

パラメトリック固有空間法による 3次元物体の認識とスポッティング

村瀬 洋 Shree K. Nayar
NTT基礎研究所 Columbia University

本論文では、2次元照合により、任意の方向を向いた3次元物体を背景から抽出し、同時に物体の識別と、物体方向の検出をする手法について述べる。2次元照合による3次元物体の認識は、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、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。この表現法による2次元照合を行えば、複雑な背景を持つ画像の中から任意の方向を向いた物体を抽出(スポッティング)することも可能となる。本論文では、この表現法による物体の認識および顔画像を例とした物体のスポッティングの実験結果について述べる。

1. はじめに

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次元特徴を抽出し、これと予め用意した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を持っているものの、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、3次元物体の見かけの2次元画像をあらかじめ全て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¹³⁾、屈折点や端点の位置を利用するもの¹⁴⁾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する。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、その物体の種類や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

Recognition and Spotting of 3D objects using
Parametric Eigenspace Representation, Hiroshi Murase,
NTT Basic Research Laboratories, and Shree K. Nayar,
Columbia University

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^{15,61}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^{17,81}、あるいはEigenface法による顔画像認識手法¹¹¹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章)。次にこのアイデアを、既に切り出された物体の認識(3章)、および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。

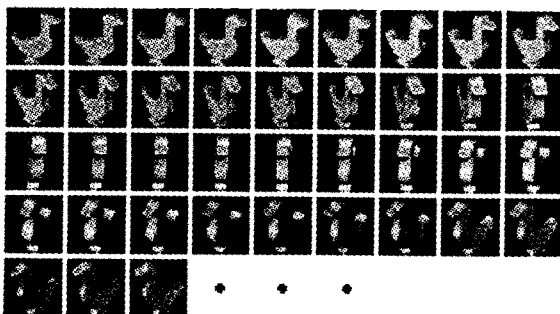


図1. 物体を回転させたときの見かけ画像の変化。

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。切り出しには4章に述べるような画像照合によるスポッティングや、背景画像との差分などにより実現される。次に物体以外の部分に0の値を代入する。その後には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。この画像をラスタースキャンし、画素値を要素とするベクトル \hat{x}

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

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 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、

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

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

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

$$\begin{aligned} & \{x_{1,1}^{(1)}, \dots, x_{R,1}^{(1)}, x_{1,2}^{(1)}, \dots, x_{R,L}^{(1)}, \\ & x_{1,1}^{(2)}, \dots, x_{R,1}^{(2)}, x_{1,2}^{(2)}, \dots, x_{R,L}^{(2)}, \\ & \dots \\ & x_{1,1}^{(P)}, \dots, x_{R,1}^{(P)}, x_{1,2}^{(P)}, \dots, x_{R,L}^{(P)} \} \end{aligned}$$

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な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。

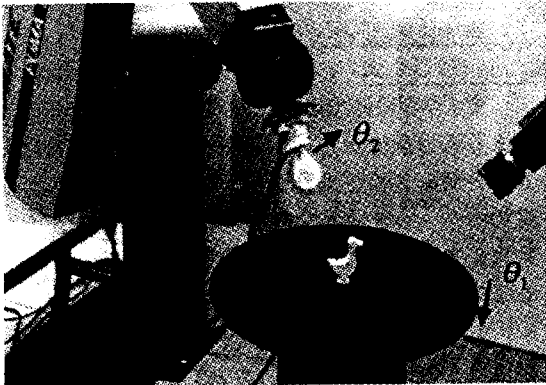


図2. 3次元物体の学習画像収集システム。

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。

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

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

($\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_N$) に対応する固有ベクトル ($e_1 \dots e_k$) 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、特異値分解やSTA法⁹⁾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。

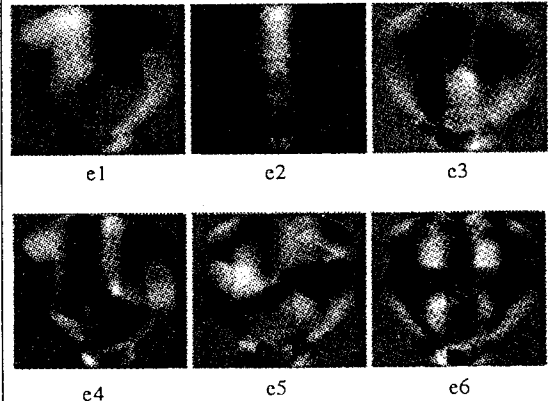


図3. 図1に示した物体の固有ベクトルの例。

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。

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

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

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。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。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。 θ_1 、 θ_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

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

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。

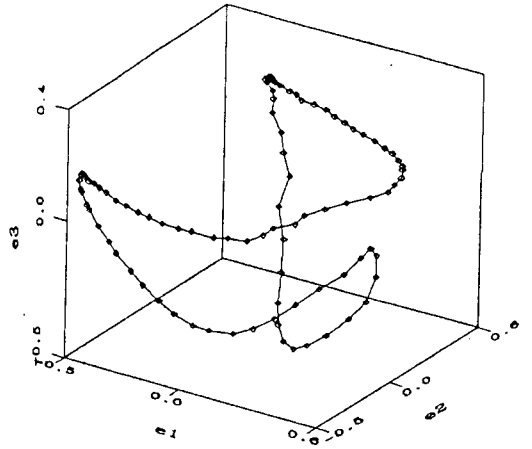


図4. パラメトリック固有空間の例(図1の物体の場合)。

3. 物体認識への応用

3.1 認識手法

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物体領域を切り出した後に大きさと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。

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

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 が識別された後にその物体のポーズを推定する。まず、入力画像 y を式

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

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。

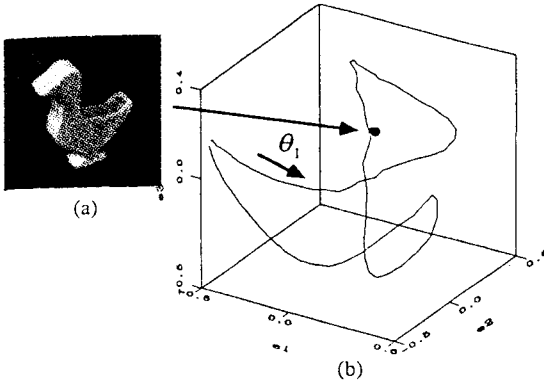


図5.(a)入力図形、(b)固有空間への投影(投影点の曲線上の位置が物体のポーズに対応)。

3.2 物体認識の実験

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)。ここでは2種類の実験を行った。

第1の実験では図6(a)に示す4種類の物体を用いた。学習用のデータとしては、4度おきに360度回転させた物体(90種類のポーズ)を、5箇所の連続的に変化する光源のもとで収集した1800枚の画像を用いた。認識評価用データとしては、学習サンプルと2度位相の異なる90ポーズの物体を3カ所の異なる光源のもとで収集した(合計1080画像)画像を用いた。それらのデータに対し2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。認識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す。平均値で1.2度のポーズ推定誤差であった。

第2の実験では、物体の種類を20種類に増した。図8に実験に使用した物体を示す。この実験では光源の位置は固定し、各物体を5度おきに回転させて撮影した画像(合計1440画像)を学習に使用した。この20種類の物体をランダムに回転させて撮影した320枚の画像に対して認識実験を行ったところ、99%以上の識別率が得られ、平均で1.5度程度のポーズの推定誤差であった。本実験により20種類程度の物体の識別に対してもパラメトリック固有空間法は有効であることを確認した。

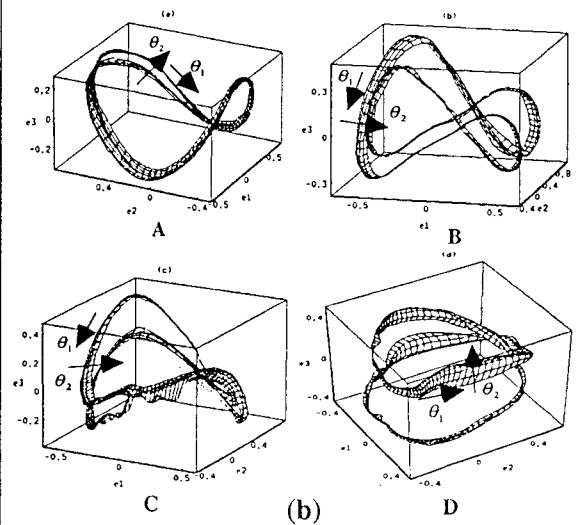
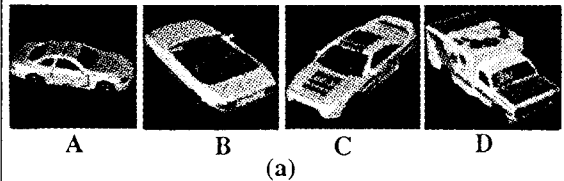


図6.(a)実験に使用した物体の例、(b)パラメトリック固有空間の例((a)の物体に対応)。

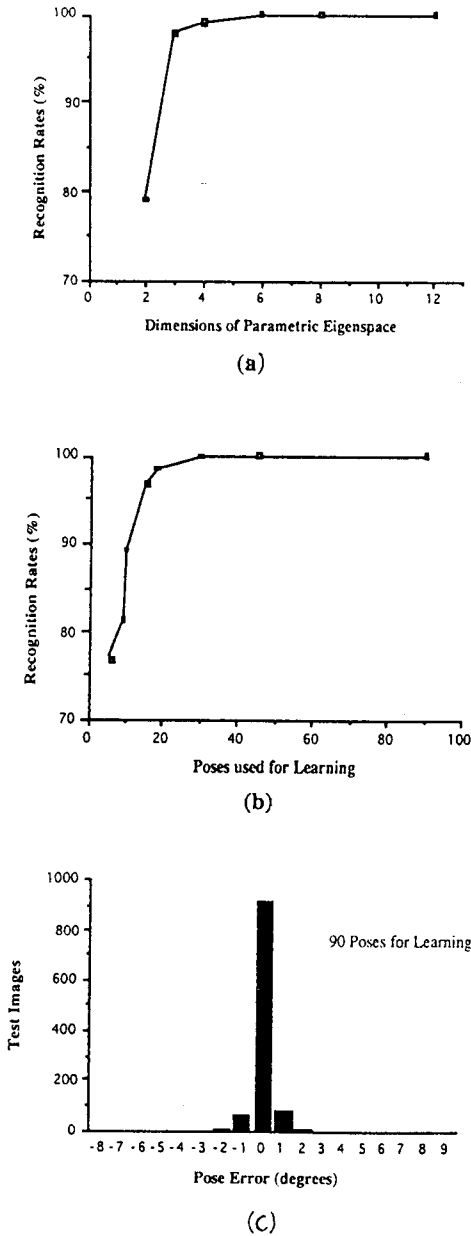


図7. (a)パラメトリック固有空間の次元と認識率との関係、(b)学習に用いたポーズ数と認識率との関係、(c)ポーズ推定誤差のヒストグラム。



図8. 実験2で使用した20種類の物体。

4、3次元物体のスポッティング

ここまでは既に切り出された物体の認識を述べてきたが、実際には物体を背景中から切り出す処理が必要である。本章では同様にパラメトリック固有空間表現を利用して物体を背景画像から切り出す手法（物体のスポッティング）について述べる。多くの実用的な応用では、背景画像の差分や時空間画像を利用することにより物体を容易に切り出すことは可能である。しかし、一枚しか画像が得られず背景パターンも複雑な状況では、物体領域の背景からの切り出しが問題となる。物体が2次元画像のような変形のすくない場合には、単純な2次元フィルタにより物体を検出することが可能であるが、本例のような任意の方向を向いた3次元物体を抽出するためには、単純な相関フィルターでは困難である。また、2次元モデルをアフィン変換して照合する手法も考えられるが、方向の変化が大きい場合にはこの手法だけでは不十分である。一方パラメトリック固有空間表現は任意の方向を向いた3次元物体の画像のような多様

な画像を効率的に表現することができるため、任意の方向に向けた物体のスポットティングにも有効である。

4.1 パラメトリック固有空間法による物体の抽出

ここでは物体の領域を複雑な背景から抽出する手法を人間の顔を例に説明する。具体的な手順を以下に示す。

学習段階では、まず図 9(a)に示すように多数の方向からの顔の画像を収集する。更に大きさの変動を許容するためにその拡大や縮小を行う(図 9(b))。大きさの変動はカメラから顔までの距離の変化に対応するためである。これらの画像の集合体から、次に、物体の方向と物体の拡大率をパラメータとしてパラメトリック固有空間を構成する。その際に、学習サンプル全てに対して背景を削除するような窓関数を設定する。窓関数は学習に利用した画像の物体領域の積集合により自動的に構成される。図 9(c)に窓関数を通して見た固有ベクトルの例を示す。

入力画像から物体の領域を自動的にスポットティングするためには、まず入力画像から上述の窓関数を用いて画像の部分領域を抽出し、これを固有空間上の点に投影する。次にこの投影点と多様体との距離を計算する。窓関数を入力画像上でスキャンすることにより、距離の小くなる点を顔の抽出点とする。また投影された点が多様体上のどの位置にあるかにより顔がどの方向を向いているかも検出可能である。

4.2 スポットティングの実験

ここで示した実験では顔の前半分(180度)を18方向から撮影したものを学習サンプルとして使用した。また顔の大きさは学習サンプルを5段階に拡大、縮小させたものを用いた。未知画像から物体をスポットティングする実験では、512x512画素の入力画像を使用し、この画像を4画素おきにスキャンしながら照合をとった。この場合計算時間はSUN SS10ワークステーションで約2分ほどであった。数種類の入力画像に対して実験を行ったところ良好な結果得られた。図 10に顔の位置、および顔の向きが検出された例を示す。

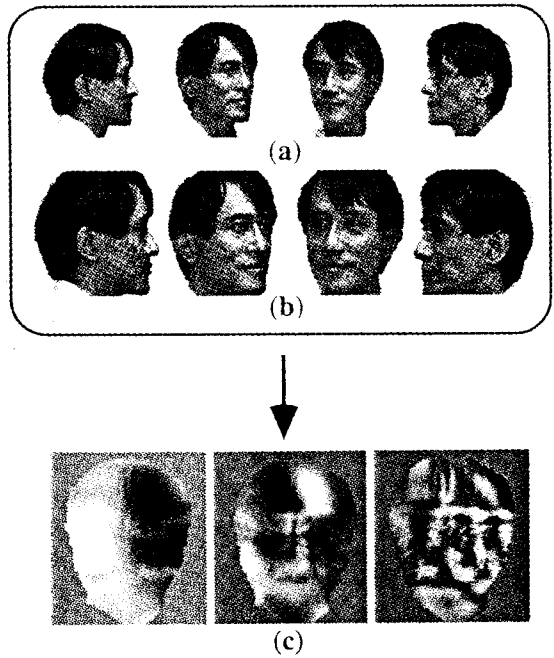


図 9. 学習段階では多数の方向から見た物体画像(a)と、それらを拡大縮小した画像(b)から固有ベクトル(c)を計算し、次にパラメトリック固有空間を構成する。

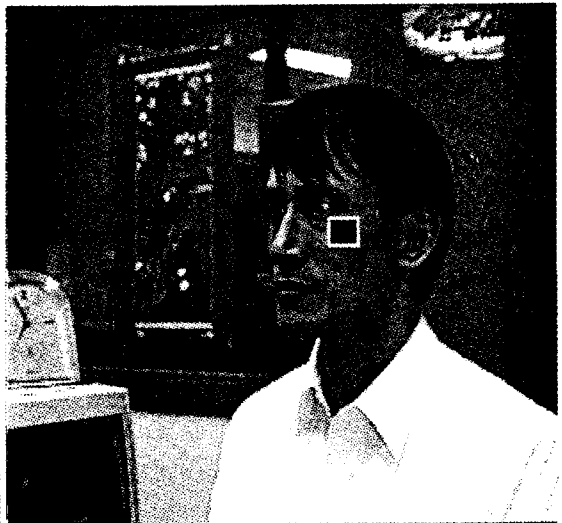


図 10. 認識段階では顔の位置をスポットティングし、更に顔の方向を検出する。(結果: ポーズ = 45度)

5、結論

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中から切り出す(スポッティング)手法、および3次元物体を識別し同時に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次元物体をスポッティング、認識、およびポーズ検出をすることが可能となった。更にこの考え方は動画像処理¹¹³⁾へも適用可能である。

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を持たれている。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¹¹²⁾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。

今回の実験では、認識では物体の1軸回転と光源の位置の2パラメータを、また物体のスポッティングでは顔物体の1軸回転と大きさの2パラメータを仮定したが、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、および同グループ員に深謝します。

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. Weinsall, A self-organizing multiple-view representation of 3D objects, Biological Cybernetics, Vol. 64, pp. 209-219, 1991.
- [13] 村瀬洋, Shree Nayar, 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法一、信学技報PRU93-120, pp31-38, 1993.