

局所特徴量の部分空間を用いた複雑背景下での 3次元特定物体認識

岡田 拓郎[†] 柏木 隆宏^{††} 黄瀬 浩一^{††}

[†] 大阪府立大学 工学部 〒599-8531 大阪府堺市中区学園町 1-1

^{††} 大阪府立大学大学院 工学研究科 〒599-8531 大阪府堺市中区学園町 1-1

E-mail: [†]okada@m.cs.osakafu-u.ac.jp, ^{††}kashiwagi@m.cs.osakafu-u.ac.jp, ^{††}kise@cs.osakafu-u.ac.jp

あらまし 3次元特定物体認識の問題の一つは、複雑な背景下での認識である。単純な背景下では高い認識率を得ることができる手法であっても、背景が複雑になると有効性が大幅に低下することもある。本稿では、我々が以前に提案した局所特徴量の部分空間を用いた認識手法を対象として、その手法が複雑背景に対してどの程度有効であるのかを実験的に検証する。認識手法は、局所特徴量の大规模な部分空間を用いて、相互部分空間法により認識するものである。背景の複雑さを変更した5種類のクエリを用いて認識実験を行い、複雑背景が与える影響について考察する。キーワード 相互部分空間法, 局所特徴量, 3次元物体認識, 近似最近傍探索, 複雑背景

1. はじめに

近年、カメラ付き携帯電話やスマートフォンなどの普及により、画像データや動画データを手軽に撮影・活用できるようになった。それに伴い、画像・動画データを用いた様々なサービスが提供され始めている。その一つとして、画像中や動画中の物体インスタンス(3次元特定物体;以後、単に物体と呼ぶ)が何であるかを認識し、それに関連したWebページなどを開くといったサービスがある。これを実現するためには、コンピュータ処理によって画像中の物体を正しく識別できなければならない。また、このとき、特別な背景を用意することなく、日常の風景の中で識別を行えば便利であると考えられる。しかし、3次元特定物体認識(以後、単に物体認識と呼ぶ)の研究は盛んに行われているが、複雑な背景下での認識は、まだ未解決の問題である。

さて、一般的に、画像認識に用いられる特徴量は、大域特徴量と局所特徴量の2つに大きく分けられる。前者は、画像全体から得る1つの特徴量であるため、認識対象以外の物体が映っている場合は、その認識は困難となる。一方、後者は1枚の画像の様々な局所領域から得る特徴量であり、認識対象物体以外の背景が映っている場合にも用いることができる。しかし、局所特徴量は、画像1枚あたり数百から数千個得られるため、その全てをデータベースに登録すると、メモリ使用量が膨大になるという問題がある。

そこで我々は以前、物体の同じ領域から抽出した局所特徴量の集合から部分空間を作成し、データベースに登録することでメモリ使用量を削減する手法を提案した[1]。この手法では、動画として与えられたクエリから、局所特徴量を抽出し、作成した部分空間をデータベースのそれと照合する相互部分空間法[2]を用いている。また、この手法は、背景なしのクエリに対して

は高い認識率を得ることができる。実際、1002物体の認識実験において、白色無地の背景下で撮影した手持ちクエリに対して90%以上の認識率を得た[1]。しかし、複雑背景下においてどの程度有効であるかの検証は、まだなされていない。

そこで本稿では、この検証を行った結果を報告する。用いた手法は、相互部分空間法において、近似最近傍探索による部分空間の絞り込みを行った後、部分空間同士の類似度の比較という、2段階の処理を行うことで、複雑背景下において物体を認識するものである。実験では、1002物体の画像データセットから、計約125万個の部分空間を作成し、これを用いて物体の認識を行った。その結果、複雑背景下において撮影したクエリに対して、最大78%の認識率を得ることができた。

2. 関連研究

2.1 3次元物体認識

ここでは、局所特徴量を用いた物体認識手法について述べる。最も単純な認識手法は、以下の通りである。まず、あらかじめ物体を様々な角度から撮影し、それらの画像から局所特徴量を抽出し、その全てをデータベースに物体IDとともに登録する。そして、クエリ画像から抽出した各局所特徴量と、データベース中のそれとを比較し、最近傍となる特徴量をもつ物体IDに投票を行う。クエリの局所特徴量すべてについてこの処理を行い、最終的に得票数が最大となった物体を認識結果として出力する。

局所特徴量を用いた3次元物体認識における初期の手法として、Loweらによってview clustering[3]が提案されている。これは、物体を様々な角度から撮影した画像をデータベースに登録する際に、似ている画像同士をクラスタリングすることで、登録する画像数を減らすというものである。また、Kimらによって提案された手法[4]では、更にデータベースを小さくす

るために、異なる物体から抽出した局所特徴量であっても、類似していれば共有して登録している。

3次元物体認識に局所特徴量を用いる場合、あらかじめ様々な視点で撮影した画像から局所特徴量を抽出し、それをデータベースに登録するため、登録する物体数が増えれば増えるほど、そのメモリ使用量は膨大なものとなる。そこで、データベースに登録する特徴量の数を削減するなどの工夫によって、メモリ使用量を小さくすることが必要となる。

2.2 部分空間法

3次元物体認識におけるデータベースのメモリ使用量を削減する手法として、部分空間を用いた手法がある[5]。この手法では、それぞれの物体を様々な視点から撮影し、それらの画像から得られる特徴量の集合を主成分分析することで、各物体を部分空間として表現し、メモリ使用量を削減している。部分空間法では、クエリ画像から得られた特徴量を、各物体の部分空間に射影することで、特徴量と部分空間との類似度を求め、この類似度を用いて認識する。

この部分空間法は、これまでに多くの手法に拡張されてきた。その代表的なものの一つに、相互部分空間法[6]と呼ばれるものがある。相互部分空間法では、データベースとクエリの双方に部分空間を用いて、両者の類似度によって物体を認識する。まず、クエリとして物体を様々な視点から撮影した複数の画像を与え、それらからデータベースと同様に部分空間を作成する。そして、その部分空間とデータベース中の部分空間との正準角によって類似度を求め、物体を認識する。これにより、クエリから抽出した特徴量そのものを用いる場合より、高精度な認識を行うことができる。

部分空間法に関連するほとんどの手法[7][8]が、画像全体から特徴量を抽出し、部分空間を作成している。そのため、データベースに登録する部分空間の数は比較的少ない。一方、我々は局所特徴量ごとに部分空間を作成するため、データベースに登録する部分空間の数は膨大なものとなる。そこで、そのような莫大な数の部分空間を扱うために、類似度の高い部分空間を高速に探索する手法の提案が必要となる。

2.3 近似探索

特徴量同士の比較においては、ANN[9]やLSH[10]のように近似を用いて高速に探索する手法が多く提案されている。しかし、これらの手法はユークリッド距離における最近傍を探索するものであり、部分空間の探索には直接的に用いることはできない。

最近、Basriらによって approximate nearest subspace search [11]と呼ばれる部分空間を近似的に探索する手法が提案された。この手法では、最も類似度の高い部分空間を求める問題を、ベクトル同士の近似最近傍探索の問題へと変換している。しかしながら、[11]でも述べられているように、この手法に必要なメモリ量は、 $O(d^2)$ (d はベクトルの次元数)と莫大である。そのため、次元数 d が大きくなるほど、必要なメモリ量は増加してしまうという問題があり、容易に用いることはできない。

本稿では、approximate nearest subspace search より単純で、部分空間の数が莫大な場合においても、高速な処理を行う

ことができる手法を用いる[1]。この手法は、部分空間の基底のみを用いた近似最近傍探索と、2次元以上の部分空間を用いた類似度計算の2段階処理を行うことで、3次元物体の認識を行う。多次元の部分空間を1次元にすることは、情報を大きく欠損することになるが、投票処理と、処理の2段階化によってその問題を解決することができる。

2.4 複雑背景下での物体認識

複雑な背景を伴った状況で物体を認識する場合、一般に背景から得られる特徴量が認識の障害となる。局所特徴量を用いる場合、基本的にはどの局所特徴量が前景(物体)のものであるのかを識別する必要がある。この目的のため、これまでにいくつかの手法が提案されている。

最も簡便な手法は、JegouらによるWGC(weak geometric consistency)である[12]。この手法は、撮影条件が変化しても、クエリ画像とデータベース画像の間で、局所特徴量のスケール比と方向差が変化しないことを利用して、照合対象となる局所特徴量を取捨選択するものである。ただし、スケール比と方向差を画像全体で推定し、そこに背景から得られた局所特徴量も混入するため、複雑背景から多数の特徴が得られている場合には動作が不安定になる。

画像をパッチと呼ばれる小領域に分けてそれを識別するというアプローチも提案されている[13]。パッチが前景か背景の一方のみに属しており、各パッチから識別に十分な情報が取り出せるならば複雑背景下でも前景のみのパッチを認識することによって物体認識が可能となる。しかし、このような条件が成り立つことは一般的に難しい。

局所特徴量が単独では十分な識別性を持たなくても、それらを複数組み合わせることによって、識別性を強化するという考えがある。Liebeらの手法[14]は、局所特徴量の照合に一般化ハフ変換の方式を導入したものである。データベース側の局所特徴量には参照点の情報が対応付けられており、検索質問の局所特徴量との照合によって検索される。照合された局所特徴量が同じ物体から得られたものであれば、参照点は共通するという点に着目し、照合結果を取捨選択している。

Martinらの手法[15]は、上記のパッチや参照点を組み合わせるとともに、局所特徴量のスケールと方向を用いた投票処理を行うことによって、より精度を高めたものである。

局所特徴量を照合する際に、どの程度の処理を用いると背景領域との誤照合の影響を軽減できるのかについては、元となる照合方式の性能や特徴量の識別性に依存する。そこで本稿では、将来的には上記のような方式を導入することを念頭においた上で、局所特徴量の部分空間が背景領域に対してどの程度ロバストであるのかを実験的に確かめる。

3. 物体認識手法

本研究で用いる物体認識手法では、物体を回転させて撮影することで取得した、連続的なフレーム画像から、物体の同じ領域にもとづく局所特徴量の集合を求め、部分空間を作成する[16]。これにより、物体が回転したときの局所特徴量の変化を部分空間上で表現し、様々な視点からの物体の認識を可能と

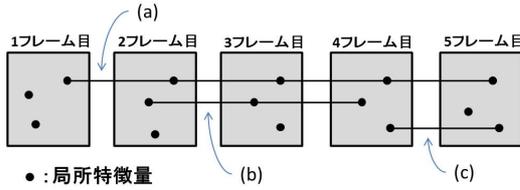


図 1 対応付けを行った特徴量の軌跡の例

している．クエリには動画像を用い，クエリからも局所特徴量の部分空間を作成し，これをデータベースに登録してある部分空間と比較し，物体を認識する．

3.1 局所特徴量の選定

撮影角度が連続して変化する複数の画像データにおいて，複数のフレーム画像に渡って一貫している局所領域に注目する．そのような部分から得る局所特徴量を用いることで，様々な撮影角度に対応する部分空間を得ることができる．

この処理は以下ようになる．まず，連続するフレーム画像において， i フレーム目の画像データの各特徴量と $i+1$ フレーム目の画像データの全ての特徴量との座標のユークリッド距離を求める．この値がある閾値よりも小さい特徴量を，対応付ける候補とする．そして，その集合から特徴量の値について最近傍の特徴量を検出する．このようにして 2 枚の画像の局所特徴量の対応付けを行う．このとき，得られた対応関係が，ノイズなどの影響により，本当は関連のない部分である可能性がある．そこで，ノイズであるかどうかの判定処理を行う．まず， i フレーム目の特徴量を f ， $i+1$ フレーム目の f と最も近い特徴量を g_1 ，2 番目に近い特徴量を g_2 とし， f と g の距離を $d(f, g)$ とする．このとき，閾値を α とすると，

$$\frac{d(f, g_1)}{d(f, g_2)} < \alpha$$

を満たす特徴量の組のみを，物体の同じ部分から得られたものであるとし，対応付けを行う．これを満たさない特徴量はノイズであると考え，除去する．これを連続するすべての画像データ間で行い，局所特徴量の軌跡を得る．以下，このような処理をトレース処理とよぶ．図 1 に，トレース処理の模式図を示す．図 1 では，黒い点が各フレーム画像における局所特徴量であり，(a)，(b)，(c) が特徴量の軌跡を表している．また，それぞれの軌跡の長さを，軌跡が続いたフレーム数から 1 引いた値であると定義する．例えば，軌跡 (a)，(b)，(c) の長さはそれぞれ，4，2，1 である．

3.2 部分空間の作成・登録処理

トレース処理によって得られた局所特徴量の集合から，KL 展開を用いて部分空間を作成する．このとき，対応付けられる軌跡の長さは，図 1 の (a)，(b)，(c) のように様々である．対応付けを行った全ての軌跡から部分空間を作成し，データベースに登録するとその数は莫大なものとなる．一方で，軌跡が短いものから作成した部分空間は，軌跡の長いものから作成したそれに比べて識別できる撮影範囲が少ないため，このような部分空間に登録するのは非効率である．そこで，局所特徴量の軌跡の長さの閾値 T を設け， T フレーム以上軌跡が続いた局所

特徴量の集合のみから部分空間を作成し，データベースに登録する．

KL 展開により局所特徴量の集合から部分空間を作成する処理を次に示す．まず，トレース処理で対応付いた多数の局所特徴量 x から，それらの自己相関行列 R を

$$R = E\{xx^T\}$$

によって計算する．そして， λ_i を R の固有値， u_i をそれに對する固有ベクトルとすると，

$$Ru_i = \lambda_i u_i$$

の固有値問題を解くことで固有ベクトルを求める．求めたい部分空間の次元数を n とすると，第 1 固有ベクトルから第 n 固有ベクトルまでの組

$$(u_1, u_2, \dots, u_n)$$

がその部分空間の基底となる．

このように作成した部分空間は，局所特徴量の変化を近似的によく表現することができる．また，KL 展開を用いることにより，部分空間を作成する際には存在しなかったフレーム間の画像の局所特徴量を補完的に表現できる．データベースには，各部分空間の基底ベクトルと物体 ID を登録する．

3.3 相互部分空間法

相互部分空間法 [2] は，クエリが動画像である場合に用いることができる．データベースを作成するときと同様に，フレーム画像から局所特徴量を抽出し，トレース処理によって，対応付けを行う．そして， T_q フレーム以上軌跡が続いた局所特徴量の集合のみから部分空間を作成する．

一般に，相互部分空間法における類似度は，部分空間同士の正準角を算出することによって求める．正準角の計算処理を次に示す．まず， M 次元部分空間 P と N 次元部分空間 Q との成す正準角を $\theta_i (i = 1, 2, \dots, M; M \leq N)$ とする．また， $u_j (j = 1, 2, \dots, M)$ ， $v_k (k = 1, 2, \dots, N)$ をそれぞれ部分空間 M および部分空間 N の正規直交基底ベクトルとし，

$$U = (u_1 \ u_2 \ \dots \ u_M)^T$$

$$V = (v_1 \ v_2 \ \dots \ v_N)^T$$

とする．このとき， $U^T V$ の第 i 番目に大きい特異値を λ_i とすると，正準角 θ_i は

$$\theta_i = \cos^{-1} \lambda_i$$

となる．このとき，正準角が小さいほど，2 つの部分空間の類似度は大きくなる．

これらの式からわかるように，正準角の計算コストは部分空間の数や次元数に大きく依存する．ゆえに，すべての部分空間との正準角を求めようとすると，その計算コストは膨大なものになる．

3.4 正準角の計算による認識処理

次に，認識処理について示す．クエリから作成した部分空間



図 2 データベースに登録した物体の一例

の数が l 個であるとき、各部分空間を $s_i (i = 1, 2, \dots, l)$ とする。 s_i とデータベースに登録してある部分空間全てとの類似度を、3.3 で述べた計算により求める。そして、最も類似度が高い部分空間にひもづく物体 ID に投票処理を行う。この処理を、クエリから作成した l 個全ての部分空間について行い、最終的に得票数が最大となった物体 ID を認識結果として出力する。

このとき、データベースに登録してある部分空間の数が物体ごとに異なるため、各得票数を物体ごとの部分空間の数で正規化する。物体 j の得票数を v_j 、データベースに登録してある物体 j の部分空間の数を N_j 、とすると、得票数を正規化した後のスコアは、 $w_j = v_j / \sqrt{N_j}$ となる。

3.5 認識処理における 2 段階化

3.3 で示したように、部分空間の数や次元数が多いと、正準角による類似度の計算は莫大になる。そこで、2 段階の認識処理を行い、高速化することを考える。まず、部分空間の 1 次元のみを用いた近似最近傍探索によって、類似度の高い k 個の部分空間を高速に求める。そして、絞り込んだ k 個の n 次元 ($n > 1$) の部分空間と、 m 次元 ($m > 1$) の部分空間 s_i との正準角を計算し、最も類似度の高い部分空間を求める。

1 段階目の処理に部分空間の基底を用いるのは、基底の距離が小さいほど、2 つの部分空間の類似度は高くなるという性質があるためである [1]。この性質により、ANN などの近似最近傍探索の手法を用いて、認識処理を高速化することができる。

上記のように、部分空間の次元数を 1 次元すると、通常、情報が大きく欠損する。しかし、投票処理によって認識を行う場合は、各部分空間の識別精度が低くても大きな問題が生じにくい。なぜなら、誤投票が分散し、正解の物体の得票数が他の物体へのそれよりも大きくなり、正しく認識できるからである。また、2 段階目の処理では、 n 次元 ($n > 1$) と m 次元 ($m > 1$) 次元の部分空間で比較しているため、近似計算の影響は小さい。

4. 実験

本稿では、ぬいぐるみやミニカー、フィギュア、日用雑貨など 1002 個の物体を用いて、複雑背景下におけるクエリに対する、従来手法の認識精度の検証実験を行った。図 2 に、実験に用いた物体の一例を示す。これらの物体の中には、互いの形状や模様が似ているものも多く、その識別は容易ではない。また、フィギュアなどには突起部分が多く、少しの視点の変化によって物体の見え方が大きく変化するものもある。

4.1 実験準備

まず、データベースについて説明する。上記の 1002 個の 3 次

表 1 局所特徴量と部分空間の 1 クエリあたりの数

	トレース処理によって得た局所特徴量の数	部分空間の数
Q1	$2.17 \times 10^3 \sim 2.15 \times 10^5$	31 ~ 1191
Q2	$2.17 \times 10^3 \sim 2.22 \times 10^5$	31 ~ 1125
Q3	$3.41 \times 10^3 \sim 2.74 \times 10^5$	47 ~ 1361
Q4	$6.11 \times 10^3 \sim 3.12 \times 10^5$	76 ~ 1441
Q5	$6.04 \times 10^3 \sim 5.58 \times 10^5$	86 ~ 2771
Q6	$1.28 \times 10^4 \sim 6.86 \times 10^5$	173 ~ 2983
	$(T_q = 50) \sim (T_q = 10)$	同左

元物体をターンテーブルで 1 回転させ、正面、上 15 度、上 30 度からウェブカメラを用いて動画を撮影し、フレーム画像 (解像度: VGA) に分割した。そして、これらの画像から SIFT 特徴量 [17] を抽出し、トレース処理を行い部分空間を作成し、データベースとした。ここで、部分空間を作成する局所特徴量の集合は、トレース処理によって得られる軌跡の長さ T が $T \geq 50$ を満たすもののみとした。登録した部分空間の数は、1 物体あたり約 1200 個で、全物体での総数は約 125 万個であった。

次に、本実験で用いたクエリについて説明する。データベースに用いた 1002 物体の中から、無作為に 50 物体を選択し、それぞれ白色無地の単純背景 (Q1) と 4 種の複雑背景 (Q2 ~ Q5) において、手で持って動画を撮影した。また、1002 物体に含まれるフィギュアから無作為に 44 物体を選択し、データベースに登録されていない他のフィギュア 20 数個を背景として、同様に動画を作成した (Q6)。Q2 ~ Q4 では、事務机の上に電源タップやペンケースなどを置き、これを背景物体とした。そして、認識対象と背景物体との相対距離を 3 段階に変えて撮影した。また、Q6 は、PC デスクの前で PC モニタやキーボード、スタンドライトなどを背景として撮影したものである。各クエリにおける、認識対象物体と背景物体との距離は以下の通りである。

(Q2) 認識対象と物体との距離が 30cm 以内。

(Q3) 認識対象と物体との距離が 30cm から 1m 程度の間。

(Q4) 認識対象と物体との距離が 1m から 5m 程度の間。

(Q5) 認識対象と物体との距離が 20cm から 60cm 程度の間。

(Q6) 認識対象と物体との距離が 20cm から 50cm 程度の間。

図 3 に作成した 6 種類のクエリをそれぞれ示す。データベースの作成時と同様に、撮影にはウェブカメラを用い、1 物体 1 背景あたり 30fps で解像度を VGA とし、フレーム長を 600 フレームとした。そして、これらの画像から SIFT 特徴量を抽出し、軌跡の長さ T_q が $T_q \in \{10, 20, 30, 40, 50\}$ フレーム以上のものから部分空間を作成し、クエリとした。表 1 に、クエリの、トレース処理によって得た局所特徴量の数と部分空間の数を示す。

本実験では、近似最近傍探索の手法として、ANN [9] を用いた。ANN では、パラメータ ϵ を用いて、探索する超球の半径を $1/(1 + \epsilon)$ 倍に縮小することで、近似の度合いを調節することができる。 ϵ の値が大きいくほど、近似の度合いは強くなり、また、処理時間は短くなる。

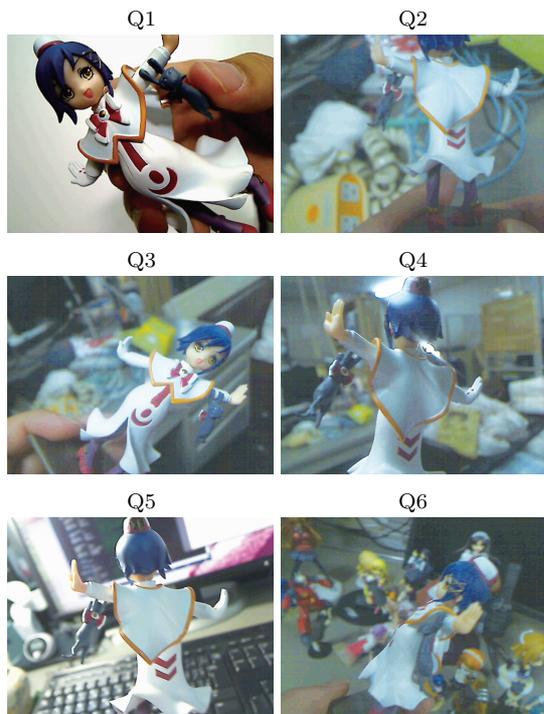


図 3 各クエリの一例

4.2 実験 1

まず、背景の複雑度合いが、認識率に与える影響を調べるため、クエリ Q1~Q5 を用いて認識実験を行った。

部分空間の次元数を $n = 5$ 、1 段階目に絞り込む部分空間の数を $k = 100$ とし、ANN のパラメータ ϵ には $\epsilon = 2$ を用いた。トレース処理の閾値 $T_q \in \{10, 20, 30, 40, 50\}$ を変化させたときの、認識率と部分空間の数を図 4 に示す。背景の種類によらず、部分空間の数が多いほど、認識率はよく、単純背景のクエリ Q1 では 98% と高精度に認識できた。一方、複雑背景のクエリでは、 T_q の値が同じとき、Q1 よりも部分空間の数は大きい、認識率は 20% 以上低下し、最大でも 78% であった。これは、背景に基づく部分空間が加わったことで、正解物体の得票数が相対的に低くなったためである。また、部分空間の数は、物体と背景との距離が遠くなるほど多くなった。

次に、部分空間の次元数を $n = 5$ 、トレース閾値を $T_q = 10$ 、1 段階目で絞り込む部分空間の数を $k = 100$ とした。ANN のパラメータ $\epsilon \in \{1, 2, 3, 4, 5\}$ を変化させて認識を行ったときの、認識率と処理時間を図 5 に示す。同じクエリにおいては、 ϵ の値を大きくしても、認識率はあまり低下することなく、処理時間が短くなった。また、背景との距離が遠くなるほど、処理時間が増加し、認識率が低下した。これは、背景が遠いほど、映りこむ物体数が増加し、データベースに登録された物体の一部に類似することが多くなったためである。

また、クエリごとに、誤認識が集中しやすい物体が存在した。図 6 に Q5 で誤認識が集中した物体を示す。図 6 の (a) の物体は、チェック模様がキーボードと類似しているために、誤投票が増えたと考えられる。実際、この物体に誤認識したクエリは、その他の物体に誤認識したクエリに比べて、キーボードが映っているフレームが多かった。

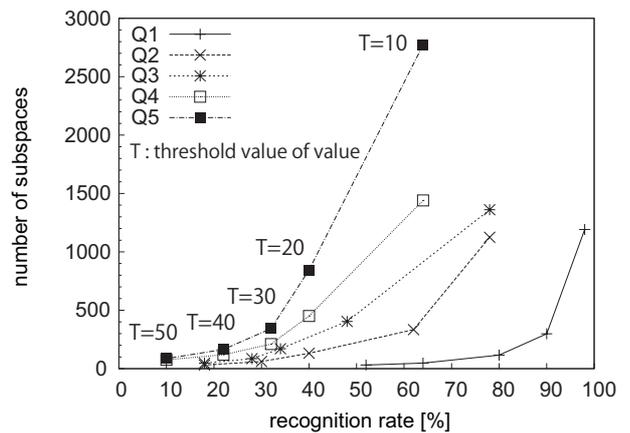


図 4 単純背景・複雑背景の認識率と部分空間の数

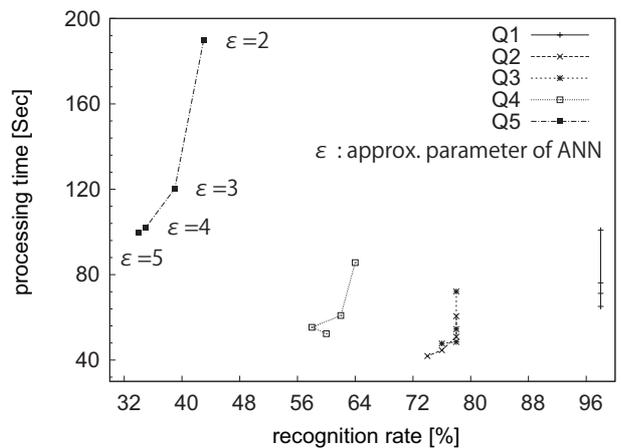


図 5 単純背景・複雑背景の認識率と処理時間



図 6 クエリ Q5 で誤認識結果となった物体の一例

Q5 において、認識率が他のクエリに比べて低いのは、照明の影響によるものである。認識できなかった物体の多くは、ライトを消した状態では正しく認識できた。しかし、表面に光沢があるものや、テクスチャの乏しいものは、照明を消しても認識に失敗した。これらは、Q2~Q3 においても同様に認識率の低い物体である。一方、色の濃淡がはっきりしているものや、全体の形状が丸い物体などは、どの背景でも正しく認識した。これらの物体は、物体から得られる特徴量の数が多く、撮影時に映りこむ背景が少ないため、認識率が高くなったと考えられる。

4.3 実験 2

次に、認識対象と似たような物体が背景にある場合の認識率を調べるため、Q6 を用いて実験を行った。

トレース処理の閾値は $T_q = 10$ とし、ANN のパラメータ ϵ には $\epsilon = 5$ を用いた。図 7 に、1 段階目に ANN によって絞り込む部分空間の数 $k \in \{10, 50, 100\}$ と、2 段階目で用いる部分

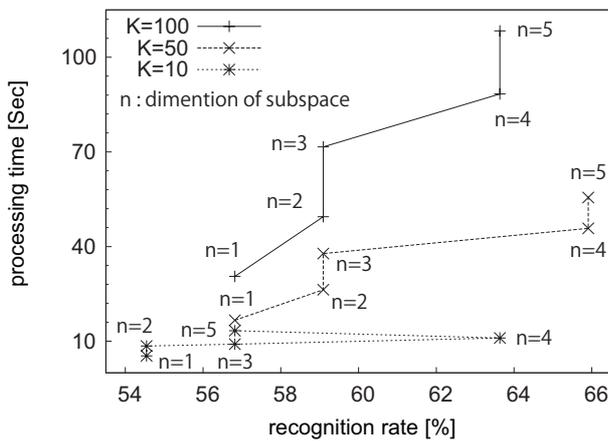


図7 背景に類似物体が含まれるクエリの認識率と処理時間

空間の次元数 $n \in \{1, 2, 3, 4, 5\}$ を変化させて認識を行った場合の、認識率と処理時間を示す。 $k = 100$ のときが必ずしも良い認識結果になるとは限らず、 $n = 4$ や $n = 5$ のときは $k = 50$ とした方が高い認識率を示した。これは、絞り込む候補が多くなることにより、背景情報から得た部分空間が特定の物体に誤対応することが多くなったためである。また、表1に示したように、クエリ Q6 の部分空間の数は、単純背景クエリ Q1 の3倍近くであった。認識対象から抽出した特徴量による部分空間の数はほとんど変化しないはずなので、この増加分の大半は、背景によるものである。特定の物体に誤認識が集中することはなかったことや、背景物体と形状のよく似たフィギュアがデータベースに沢山登録されていることから、Q6 の認識率が低くなったのは、複数の背景物体の様々な部分が誤対応したことが原因である。背景のみを撮影したクエリを数個作成し、認識実験を行ったところ、フィギュアであると誤認識することが殆どであった。 $k = 10$ において、 $n = 5$ のときに、認識率が低下しているのも同様に、解の候補を広げることで誤認識が増加したためである。

また、処理時間は、 $k = 10$ においては、 n を大きくしてもほとんど変わらなかった。これは、絞り込む候補が少なく、2段階目の処理に要する時間が少ないためである。部分空間の正準角を求める際の計算コストは、部分空間の数や次元数に大きく依存するため、 $k = 50$ や $k = 100$ においては、 n の値が大きくなるのに伴って、処理時間が増大した。

5. まとめ

本稿では、我々が先に提案した、局所特徴量の部分空間を用いて相互部分空間法によって物体を認識する手法が、複雑背景下でどのような振る舞いをするのかを実験的に確かめた。その結果、以下のことが明らかとなった。(1) 色の濃淡がはっきりしていて、表面に光沢の少ない物体は、比較的高精度に認識できる一方、テクスチャに乏しいものや、エッジが不明瞭な物体は、背景の種類によらず、認識に失敗しやすい。(2) 物体と背景との距離が遠くなるほど、映りこむ背景物体の数が増えるため、データベースに登録された物体に類似した模様や形状の物体が背景に含まれる可能性が高まり、誤認識が増加する。(3) 物体

の近くに蛍光灯のような光源があると、物体表面のてかりなどが生じ、認識率が低下する。(4) 背景に類似物体が存在すると認識がさらに困難となる。

これらの点は、局所特徴量の部分空間を単独で照合する認識方式の限界を表すものといえる。この問題を解決するためには、2.4 で述べたような局所領域の共起や配置に関する情報を部分空間法にも導入し、部分空間の識別性を高める必要があると考えられる。今後の課題は、このような処理を、部分空間法の特徴を活かした形で実現することである。

謝辞 本研究の一部は科学研究費補助金基盤研究 (B) (22300062) の補助による。

文 献

- [1] 柏木, 黄瀬: “局所特徴量の多数の部分空間を用いた3次元物体認識とその高速化・高精度化”, 電子情報通信学会技術研究報告, **111**, 193, pp. 133-138 (2011).
- [2] 前田, 山口: “相互部分空間法による三次元パターンマッチングにおける正準角の性質”, 情報処理学会研究報告, 2007-CVIM-160.
- [3] D. G. Lowe: “Local feature view clustering for 3D object recognition”, CVPR2001 (2001).
- [4] S. Kim, I. S. Kweon: “Scalable representation for 3D object recognition using feature sharing and view clustering”, Pattern Recognition, **41**, pp. 754-773 (2008).
- [5] S. Watanabe, N. Pakvasa: “Sunspace method in pattern recognition”, Proc. 1st IJCPR (1973).
- [6] 福井, 山口: “部分空間法の理論拡張と物体認識への応用”, 情報処理学会論文誌, **46**, 15, pp. 21-34 (2005).
- [7] 村瀬: “パラメトリック固有空間法による画像認識”, 電子情報通信学会誌, **85**, 4, pp. 252-258 (2002).
- [8] 福井, 山口: “カーネル非線形制約相互部分空間法による物体認識”, 電子情報通信学会論文誌, **88**, 8, pp. 1349-1356 (2005).
- [9] R. S. S. Arya, D. M. Mount and A. Y. Wu: “An optimal algorithm for approximate nearest neighbor searching”, Journal of the ACM, **45**, 6, pp. 891-923 (1998).
- [10] A. Andoni, P. Indyk: “Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions”, Comm. of the ACM, **51**, 1, pp. 117-122 (2008).
- [11] R. Basri, T. Hassner, L. Zelnik-Manor: “Approximate nearest subspace search”, IEEE Trans. PAMI, **33**, 2, pp. 266-278 (2011).
- [12] H. Jegou, M. Douze, C. Schmid: “Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search”, Proc. of ECCV, pp. 304-317 (2008).
- [13] N. Plath, M. Toussaint, S. Nakajima: “Multi-Class Image Segmentation Using Conditional Random Fields and Global Classification”, Proc. of ICML, pp. 817-824 (2009).
- [14] B. Leibe, A. Leonardis, B. Schiele: “Combined Object Categorization and Segmentation with an Implicit Shape Model”, Proc. ECCV workshop on statistical learning in computer vision, pp. 17-32 (2004).
- [15] M. Klinkigt, K. Kise: “From Local Features to Global Shape Constraints: Heterogeneous Matching Scheme for Recognizing Objects under Serious Background Clutter”, Proc. ACCV, pp. 64-75 (2010).
- [16] 柏木, 外山, 黄瀬: “局所特徴量の部分空間を用いた3次元物体認識におけるメモリ削減手法”, 電子情報通信学会技術研究報告, PRMU2010-19, **110**, 27, pp. 97-102 (2010).
- [17] D. Lowe: “Distinctive image features from scaleinvariant keypoints”, International Journal of Computer Vision, **60**, pp. 91-110 (2004).