な倍率 α に制限し，多重解像度を用いて階層的にこれを適用することにする．つ

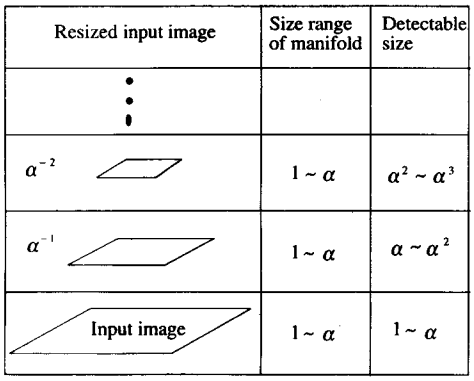


図6 多重解像度表現と多様体表現によるサイズ変動吸収
Fig. 6 Hierarchical scaling of the input image for arbitrary size.

まり，大きさの少ない変動は多様体により吸収し，大きい変動は多重解像度により吸収することにする．いま仮に第 k 階層目の画像を $i_k(x,y)$ で表現すると第 $k+1$ 階層目の画像は $i_{k+1}(x,y) = i_k(\alpha x, \alpha y)$ で表現できる．これと多様体が表現する物体の大きさの変動範囲とを結合することにより，連続的な大きさの変化に対応できることになる．この α の値として今回は1.5を設定した．その理由はこの程度の大きさの中に一般的に特徴となるパターンが十分入っているためにこの値を選んだ．この範囲で階層的に注視領域を操作して目標となる物体をイメージスポッティングする．実際には注視領域のサイズは変えずに，画面全体の大きさを再帰的に変化させながら走査していく．**図6**にその概念図を示す．

5. 物体の抽出実験

5.1 イメージスポッティングの実験

物体の例として今回3つの物体（招き猫，ジュースの缶，顔）を対象として，合計20種類の画像を実験サンプルとして収集した．その物体の例を**図7**(a)に示し，評価用に使用した画像の1例を図7(b)に示す．図7(c)に実際の多様体からの距離を図表現する．距離が小さいほど白くなっている．この極値を求めることにより，ターゲットとなる場所が検出される．また多様体上での投影点の位置により大きさと同時に物体の方向も検出できるわけである．**図8**に様々な画像に対して対象物体の領域を抽出した例を示す．白い枠はイメージスポッティングした結果である．

5.2 次元を減らしたときの効果

本手法はテンプレートマッチングの総当たり法を，固有空間による次元圧縮により効率化していると考えられる．本実験の場合，注視マスクの次元が約10,000次元であり，それが10次元の固有空間で表現されるため，約3桁のデータ圧縮になっていると考えられる．これにより記憶容量，計算時間ともに，総当法に比較して約3桁効率化されることになる．一方，次元数が減少すれば当然その表現精度が下がる．ここでは次元数を変化させた際のイメージスポッティングの能力について検討した．図7(b)の入力画像に対して白線に沿って注視領域マスクを走査した際の，その領域に対する固有空間上の点と，多様体との距離をプロットし，それを**図9**に示す．その際に図9(a)は2次元の固有空間の場合，図9(b)は6次元の場合，(c)は16次元の場合を示す．矢印で示した場所が正しい場所である．固有空間の次元が16の場合には正しい場所で距離が小さくなっているため正しくスポッティングができることになる．一方，次元数を減少させると表現能力が低下する．そのため本来異なるようなパターンに対しても小さい距離を出す．つまりパターンの分離能力が低下する．6次元の場合にはすでに $x = 0.7$ 近辺で正しい場所よりも小さい値をとっていることがわかる．つまりこの領域で誤りが発生することになる．もっと極端の場合として2次元の場合ではかなり多数の場所で距離が小さくなり，もはやイメージスポッティングはできないほどとなっている．さまざまな入力対象に対して実験を行ったところ一般的には固有空間の次元数は10次元程度は必要であることが分かった．

どの程度の次元が必要であるかを示す一つの尺度が次式で表される寄与率 c である．

$$c = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{K} \lambda_i}$$

ここで，k は固有空間の次元を，K は注視領域の画素数（次元）を表す．寄与率は画像を再構成したときにどの程度の2乗誤差で近似的に再構成できるかを表した指標である．つまり寄与率が高ければその部分空間で物体を表現する能力が高いことになる．**図10**は固有空間の次元に対する寄与率（固有値の和）をグラフにしたものである．このグラフの形状は対象に依存し，もし画像の変動が大きければある次元での寄与率は低くなり，一方変動が少なければ寄与率は高くなる．今回の対象の場合には10次元程度は必要であることが分かったが，その際の寄与率は0.7程度である．

5.3 雑音の影響

本手法では，構造的な特徴を使用していないため，重畳雑音のような外乱に対してはかなり強い．**図11**

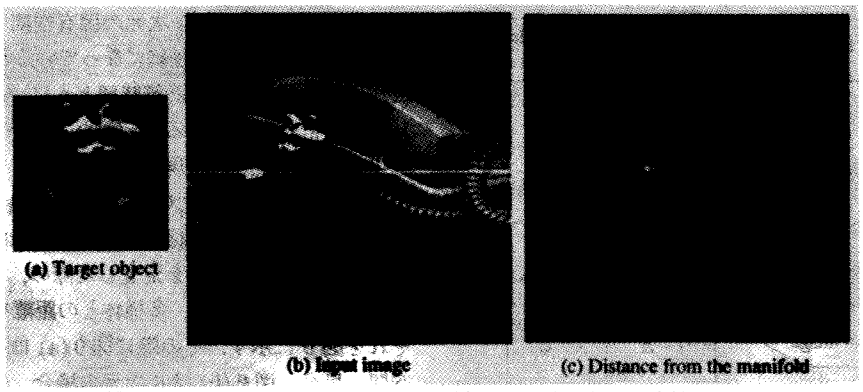


図7 イメージスポッティングの例．(a) 対象，(b) 入力画像，(c) 多様体との距離
Fig. 7 An example of image spotting. (a) A target object, (b) An input image, (c) Distance map from the manifold.

図8 イメージスポッティングの結果
Fig. 8 Results of image spotting.

は図7(b)に示した画像に対してS/N（信号雑音比）を20 dBとなるように白色雑音を加えた画像である．これらの雑音の乗った画像に対しても正しくイメージスポッティングが可能である．5.2節で示したと同じ線上の距離の分布（16次元の固有空間を使用）は図9(d)になる．距離値の分布は，雑音のない場合（図9(c)）に比較してもあまり変化がなく，本手法は雑音に対して影響を受けにくい性質を持っていることがわかる．それは固有ベクトルによるフィルタが平滑フィルタの機能を果たし，雑音成分を低減させる作用があるためである．

5.4 ポーズ推定の精度

今回用いた物体のポーズ推定の精度を図**12**に示す．本手法は注視マスクにより物体の境界の情報を削除している．これはスポッティングの際に外乱となる背景領域を削除するという重要な働きを持つ．しかし一般的には背景と物体の境界情報は認識やポーズ推定にとって重要な情報を持っている．ポーズ推定の精度を注視マスクを使用した場合と使用しない場合で比較した．実験には学習データとして45方向から撮影した招き猫の画像を，ポーズ推定にはそれとは異なる方向から撮影した45方向の画像を使用した．図12に境界情報を使用した場合の誤差のヒストグラム（平均誤差0.4度）と，注視マスクを使用して物体の内部の模様だけを手がかりにポーズ推定をした場合の結果（平均誤差1.3度）を示す．ポーズ推定精度が0.9度程度低

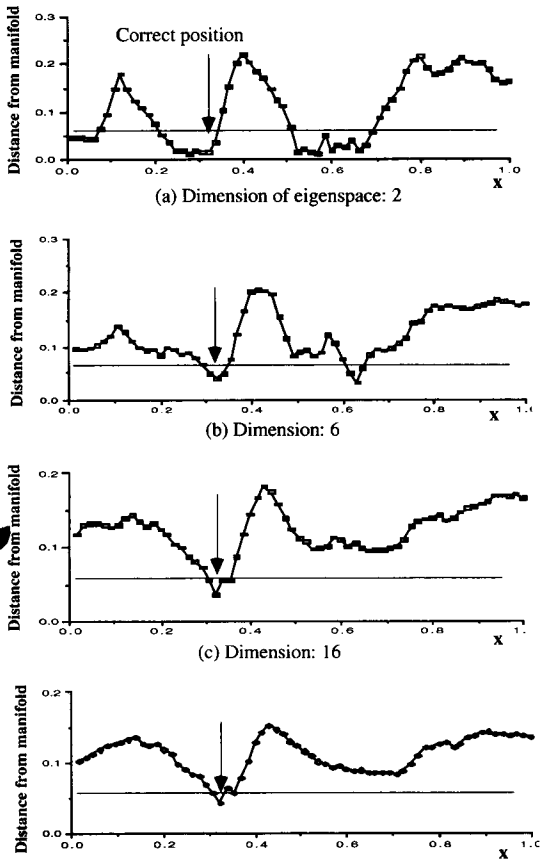


図9 図7(b)白線上の各位置での多様体との距離．(a)固有空間：2次元，(b)6次元，(c)16次元，(d)16次元，ただし白色雑音を重畳（SNR=20 dB）

Fig. 9 Distance from the manifold for each position on the white line in figure 7(b): (a) Using 2-dimensional eigenspace, (b) Using 6-dimensional eigenspace, (c) Using 16-dimensional eigenspace, (d) Using 16-dimensional eigenspace for an image with white noise (SNR=20 dB).

下していることがわかる．もし，ポーズ推定の精度が重要であるような応用に対しては，一度領域を抽出した後に，領域全体の情報を利用してポーズ推定を再度行う処理を後段に加えれば，精度の向上は期待できる．そのような後処理については今後の検討事項とする．

6. 結　　論

本論文では，任意の方向を向いた3次元物体を2次元照合により複雑な背景から抽出する手法について述べた．3次元物体は見る方向により見かけ画像が大きく変化し，更に画像中に占める物体の大きさが不明の場合には，単純なテンプレートマッチングやフィルタリングの手法では物体の抽出が困難である．ここで提案したパラメトリック固有空間法は，連続的に変化する画像系列を固有ベクトル空間上での多様体で表現する手法である．これにより少ない記憶容量で3次元物体を2次元画像の集合体として記憶することができるようになった．このパラメトリック固有空間法を階層的に適用することにより，従来困難であったエッジや

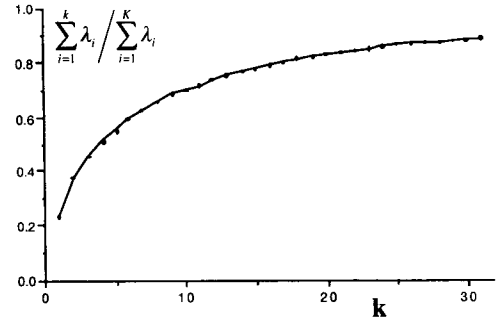


図10 累積寄与率

Fig. 10 Sum of the contribution ratio.

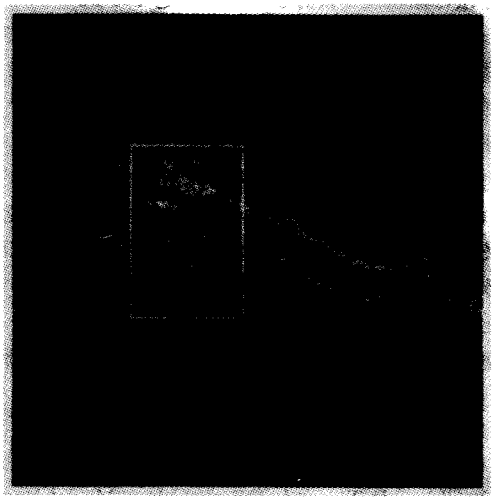


図11 白色雑音を加えた場合の入力画像（SNR=20 dB）

Fig. 11 An input image with white noise (SNR=20 dB).

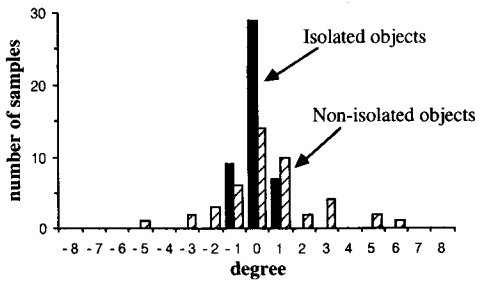


図12 ポーズ推定エラーのヒストグラム

Fig. 12 Histogram for pose estimation error.

表面形状などの3次元構造を抽出することなく，多様な見え方を持つ3次元物体を複雑な背景画像から抽出することが可能となった．また，その時のパラメータを読み取ることにより同時に物体のポーズを推定をすることも可能となった．また，物体領域の大きさの変動に対しても，小さな変動に対しては固有空間中の多様体表現により，大きな変動に対しては多重解像度表現により，正しく物体の切り出しが可能となる．今回の実験では，物体の1軸回転の場合を仮定したが，物体の任意のポーズ等を考えると更にパラメータ数が増える．今後はよりパラメータが増えた場合や，一部しか見えていないような物体の抽出などの場合に対して，本手法の拡張性を検討して行く予定である．

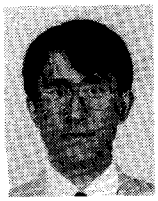
一方，心理学の分野においても，人間が3次元物体を認識する際に，3次元特徴を利用しているのか，2次元照合を利用しているのか興味が持たれている[18),19)]．たぶん人間はその両者を場合によって使い分けていると考えられ，両者の側面から対比されながら研究が行われている．本研究は2次元照合の立場からの研究であるが，今後，マシンビジョンにおいても，両者の立場から良い点悪い点を比較しながら研究を進めていくことが必要であろう．

謝辞 日頃，御指導いただくNTT池上基礎総研所長，石井科学部長，内藤リーダに深謝します．

参考文献

1) Chin, R.T. and Dyer, C.R.: Model-Based Recognition in Robot Vision, *ACM Comput. Surv.*, Vol.18, No.1, pp.67–108 (1986).
2) Besl, P.J. and Jain, R.C.: Three-Dimensional Object Recognition, *ACM Comput. Surv.*, Vol.17, No.1, pp.75–145 (1985).
3) Ullman, S. and Basri, R.: Recognition by Linear Combination of Models, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.13, No.10, pp.992–1006 (1991).
4) Poggio, T. and Edelman, S.: A Networks That Learns to Recognize Three-Dimensional Objects, *Nature*, Vol.343, pp.263–266 (1990).
5) Tanimoto, S.L.: Template Matching in Pyramid, *Computer Vision, Graphics and Image Processing*, Vol.16, pp.356–369 (1981).
6) Liu, Z.Q. and Caelli, T.M.: Multiobject Pattern Recognition and Detection in Noisy Background Using a Hierarchical Approach, *Computer Vision, Graphics and Image Processing*, Vol.44, pp.296–306 (1988).
7) Rosenfeld, A. and Vanderbrug, G.J.: Coarse-fine Template Matching, *IEEE Trans. on System, Man, and Cybernetics*, Vol.7, No.2, pp.104–107 (1977).
8) Weng, J.J., Ahuja, N. and Huang, T.S.: Learning Recognition and Segmentation of 3D Objects from 2D Images, *IEEE ICCV*, pp.121–128 (1993).
9) Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, Academic Press, London (1990).
10) Oja, E.: *Subspace Methods of Pattern Recognition*, Research Studies Press, Hertfordshire (1983).
11) 村瀬 洋，木村文隆，吉村ミツ，三宅康二：パターン整合法における特性核の改良とその手書き平仮名文字認識への応用，信学論(D)，Vol.J64-D, No.3, pp.276–283 (1981).
12) Murase, H. and Lindenbaum, M.: Spatial Temporal Adaptive Method for Partial Eigenstructure Decomposition of Large Images, *NTT Technical Report*, No.6527, March (1992) (IEEE Transaction on Image Processing に掲載予定).
13) Murase, H. and Nayar, S.K.: Learning Object Models from Appearance, *AAAI-93*, pp.836–843, American Association for Artificial Intelligence (July 1993).
14) Murase, H. and Nayar, S.K.: Illumination Planning for Object Recognition in Structured Environments, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (June 1994).
15) Sirovich, L. and Kirby, M.: Low Dimensional Procedure for the Characterization of Human Faces, *Journal of Optical Society of America*, Vol.4, No.3, pp.519–524 (1987).
16) Turk, M.A. and Pentland, A.P.: Face Recognition using Eigenfaces, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.586–591 (June 1991).
17) Press, W., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T.: *Numerical Recipes in C*, Cambridge University Press, Cambridge (1988).
18) Tarr, M. and Pinker, S.: Mental Rotation and Orientation-Dependence in Shape Recognition, *Cognitive Psychology*, Vol.21, pp.233–282 (1989).
19) Edelman, S. and Weinshall, D.: A Self-Organizing Multiple-View Representation of 3D Objects, *Biological Cybernetics*, Vol.64, pp.209–219 (1991).

(平成6年11月4日受付)
(平成7年1月12日採録)



村瀬 洋（正会員）

昭和30年生．昭和55年名古屋大学工学研究科電子工学専攻修士課程修了．同年日本電信電話公社（現在のNTT）入社．以来，文字・図形認識，コンピュータビジョンの研究に従事．平成4年から1年間米国コロンビア大学計算機学科に研究員として滞在．現在，NTT基礎研究所情報科学部主幹研究員と同時にNTT特別研究員．工学博士．昭和60年度電子情報通信学会学術奨励賞受賞．平成6年IEEEのCVPR会議にて最優秀論文賞受賞．IEEE，電子情報通信学会，AVIRG各会員．

シュリー K. ナイヤー

1990年カーネギーメロン大学計算機学科にてPhDを取得．1991年からコロンビア大学計算機科学科の助教授．コンピュータビジョン，特に物理ベースビジョン，ロボットビジョンの研究に従事．1990年のIEEEのICCV会議でMarr賞を受賞．