

スパース表現に基づく識別手法は非線形クラス部分空間に対して頑健か？

～見えに基づく物体認識における検証～

伊東 隼人[†] 酒井 智弥^{††} 井宮 淳^{†††}

[†] 千葉大学大学院 融合科学研究科

^{††} 長崎大学工学部

^{†††} 千葉大学 総合メディア基盤センター

E-mail: [†]hayato-itoh@graduate.chiba-u.jp, ^{††}tsakai@t.nagasaki.jp, ^{†††}imiya@faculty.chiba-u.jp

あらまし 質問データを少数の訓練データで簡潔に線形合成するスパース表現に基づく識別の手法は、特徴量の欠損やノイズに対して頑健な多クラス識別や多重ラベリングを実現できる。この識別手法はクラス部分空間が線形であるパターン認識の問題において有効性が確認されているが、クラス部分空間が非線形であっても、テストデータを線形合成する少数の訓練データを発見できる場合には識別に成功する可能性がある。本研究ではこの可能性を検証するため、見えに基づく物体認識を例にスパース表現に基づく識別手法の性能を評価した。物体の画像データは姿勢・視点の変化に関する非線形多様体を構成するが、訓練データ数が十分多い場合はテストデータの姿勢・視点に近い訓練データの部分集合を頑健に同定できることを実験的に確認した。

キーワード 多クラス分類, 貪欲アルゴリズム, 固有空間法

1. はじめに

画像の見えに基づく物体認識では、識別結果が物体の見えや撮像環境の変化に不変であることが求められる。物体の姿勢や視点位置の変化、照明変動によって物体の見えは変化するため、様々な見えの画像の集合でもって物体のクラスを表現する必要がある。

古典的な部分空間法では画像の見えの変化が張る線形部分空間で物体クラスが表現される [1] [2] [3]。パラメトリック固有空間法 [4] では画像の変化が固有空間に埋め込まれた低次元多様体で表現される。

クラスが既知である訓練データを用いて属すべきクラスが未知の質問データのクラスを決定する問題は、教師あり識別問題と呼ばれる。一般的な識別器は質問データが与えられたとき、すべてのクラスに対して類似度を測り類似度が最大のクラスのラベルを付与する。

一方、質問データを少数の訓練データで簡潔に線形表現することに基づく識別が提案されており、ノイズや特徴量の欠損、オクルージョンに対して高い頑健性が示されている [5] [6] [7]。高い頑健性は、スパース正則化によって質問データに関係する訓練データが特定されるため、無関係なクラスが強く除外されること、ノイズや遮蔽部分は訓練データで簡潔に合成できないことに起因する。

一般に見えの変化に対しては画像の部分空間は線形にならない。部分空間の非線形性に対してカーネル部分空間法は識別率の向上に成功している [8] [9]。

それでは、疎表現に基づく手法もカーネルトリックを利用する核関数化で性能の向上を図るべきであろうか。

クラス部分空間が非線形であっても、質問データを線形に合成できる少数のデータが訓練データに含まれていれば、核関数化せずとも既存の疎表現に基づく手法でこれを見つけ出せることが期待できる。この可能性を検証するため、本研究ではクラス部分空間が部分的に線形近似できる場合の、スパース表現に基づく識別手法の性能を実験にて評価する。

2. クラス部分空間の非線形性

訓練データ、質問データが共に高次元ベクトルで表されるベクトル空間モデルを考える。クラスの要素であるデータはクラス部分空間を張っている。図 1 のような v_1, v_2, v_3, v_4 という 4 つのベクトルを考える。これらのベクトルはそれぞれ高次元ベクトルを表しているものとする。 v_1, v_2, v_3 が同一クラスであるとき、3 つのベクトルのうち 1 つを残りのベクトルの線形結合で表せる。このように、クラスの要素の線形結合も同一クラスの要素であり、同一クラスのデータが 1 つの線形部分空間上にある性質をクラス部分空間の線形性という。一方 v_1, v_2, v_3, v_4 が同一クラスのデータであるとき、 v_4 を残りの 3 つのベクトルの線形結合で表すことはできない。クラスの要素であるデータの線形結合で同一クラスの要素を表わせるとは限らない、同一クラスのデータが 1 つの線形部分空間上にない性質をクラス部分空間の非線形性という。

物体認識では姿勢・視点位置が変化を含む場合、クラス部分空間が非線形になる。姿勢・視点位置が変化することで画像の見えは大きく変わり、同一クラスのデータ

の線形結合で姿勢・視点位置の異なる他の要素のデータを合成することはできないためである。村瀬 [4] が示したように、物体が画像中で回転している場合、画像を画素を座標軸とした空間では回転運動の自由度に一致する次元の非線形多様体として近似されることが知られているなど、物体姿勢、視点位置の変化を含むデータは図 2 のように、ある非線形多様体を作ると考えられる。照明変化のみを含む物体認識の場合はクラス部分空間が線形になることが知られている [10]。また、物体認識ではないが、文書解析に用いられる単語の出現回数などのデータも非線形性を持つ。

部分空間法ではクラス部分空間を張る基底を主成分分析を用いて求め、質問データと求めた基底の張る空間との近さを計って識別を行う。クラス部分空間が線形である場合はクラス部分空間を張る少数の適切な基底を求めることができるが、非線形である場合は主成分分析でクラス部分空間を張る適切な基底を見つけることができないことがあり、認識率の低下を生じる。認識率が低下する可能性があるものの、部分空間法は求めた各クラス部分空間を張る基底と質問データが近ければ識別は可能である。しかし、クラス部分空間が線形部分空間であり、質問データを少数の訓練データで合成することを仮定をしているスパース表現を用いた識別法にとってクラス部分空間が非線形であるデータは不向きである。

部分空間の非線形性に対してカーネル部分空間法は識別率の向上に成功している [8] [9]。核関数を利用したカーネル主成分分析 [11] を利用することで非線形クラス部分空間を張る適切な基底を見つけることができるためである。データを表す高次元ベクトル空間よりさらに高い、無次元に近い空間にデータを射影を考えると、射影前の空間では非線形なクラス部分空間であったとしても、無次元に近い高次元ベクトル空間では線形なクラス部分空間となる。核関数を用いるカーネルトリックを利用することで、実際には無次元に近い高次元への射影を行わずに射影後のベクトル空間での主成分分析が行える。

Wright らは質問データを少数の訓練データの線形結合に分解することに基づく顔画像認識を提案し、高い認識率と頑健性を示した [5]。また、質問データが張る部分空間を少数のクラス部分空間に分解するパターン識別の枠組みを酒井が提案し [6]、伊東らは部分空間の合成を利用したスパース表現を用いることで、Wright らの識別法よりも高い認識性能と頑健な認識が可能であることを示した [7]。

Wright ら、伊東らの実験では cropped version of The Extended Yale Face Database B [12] が用いられている。このデータベースは図 4 に示されるような、38 人の顔を正面から照明の向きや明るさを変えながらそれぞれ 60 枚撮影した画像ものである。すなわち照明変化のみを含むデータ集合であり、各人物のクラス部分空間は線形部分空間になっている。スパース表現を用いたパターン

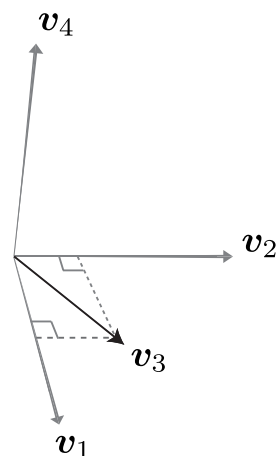


図 1 クラス部分空間の線形・非線形性。 v_1, v_2, v_3 が同一クラスならクラス部分空間は線形である。しかし、 v_4 が同一クラスに含まれる場合、クラス部分空間は非線形である。

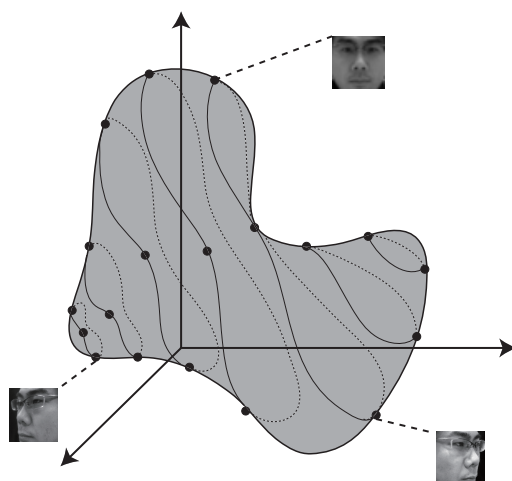


図 2 同一クラスのデータが属する非線形多様体。1つの点が1つのデータを表している。

認識では、テストデータを訓練データの線形結合で表せることを仮定した手法であり、各クラス部分空間が線形部分空間である YaleB を用いた実験は、これらの手法にとって望ましい条件を備えたデータである。そして、スパース表現を用いた識別法の非線形多様体を作るデータに対する有効性は未知である。

しかし、十分な訓練データがあり、かつその中からテストパターンを簡潔に合成できる少数の訓練データを発見できる場合には識別に成功する可能性がある。図 3 の示すような多様体に接する超平面を考えると、局所的には線形部分空間になっていると考えられるためである。本研究では物体認識の問題を対象に、クラス部分空間が線形部分空間でない場合におけるスパース表現を用いた識別法の識別性能を評価し、正しく識別できるのならば、その場合の条件を検証した。

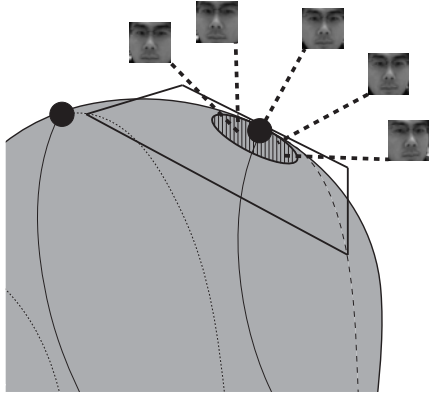


図 3 多様体の局所的な線形性．多様体の一部は接している超平面上にあると考えられる．斜線部は接している超平面上にあることを意味する．



図 4 cropped version of The Extended Yale Face Database B．38 人ごとに照明の向きと明るさを変えた画像を約 60 枚ずつ含む．

3. スパース表現に基づく識別法

3.1 Sparse representation-based classification

疎表現に基づく識別法 (Sparse representation-based classification, 以下 SRC) は, 質問データを簡潔に表現する少数の訓練データに基づきクラスを特定する [5]. 十分な訓練データがあるとき, 質問データを合成する訓練データの組み合わせは一意ではない. しかし, 簡潔な表現が存在するとき, 質問データの線形合成に用いる訓練データ数が少ないことを奨励する疎正則化 [13] を利用することで識別が可能となる.

3.1.1 質問データの疎分解

パターンのデータを d 次元列ベクトルで表し, k ($= 1, \dots, C$) 番目のクラスの訓練データを列に並べた行列を $S_k \in \mathbb{R}^{d \times n_k}$ とする. 各クラスの訓練データ行列を順に並べたものをひとつの訓練データ行列 $S = [S_1, \dots, S_C] \in \mathbb{R}^{d \times N}$ とし, 質問データを $q \in \mathbb{R}^d$ と表す. 訓練データは十分にあるとし, $d < N$ である. ここで,

$$N := \sum_{k=1}^C n_k, \quad (1)$$

は各クラスの訓練データ数の和を表す. 質問データを訓練データで線形に表現することは, 次の線形方程式から線形結合係数 α を求めることに相当する.

$$q = S\alpha \quad (2)$$

これは, $d < N$ の不定方程式であり, 一意な解を求めることはできない. 質問データ q に対しては, 属すべきカテゴリの学習データだけが簡潔な説明をすることができると期待できるため, 次のような疎正則化を行うことで一意な解を得られる.

$$\min \|\alpha\|_0 \text{ subject to } q = S\alpha \quad (3)$$

このように線形結合係数 α が疎なベクトルになるような線形結合を疎表現と呼ぶ. また, α の要素で非ゼロである成分が高々 s 個である場合, これを s -sparse であるという. ℓ_0 ノルムで正則化を行った式 (3) は NP 困難であり, 実時間で解を得ることは出来ない. しかし貪欲法を用いることで, 次のような緩和を行った場合でも高確率で式 (3) の解と一致することが知られている [13].

$$\min \|\alpha\|_1 \text{ subject to } q = S\alpha \quad (4)$$

3.1.2 次元削減と解の保証

十分なデータ数があるという仮定より, 疎正則化は $d < N$ を前提としているが, 次元 d が高すぎる場合は行列 $R \in \mathbb{R}^{k \times d}$ を用いて低次元へ射影する.

$$q' = Rq \quad (5)$$

式 (5) を式 (4) に代入すると次式のようになる.

$$\min \|\alpha\|_1 \text{ subject to } q' = RS\alpha \quad (6)$$

圧縮センシングの数理によれば, 式 (6) の解の一意性が保証される次元は次のように与えられる.

$$k \geq C\mu(R, S)s \log d \quad (7)$$

C は正定数である. 次元 k が式 (7) を満たすとき解の一意性が保障される. $\mu(R, S)$ はコヒーレンスと呼ばれ, R と S の各列の最大相関を表す [13].

$$\mu(R, S) := \sqrt{d} \max_i |R_i^\top S_j| \quad (1 \leq i, j \leq d) \quad (8)$$

コヒーレンスが小さいほど低い次元で一意な解が得られる. R がランダム行列の場合, その要素がランダムであることから, いかなる行列 S に対してもコヒーレンスは小さい. R がランダム行列であり, $s \ll d$ であるときの s -sparse な解の一意性が保証される次元の下限 k_0 は次式を満たすことが知られている [14].

$$k \geq k_0 = 2s \log \left(\frac{d}{k_0} \right) \quad (9)$$

識別に用いる次元を選択する 1 つの指標として利用できる.

アルゴリズム 1: 疎表現に基づく識別法

入力: 質問データベクトル $q \in \mathbb{R}^d$,
 N 個の訓練データからなる C クラスの
 訓練データ行列 $S = [S_1 \dots S_C] \in \mathbb{R}^{d \times N}$,
 各訓練データ数のリスト \mathcal{N}

出力: 付与するラベル \mathcal{L}

1. q, S の次元 d が大きすぎる場合はランダム射影で次元削減する
 2. S の各列を 2 ノルムで正規化する
 3. S に対する q の疎表現を探す
 4. 付与するクラスのラベル $\mathcal{L} = \arg \min_k r_k(q; \alpha)$ を探す
-

3.1.3 アルゴリズム

ここで,

$$r_k(q; \alpha) = \|q - S_k \alpha_k\|_2 \quad (10)$$

とし, $\mathcal{N} = \{n_1, \dots, n_C\}$ は各カテゴリごとの訓練データ数を表す. \mathcal{N} に従い, α をカテゴリごとの係数にブロック化したものが α_k である.

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_C \end{pmatrix} \in \mathbb{R}^N \quad (11)$$

得られた線形結合係数を元に, 最も残差が小さい合成を行っているカテゴリを探すことで識別を行う.

3.2 クラス部分空間を考慮したスパース表現

3.2.1 部分空間の疎分解

識別において質問データは, いくつかのクラスに属していると仮定する. 質問データ集合が少数のクラスに属すデータから構成されている場合, これを簡潔に表現できる訓練データのクラスも少数であると仮定できる. このことから質問データをカテゴリの部分空間の和空間であると考え, 質問データが張る部分空間を少数のクラス部分空間に分解することで識別を達成できる. 本研究では, この分解を部分空間の疎分解と呼ぶ. 質問データを最も簡潔に表現している部分空間が, 質問データに割り当てるラベルを持ったクラスに相当する.

パターンのデータを d 次元列ベクトルで表し, k ($= 1, \dots, C$) 番目のクラスの訓練データを列に並べた行列を $S_k \in \mathbb{R}^{d \times n_k}$ とする. 訓練データが張る部分空間

$$S_k := \text{span } S_k \quad (12)$$

を第 k クラスの部分空間の近似と見なす. すべての訓練データが張る部分空間は, 各クラスの訓練データ行列を連結した $S = [S_1, \dots, S_C] \in \mathbb{R}^{d \times N}$, を用いて

$$S := \bigcup_{k=1}^C S_k = \text{span } S, \quad (13)$$

と表すことができる. 過完備な基底として \mathbb{R}^d を張るに十分な訓練データ集合が与えられていると仮定すると, 任意の質問データ $q \in \mathbb{R}^d$ は訓練データの線形結合で表現できる.

$$q = \sum_{k=1}^C S_k \alpha_k = S \alpha \quad (14)$$

質問データが複数のベクトルの集合として行列

$$Q := [q^{(1)}, \dots, q^{(n)}], \quad (15)$$

で与えられたとき,

$$Q = SA, \quad (16)$$

を解くことで, 質問データの集合をクラスの部分空間の和空間で表せる. ここで,

$$A := [\alpha^{(1)}, \dots, \alpha^{(n)}] \in \mathbb{R}^{N \times n}, \quad (17)$$

は未知の係数行列であり, 第 j 列

$$\alpha^{(j)} := \begin{pmatrix} \alpha_1^{(j)} \\ \vdots \\ \alpha_C^{(j)} \end{pmatrix} \in \mathbb{R}^N, \quad (18)$$

は j 番目の質問データを合成する訓練データの係数ベクトルの連結である. 行列 A は次のようにも表せる.

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_C \end{pmatrix} \quad (19)$$

ここで,

$$A_k := [\alpha_k^{(1)}, \dots, \alpha_k^{(n)}] \in \mathbb{R}^{n_k \times n}, \quad (20)$$

とした. しかし, 訓練データが十分にあるという仮定から, 式 (14), (16) は $d < N$ の不定方程式であり, 解は一意に定まらない. そのため, SRC と同様にスパース正則化を行う.

解である係数行列を式 (19) のように区分的に表現した場合, 質問データを表せるいくつかの区分係数行列 A_k のみが非ゼロ要素を持つ. 本稿ではこのことを区分疎性と呼ぶ. 区分疎性を利用するにあたり, $f_{\mathcal{N}}$ という写像を次式のように定義する.

$$f_{\mathcal{N}} : \begin{pmatrix} X_1 \\ \vdots \\ X_C \end{pmatrix} \rightarrow \begin{pmatrix} \|X_1\|_{\mathbb{F}} \\ \vdots \\ \|X_C\|_{\mathbb{F}} \end{pmatrix} \quad (21)$$

ここで \mathcal{N} は各区分のサイズのリストである. また, $\|\cdot\|_{\mathbb{F}}$ はフロベニウスノルムである. これらの表記を利用して区分疎性は次のように定量化できる.

$$\|X\|_{0, \mathcal{N}} := \|f_{\mathcal{N}}(X)\|_0 \quad (22)$$

アルゴリズム 2 疎部分空間法

入力: n 個の質問データからなる質問データ行列 $Q \in \mathbb{R}^{d \times n}$, N 個の訓練データからなる訓練データ行列 $S \in \mathbb{R}^{d \times N}$,各訓練データ数のリスト \mathcal{N} 出力: 付与するラベルの集合 \mathcal{L} 1. Q, S の d が大きすぎる場合は次元削減する2. S の各列を 2 ノルムで正規化する3. S に対して Q を疎分解する4. 付与するクラスのラベル $\mathcal{L} = \arg \min_k r_k(Q; A)$

あるいは

$$\mathcal{L} = \{\arg \min_k r_k^{(1)}(Q; A), \dots, \arg \min_k r_k^{(n)}(Q; A)\}$$

を探す

質問データの張る部分空間をなるべく少ないクラス部分空間に分解することは, 式 (22) で定量化した区分疎性の最小化問題に帰着する.

$$\min \|\alpha\|_{0, \mathcal{N}} \quad \text{subject to } q = S\alpha \quad (23)$$

$$\min \|A\|_{0, \mathcal{N}} \quad \text{subject to } Q = SA \quad (24)$$

Wright らは疎性を利用した顔認識 (SRC) において, ランダム射影 [15] などの線形変換を用いてパターンを表す次元削減しても認識率が損なわれないことを示した. 特に, ランダム射影はパターン認識において汎用的な次元削減手法であると考えられるため [6], パターンの次元が高すぎる場合, 本研究ではランダム射影を用いて次元削減を行う.

3.2.2 疎部分空間法

疎分解から得られた線形結合係数を用いて, 各クラスごとに質問データを合成する. 質問データがある 1 つのクラスに属すると考える場合, 質問データを簡潔に表現できるのはある 1 つのクラス部分空間である. 従って, 疎分解から得られた線形結合係数行列と各クラスの訓練データ積から合成した質問データと実際の質問データの残差を計算し, 最も良く質問データを表現するクラスを探す. 最も小さい残差を与えるクラスが, 質問データの属すべきクラスであり, 次式で求められる.

$$\arg \min_k r_k(Q; A) := \arg \min_k \|Q - S_k A_k\|_F \quad (25)$$

疎部分空間法のアルゴリズムはアルゴリズム 2 のようにまとめることができる. アルゴリズム 2 で最も重要かつ計算量が集中するのはステップ 3 の疎分解である. 次節で疎分解のアルゴリズム, 計算量等について述べる.

3.2.3 疎分解のアルゴリズム

ROMP (Regularized Orthogonal Matching Pursuit) [18][19] に基づき区分的に疎な解を得ることで部分空間を疎分解するアルゴリズムをアルゴリズム 3 に示す. 疎分解のアルゴリズムには貪欲法の 1 つである ROMP を採用した.

アルゴリズム 3 部分空間の疎分解

入力: 質問データ行列 $Q \in \mathbb{R}^{d \times n}$,訓練データ行列 $S \in \mathbb{R}^{d \times N}$,列ブロックのサイズのリスト \mathcal{N} 疎の度合い M_0 ;出力: 区分的に疎な行列 A 非零な区分行列のインデックス \mathcal{I} 1. リストを $\mathcal{I} = \emptyset$, 残差を $R := Q$ として与える

2. repeat

3. $U := S^T R$ 4. $\gamma := f_{\mathcal{N}}(U)$ 5. \mathcal{J} は γ の成分の中で上から順に大きな値 M_0 個のインデックスの集合, あるいはどの成分も小さい場合はすべての非零成分のインデックスの集合とする6. \mathcal{J} を γ の成分の降順に並び替える7. すべての $i < j \in \mathcal{J}$ に対して $\gamma_i \leq 2\gamma_j$ となる部分集合 $\mathcal{J}_0 \subset \mathcal{J}$ の中から, 最大エネルギー $\|\gamma|_{\mathcal{J}_0}\|_2^2 := \sum \gamma_k^2$ をとる \mathcal{J}_0 を探す8. $\mathcal{I} := \mathcal{I} \cup \mathcal{J}_0$ 9. for each j do10. $\alpha^{(j)} := \arg \min_{\alpha} \|q^{(j)} - \sum_{k \in \mathcal{I}} S_k \alpha_k\|$

11. end for

12. $R := Q - \sum_{k \in \mathcal{I}} S_k A_k$ 13. until $\|R\|_F = 0$ or $\text{card } \mathcal{I} \geq 2M_0$

質問データ行列 Q , 訓練データ行列 S , 訓練データの列ブロックのサイズのリスト \mathcal{N} , そして疎の度合いを表す M_0 を入力して与える. ROMP による計算の結果に出力されるものは区分的に疎な行列 A と非ゼロな区分行列のインデックス \mathcal{I} をである. ROMP の反復は大きく 3 つのステップ X, Y, Z に分けることができる. ステップ X の行列 R の各列は, 対応する列の質問データについて未だ疎分解されていない成分を表す残差である. ステップ Y では訓練データ行列 S と残差 R の相関 U を測り, \mathcal{N} に従って U のブロックごとのエネルギー γ を計算する. さらに γ の主要なエネルギーを構成する訓練データ行列のブロックのインデックスを \mathcal{J}_0 として定める. ステップ Z では, 選択したブロックのみを用いて各質問データを合成するための線形結合係数を求めている. ブロックに対応するクラス部分空間の和空間は L2 空間であるので, 係数は最小 2 乗法で計算できる.

アルゴリズム 3 においてもっとも計算量が多いのはステップ 10 の最小 2 乗の問題であり, $\mathcal{O}(nM_0^2 d)$ を必要とする. 最小 2 乗の問題は共役勾配法を用いることで $\mathcal{O}(nM_0 d)$ に計算コストを抑えることができる. アルゴリズム 3 の総計算コストは $\mathcal{O}(nM_0^2 Nd)$ または $\mathcal{O}(nM_0 Nd)$ である.

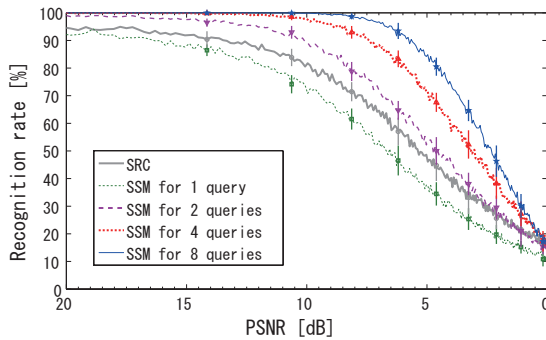


図 5 cropped version of extended Yale B に対する認識率．横軸は加えたガウシアンノイズの強さを表す．

4. 頑健性の検証

4.1 線形性に対する頑健性

図 4 に示した cropped version of extended Yale face database B [12] は照明変動のみを含む正面顔画像集合であり，クラス部分空間が線形部分空間を張るデータである．このデータを 32×32 ピクセル (=1024 次元) にランダム射影し，半分を訓練データ，残り半分の中からランダムに選択したうえでガウシアンノイズを加えたものを質問データとし，SRC および疎部分空間法（以下 SSM:=Sparse Subspace Method と表記）の認識率を測定した．

図 5 は SRC, SSM のガウシアンノイズに対する認識率を表す．横軸がノイズの強さを表し，右に行くほどノイズが強くなり認識率は下がっていく．SSM は，質問データを 1 つ使った認識では 15 [dB] 付近で認識率が 90% を下回り，10 [dB] 付近では 80% を下回る．一方，質問データ数 $n = 8$ では，9 [dB] 付近でも認識率が 100% であり，80% を下回るのは 4 [dB] 付近である．質問データを多く使用することでノイズに対して頑健な認識が行えていることが確認できた．SRC は SSM のように複数の質問データによる頑健化は行えないが，15 [dB] ほどまでは質問データ 1 つの場合の SSM よりも頑健な識別が行えている．

4.2 非線形性に対する識別の可否

まず，非線形クラス部分空間を持つデータに対する検証として，Web カメラと Haar-like な顔検出 [16] を用いて得た図 6 のように，顔を上下左右に細かい刻みで動かして撮影したデータに対する識別を考える．3 次元空間中にある顔を様々な方向から撮像した画像のクラス部分空間は線形部分空間ではない．しかし，ある位置を中心に小さな位置変動で撮影した画像の集合は部分的に線形部分空間を張っているとみなせる．また，疎部分空間法はノイズに頑健であることが確認できたため，合成しきれない残差をノイズとみなせば，多少の残差があったとしても最適なクラスを選択することができると予想され



図 6 視点位置の変化の刻みが小さなデータ集合

る．図 6 のように姿勢位置の変化を小さい画像 32 枚訓練データを cropped version of extended Yale B の訓練データに加え，Web カメラから得た同一人物の，訓練データと姿勢位置の変わらない常態での顔画像をテストデータとし，データの次元は 32×32 となるようランダム射影を行った上で同一人物であると識別できるかを実験した．

結果，SRC, SSM とともに同一人物であるという識別が行えた．クラス部分空間が非線形部分空間であったとしても，テストデータを簡潔に表現する組み合わせを見つければ識別が可能であることがわかった．

4.3 非線形性に対する頑健性

次に，図 7 に示した ETH80 [17] を用いて識別を行った．ETH80 は 80 クラスの 3 次元物体を視点を図 8 のように縦・横方向に変化させながら撮影した画像集合を含むデータベースであり，各クラス部分空間は非線形な部分空間を張っている．図 6 と異なり，ETH80 の視点位置の変化の刻みは大きく，テストデータを簡潔に表現できる組み合わせを探すことができるかは未知である． 15×15 ピクセル (=225 次元) にダウンサンプリングした画像データを用い，データの半分を訓練データとし，残り半分の中からランダムに質問データを選択し，SRC の認識率を測定した．結果，属すべきクラスに識別できたのは全試行回数 5000 回のうち，54.5% であった．

SSM の認識率を測定する際には，質問データデータをランダムに選択するか，あるいは視点位置の近い物を連番で選択するか，という 2 通りの質問データの選択方法を用いて認識率を測定した．視点位置の近いものだけを質問データとして与えた方が画像の見えの変化が少ない．そのため，質問データの張る部分空間が線形部分空間に近づき，疎分解も行いやすいのではないかと考えられるためである．また，非線形クラス部分空間に対する次元削減の影響も未知であるため，ランダム射影後の次元が 225, 512 次元である，2 通りの次元に対して測定を行った．

図 9 は横軸が使用した質問データ数，縦軸が認識率を表している．データの次元数・質問データの選択法によらず，認識率は 80% 以下であった．線形性をもつデータに対する認識率測定の際は質問データ数を増やすことによって識別の頑健が行えた．しかし，非線形性をもつクラスのデータの場合，質問データの数を増やしても広い

線形な部分空間を張ることができず、識別の頑健化が行えない。

SRC, SSM の認識率が低い原因について、疎分解の順序を元に説明する。疎分解のアルゴリズムでは、はじめにテストデータと訓練データの相関を計り、相関の大きなクラスとなるべく少ない訓練データで質問データを合成することを試みる。そして、合成したデータと質問データの残差が小さくなるまで処理を繰り返す。このとき、一度目の処理では質問データの属すべきクラスのデータを用いた合成の係数が得られているが、反復では属すべきクラスのデータで表現できない残差を表す組み合わせを探ることになる。点位置の変位の大きな画像間では見えの変化が大きいため 1 度目の処理後に残る残差が大きい。その大きな残差を表現するために、本来属すべきでないクラスの訓練データを用いた組み合わせを探すために識別性能が下がると考えられる。SSM の場合の図 5 と図 9 を比較すると、図 9 の質問データ数が 4 のときの認識率は図でノイズの強さが 5dB の場合に相当する。つまり、訓練データで表現しきれない残差が 5d のノイズに相当すると考えられる。

この結果から、クラスの訓練データにより小さい視点位置変化の画像を追加すれば、質問データを属すべき訓練データで表現しきれない残差は小さくなり、誤認識も避けられると予想される。

最後に、スパース表現を用いた識別法で物体認識を行う際、視点位置の変化刻みがどの程度で撮影された訓練データがあれば、正しく識別できるかどうかを検証した。図 10 のように、車のラジコンの正面画像を横方向に視点位置を回転させて撮影した画像集合を ETH80 の訓練データに追加した。撮影した角度の刻みは 4 度、計 90 枚の画像である。追加するクラスの訓練データは、視点位置の刻みを 8 度、12 度、16 度・・・と変えながら加えた。撮影した画像のうち、訓練データとして選択しなかったものの中からランダムに質問データを選択し、認識率を測定した。この際、ランダム射影で圧縮されたパターンの次元は 225 次元とした。

SRC の認識率を示したものが図 11 であり、縦軸が認識率、横軸は訓練データに加える際の視点位置変位の刻みである。車の正面画像の場合、撮影画像の刻みが 100 度以下ならばほぼ 100% 識別できていた。

SSM にも SRC と同様の方法で選択した訓練データを加え認識率を測定した。ただし、質問データの選択の仕方は視点の近いものを連番で与えるか、ランダムに選択するかで場合分けを行った。

図 12, 13 がそれぞれ、質問データを連番で与えたときの認識率とランダムに質問データを与えたときの認識率である。変位の刻みを 20 度以下で質問データを一度に 2 枚以上与えた場合は 100% の識別が行えることがわかった。

SRC, SSM どちらの場合も、視点位置の変化の刻みを

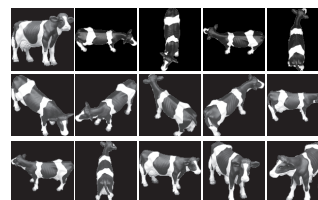


図 7 視線変化のみのデータ集合。ETH80 に含まれるクラスのうちの 1 つ。

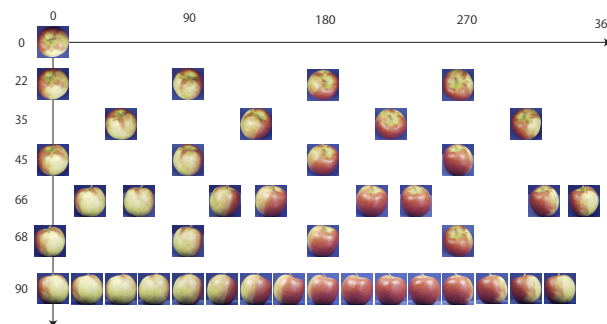


図 8 ETH80 の撮影視点

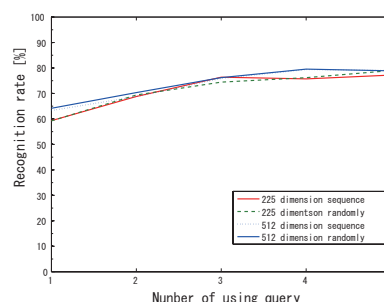


図 9 ETH80 に対する SSM の認識率



図 10 ETH80 に追加した物体の画像例。縦方向を 90° の位置に固定、横方向 360° を 4° 刻みに撮影しクリッピングした。

細かく、訓練データ数を増やせば認識率は高くなり、刻みを荒くしていくと認識率が低下していくのが見て取れる。SRC の場合は、今回は車の正面画像を横方向に回転させただけであるため、たまたま画像の相関が高くなる訓練データがある場合のみは、訓練データの視点位置の変化が大きく、データ数が少ない場合でも識別に成功する場合があったと考えられる。

5. おわりに

スパース表現を用いた識別の、クラス部分空間の非線形性に対する振る舞いを検証した。結果、非線形クラス部分空間をなすデータに対しても、対象の訓練データ次

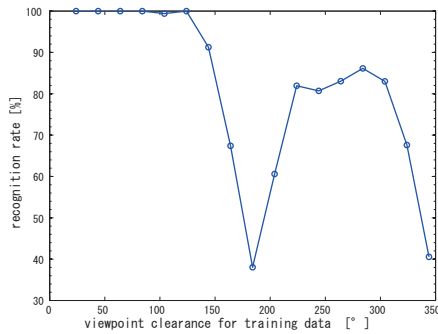


図 11 SRC の認識率 .

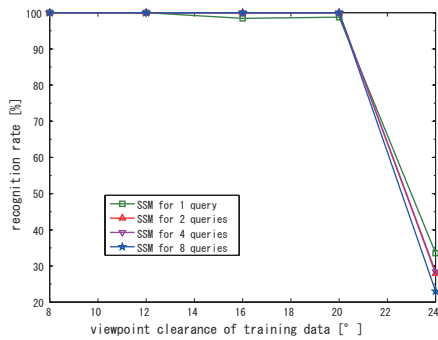


図 12 テストデータを連番で与えた場合のSSMの認識率 .

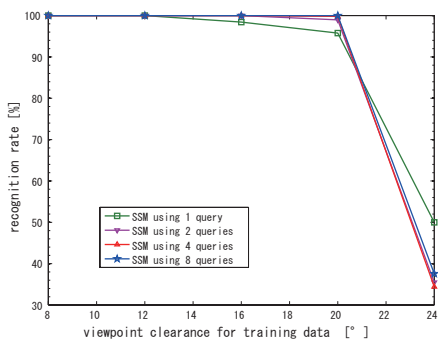


図 13 テストデータをばらばらに与えた場合のSSMの認識率

第でテストデータを簡潔に表現する線形空間を探ことができ、正しい識別が行えることを確認した。また、物体認識を行うためにはどの程度の視点位置変化で撮影された訓練データが必要であるかが判明した。しかし、非線形クラス部分空間を持つデータを識別するためには大量のデータが必要であり、かつ画像以外の文書・バイナリデータはどれだけ大量のデータを用意しても近似的に線形なデータとすることはできない。これらの問題を解決するため、今後はカーネル関数を用いたカーネルトリックを利用し無限に近い高次元で疎分解を行うことを考えていく。

謝 辞

本研究の一部は、科学研究費補助金（課題番号22700163）の補助を受けた。

文 献

[1] S. Watanabe and N. Pakvasa, "Subspace method in

pattern recognition" Proc. 1st Int. J. Conf on Pattern Recognition, pp. 2-32, Feb. 1973

[2] 飯島泰蔵, "パターン認識" 電気・電子工学大系 43, コロナ社, 1973.

[3] 福井 和広, 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. コンピュータビジョンとイメージメディア, vol. 46, num. 15, pp. 21-34 2005

[4] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance" Int.J.Computer Vision, vol. 14, pp. 5-24, 1995

[5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 210-227, February. 2009.

[6] T. Sakai, "Compressed Pattern Recognition A Framework of Pattern Measurement and Analysis Exploiting Sparsity" Technical Report of IEICE, vol. 109, num. 344, pp. 43-48, 2009

[7] 伊東 隼人, 酒井 智弥, 井宮 淳, "部分空間の疎分解に基づくパターン識別法" IPSJ SIG Technical Report, vol. 2010-CVIM-172, no. 33, pp. 1-7, 2010.

[8] 坂野鋭, 武川直樹, 中村太一, "核非線形相互部分空間法による物体認識" 信学論 (D-II), vol.J84-D-II, no. 8, pp. 1549-1556, Aug. 2001

[9] 福井和広, 山口修, "カーネル非線形制約相互部分空間法による物体認識" 信学論 (D-II), vol.J88-D-II(8), pp. 1349-1356, Aug. 2005

[10] P. N. Belhumeur and D. J.Kriegman, "What Is the Set of Images of an Object Under All Possible Illumination Conditions?" Int.J.Computer Vision, vol. 28, pp. 245-260,1998

[11] B. Scholkopf, A. Smola and K. R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem" Neural Computation, vol. 10,num. 5, pp. 1299-1319, 1998

[12] A. S. Georghiades and P. N. Belhumeur, "From few to many: illumination cone models for face recognition under variable lighting and pose" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23(6), pp. 643-660, 2001.

[13] E. J. Candes and M. B Wakin, "An Introduction To Compressive Sampling." IEEE M SP, vol. 25, num. 2, pp. 21-30, March. 2008

[14] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when then projection radically lowers dimension" J. Amer. Math. Soc, 2009

[15] S. S. Vempala, The Random Projection Method (DIMACS Series in Discrete Mathematics and Theoretical Computer Science), Amer Mathematical Society, 2004

[16] P. Viola and M. J. Jones, "Robust Real-Time Face Detection" Int.J.Computer Vision, Vol. 57, num. 2, pp. 137-154-154, 2004.

[17] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization" proc. of CVPR'03, pp. 409-15, 2003

[18] J. A. Tropp, "Greed is good: Algorithmic result for sparse approximation with application"IEEE Trans. Information Theory, vol. 50, num. 10, pp. 2231-2242, 2004

[19] D. Donoho, M. Elad and V. Temlyaykov, "Stable recovery of sparse overcomplete representations in the presence of noise" IEEE Trans. Information Theory, vol. 52, num. 1, pp. 6-18, 2006