社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
PRU93-120(1994-01)

2次元照合による3次元物体の認識とその学習

- パラメトリック固有空間表現 -

村瀬 洋　　Shree K. Nayar

NTT基礎研究所　　コロンビア大学
180 東京都武蔵野市緑町3-9-11
murase@siva.ntt.jp　　nayar@cs.columbia.edu

あらまし　本報告では、2次元照合により、任意の方向を向いた3次元物体を識別し、その物体の方向を検出する手法について述べる。2次元照合による認識は、3次元特徴の抽出が不要である等の特長はあるが、見る方向や光源の位置により複雑に変化する画像をあらかじめ記憶すること、およびそれと入力画像とを照合することが記憶容量、計算量の点で困難であると考えられ、従来あまり試みられていなかった。ここでは、変形する画像系列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の提案により、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。その結果、困難な3次元特徴の抽出なしで3次元物体の認識とポーズ推定をすることが可能となった。本報告では、この手法の提案とともに他の2次元照合的な手法との比較実験結果についても述べる。

キーワード　物体認識、ポーズ推定、主成分分析、視覚学習

Learning and Recognition of 3D Object from Appearance

- Parametric Eigenspace Representation -

Hiroshi Murase　　Shree K. Nayar

NTT Basic Research Labs
3-9-11 Midoricho
Musashinoshi, Tokyo 180
Japan

Computer Science Dept.
Columbia University
New York, NY10027
USA

Abstract　We address the problem of learning object models for recognition and pose estimation. We formulate the recognition problem as one of matching visual appearance rather than shape. A new compact image representation called parametric eigenspace is proposed. The image set is compressed to obtain a low-dimensional subspace in which the object is represented as a hypersurface parametrized by pose and illumination. The recognition system projects the image onto the eigenspace. The object is recognized based on the hypersurface it lies on. The position of the projection on the hypersurface determines the object's pose. We have conducted experiments using several objects with complex appearance characteristics.

Key words　Object recognition, pose estimation, principal component analysis, visual learning

1、まえがき

3次元物体をその2次元画像から識別し、その物体の向きまでも検出する3次元物体認識技術は産業用ロボットの要素技術として、また一般環境内での物体の監視など幅広い応用があり、これまでその実現のために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、モデルと入力特徴との照合に3次元モデルを利用する手法と2次元モデルを利用する手法に分類される。

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ーなどの幾何学的特徴や、表面の3次元的特徴を抽出し、これと予め用意してある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、モデルが3次元の完全な記述を持っているため、回転などに容易に対処できる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もしくは復元する処理があまり容易でなく、現在も研究レベルにとどまっている。

一方、2次元モデルを利用する方法として、3次元物体の見かけの2次元画像を予め全て記憶しておき入力画像とこれとを比較する手法も考えられるが、これでは膨大な画像データ量となるため、記憶量、計算量の観点から、あまり実際的ではない。また、エッジの2次元位置特徴を利用するものや[3]、屈折点や端点の位置を利用するもの[4]もあるものの、これらは特徴点の照合であり、2次元のパターンを積極的に照合に利用するものではなかった。本論文では2次元モデルとの画像信号レベルでの2次元照合により3次元物体を認識する手法について述べる。

ここでは、3次元物体の向きや光源の変化に対応して連続的に変動する2次元画像の変化を、画像の固有ベクトルから構成される部分空間（固有空間）上での多様体で表現するパラメトリック固有空間法を提案した。学習段階では物体の画像集合から、固有空間を計算し、その上で多様体を構成する。認識段階では入力画像を一旦この固有空間上の点に投影し、その点に最も近い多様体上での点の位置を検出することにより、その物体の種類と物体の向きなどを検出する。本手法では、物体の例を与えるだけで自動的にその物体を学習することができる。また、認識段階では、入力画像中の3次元物体を認識すると同時に、その物体のポーズを検出することも可能である。

本手法は部分空間法によるパターン認識[5,6]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[7,8]、あるいはEigenface法による顔画像認識手法[11]などが上げられる。しかし、これらはいずれもパターンの分類に主眼においたものであり、本手法のように物体の向きなどのパラメータを検出したり、3次元物体を表現しようとするものではなかった。

本報告では、2次元照合による物体認識のためのパラメトリック固有空間法について述べる。更に、本手法と他の2次元照合的な手法との比較を、認識実験により述べる。

2、パラメトリック固有空間法による学習

3次元物体の見かけの画像は、その物体の方向や照明の位置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。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することは画像符号化の目的と同じである。そこで、ここでは符号化を基本とした画像の表現法としてパラメトリック固有空間法と言う新しい考え方を提案する。

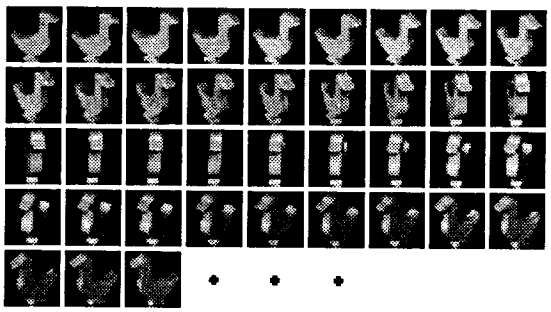


図1．物体を回転させたときの見かけ画像の変化.

画像の学習段階では、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。この段階は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、部分空間上の多様体（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。物体の種類が複数の場合にはその数だけ多様体が構成される。認識段階では、まず入力画像を固有空間に投影し、次にこの点と固有空間上の多様体との位置関係により、認識および物体の方向を検出する。

2.1 画像の正規化

入力画像から、まず物体部分を切り出す。ここでは、しきい値処理や背景との差分処理により切り出しを行った。次に物体以外の部分に0の値を代入する。その後に物体を正方形に接するように、物体の縦横比を一定のまま大きさの正規化を行う。この画像の画素値をラスター上にスキャンし、その画素値を要素とするベクトル $\hat{x}$

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

でもとの画像を表現する。ここでNは画素数である。

次に、センサー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x

$$x=[x_1,x_2,\cdots,x_N]^T$$

とすると、ここではベクトルxの大きさ、つまり画像のエネルギーが1（$\|x\|=1$）になるように式、

$$x_n=\frac{1}{\sigma}\hat{x}_n \quad 、 \quad \sigma=\sqrt{\sum_{n=1}^{N}(\hat{x}_n)^2}$$

により正規化する。

3次元物体の見かけ画像は、物体の向きと光源によって変化する。ここでは仮に物体の向きが1軸の周りで回転変化し、照明は移動する点光源に背景光が重畳された場合を考える。これは工場内などの環境の中では不自然な設定ではない。また、物体の任意のポーズを扱うためには、パラメータ数を増やすことにより拡張できる。

ここでP種類の物体を学習する場合を考える。p番目の物体を1回転し、且つ光源の向きを変化させて収集した画像の集合を

$$\{x_{1,1}^{(p)},\cdots,x_{R,1}^{(p)},x_{1,2}^{(p)},\cdots,x_{R,L}^{(p)}\}$$

で表現する。ここでRは回転方向の刻み数、Lは光源の方向数を表わす。これを第p物体の画像集合と呼ぶ。またすべての物体に対する画像集合を、

$$\{x_{1,1}^{(1)},\cdots,x_{R,1}^{(1)},x_{1,2}^{(1)},\cdots,x_{R,L}^{(1)},$$
$$x_{1,1}^{(2)},\cdots,x_{R,1}^{(2)},x_{1,2}^{(2)},\cdots,x_{R,L}^{(2)},$$
$$\cdots$$
$$x_{1,1}^{(P)},\cdots,x_{R,1}^{(P)},x_{1,2}^{(P)},\cdots,x_{R,L}^{(P)}\}$$

で表現し、これを全物体の画像集合と呼び、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と、光源方向の制御が可能はロボットアームを用いた。その様子を図2に示す。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。

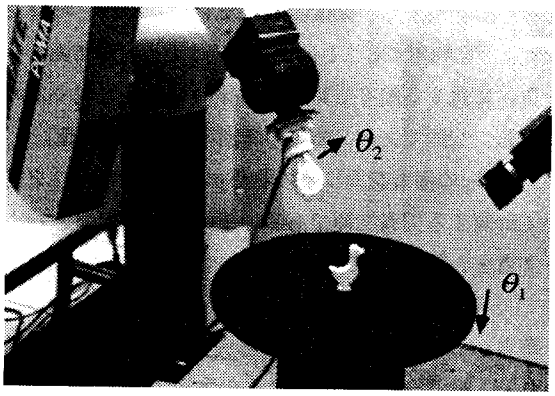


図2． 3次元物体の学習画像収集システム.

2.2 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。ここでは全物体の固有空間と物体pの固有空間の2つの固有空間を計算する。

まず、全物体の固有空間を計算する。全物体の画像集合の平均c

$$c=\frac{1}{RLP}\sum_{p=1}^{P}\sum_{r=1}^{R}\sum_{l=1}^{L}x_{r,l}^{(p)}$$

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Xを作る。

$$X \equiv \left[x_{1,1}^{(1)} - c, \cdots, x_{R,1}^{(1)} - c, \cdots, x_{R,L}^{(P)} - c\right]$$

画像集合の共分散行列Qは、

$$Q \equiv XX^T$$

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

$$\lambda_i e_i = Q e_i$$

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（$e_1 \cdots e_k$）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16384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、特異値分解などを利用することにより解くことが可能である。全物体に対する固有空間は全物体の集合を表現するのに適した空間であり、物体の識別の際に利用する。

物体pの固有空間はその物体pの画像集合だけを用いて計算する固有空間である。その共分散行列を $Q^{(p)}$ とすると、第p物体の固有空間は

$$\lambda_i^{(p)} e_i^{(p)} = Q^{(p)} e_i^{(p)}$$

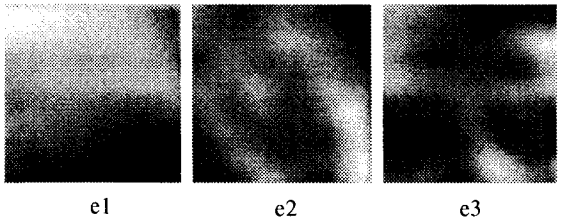
を解き、その固有ベクトルを基底とすることにより得られる。物体pの固有空間はその物体を表現するのに適した空間であるため、物体の名前が識別された後に、その物体のポーズを推定する際に利用される。

図3(a)に風景画像などの多種な画像から作成した固有ベクトルの例を、図3(b)には図1の画像から作成した特定物体の固有ベクトルの例を示す。図3(a)はより一般の画像を表現できるが、表現効率は高くない。一方、図3(b)は一般の画像は表現できないが、特定の画像集合を少ない基底で効率良く表現できる。物体の識別後のポーズ推定などで、その物体の固有空間を使う理由はここにある。

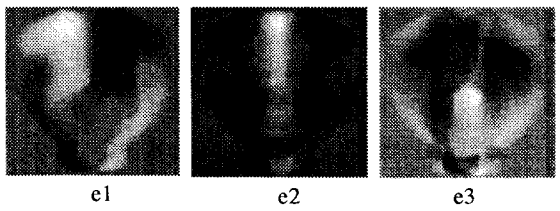
2.3 見かけ画像のパラメトリック固有空間表現

次に物体のポーズや、光源の位置の変化により連続的に変化する3次元物体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

$$g_{r,l}^{(p)} = [e_1, e_2, \cdots e_k]^T (x_{r,l}^{(p)} - c)$$



(a) 一般風景画像の固有ベクトル



(b) 特定物体の固有ベクトル

図3． 固有ベクトルの例.

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。それは、一般的に物体のポーズの変化が少ない場合には画像の変化も少ないため相関が強く、また強く相関を持った画像は固有空間上で近い位置に投影されるためである。例えば図1に示す図形の固有空間上での系列は、図4に示すようになる。実際には多次元空間での点列であるが、表示の都合上3次元で表示した。これらの点列は補間により連続的な変化として表現する。補間にはここではキュービックスプラインを用いた。

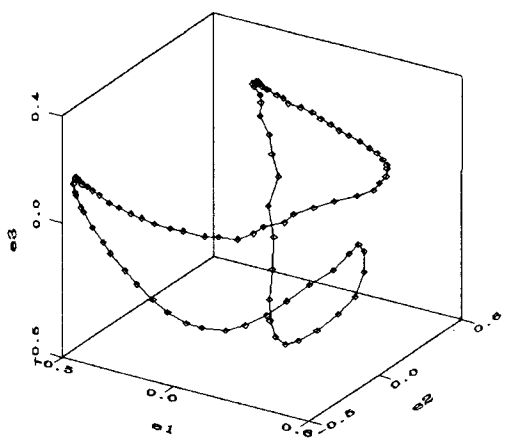


図4． パラメトリック固有空間の例

更に光源の位置を変化させた画像も同様に固有空間上に投影すると、物体のポーズと光源の位置の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$g^{(p)}(\theta_1,\theta_2)$で表現する。$\theta_1$、$\theta_2$はそれぞれ回転と光源の位置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光源の位置（中間の方向や光源の位置）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。曲面は物体の種類の数だけ構成される。

同様に物体pの固有空間に対して物体pの曲面を構成する。上記と同様に学習サンプルを物体pの固有空間に物体pの画像集合を

$$f_{r,l}^{(p)} = [e_1^{(p)}, e_2^{(p)}, \cdots e_k^{(p)}]^T (x_{r,l}^{(p)} - c^{(p)})$$

により投影し、補間処理により曲面を構成する。ここで$c^{(p)}$は物体pの学習サンプルの平均である。そして、補間した表現を$f^{(p)}(\theta_1,\theta_2)$で表わす。

3、認識

まず学習段階で用いたと同様の前処理を行う。つまり入力画像から、いき値処理などを用いて物体領域を切り出す。切り出した後に大きさの正規化を行い、明るさの正規化を行う。正規化後の入力画像のベクトルをyとする。次にこのベクトルを次式により全物体の固有空間上の点zに投影する。cは前述の平均画像である。

$$z = [e_1, e_2, \cdots e_k]^T (y - c)$$

認識はこの投影された点zがP種類ある曲面のどこに乗っているかを調べることになる。つまり、点zと曲面$g^{(p)}(\theta_1,\theta_2)$との距離

$$d_1^{(p)} = \min_{\theta_1,\theta_2} \| z - g^{(p)}(\theta_1,\theta_2) \|$$

を最小とするpを求めることにより実現できる。

物体名pを識別した後に、次段階としてその物体のポーズを推定する。まず、入力画像yを式

$$z^{(p)} = [e_1^{(p)}, e_2^{(p)}, \cdots, e_k^{(p)}]^T (y - c^{(p)})$$

により物体pの固有空間に投影する。ポーズを検出するということは、点$z^{(p)}$が曲面上のどこに位置しているかに対応している。そこで距離

$$d_2^{(p)} = \min_{\theta_1,\theta_2} \| z^{(p)} - f^{(p)}(\theta_1,\theta_2) \|$$

を最小とするθ_1を見つける。図５に入力画像を固有空間上の点に投影し、その点に最小距離をもつθ_1を見つける様子を示す。この例では１パラメータ（曲線）の例で示すが、実際には２パラメータの曲面上の探索になる。パラメータの数が増えれば一般的には多様体となる。

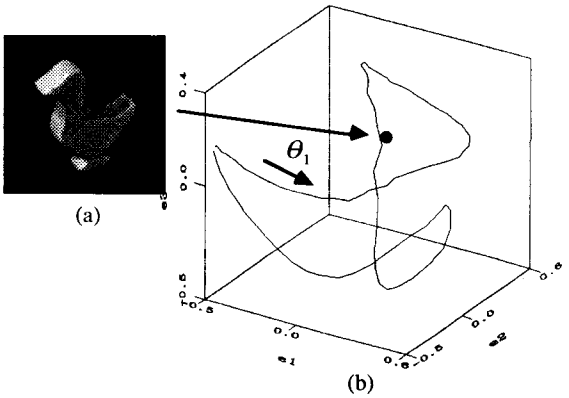


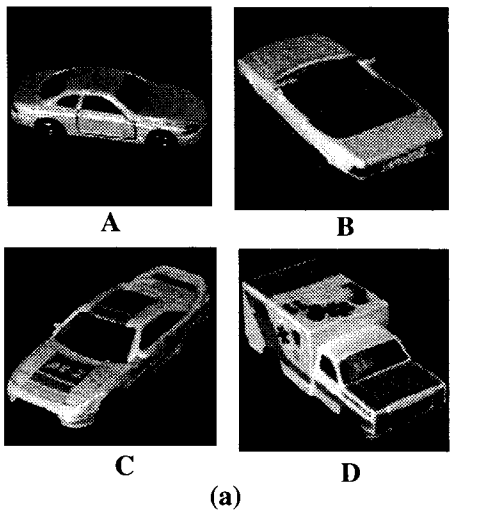
図５.(a)入力図形、(b)固有空間への投影（投影点の曲線上の位置が物体のポーズに対応）.

4、認識実験

学習段階では物体を計算機制御の可能なターンテーブルの上に乗せ、学習画像サンプルを収集した（図１）。今回の実験では図６(a)に示す４種類の物体を用いた。学習サンプルとして４度おきに３６０度回転させたもの、つまり９０種類のポーズの画像を収集した。光源の位置は連続的に変化する５ヵ所とした。つまり、全体で学習サンプル数は１８００(4x90x5)となる。認識実験用のデータとして、学習サンプルと２度位相の異なる９０種類のポーズで光源位置を３箇所移動して収集した（合計１０８０画像）。３章で示した学習処理をほどこし、それぞれの物体について曲面を計算した例を図６（b）に示す。次に４章で示した認識処理を適用し、認識実験を行った。

その実験結果を以下に示す。図７(a)は固有空間の次元に対する認識率を示す。８次元程度でほとんど認識率は飽和していることがわかる。本手法では学習サンプルに存在しないのポーズの物体をもある程度、認

識することが可能である。そこで限られたポーズ数でどの程度認識できるかを調べた。図7(b)学習に利用するポーズを減らした際の認識率の変化を示す。今回示した程度の複雑さの物体では学習には15ポーズ程度で十分であることがわかる。図7(c)にはポーズ推定の精度を示した。平均値で1.2度のポーズ推定誤差であった。



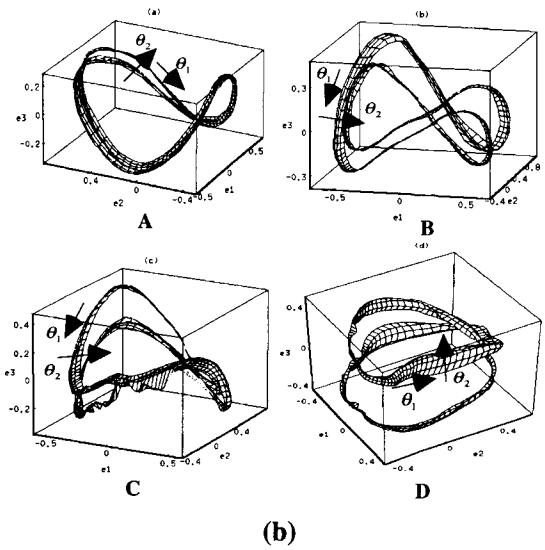


図6.(a)実験に使用した物体の例、(b)パラメトリック固有空間の例（(a)の物体に対応）.

5、検討

5.1 他手法との比較

本手法は基本的に2次元照合による3次元物体の認識である。従来、文字等の2次元図形認識用に開発された認識法がはたして3次元物体認識にどの程度適応可能かは興味深い。ここでは本手法と他の代表的な3手法と比較を行った。手法1は単純相関法、手法2はTurk[14]らが顔画像認識に用いたEigenface法、手法3は村瀬[9]らが文字認識で行った投影法である。手法の詳細はそれぞれの文献を参照されたい。

(1) 手法1： 単純相関法

学習段階では物体p毎に、学習用の画像集合の平均画像ベクトル$c_1^{(p)}$を計算する。認識段階では入力画像yと平均画像ベクトルとの相関値$d_3^{(p)} = y^T c_1^{(p)}$を計算し、それが最大となるpを認識結果とする。

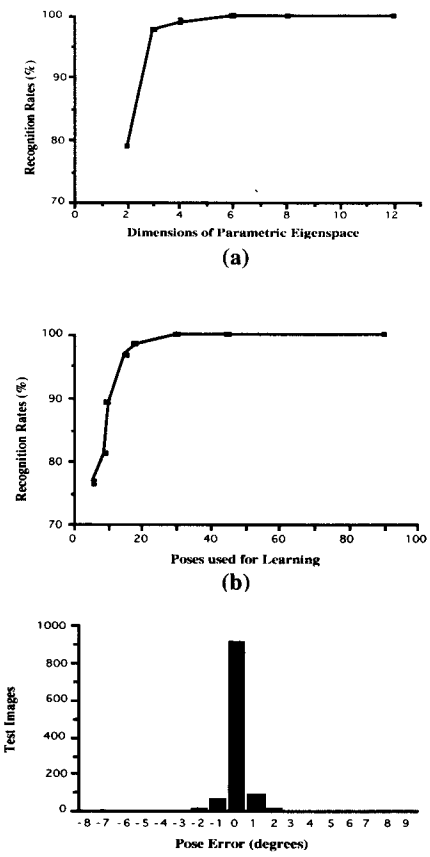


図7.(a)パラメトリック固有空間の次元と認識率との関係、(b)学習に用いたポーズ数と認識率との関係、(c)ポーズ推定誤差のヒストグラム.

(2) 手法２： Eigenface法（Turk他[11]）

学習段階では画像集合の共分散行列の固有ベクトルを計算し固有空間（仮に８次元とした）をつくる、更に、学習サンプルをそこに投影し、その平均値 $c_2^{(p)}$ を物体pごとに計算する。認識段階では、入力画像を固有空間に投影し、その点と平均値との距離 $d_4^{(p)} = \left\| z^T - c_2^{(p)} \right\|$ を計算し、それが最小となるpを認識結果とする。

(3) 手法３： 投影法（村瀬他[8]）

学習段階では物体pごとに画像集合の相関行列の固有ベクトル $[e_1^{(p)}, e_2^{(p)}, \cdots, e_k^{(p)}]$ を計算し、各物体pの固有空間をつくる。認識段階では、入力画像yを各物体の固有空間（仮に８次元とした）に投影し、その投影エネルギー $d_5^{(p)} = \sum_{k=1}^{K} (y^T e_k^{(p)})^2$ が最大となるpを認識結果とする。

実験には図１で示した４種類の物体を含む８種類の物体を用いて行った。光源位置は固定とした。学習には各物体につき１８の異なるポーズの画像（合計１４４枚）を用い、認識では９０の異なるポーズの画像（合計７２０枚）を用いた。それぞれの認識率を表２に示す。手法１、手法２は変形の少ない２次元的な図形に対しては効果があるが、本手法のように見かけの変化の大きい３次元物体にはあまり精度が得られないことがわかる。手法３は３次元物体に対しても比較的精度は得られものの、物体のポーズの検出はできない。本提案の手法はこれらの手法に比較して高い精度が得られることを確認した。また、同時に物体のポーズを検出できる点も優れている。

表１.手法間の比較実験結果

	手法１	手法２	手法３	本手法
認識率	68.6%	66.8%	98.7%	99.8%
ポーズ検出	不可能	不可能	不可能	可能

5.2 顔画像への応用

本手法は図８に示すような顔画像へもそのまま適用可能である。本手法によれば、任意の方向を向いた顔画像を認識すると同時に、その顔の方向（ポーズ）を自動的に検出することが可能となる。顔の方向を検出した実験結果（１８方向の顔画像を学習に使用）の例を図8に示す。

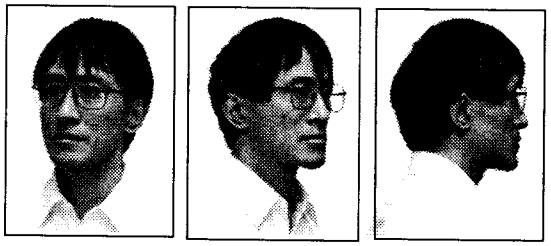
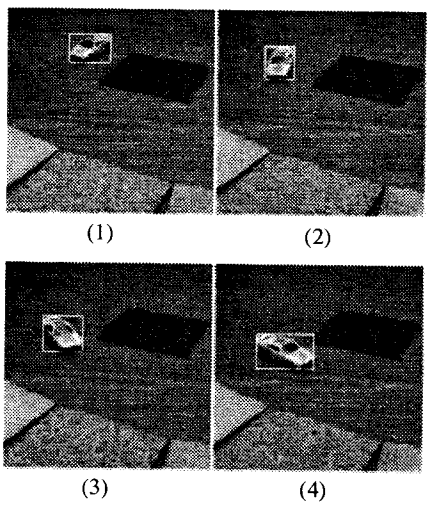


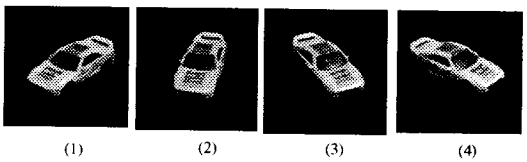
図８． 顔画像認識への応用。（正面を０度とした時の顔の方向の検出結果を示す。）

5.3 動画像への応用

本手法を動画像に適用した例を示す。この例では、入力画像から背景との差分により物体を切り出し（図9(a)）、次に物体のポーズを検出している。図9(b)は最も近いポーズの学習サンプルを示した。



(a)切り出し



(b) 推定したポーズのパターン

図９.パラメトリック固有空間法を動画像に適用した例

5.4 人間の3次元物体認識法との対比

心理学の分野でも、人間が3次元物体を認識する際に、はたして2次元照合を利用しているか3次元照合を利用しているか興味が持たれている。Edelman等はメンタルローテーションの心理実験により以下の知見[12]を示した。ある物体がその個人にとってあまり見慣れていないような場合には、3次元構造を考慮しながらモデル物体との照合をとるが、良く見慣れた物体については2次元照合を行っている。これは、日常良く見る出現頻度の高い物体については、人間も処理の単純な2次元照合を利用していることを示している。

6、結論

本論文では、任意の方向を向いた3次元物体を2次元照合により識別し、同時にその物体のポーズを検出する手法について述べた。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（例えば曲面）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。その結果、従来困難であったエッジや表面形状などの3次元構造を抽出することなく、2次元画像例から物体を学習し、2次元照合により3次元物体を認識し、ポーズを検出することが可能となった。今回の実験では、物体の1軸回転と光源の位置の2パラメータ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、物体の種類が増えた場合などについて本手法の拡張性を検討して行く予定である。

謝辞 日頃ご指導頂く、木村NTT基礎研所長、中津科学部長、内藤リーダに深謝します。

参考文献

[1] R. T. Chin and C. R. Dyer, Model-Based Recognition in Robot Vision, ACM Computing Surveys, Vol. 18, No. 1, pp. March 1986.

[2] P. J. Besl and R. C. Jain, Three-Dimensional Object Recognition, ACM Computing Surveys, Vol. 17, No. 1, pp. 75-145, 1985.

[3] S. Ullman and R. Basri, Recognition by Linear Combination of Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 10, pp. 992-1006, October 1991.

[4] T. Poggio and S. Edelman, A networks that learns to recognize three-dimensional objects, Nature, Vol. 343, pp. 263-266, 1990.

[5] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, London, 1990.

[6] E. Oja, Subspace methods of Pattern Recognition, Research Studies Press, Hertfordshire, 1983.

[7] 飯島泰蔵、文字読み取り装置ASPET/71, TV学会誌、27,3, pp.157-164, 1973.

[8] 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), J64-D, 3, pp.276-283, 1981.

[9] H. Murase and M. Lindenbaum, Spatial Temporal Adaptive Method for Partial Eigenstructure Decomposition of Large Images, NTT Technical Report No. 6527, March 1992.

[10] H. Murase and S. K. Nayar, Learning object models from appearance, AAAI-93, American Association for Artificial Intelligence, pp. 836-843, July, 1993.

[11] M. A. Turk and A. P. Pentland, Face Recognition Using Eigenfaces, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, June 1991.

[12] S. Edelman and D. Weinshall, A self-organizing multiple-view representation of 3D objects, Biological Cybernetics, Vol. 64, pp. 209-219, 1991.