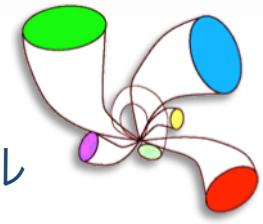
部分空間法研究会 2008

Subspace2008 in conjunction with MIRU2008

2008 年 7 月 28 日　軽井沢プリンスホテル



MIRU2008 サテライトワークショップ

主催　部分空間法研究会　実行委員会

共催　画像の認識・理解シンポジウム (MIRU2008)

序文

部分空間法研究会2008にようこそ． 当研究会は2006年にはMIRU2006の2007年にはACCV2007のサテライトとして開催され，2008年は再びMIRUのサテライトにもどって参りました．

我々は部分空間法こそ日本発の，しかも実用性が極めて高い技術であるという矜持からこの研究会を組織, 運営して参りました. その努力は今のところ報われています. 投稿される論文は毎回高いレベルを保っており, 聴衆も大学などのアカデミアから実用化研究に従事する企業の研究者まで幅広いスペクトラムを見せています．

このことは我々の活動がアカデミックなレベルの高さを追求するだけではなく広く世に役立つ技術を出していくことの助けになっていることの証左と捉え，今後も「役に立つ」研究会であることを期待されているものと考えています．

2009年度には京都でIEEE ICCV(International Conference of Computer Vision)が開催されます．当研究会は昨年のACCVに引き続きICCVのサテライトとして開催することを計画しています. これから来年度にかけては部分空間法が世界に出ていくための重要な年になります．

今回の研究会を契機に大きな仕事が現れることを実行委員一同祈っております．

ではお楽しみください．

2008 年 7 月 28 日

部分空間法研究会Subspace2008実行委員長 坂野鋭
実行委員 天野敏之，大町真一郎，佐藤敦
玉木徹，福井和広，堀田政二，牧淳人
顧問 前田賢一

プログラム

7 月 28 日（月曜日）

09：00 開会の挨拶

坂野 鋭 委員長

09:10-10:10 オーラルセッション 1： 座長 大町真一郎

1-1 市野将嗣(早大)，坂野鋭(NTT データ)，小松尚久(早大)，話者認識における核非線形相互部分空間法の適用と有効性に関する一考察

1-2 呉 嘉寧，福井 和広(筑波大)，局所部分空間集合を用いたアンサンブル識別に基づく３次元物体認識

1-3 Atsunori Kanemura, Shin-ichi Maeda, Shin Ishii (Kyoto Univ.), Subspace Selection for Resolution Synthesis

1-4 Yen-Wei Chen, Rui Xu (Ritsumeikan Univ.)， Generalized N-Dimensional Principal Component Analysis and Its Application to Medical Volumes

10：10-10：20：休憩

10:20-11:20 オーラルセッション 2： 座長 坂野鋭

2-1 Tomoya Sakai (Chiba Univ.), Dimension-Incremental Subspace Learning for High-Dimensional Data Classification

2-2 鷲沢嘉一（理研），抑制付きカーネル部分空間法

2-3 堀田政二（農工大），線型多様体間距離に基づくパターン識別と学習

2-4 山下幸彦（東工大），重み付き相関行列による局所部分空間法

11：20-11：30 休憩

11：30-12：00 オーラルセッション 3：玉木徹

3-1 藤木淳，赤穂昭太郎(産総研)，球面最小二乗法による球面上の曲線あてはめ

3-2 Sang-Woon Kim, Jian Gao (Myongji Univ.)， Computational Complexities of Dimensionality Reduction Schemes for Dissimilarity-Based Classification

12：00-13：00 チュートリアル講演：座長 福井和広

T-1 堀田政二（農工大），天野敏之（奈良先端），玉木徹（広島大），使ってみよう部分空間法 - 部分空間法体験実習 -

13：00-16：00

ポスターセッション，部分空間法相談デスク（実行委員会）

16：00-17：00 特別講演： 座長 前田賢一

S-1 小川英光（東京福祉大，東京工業大学名誉教授），What と How：部分空間法の歴史から学ぶもの

目次

話者認識における核非線形相互部分空間法の適用と有効性に関する一考察

市野　将嗣†　　坂野　鋭*††　　小松　尚久†

† 早稲田大学理工学術院　〒 169–8555 東京都新宿区大久保 3–4–1
†† （株）NTT データ　〒 135–8671 東京都江東区豊洲 3–3–9 豊洲センタービルアネックス
*現在，日本電信電話株式会社　コミュニケーション科学基礎研究所
E-mail: †{ichino,komatsu}@kom.comm.waseda.ac.jp, ††keen@cslab.kecl.ntt.co.jp

あらまし　本稿では核非線形相互部分空間法を用いた音声に基づく話者認識のアルゴリズムを提案し, 実験的に有効性を示す. 従来より音声による話者認識において，混合ガウス分布モデルが用いられている．音声による個人認証は，連続的な音声入力を仮定しているにもかかわらず，混合ガウス分布モデルはこれを単一の音声認識問題の連続したものとして扱っている．そこで我々は，音素の連続する軌跡の形状を比較するアプローチにより連続的な入力音声を積極的に利用し，さらに高性能な認識系を構成することを目指す．さらに音声データには非線形性の存在が予想されることを踏まえ，非線形アルゴリズムのひとつである，核非線形相互部分空間法を認識アルゴリズムとして適用することで，従来より用いられている混合ガウス分布モデルに比較して高い識別率が得られることを示す.
キーワード　音声，テキスト指定型話者認識，核非線形相互部分空間法

A study on application and effectiveness of the Kernel Mutual Subspace Method in speaker recognition

Masatsugu ICHINO†, Hitoshi SAKANO*††, and Naohisa KOMATSU†

† Faculty of Science and Engineering,Waseda University,Okubo 3–4–1, Shinjuku-ku, Tokyo,169–8555 Japan
†† NTT Data Corporation,Toyosu Center Bldg Annex, 3–3–9 Toyosu, Koto-ku, Tokyo 135–8671, Japan
*The author's affiliation is NTT Comunication Science Laoratories now
E-mail: †{ichino,komatsu}@kom.comm.waseda.ac.jp, ††keen@cslab.kecl.ntt.co.jp

Abstract　We propose a method of speaker recognition based on voice by using the kernel mutual subspace method. The Gaussian Mixture Model has already been developed as an algorithm for person authentication using voice. However, it is not sufficiently discussed the verification of the algorithm using the distribution of voice data. Voice data is obtained from an audio stream. However, GMM produces continuous static-voice matching problems when used for speaker recognition. Developing a way to use continuous observation might improve the accuracy of speaker recognition by reducing noise and enabling the extraction of invariant features from voice streams. We propose a method of speaker recognition based on voice by using the kernel mutual subspace method. We experimentally demonstrate the proposed method's effectiveness with simulation results and show that the method achieved higher accuracy than that of using the Gaussian Mixture Model.
Key words　voice, text indicated speaker recognition, kernel mutual subspace method

1. ま え が き

本稿では核非線形相互部分空間法を用いた音声に基づく話者認識のアルゴリズムを提案し, 実験的に有効性を示す.

音声による個人認証はマイクロフォンなどで実装できるため，携帯電話や PC に備え付けのマイクなど，身近なところで利用できる．また，発話は人間が普段から行っている自然な動作であるため，指紋などによる個人認証に比べ，ユーザの心理的な負担が少ない．そのため，音声による個人認証は高レベルのセキュリティを必要としない場面での簡便な個人認証方式として

注目されている．

音声による個人認証のためには特定のパスワードを用いることを前提としたテキスト固定型，何を話してもよいテキスト自由型，認証システムが話す言葉を指定するテキスト指定型の 3 つの方法が提案されている．

本研究では，個人認証方式としてテキスト指定型を採用する．認証を行う際，本人の特徴の現れやすいテキストを提示することによって認証精度の向上が期待できる．また，事前に録音をしておくことが困難であるため，なりすましの問題が軽減される．それに対してテキスト固定型，テキスト自由型は録音装置から出力された音声を用いるなりすましに弱いことが指摘されている．

従来より音声による個人認証の研究は記憶した音声情報との照合の問題として捉えられ，音素の類似性を比較するアプローチが行われてきた．

その中でも，隠れマルコフモデル（Hidden Markov Model, 以下 HMM）や動的計画法 (dynamic programing，以下 DP) による認識が試みられてきた [1] [2]．これらは特定の情報を学習・認識するものであり，テキスト固定型の音声認証アルゴリズムとの親和性が高い．しかし，HMM や DP などに基づく音声認識手法では，時系列情報を利用するため，明らかにテキスト固定型以外への応用が困難である．

テキスト指定型が可能な方法として，部分空間法 [3] [4] が認識アルゴリズムとして用いられた [5]．部分空間法は時系列情報を利用しないためテキスト指定型，テキスト自由型への適用が可能である．

著者らは既に音声において非線形性が存在することを確認している [6]．音声データの分布に非線形構造が認められる場合，部分空間法のような線形性を仮定したアルゴリズムでは分布を十分正確に近似することが出来ず，認識精度の低下を引き起こすことになる．

音声データの分布の非線形性を考慮した話者認識の研究として，混合ガウス分布モデル (Gaussian Mixture Model, 以下 GMM) が広く用いられている [7]．GMM は時系列情報を利用しないためテキスト自由型，テキスト指定型の音声認証アルゴリズムとも共用することが可能である．

GMM は話者認識のアルゴリズムとして広く用いられているにもかかわらず，音声データの分布によるアルゴリズムの妥当性の検証は十分には行われていない．音声データの分布の様子や特徴を考慮して話者認識のアルゴリズムを選択することにより，さらに高性能な認識系が構成できる可能性がある．

また，部分空間法を用いた話者認識を非線形に拡張した研究として，核非線形部分空間法 [8] [9] を適用した例がある [10]．そこでは，核非線形部分空間法を適用した際の識別性能は，GMM を用いた際の識別性能に匹敵することが示されている．

音声による個人認証は，連続的な音声入力を仮定している．しかし，GMM あるいは核非線形部分空間法では，これを単一の音声認識問題の連続したものとして扱っている．

本研究では，音素の連続する軌跡の形状を比較するアプローチにより連続的な入力音声を積極的に利用し，さらに高性能な認識系を構成することを目指す．

こうした話者認識装置において重要なのは音素の連続する軌跡がどのような形状を取るかである．

音声データの特徴抽出系として LPC ケプストラムやメルケプストラムが話者認識でよく用いられている．これらの特徴量は本来，音声認識のために開発された特徴抽出系であり，個人性よりも音声信号の特徴を残している可能性がある．各個人の同一音声が個人の別の音素より離れて分布している可能性がある．つまり，各個人の音素の連続する軌跡は非線形に絡み合って存在する可能性がある．軌跡の形状を正確にモデル化できない場合には，他のカテゴリに誤認識してしまう．

以上を踏まえて，本稿においては，非線形分布が存在し，連続的な入力が仮定できる場合に強力な識別アルゴリズムとして知られている核非線形相互部分空間法を用いることを提案する．以下，**2.** では，非線形, 連続分布条件下での話者認識アルゴリズムを提案する．また，**3.** では話者認識の特徴抽出系でよく利用される LPC ケプストラムとメルケプストラムに関して実験を行い提案手法の有効性を示す．さらに，**4.** ではまとめと今後の課題である．

2. 核非線形相互部分空間法

2.1 相互部分空間法

相互部分空間法（Mutual Subspace Method, 以下 MSM) [11] [12] は，認識対象の入力として複数データが利用できる場合に適用され，入力ベクトル集合も主成分分析を用いて部分空間で表現し，テンプレートの部分空間との間の角度を類似度として識別を行う．この角度に基づいた識別は正準角の概念を用いる．

2 つの部分空間 V, W のなす正準角 θ の余弦は，次のように計算される [11].

テンプレートの部分空間を V，入力された時系列データに対する部分空間を W とする．V の部分空間の次元を M，W の部分空間の次元を N とし，$\phi_m(m = 1, \cdots, M), \psi_i(i = 1, \cdots, N)$ を各部分空間 V, W における正規直交基底ベクトルとする．次いで，式 (1) で表される行列 X の固有値問題を解き，その最大固有値を第 1 正準角に対応する類似度 S_{mutual} とする．また，$N \leqq M (1 \leqq i, j \leqq N)$ とする．

ここで，

$$X = (x_{ij}) \tag{1}$$

$$x_{ij} = \sum_{m=1}^{M} (\psi_i \cdot \phi_m)(\phi_m \cdot \psi_j) \tag{2}$$

とおくと，

$$XU = \Lambda U \tag{3}$$

となる．ここで U は X の固有ベクトル，λ_{max} は固有値 Λ の最大値である．故に，第 1 正準角に対応する類似度 S_{mutual} は，

$$S_{mutual}(V, W) = \lambda_{max} = \cos^2 \theta \tag{4}$$

となる．

さらに，式 (1) の固有値問題の第 j 固有値を第 j 正準角に対応する類似度として扱うことができる [13].

MSM では，学習データと入力データの変動の少ない統計量同士を比較することになり，扱う対象に非線形性がない場合には強力な物体認識手法になる．

2.2 核非線形相互部分空間法

前節で導入した MSM は，扱う対象に非線形性がある場合には，十分な精度を達成できないという問題がある．このような非線形性の問題を解決するために，坂野らによって相互部分空間法と核非線形主成分分析 (Kernel Principal Component Analysis，以下 KPCA) [14] を融合した核非線形相互部分空間法 [15](Kernel Mutual Subspace Method, 以下 KMS) が提案されている．

Schölkopf により提案された強力な非線形 PCA である KPCA では，非線形な関数表現を考えるために関数空間(注1)への非線形写像

$$\Psi : \mathcal{R}^n \to \mathcal{F}, \vec{x} \to \vec{X} \tag{5}$$

を考える．ただし，$\mathcal{F}$ は極めて高次元もしくは無限次元の関数空間である．

次の $m \times m$ 行列

$$K_{ij} = (\Psi(\vec{x}_i) \cdot \Psi(\vec{x}_j)) \tag{6}$$

を定義し，

$$m\lambda\alpha = \alpha K \tag{7}$$

なる固有値問題を解き，特異値分解の公式に従い，

$$V = \frac{1}{\lambda}\alpha\Psi(\vec{x}) \tag{8}$$

の形で基底ベクトルを計算する．$\Psi(\cdot)$ の選択のためには，

$$k(\vec{x}, \vec{y}) = (\Psi(\vec{x}) \cdot \Psi(\vec{y})) \tag{9}$$

を満たすような写像を選ぶ．このような写像が選択できた場合には，$\Psi(\vec{x}) \cdot \Psi(\vec{y})$ は単に関数 $k(\vec{x}, \vec{y})$ を計算することに帰着される．これより，

$$V \cdot \Psi(\vec{x}) = \frac{1}{\lambda}\sum_{i=1}^{m} \alpha_i k(\vec{x}_i, \vec{x}) \tag{10}$$

のようになり，高次元の固有ベクトル V をあらわに求めないでも，その写像を計算できるようになる．

KMS では，辞書側と入力側の部分空間の角度を類似度として扱う．ここで，V を辞書側部分空間の基底ベクトル，W を入力側部分空間の基底ベクトルとする．V，W はそれぞれ辞書側，入力側データから KPCA によって計算された部分空間の基底ベクトルであり，ノルムが正規化されていると仮定すると，辞書登録された m 個の音声群と m' 個の入力された音声系列の類似度は，$(V \cdot W)$ の大きさで評価される．$(V \cdot W)$ は $\mathcal{F}$ 上の内積であるから，この値をあらわに有限の時間で計算することができない．しかし，V, W を

$$V = \sum_{i=1}^{m} \alpha_i \Psi(\vec{x}_i) \tag{11}$$

$$W = \sum_{j=1}^{m'} \alpha'_j \Psi(\vec{x}'_j) \tag{12}$$

と表現することにより，$(V \cdot W)$ の表式は

$$V \cdot W = \sum_{i=1}^{m} \alpha_i \Psi(\vec{x}_i) \cdot \sum_{j=1}^{m'} \alpha'_j \Psi(\vec{x}'_j) \tag{13}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m'} \alpha_i \alpha'_j (\Psi(\vec{x}_i) \cdot \Psi(\vec{x}'_j)) \tag{14}$$

となる．

ここで，式 (14) に式 (9) を代入すると，

$$V \cdot W = \sum_{i=1}^{m}\sum_{j=1}^{m'} \alpha_i \alpha'_j k(\vec{x}_i, \vec{x}'_j) \tag{15}$$

となり，有限の時間で類似性が評価できることがわかる．実際に類似度を評価するには式 (1)，(2)，(3) に式 (15) を代入すればよい．

3. 認 識 実 験

本節では, 音声による話者認識に KMS を用いた際の実験結果を示す. 最初に実験系の概要を示し, 認識実験結果を報告し, 結果に関する考察を述べる．

3.1 実験系の概要

3.1.1 特 徴 抽 出

本研究で用いる話者認識装置では音声のうち有音のみを対象とした．発話動作による連続的な音声情報を解析するためである．

特徴抽出系として話者認識でよく用いられている LPC ケプストラムとメルケプストラムの 2 つの特徴量に関して実験を行った. LPC ケプストラムは線形予測分析により求まるケプストラムであり，ピークを重視したスペクトル包絡になる．また，メルケプストラムは周波数軸を人間の聴覚の特性を考慮したメルスケールに変換してからケプストラム分析を行うことにより抽出される．

KMS は，データを部分空間で表現する際，サンプル数の 3 乗に比例する処理量を有する．また，認識時の類似度を計算する際には，すべての学習データ，認識対象データの間での核関数を計算することになり，処理量が大きい．本研究では，計算時間の発散(注2)を防ぐために, 学習データのサンプル数を削減することにより処理量削減を行った．具体的には，学習データに対して k-平均法を行った際に求められるクラスタ中心を学習

(注1)：関数空間のベクトルについても有限次元のベクトル空間の記号・用語 (「行列」，転置の記号) を用いることとする．

(注2)：音声個人認証の場合，数 kHz〜数十 kHz でサンプリングしたときの数秒〜数十秒の音声データに対して前処理を行い，認証する．そのため，音声データとして数千サンプルを扱うことになる．

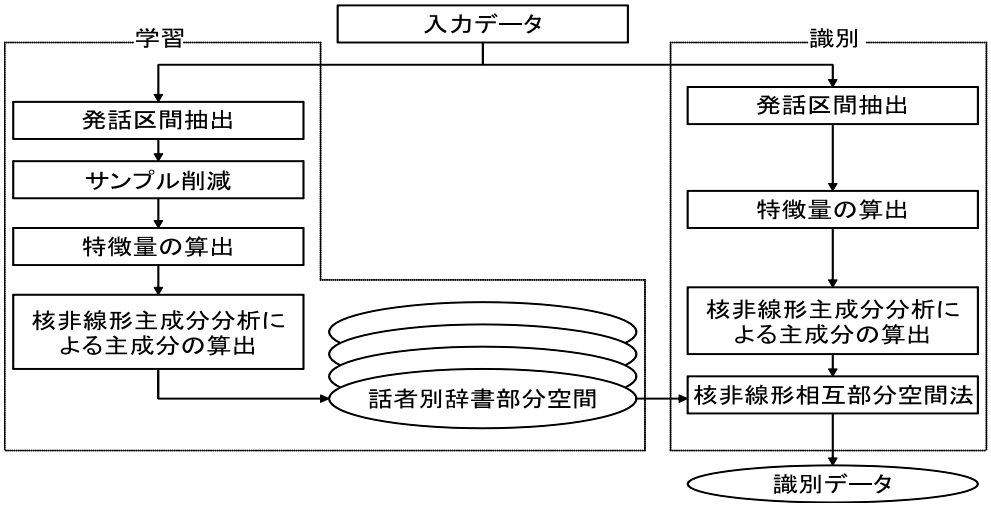


図 1　実験系の概要

データとして用いた [16].

この特徴抽出系と KMS を用いた個人認証方式の概要を図 1 に示す.

3.1.2　実験データ

実験データとして，大規模データベース XM2VTS [17] を使用した．XM2VTS は被験者を一ヶ月間隔で 4 回撮影した (1 回ごとに 1 時期とし，合計 4 時期あるとする) データベースである．発話した際の動画像 (顔画像と音声) が収録されている．

今回の実験データとして，40 人分 (男性 26 人，女性 14 人，4 時期分) を用いた．本研究では，あらかじめ「0,1,2,3,4,5,6,7,8,9」(英語) の 10 個の数字を発話した際の音声のデータをシステムに登録し，認証時に毎回数字の並びを変化させた数字列を提示し，その数字を連続で読み上げることにより認証するというテキスト指定型話者認識を想定している．そのため今回の実験では，学習データは「0,1,2,3,4,5,6,7,8,9」(英語) の 10 個の数字を連続で読み上げたデータを 1 つのデータとし，テストデータは「5,0,6,9,2,8,1,3,7,4」(英語) の 10 個の数字を連続で読み上げたデータを 1 つのデータとして評価した．そして，4 時期中の 2 時期分を学習データ，テストデータとして用いた．ただし，学習データとテストデータに関して，同一時期のデータを選んでいない．また，データの選び方に関してクロスバリデーションを行い，2 回分をあわせて実験結果としている．1 時期につき同じテキストを 2 回発話しているので 1 回の実験につき学習データは 1 人あたり 4 データ，テストデータは 1 人あたり 4 データを用いた．結局 2 回の実験でテストデータとして 40×4×2=320 データを用いた.

1 人あたり 4 データ (あわせて 16 秒前後) を用いて辞書部分空間を表現し，1 人あたり 1 データずつ (4 秒前後) を用いて入力部分空間を表現した.

発話区間抽出では，実験で使用するフレームを選択する．実験で使用するフレームとして，音声波形の有音に対応するフレームのみを手動で切り出した結果を用いた．その後，LPC ケプストラムやメルケプストラムを算出し，特徴ベクトルとして用いた.

実験諸元を表 1 に示す.

表 1　実 験 諸 元

サンプリング周波数	32[kHz]
フレーム長	32[ms]
フレーム周期	8[ms]
特徴抽出	LPC ケプストラム，メルケプストラム
特徴ベクトル次元数	24

3.2　実 験 結 果

音声に関して，KMS を適用することの有効性を確認するために，話者認識において広く用いられている GMM との比較を行った.

KMS については大幅な処理時間の削減と部分空間次元数によらず安定した識別率を得ることを考慮して，学習サンプルを 80%削減した．GMM についてはサンプル削減を行わない.

GMM は複数の入力を仮定した方法ではないため，複数回の認識処理の類似度の平均を類似度として用いた.

LPC ケプストラムを特徴量として用いる際には，KMS は第 1 正準角から第 9 正準角まで考慮した類似度の平均を類似度とした．メルケプストラムを特徴量として用いる際には，KMS は第 1 正準角から第 8 正準角まで考慮した類似度の平均を類似度とした.

核関数として，ガウス型動径基底関数

$$k(\vec{x},\vec{y}) = \exp\left(\frac{-||\vec{x}-\vec{y}||^2}{2\sigma^2}\right) \tag{16}$$

を用い，予備実験により適切と思われる σ を設定した.

LPC ケプストラムを特徴量として用いた際の識別結果を表 2，メルケプストラムを特徴量として用いた際の識別結果を表 3 に示す．LPC ケプストラム，メルケプストラムともに KMS が GMM より高い識別率を示すことがわかる.

図 2，3 に累積識別精度特性 (Cumulative Match Characteristic Curve) を示す．図において，縦軸が累積識別率 (Cumulative

Match Rate, 以下 CMR)[注3]，横軸が順位を表す．これらの図より LPC ケプストラム，メルケプストラムともに KMS は GMM と比較して高い CMR を示すことがわかる．また，KMS を用いて誤識別した場合においても識別候補の上位に本人が判定される．つまり，唇動作などの他のモダリティと統合した場合にも KMS は GMM より高精度化できる可能性があると考えられる．

また，図 4，5 に照合精度特性 (Receiver Operating Characteristic Curve, 以下 ROC 曲線) を示す．図において横軸が FRR(False Reject Rate)，縦軸が FAR(False Accept Rate) を表す．図 4，5 より LPC ケプストラム，メルケプストラムともに KMS が GMM と比較して高い認識性能を示すことが確認できた．

表 2　識別結果 (LPC ケプストラム)

識別手法	KMS	GMM
識別率（%）	90.3	80.0
σ	2.5	-
辞書次元数	13	-
入力次元数	10	-

表 3　識別結果 (メルケプストラム)

識別手法	KMS	GMM
識別率（%）	98.1	96.9
σ	2.0	-
辞書次元数	16	-
入力次元数	8	-

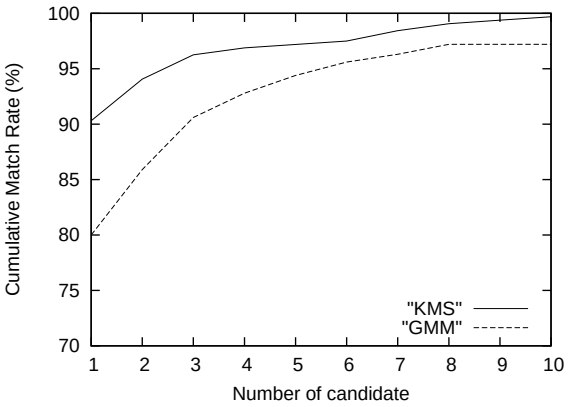


図 2　累積識別精度特性 (LPC ケプストラム)

3.3 考　　察

今回の実験において，KMS が GMM より高い識別性能を示した理由について考察する．表 2，3 に示したように LPC ケプストラムのほうがメルケプストラムより KMS の識別率と GMM の識別率の差が大きかったので，以下，LPC ケプストラムの場合に関して考察する．

KMS が GMM より高い識別性能を示した理由として次の 3 つを考えた．

- 雑音除去の効果
- 細かい特徴抽出・認識
- モデル同士を比較することにより認識

以下，これらについて説明する．

(注3)：識別アルゴリズム，識別装置あるいは個人識別システムが，同一の音声同士の識別判定で，与えられた順位以内の候補として選択する確率

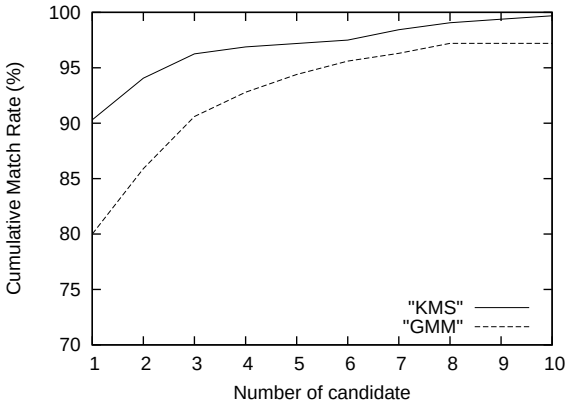


図 3　累積識別精度特性 (メルケプストラム)

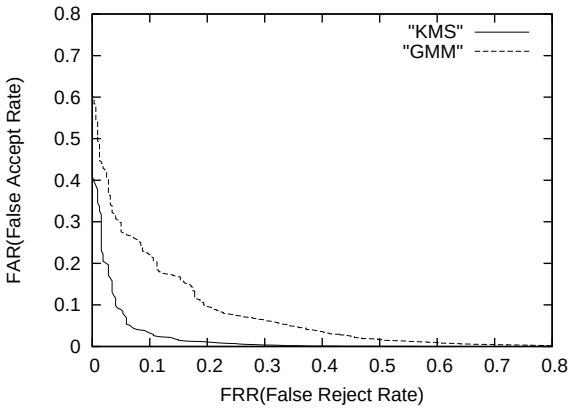


図 4　ROC 曲線 (LPC ケプストラム)

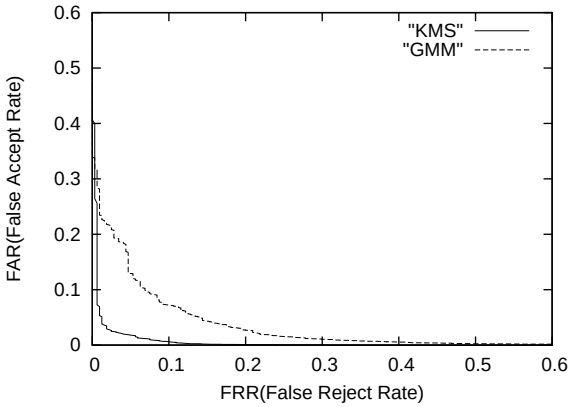


図 5　ROC 曲線 (メルケプストラム)

3.3.1 雑音除去の効果

今回の実験では，音声データの記述に LPC ケプストラムを使用した．LPC ケプストラムは本来, 音声認識のために開発された特徴抽出系でありピークを重視したスペクトル包絡を表すため，個人性よりも音声信号の特徴を残していると考えられる．つまり, 各個人の同一音声が個人の別の音素より離れて分布し

ていると考えられる．

さらに英語の音素は 44 個あるため，音声データとして取り得る状態が多く，複雑な分布になると考えられる．

つまり本人と他人の音声データは複雑に絡み合って存在していると考えられる．

それを踏まえ，次の仮説を立てた．

本人のサンプル (フレームに相当) であるにもかかわらず他人分布に近いサンプルが存在する可能性がある．このようなサンプルが多く存在する場合には 1 サンプルの入力を対象とした識別アルゴリズムでは識別精度が悪くなると考えられる．

GMM(K-平均法で考える(注4)) を用いる場合，クラスタ中心と 1 サンプルごとの距離の平均を類似度とする．他人分布に近いサンプルが多く存在する場合には本人のクラスタ中心とサンプルの距離が大きくなる．つまり距離の平均値が大きくなるため誤識別すると考えられる．ただし，他人分布に近いサンプルが多く存在していてもクラスタ中心と 1 サンプルごとの距離の平均 (以下，平均距離) が小さい値を示す場合には正しく識別できると考えられる．

KMS は学習，入力データの両方を KPCA を用いて主成分を計算し，その主成分を用いて類似度を求める方法である．つまり学習と入力の双方の変動の少ない部分同士を比較する手法である．そのため，他人分布に近いサンプルが多少存在しても影響をほとんど受けないと考えられる．本人分布に近いサンプルが多く存在していても本人分布との平均距離が他人分布との平均距離より大きい値を示す場合には本人テンプレートの分布と入力のデータの分布の形状が異なることが予想され誤識別すると考えられる．ただし，音声データの性質を考慮するとこのようなデータは多くないと考えられる．

以上より，GMM より KMS のほうが高い識別性能を示したと考える．このことを検証するために以下の実験を行った．

入力データのみに雑音を付加し，その影響を KMS と GMM の識別率を比較することにより調べた．音声データは複雑に絡み合って存在しているため，雑音を付加することによりさらに本人分布からサンプルを離し，他人分布に近づけることができる．そのとき，識別率の低下が少ないほうが雑音の変動の影響を受けずに主成分を計算できたと考えられる．

そのことを確認するために，電子協騒音データベース (人混み) [18] を使用して音声のテストデータに SN 比が 15dB,10dB,5dB になるように雑音付加の割合を変えて実験を行った．学習データは **3. 1. 2** と同じ条件のもの (雑音を付加しないデータ) である．

識別結果を表 4 に示す．この結果より雑音を強くすることに伴う識別率の低下が GMM に比べ KMS のほうが少ないことがわかる．つまり KMS のほうが GMM より変動を吸収していると考えられる．

3. 3. 2　細かい特徴抽出・認識

KPCA により求まる主成分に着目すると，低次元の主成分は分布の概形を表現し，次元数が上がるにつれて主成分は細かい情報を表す．

(注4)：要素となるガウス分布すべての分散共分散行列が等しく単位行列で，かつすべての混合重み等しい場合の EM アルゴリスムは，K-平均法とほぼ同じ振る舞いをする

表 4　雑音付加時の識別結果

	clean	15dB	10dB	5dB
KMS (%)	90.3	90.6	91.3	85.6
GMM (%)	80.0	79.7	77.2	73.1

一方，KMS の類似度に着目すると，学習部分空間と入力部分空間の間には，部分空間の複数の基底を用いると複数の正準角を定義することができる．

LPC ケプストラムは音韻性を残していると考えられるため各個人の音素の連続する軌跡の概形は似た形状になると考えられる．つまり個人を認識するためには軌跡の細かい情報も必要であると考えられる．この情報を KMS では主成分で表現しているため高い識別精度を得ることができたと考える．

このことを確認するために以下の調査を行った．

今回の実験における識別率と部分空間次元数の関係を図 6 に示す．図 6 より部分空間次元数が大きくなるにつれて識別率が高くなることが分かる．そして部分空間次元数が 6 のとき GMM より識別率が高くなり，部分空間次元数が 13 のとき識別率が最も高い．つまり分布の概形を表す主成分に加えてより細かな情報を記述した主成分が個人の識別に有効に作用していることが分かる．

使用する正準角の数と識別率の関係を表 5 に示す．使用する正準角を増やすことにより識別率が向上する．そして正準角を 7 個使うとき GMM の識別率より高くなり正準角を 9 個使うとき識別率が最も高い．つまりより細かな情報を表す正準角が識別に有効に作用していることが分かる．

KMS は **3. 3. 1** の雑音除去効果を備えながら分布の細かい部分も用いて比較することができるため高い識別性能を得ることができたと考えられる．

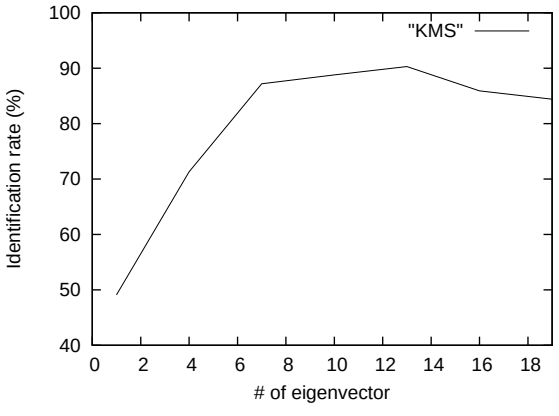


図 6　識別率と部分空間次元数の関係

3. 3. 3　モデル同士の比較

本人，他人のデータ分布が非線形に絡み合って存在している場合，音声データの分布の重なりが大きくなるため GMM のような 1 フレームのみを用いた識別は難しい．

それに対して KMS は学習，入力データを KPCA を用いて

表 5　使用する正準角の数と識別率の関係

使用する正準角の数	1	2	3	4	5	6	7	8	9	10
識別率 (Mean)(%)	35.6	51.3	64.1	70.3	75.3	79.1	84.7	88.4	90.3	82.8

主成分を計算し，モデル同士を比較する手法である．つまり，分布の重なりが大きい場合においても全体をみて一致，不一致を認証するため安定して認証が行えると考えられる．

学習，入力データを KPCA を用いて主成分を計算し，モデル同士を比較することの有効性を確認するために 1 サンプル対 1 サンプルの認識アルゴリズムである MPFC(Multiple Potential Function Classifier) [19] と 1 サンプル対 N サンプルの認識アルゴリズムである GMM と核非線形部分空間法 (Kernel based Nonlinear Subspace method，以下 KNS) との比較実験を行った．KMS と KNS の違いは入力データを KPCA を用いて主成分を計算して類似度の算出に利用しているかどうかである．MPFC，GMM，KNS と KMS の識別性能を比較することにより学習，入力データを KPCA を用いて主成分を計算し，分布同士を比較することの有効性を示すことができると考えられる．

実験データは **3. 2** と同じ条件のものを用いる．

核関数は式 (16) の動径基底関数を使用した．また，KNS に関して複数回の認識処理の類似度の平均を類似度として用いた．KMS に関して表 2 と同様に第 1 正準角から第 9 正準角までを考慮した類似度の平均を類似度とした．

識別結果を表 6 に示す．

表 6　識別結果 (KMS と MPFC，GMM，KNS の比較)

識別手法	KMS	GMM	KNS	MPFC
識別率（%）	90.3	80.0	77.5	32.8
σ	2.5	-	2.5	0.3
辞書次元数	13	-	16	-
入力次元数	10	-	-	-

この結果より 1 サンプル対 1 サンプルの認識 (MPFC) よりも 1 サンプル対 N サンプルの認識 (GMM,KNS) のほうが識別率が高いことがわかる．さらに 1 サンプル対 N サンプルの認識よりも N サンプル対 N サンプルの認識 (KMS) のほうが識別率が高く，モデル同士を比較することの有効性を示すことができたと考えられる．

さらに KMS と GMM に関して細かく調べるために本人データ，他人データそれぞれを K-平均法を用いてクラスタリングして求まるクラスタ中心からの距離分布と識別結果の関係を調べた．図 7 に示すように 1 サンプル (フレームに相当) ごとに本人データから求まるクラスタ中心との最小距離，他人データから求まるクラスタ中心との最小距離を比較した．

他人分布に近いサンプルが多少多く存在する場合でも，KPCA を用いて主成分を計算することにより変動の少ない成分 (主成分) を抽出し主成分同士を比較するため，KMS では正しく識別できると考えられる．

以下の 4 つの場合について調べた．

a)　KMS でも GMM でも正しく識別できている人物について (320 データ中 247 データ，いわゆる識別しやすいデータ)

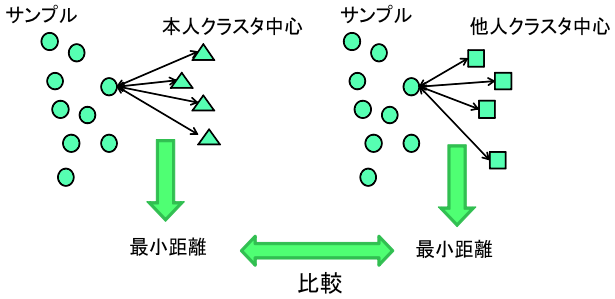


図 7　クラスタ中心からの距離分布

サンプル数 471(1 人分発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いサンプルは 367(471 サンプル数に対して 78%) であった．471 サンプル中，本人のクラスタ中心に近いサンプルが 104 サンプルあり (本人のクラスタ中心に近いサンプルが一番多い)，他人のクラスタ中心に近いサンプルに比べ多い結果を得た．平均距離が小さい傾向にあった．

b)　KMS でも GMM でも正しく識別できない人物について (320 データ中 22 データ，いわゆる識別しづらいデータ)

サンプル数 405(1 人分発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いサンプルは 389(405 サンプル数に対して 96%) であった．405 サンプル中，本人のクラスタ中心に近いサンプルが 16 サンプルであり (他人一人一人のクラスタ中心に近いサンプルのほうが多い)，他人のクラスタ中心に近いサンプルに比べ少ない結果を得た．平均距離が大きい傾向にあった．

c)　GMM では正しく識別できないが KMS では正しく識別できている人物について (320 データ中 42 データ，いわゆる a) の場合と b) の場合の中間に位置するデータ)

サンプル数 433(1 人分発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いサンプルは 383(433 サンプル数に対して 88%) であった．433 サンプル中，本人のクラスタ中心に近いサンプルが 50 サンプルあり，他人のクラスタ中心に近いサンプルと本人のクラスタ中心に近いサンプルの差は a) の場合に比べ小さい．この場合，本人のクラスタ中心に近いサンプルが一番多い場合もあれば 2 番目や 3 番目に多い場合もあった．本人分布との平均距離が他人分布との平均距離より小さな値を示す傾向にあった．

d)　GMM では正しく識別できるが KMS では正しく識別できない人物について (320 データ中 9 データ)

サンプル数 316(1 人分の発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いデータは 264(316 サンプル数に対して 83%) であった．基本的には a) の場合に分類されるはずである．

そこで 1 フレームごとの距離の変動を調べた．1 人分に関し

ての本人のクラスタ中心と入力サンプルとの距離の変動を図 8 に示す．図において，横軸がフレーム番号，縦軸が距離をあらわす．この結果より全体としては距離が小さいが，極端に距離の大きいフレームが存在していることがわかる．KMS では共分散行列を計算し固有ベクトルを求めるのでそのようなフレームがある場合，影響を及ぼす．そのため KMS では誤識別したと考えられる．それに対して GMM では平均距離を類似度とするため，そのようなフレームが存在しても平均距離が小さければ識別できると考えられる．

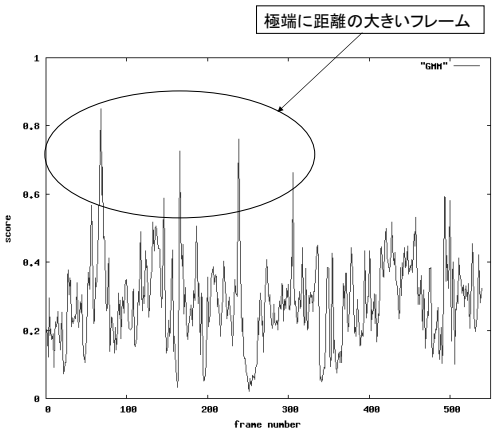


図 8　1 フレームごとの距離の変動

d) のデータよりも c) のデータが多いため GMM より KMS のほうが識別率が高かったと考えられる．

このように他人のクラスタ中心に近いサンプルが多少多い場合でも KMS は識別できていることがわかる．つまり GMM では学習データの分布と入力データの分布がぴったり張り付いていないと正しく認識できないが，KMS ではぴったり張り付いていない場合でも正しく認識できることがわかる．

KMS は **3. 3. 1** で示したように変動の少ない成分を **3. 3. 2** で示したように細かな部分も含めてモデルとして利用してモデル同士を比較することにより認識を行っているため KMS が高い認識性能を示したと考えられる．

以上より，今回の実験において，KMS が GMM より高い識別性能を示したと考えられる．

4. む す び

テキスト指定型話者認識に適用可能な音声による話者識別のアルゴリズムを提案した．

各個人の音素の連続する軌跡は非線形に絡み合って存在すると考えられるため，本研究では，音素の連続する軌跡の形状を比較するアプローチにより連続的な入力音声を積極的に利用することを考えた．具体的には認識アルゴリズムに KMS を用いることを提案し，識別実験を通して，提案手法が有効であることを示した．

今後は，より大規模なデータで再検証を行い，KMS を用いることの有効性を確認していく予定である．

さらに，より認証精度を高めるために本人と他人の部分空間の正準角を広げるもしくは本人同士の部分空間の正準角を狭めるようなアプローチを考えていきたい．

謝辞　本研究の一部は，日本学術振興会科学研究費補助金 (特別研究員奨励費) の補助による．

文　　献

[1] J.P. Campbell,“Speaker recognition: A tutorial,” Proc.IEEE, vol.85, no.9, pp.1437-1462, 1997.

[2] K.Yu, J.Mason and J.Oglesby,“Speaker recognition using hidden Markov models, dynamic time warping and vector quantisation,” Vision, Image and Signal Processing, IEE Proceedings- vol.142, issue 5, pp.313 - 318, October 1995.

[3] S.Watanabe and N.Pakvasa, “Subspace method of pattern recognition,”Proc. 1st Int.J.Conf. on Pattern Recognition, 1973.

[4] E. Oja,“Subspace Methods of Pattern Recognition,”Research Studies Press, 1983.

[5] J.B. Attili, M. Savic, and J.P. Campbell,“A TMs32020-based real time, text-independent, automatic speaker verification system,“ IEEE International Conference ICASSP’88, pp. 599-602, April 1988.

[6] 市野将嗣，高倉大樹，坂野　鋭，小松尚久，“母音音素分布の非線形性について，” 信学総大，D-14-14，March 2004.

[7] R.A. Reynolds and R.C. Rose,“Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models,” Speech and Audio Processing, IEEE Transactions on , Volume:3, Issue:1, pp.72-83, January 1995.

[8] 津田宏治，“ヒルベルト空間における部分空間，” 信学論 (D-II)，vol.J82-D-II，no.4，pp.592-599，April 1999.

[9] 前田英作，村瀬　洋，，“カーネル非線形部分空間法によるパターン認識．” 信学論 (D-II)，vol.J82-D-II，no.4，pp.600-612，April 1999.

[10] 浜崎　武，野田秀樹，河口英二，“ヒルベルト空間で部分空間法を用いた話者識別，” 信学技報，PRMU2000-127,2000 年.

[11] 前田賢一，渡辺貞一，“局所構造を導入したパターン・マッチング法，” 信学論（D），vol.J68-D，no.3，pp.345-352，March 1985.

[12] 山口　修，福井和広，前田賢一，“動画像を用いた顔認識システム，” 信学技報，PRMU97-50，June 1998.

[13] 福井和広，山口　修，鈴木　薫，前田賢一，“制約相互部分空間法を用いた環境変動にロバストな顔画像認識–照明変動の影響を抑える制約部分空間の学習–，” 信学論（ D-II），vol.J82-D-II，no.4，pp.613-620，April 1999.

[14] B.Schölkopf *et al.*,“Nonlinear component analysis as a Kernel eigenvalue problem,” Neural Computation, vol.10, pp.1299-1319, 1998

[15] 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，August 2001.

[16] 市野将嗣，坂野　鋭，小松尚久，“クラスタリングを用いた核非線形相互部分空間法の処理量削減手法，” 信学論 (D)，vol.J90-D，pp.2168-2181，August 2007.

[17] K.Messer, J.Matas, J. Kittler, J. Luettin and G. Maitre, “XM2VTSDB: The extended M2VTS database,” in Proc. of Int. Conf. on Audio and Video based Biometric Person Authentication, Washington, USA,1999.

[18] 板橋秀一，“騒音データベースと日本語共通音声データ DAT 版，” 音響誌, vol.47, no.12, pp.951-953, 1991.

[19] H. Sakano, T. Suenaga, “Classifiers under continuous observation,” In Ed. T. Caeli, et al., Structual, Syntactic, and Statistical Pattern Recognition 2002, Lecture Note on Computer Science 2396, pp.798-805, 2002, Springer-Verlag Berlin Heidelberg 2002 Proc. in IAPR Intl. Workshop. SPR 2002

局所部分空間集合を用いたアンサンブル識別に基づく3次元物体認識

呉　嘉寧† 　福井　和広†

† 筑波大学大学院
システム情報工学研究科コンピュータサイエンス専攻
〒 305-8577　茨城県つくば市天王台 1-1-1
E-mail: †lacarte@viplab.is.tsukuba.ac.jp, ††kfukui@cs.tsukuba.ac.jp

あらまし　3 次元物体認識に複数視点を導入することで識別性能を向上させることが可能である．一般に複数視点から得られた特徴ベクトルの分布が非線形性を持つため，カーネルトリックを用いた非線形識別手法がよく用いられる．この非線形識別手法は特徴ベクトルを高い次元の空間に写像し非線形性を弱めることで多くの応用で有効性が示されているが，その計算量は膨大である．そこで本論文ではカーネルを用いた手法と同等の性能を非線形写像なしに達成する手法を提案する．具体的にはまず非線形分布を複数の局所部分空間で近似する．次に局所部分空間の集合をクラスタリング条件と各部分空間の次元の様々な組み合わせに対して生成し，アンサンブル識別を行う．提案法の有効性は公開データセットによる識別実験によって検証する．

キーワード　アンサンブル識別, 相互部分空間法, 平均学習部分空間法, 物体認識

3D Object Classification Based on Ensemble Classification with Local Subspaces

Jianing WU† and Kazuhiro FUKUI†

† Department of Computer Science, Graduate School of Systems and Information Engineering,
University of Tsukuba
Tennodai 1-1-1, Tsukuba-shi, Ibaraki, 305-8577 Japan
E-mail: †lacarte@viplab.is.tsukuba.ac.jp, ††kfukui@cs.tsukuba.ac.jp

Abstract　Classification performance of 3D object classification can be improved by multiple view points. Kernel-based methods are often introduced to handle the nonlinearly distributed feature vectors obtained from multiple view, by transforming the distribution to a higher dimensional space. However, this nonlinear mapping makes their computation to be complex. We aim to construct a comparable method with the kernel-based methods without using nonlinear mapping. Firstly we approximate a distribution of feature vectors with multiple local subspaces. Secondly we consider these local subspaces as weak classifiers and apply ensemble classification algorithm. We will evaluate the proposed method by classification experiments using a public data set.

Key words　Ensemble Classification, Mutual Subspace Method, Averaged Learning Subspace Method, Object Classification

1. はじめに

画像パターンに基づくビューベース物体認識を行うに当たって，単一方向の画像だけではなく複数方向から見た画像を採用することは自然である．このような画像は複数視点カメラの採用や単一カメラで回転などの運動を行う物体を撮影することで得られる．対象画像を特徴抽出，あるいは単に画像を輝度値の 2 次元配列とみなし，その行を展開してベクトル化する．複数の入力画像から得られる特徴ベクトル集合を用いた識別手法として，部分空間法の拡張である相互部分空間法 [1] が注目されている．

部分空間法は学習において識別対象となる物体クラスごとに複数の特徴ベクトルを取得し，クラスごとに主成分分析 (KL 展開) を行い，各クラスの部分空間を生成する．識別では一本の入力特徴ベクトルと各クラスの部分空間の角度を算出する．入力ベクトルはもっとも角度の小さいクラスに識別される．相

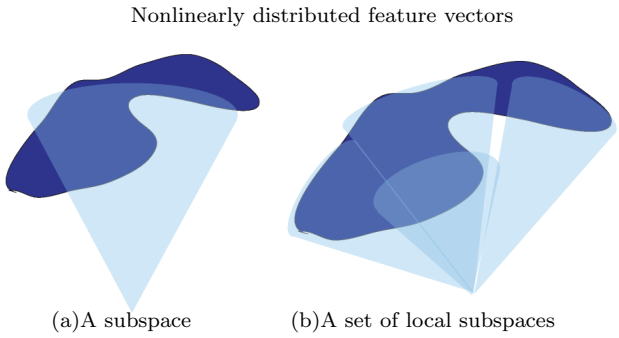


図 1　単一の線形部分空間による近似と複数の局所部分空間による近似の比較

Fig. 1　Comparison between approximation using a linear subspace and a set of local subspaces .

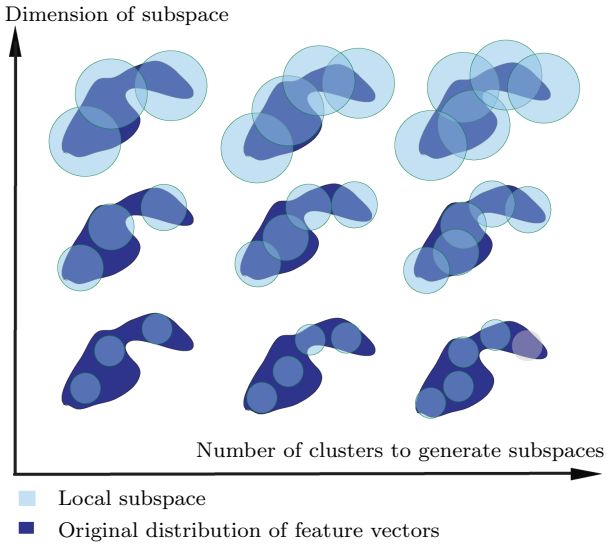


図 2　クラスタ数と局所部分空間の次元の関係

Fig. 2　The relation between the number of clusters and the dimension of the local subspaces.

互部分空間法 [1] では識別における入力特徴ベクトルを複数にし，これらを主成分分析して入力部分空間する．さらに学習した各クラスの部分空間と入力部分空間の類似度 (正準角，後述) を算出する．入力部分空間は類似度のもっとも高いクラスに識別される．相互部分空間法は複数視点の入力に基づく識別であり，3 次元物体の識別に有効な枠組みである．

しかし，複数視点画像では物体の見え方が大きく変化するため，得られる特徴ベクトルのパターン分布は複雑で非線形的な広がりを持つ可能性が高い．特徴ベクトルのパターン分布を主成分分析によって単一の線形部分空間で近似する相互部分空間法では図 1 の (a) に示すように表現できず，識別性能が低下する．

非線形的なパターン分布の識別において有効な手法として，核非線形相互部分空間法 [2] が知られている．この手法はパターン分布を非線形写像を用いて高い次元の特徴空間に写像し，写像先の特徴空間において線形分離可能なパターンに変換して識別するものであり，唇の認識などで高い性能を示している [2]．しかし，学習データ数 n の増加に伴い n^2 の計算量が生じ，実時間システムへの応用が難しい．

本論文では 3 次元物体の識別に伴う非線形的なパターン分布に有効で，かつ非線形手法より少ない計算量で実現できる識別手法を提案する．提案法の主なアイディアは以下の二つである．

まず，単一の線形部分空間での近似が困難なパターン分布に対して，図 1 の (b) に示すようにクラスタリングでこれを分割し，それぞれのクラスタに対して主成分分析を行って部分空間近似を行い，パターン分布を線形部分空間 (以後局所部分空間と呼ぶ) の集合で近似する．入力部分空間とあらかじめ学習した各クラスの局所部分空間集合の正準角に基づく類似度を計ることで識別を行う．この際，分割数や次元などのパラメータを最適に定めることが難しいため，図 2 に示すように分割数と局所部分空間の次元を変えながら，それぞれの場合を弱識別器としてアンサンブル学習を行うことでこのパラメータ選定の難しさに対応する．

また，前述の弱識別器を構成する過程ではクラス内の分布を再現できるよう構成されているが，クラス間の分離について考慮したとは言えない．そこで，学習部分空間法 [3] のフレームワークに基づき，学習フェーズにおいて各局所部分空間にクラス間分離を考慮した補正を適用する．

以下本論文ではまず第 2 章で従来手法について説明し，次に第 3 章で提案手法のアルゴリズムや識別の流れを説明する．第 4 章では公開データベースを用いた従来法，提案手法および非線形手法の比較実験を行い，本手法の有効性を示す．第 5 章でまとめと今後の課題について述べる．

2. 従 来 手 法

2.1　部分空間の正準角に基づく類似度

本研究の基礎となる相互部分空間法 [1] では部分空間同士の類似度として，2 つの部分空間の成す角度である正準角を用いる．N 次元の部分空間 P と M 次元の部分空間 Q の間には N 個の正準角が定義できる ($N \leqq M$)．P と Q が完全に一致している場合はすべての正準角が 0 度であり，完全に直交している場合はすべての正準角が 90 度となる．本論文では最小正準角を類似度計算に用いる．P と Q の最小正準角を θ として，その類似度 $Ang(P,Q) = \cos^2\theta$ は式 (1) で定義する．

$$\cos^2\theta = \max_{\substack{u\in P, v\in Q \\ \|u\|\neq 0, \|v\|\neq 0}} \frac{|(u,v)|^2}{\|u\|^2\|v\|^2} \tag{1}$$

$\cos^2\theta$ は以下の行列 X の最大固有値となる．

$$\mathrm{X} = (x_{mn}) \quad (m,n = 1...M) \tag{2}$$

$$x_{mn} = \sum_{l=1}^{N} (\psi_m, \phi_l)(\phi_l, \psi n) \tag{3}$$

ここで ψ_m, ϕ_l は部分空間 P と Q の 第 m, l 基底ベクトル，(ψ_m, ϕ_l) は ψ_m と ϕ_l の内積を表す．

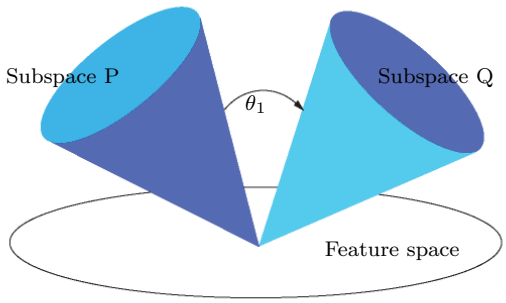


図 3　相互部分空間法の概念図
Fig. 3　Concept of mutual subspace method .

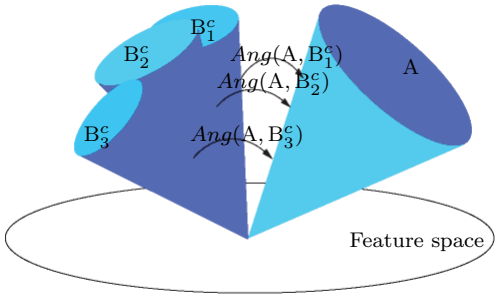


図 4　提案法の模式図
Fig. 4　Concept of the proposed method .

2.2　相互部分空間法の非線形拡張

核非線形相互部分空間法 [2] は相互部分空間法の非線形拡張である．パターン分布を高次元の特徴空間へ写像し，特徴空間における正準角に基づいて識別を行う．非線形写像 $\Psi : R^N \to F$ を定義すると，特徴ベクトル u を特徴空間 F に写像した $\Psi(u)$ が得られる．F は非常に高次元もしくは無限次元であるため，直接 $\Psi(u)$ と $\Psi(v)$ の内積を求めることは困難であるが，カーネルトリックを用いると F における内積が求まる．これにより，特徴空間 F における部分空間はカーネル PCA [2] を用いて求まる．クラス P' と Q' の学習ベクトルを u と v とすると，P' と Q' にカーネル PCA を適用して得られる基底 V と W の内積は式 (4) で求められる．

$$(V, W) = \sum_{i=1}^{M} \sum_{j=1}^{M'} a_i a_j k(u_i, v_j) \quad (u_i \in P', v_j \in Q') \qquad (4)$$

a_i，a_j は係数であり，$k(\cdot, \cdot)$ はカーネル関数である．式 (4) を式 (3) に代入すると F における正準角が計算できる．

3.　提 案 手 法

3.1　局所部分空間集合によるパターン分布の近似

相互部分空間法では特徴ベクトルのパターン分布を単一の線形部分空間で近似したが，第 1 章で述べたように複数視点から得られるパターン分布は非線形的である可能性が高く，単一の線形部分空間では精度の高い近似ができない．そこで提案法では対象パターン分布をユークリッド距離に基づいて複数のクラスタに分割し，各クラスタをそれぞれ線形部分空間で近似する．

具体的にはまずパターン分布に対し k-means 法でクラスタリングを行う．そして得られた各クラスタに対して PCA を適用して各局所部分空間の基底ベクトルを求める．

3.2　局所部分空間集合による類似度の定義

入力パターン分布を近似した部分空間 A とクラス c のパターン分布を近似した局所部分空間の集合 B^c，B^c に含まれる k 個の局所部分空間を $\mathrm{B}_i^c (i = 1, \sim, k)$ とすると，A と B_i^c の類似度 $Ang(\mathrm{A}, \mathrm{B}_i^c)$ は式 (1) から求まる．

この部分的な類似度 $Ang(\mathrm{A}, \mathrm{B}_i^c)$ を用いて，入力部分空間 A とクラス c の局所部分空間集合との類似度 $Sim(\mathrm{A}, \mathrm{B}^c)$ は式 (5) のように定義できる．

$$Sim(\mathrm{A}, \mathrm{B}^c) = \frac{1}{k} \sum_{i=1}^{k} Ang(\mathrm{A}, \mathrm{B}_i^c) \qquad (5)$$

入力パターン分布 A と全クラスの局所部分空間集合の類似度を算出し，最終的に入力パターン分布 A は最も高い類似度を示したクラスに識別される．

3.3　アンサンブル学習

パターン分布を分割する分割数 k と各局所部分空間の次元，クラスタリングの初期配置などのパラメータを事前に最適決定することが困難である．そこで本論文では図 2 に示すように分割数 k と局所部分空間の次元 j を変えながら，それぞれの場合を弱識別器としてとらえて，これらを統合するアンサンブル学習を用いることで性能向上を図る．

分割数 k と局所部分空間の次元 j を変えながら類似度 $Sim(\mathrm{A}, \mathrm{B}^c)$ を算出し，それらの総和を次式のように定義し，最終的な類似度とする．

$$Sim_{total}(\mathrm{A}, \mathrm{B}^c) = \sum_{j} \sum_{k} Sim(\mathrm{A}, \mathrm{B}^c) \qquad (6)$$

3.4　局所部分空間の補正

局所部分空間は該当クラスのパターンを最適に表現できても，クラス間の関係は考慮されていない．そこで，各局所部分空間にクラス間分離を考慮した変換を行うことで性能向上を試みる．クラス間分離を逐次的に補正する手法として，平均学習部分空間法 [3] が知られており，文字認識などにおいて有効性が確かめられている．

平均学習部分空間法では識別に先立ち，学習に使用した特徴ベクトルで識別を行い，その識別結果に基づいて各クラス部分空間の関係を補正する．補正は次のような規則で行う．各部分空間について誤識別を起こした場合を (1) ベクトルと部分空間が同クラスであるにも関わらず識別結果が他クラスとなった場

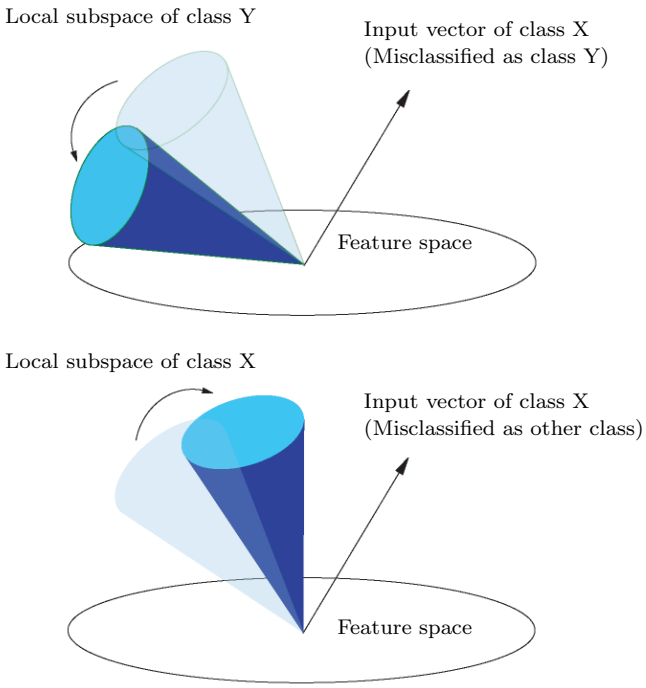


図 5　局所部分空間の補正.
Fig. 5　Modification of local subspace.

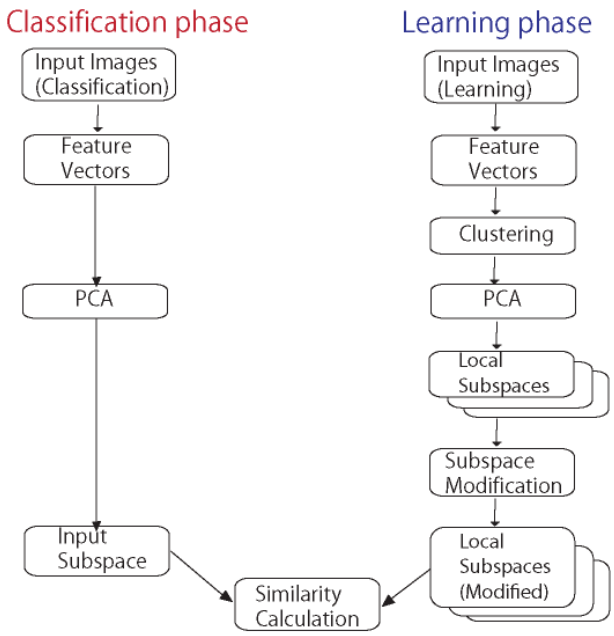


図 6　識別の流れ
Fig. 6　Flow chart of classification process.

合，(2) 逆にベクトルと部分空間が異なるクラスであるにも関わらず識別結果が自クラスとなった場合に分ける．該当する特徴ベクトルについて，どの場合で誤識別したかをマーキングする．すべてのベクトルについて識別が終了した後，各部分空間について，その部分空間が誤識別した特徴ベクトルのマーキングに応じて，図 5 に示すように (1) の場合は該当部分空間をベクトルに近づけるように回転し，(2) の場合は該当部分空間をベクトルから離すように回転する．提案手法では平均学習部分空間法と同様に誤識別を起こした局所部分空間を補正する．

テスト入力ベクトル x と局所部分空間 B_i^c の部分空間法による識別が誤ったとき B_i^c を補正する．$S_{\mathrm{B}_i^c}(k)$ は局所部分空間 B_i^c の共分散行列であり，k 回目の補正は下式のように行われる．

$$S_{\mathrm{B}_i^c}(k+1) = S_{\mathrm{B}_i^c}(k) + \alpha \sum_{x_l \in C_{\mathrm{B}_i^c}} x_l x_l^T - \beta \sum_{x_l \in D_{\mathrm{B}_i^c}} x_l x_l^T \quad (7)$$

$S_{\mathrm{B}_i^c}(0)$ は補正前の B_i^c の共分散行列である．ベクトルの集合 $C_{\mathrm{B}_i^c}$ にはクラス c であるにも関わらず他のクラスに識別されたベクトルが含まれ，ベクトルの集合 $D_{\mathrm{B}_i^c}$ には他クラスであるにも関わらずクラス c に識別されたベクトルが含まれる．また，α と β は補正の度合いを調整するパラメータであり，予備実験に定める．

3.5　識別の流れ

提案手法は図 6 に示すように大きく学習フェーズと識別フェーズに分かれる．

まず入力画像をラスター操作してそのまま特徴ベクトルと見なす．この過程は両フェーズに共通である．学習フェーズでは分割数を多段階に変化させて，入力される各クラスの特徴ベクトルを k-means 法によりクラスタリングする．各クラスタに PCA を適用し，局所部分空間で近似する．次に学習に使用した特徴ベクトルをテスト入力として各局所部分空間と部分空間法を行い，誤った局所部分空間を補正する．

識別フェーズでは入力される特徴ベクトルを主成分分析して部分空間とし，次元を変えながら各クラスの局所部分空間集合と式 (7) で定義した類似度を算出し，識別結果とする．

4.　識 別 実 験

4.1　比 較 実 験

従来手法である MSM，核非線形相互部分空間法 (KMSM) および提案手法の性能と計算量を評価する．さらに，提案手法で採用したアンサンブル学習および局所部分空間補正の有効性を確かめる．評価には 41 視点から撮影した物体からなる公開データセット ETH 80 Image Set [11] を用いる．このデータセットには図 7 に示すように，8 クラス，各クラスについて 10 個体の物体の背景を除去した画像が含まれる．各物体については 41 視点から撮影された画像があり，図 8 にその一例を示す．画像はグレースケールの 320×320 ピクセルであるが，評価実験ではこれを 16×16 ピクセルに縮小し，ラスター操作して 256 次元の特徴ベクトルとする．

評価実験を次のように行った．各クラスからランダムに 5 物体を選び，そのうち学習データとして 4 物体の画像を用いて，テスト用として残り 1 物体の画像を用いる．テスト用とした物体の画像から連続した 10 フレーム (視点) を入力とし，1 フ

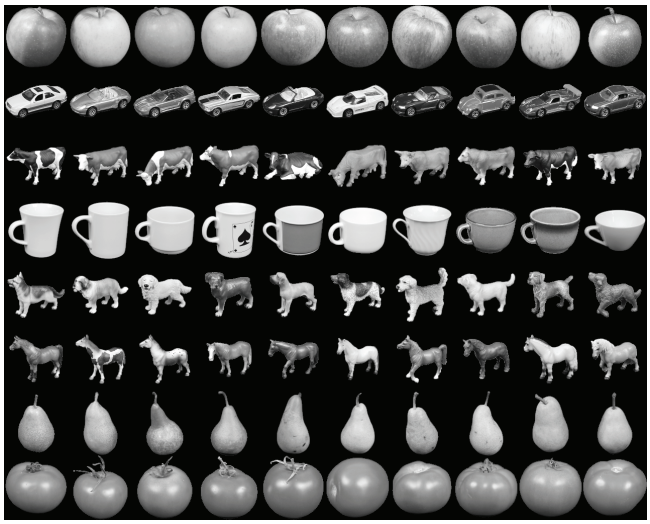


図 7　The ETH-80 Image Set.
Fig. 7　The ETH-80 Image Set.

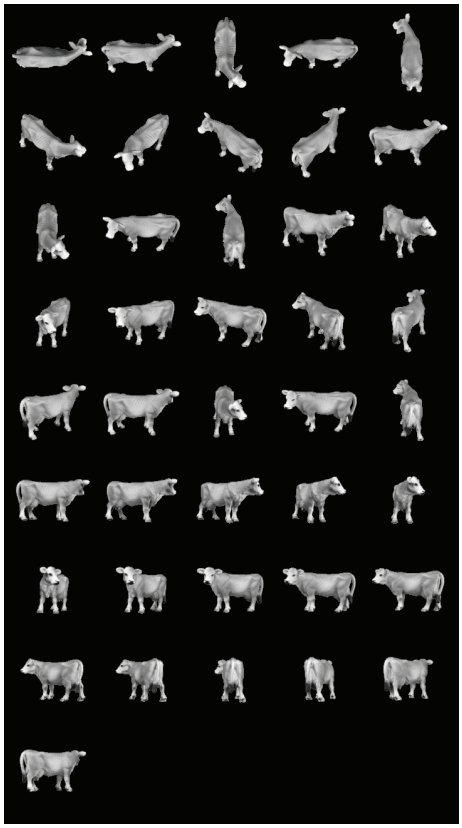


図 8　データセットに含まれる画像のサンプル．対象物体を 41 視点から撮影した画像である．
Fig. 8　Sample of images contained in the data set. These are images from 41 viewpoints of an object contained in the data set.

レームずつずらしながら 41 回の識別実験を行う．8 クラス ×41 回 = 328 回の実験を行った後，学習とテスト用のデータを入れ替えながら識別実験を 328×5 = 1640 回繰り返す．

提案法の分割数は 2 から 5 とし，局所部分空間の補正に使用するパラメータ α と β はそれぞれ予備実験から 3 と 1 とした．MSM は入力，辞書ともに予備実験より 3 次元に固定する．提

表 1　各手法の識別結果
Table 1　Result of each method.

手法	識別率 (%)	分離度	EER(%)
MSM	69.5	0.34	20
KMSM	87.2	0.41	15
提案法 補正前	86.5	0.44	14
提案法 補正後	91.0	0.51	10

表 2　各手法の 1 入力 (10 フレーム) あたりの計算時間
Table 2　Calculation time per input (10 images) of each method.

手法	計算時間 (秒)
MSM	0.1
提案法	0.4
KMSM	3.1

表 3　各弱識別器と提案法の性能比較
Table 3　Performance comparison of weak classifiers and the proposed method.

分割数 k	次元	識別率 (%)	分離度	EER(%)
2	dim 1	69.5	0.37	15
2	dim 2	75.6	0.39	12
2	dim 3	70.6	0.43	14
3	dim 1	68.3	0.37	18
3	dim 2	72.1	0.37	15
3	dim 3	67.5	0.41	14
4	dim 1	74.6	0.40	15
4	dim 2	73.1	0.39	15
4	dim 3	76.2	0.43	13
5	dim 1	72.1	0.39	15
5	dim 2	74.6	0.40	13
5	dim 3	70.8	0.40	14
2 ∼ 5	dim1 ∼ dim3	86.5	0.44	14

案法は入力と辞書の次元は寄与率によって変動する．アンサンブル学習に使用した組み合わせは累積寄与率 98% を達成する次元を dim1，dim1 より 1 次元減らした次元を dim2，dim2 より 1 次元減らした次元を dim3 とした．

MSM，提案手法，KMSM の識別結果は表 1 に示すとおりとなった．提案法 補正後の実験結果はすべての弱識別器を補正した場合のものである．分離度は 1 に規化されており，値が高いほど高い識別性能を示す．EER(Equal Error Rate) は FAR(False Accept Rate) 曲線と FRR(False Reject Rate) 曲線の交点であり，低いほど高い性能を示す指標である．

MSM，提案手法，KMSM の識別フェーズにおける計算時間が表 2 である．

アンサンブル学習の効果を確かめるため表 3 に補正前の分割数 k や次元を固定した各弱識別器の識別結果とそれらをアンサンブル学習した補正前の提案法の識別結果を分割数 2∼5，次元 dim1∼3 として示した．

4.2 考　　察

表 1 に見られるように，提案手法は従来の MSM より高い識別性能を示しており，17% の識別率向上を示している．非線形

手法である KMSM と同等の性能が得られる．また，提案法の局所部分空間補正の適用前と適用後の結果を比較すると，適用後がすべての指標において高い性能を示しており，局所部分空間の補正は有効であるといえる．

表 2 では MSM，提案法と KMSM の計算量を比較している．提案手法は MSM より計算量が増えているものの，非線形手法である KMSM の 13% 程度で識別を行うことができる．また，学習データの増加に対しては提案法と MSM がともにオーダー n の計算量増加であり，KMSM はオーダー n^2 である．したがって学習データの増加に伴って提案法は KMSM と比べて計算量の観点でさらに有利になると考えられる．以上の結果によって，提案法が非線形手法より少ない計算量で同程度の識別性能を示していると言える．

アンサンブル学習に採用した各弱識別器の識別結果が表 3 である．いずれの場合もアンサンブル学習を行った場合より低い性能であり，アンサンブル学習が識別性能の向上に寄与していることがわかる．

5. むすび

本論文では局所部分空間集合によって非線形パターン分布を近似し，アンサンブル学習を適用した 3 次元物体認識手法を提案した．各クラスのパターン分布を分割して主成分分析を適用し，それぞれ局所部分空間とする．さらに様々な条件で生成した局所部分空間集合を用いた各識別器を弱識別器と捉えアンサンブル学習を用いることで識別性能を向上させる．比較実験を通して提案法が従来手法より性能向上し，非線形手法と同程度の識別性能がより少ない計算量で得られた．また，局所部分空間の補正によって識別性能が向上することを確かめた．

今後の課題として，識別対象に応じた特徴抽出を採用し，提案手法と組み合わせることで更に識別性能の向上を図ることが挙げられる．

文　献

[1] O. Yamaguchi, K. Fukui and K. Maeda: Face recognition using temporal image sequence, Proc. IEEE Third International Conference on Automatic Face and Gesture Recognition, pp.318-323, 1998.

[2] H. Sakano, N. Mukawa: Kernel mutual subspace method for robust facial image recognition, Proc. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, Vol.1, pp.245-248, 2000.

[3] J. Laaksonen, E. Oja: Subspace dimension selection and averaged learning subspace method in handwritten digit classification, Proceedings of the 1996 International Conference on Artificial Neural Networks, p.227-232, 1996.

[4] L. Wolf, A. Shashua: Learning over sets using kernel principal angles, Journal of Machine Learning Research, 4(10), pp.913-931, 2003.

[5] Y. Freund: Boosting a weak learning algorithm by majority, Proc. Third Annual Workshop on Computational Learning Theory, 1990.

[6] Y. Freund, R. E. Schapire: A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, No.55, 1997.

[7] 前田賢一, 渡辺貞一: ”局所的構造を導入したパターン・マッチング法 ”, 電子情報通信学会論文誌 (D), Vol.J68-D, No.3, pp.345-352, 1985.

[8] E. Oja: Subspace methods of pattern recognition, Research Studies Press, 1983.

[9] 黒沢 由明: 球面ガウス分布から導出される部分空間法, Trans. IEICE, Vol.J81-D-2, No.6, pp.1205-1212, 1998.

[10] 杉山 善明, 有木 康雄: 多重部分空間法に基づくテレビスポーツニュース映像の自動分類, Trans. IEICE, Vol.J81-D-2, No.9, pp.2112-2119, 1998.

[11] B. Leibe and B. Schiele: Analyzing appearance and contour based methods for object categorization, Proc. IEEE Conference on Computer Vision and Pattern Recognition, Vol.2, pp.409-415, 2003.

Subspace Selection for Resolution Synthesis

Atsunori KANEMURA†, Shin-ichi MAEDA†, and Shin ISHII††

†, †† Graduate School of Informatics, Kyoto University
Gokasho, Uji, Kyoto 611–0011, Japan
E-mail: †{atsu-kan,ichi}@sys.i.kyoto-u.ac.jp, ††ishii@i.kyoto-u.ac.jp

Abstract Resolution synthesis (RS) is a framework for expanding a given image using an interpolator *trained in advance* with a training dataset. We address how to determine the optimal size of the support for RS using a sparse Bayesian formulation. Experiments show that compact supports can be automatically learned by our Bayesian RS.
Key words Image expansion, resolution synthesis, sparse Bayesian estimation, subspace selection

1. Introduction

Resolution synthesis (RS) [1,2] is a framework for expanding a given image using an interpolator *trained in advance* using a training dataset. Prior training is the characterizing feature of RS and it differentiates RS from classical image expansion methods such as bilinear interpolation and splines. When determining the value of a pixel in a high-resolution image, the bilinear interpolation filter uses at most four low-resolution pixels around the pixel of interest. In contrast, RS in principle can use a support of arbitrary size. Atkins' original RS [1,2] used a 5×5 window without providing logical justification to this choice. The supports should be simple for efficient processing of images and also for preventing overfitting, whereas those that are too simple will deteriorate the expansion performance. We address the problem of determining an optimal support by formulating RS from a viewpoint of sparse Bayesian estimation.

Let r be an integer magnification factor. The purpose here is to estimate an $rM \times rN$ expanded image $\hat{\boldsymbol{\xi}}$ from a given $M \times N$ image $\boldsymbol{\xi}$. In RS, an interpolator called a resolution synthesizer (RSer) expands the image by replacing each one pixel in the given image by an $r \times r$ high-resolution patch. To estimate the high-resolution patch, RS uses the low-resolution pixel patch surrounding the low-resolution pixel to be replaced (Fig. 1). This local interpolation is repeated for every pixel in the given image and the expanded image is constructed by tessellating the high-resolution patches.

In Section 2, we describe the classical maximum likelihood RS (MLRS). Section 3 presents a Bayesian modeling of RS (BayesRS), and we derive an iterative algorithm to find the optimal BayesRSer in Section 4. Experimental results are given in Section 5. Section 6 summarizes this article.

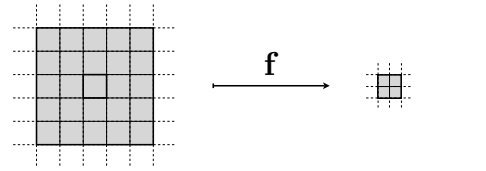


Fig. 1 Resolution synthesis uses $m \times m = Q$ low-resolution pixels to estimate $r \times r = D$ high-resolution pixels.

2. Resolution Synthesis

In advance of real image expansion jobs, we train a RSer using a training dataset. The dataset consists of a large number of low- and high-resolution patches, and the RSer learns the relationship between the low- and high-resolution patches. Let $\mathbf{z}_n$ be the $m^2 = Q$-dimensional vectors of low-resolution patches, $\mathbf{x}_n$ be the $r^2 = D$-dimensional vectors of high-resolution patches, and $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{z}_n)\}_{n=1}^N$ be the dataset consisting of N pairs of the patches. We stack the vectors column-wise and obtain matrices $\mathsf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$ and $\mathsf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$.

We assume a linear relationship between $\mathbf{x}_n$ and $\mathbf{z}_n$:

$$\mathbf{x}_n = \mathsf{W}\mathbf{z}_n + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_n, \tag{1}$$

where W is a $D \times Q$ filtering matrix, $\boldsymbol{\mu}$ is a D-dimensional bias vector, and $\boldsymbol{\varepsilon}_n$ is isotropic Gaussian noise with precision (inverse variance) β. Let $\mathbf{w}_d$ be the dth row of W. Then $\mathbf{w}_d$ is the filtering kernel to estimate the dth pixel of the high-resolution patch. Therefore we regard W as a matrix built by stacking D filters. This model leads to the probability distribution of $\mathbf{x}_n$, or the likelihood,

$$p(\mathbf{x}_n|\mathbf{z}_n, \mathsf{W}, \boldsymbol{\mu}, \beta) = \mathcal{N}(\mathbf{x}_n|\mathsf{W}\mathbf{z}_n + \boldsymbol{\mu}, \beta^{-1}\mathsf{I}_D), \tag{2}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = |2\pi\Sigma|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$ and I_D is the D-dimensional identity matrix.

MLRS estimates the parameters via the maximum likelihood rule

$$(\mathsf{W}^*, \boldsymbol{\mu}^*) = \arg\max_{(\mathsf{W},\boldsymbol{\mu})} \Big(\sum_{n=1}^{N} \ln p(\mathbf{x}_n|\mathbf{z}_n, \mathsf{W}, \boldsymbol{\mu}) \Big), \tag{3}$$

whose solution can be easily found as

$$\widetilde{\mathsf{W}}^* = (\widetilde{\mathsf{Z}}\widetilde{\mathsf{Z}}^{\mathrm{T}})^{-1}\widetilde{\mathsf{Z}}\mathsf{X}^{\mathrm{T}}, \tag{4}$$

where $\widetilde{\mathsf{W}}$ and $\widetilde{\mathsf{Z}}$ are extended matrix and vector, respectively, to include $\boldsymbol{\mu}$ and defined as

$$\widetilde{\mathsf{W}} = \begin{bmatrix} \mathsf{W} & \boldsymbol{\mu} \end{bmatrix}, \quad \widetilde{\mathsf{Z}} = \begin{bmatrix} \mathsf{Z} \\ \mathbf{1}^{\mathrm{T}} \end{bmatrix}. \tag{5}$$

MLRS then estimates a high-resolution patch $\mathbf{x}$ from a given $\mathbf{z}$ by the following filtering equation:

$$\mathbf{x} = \widetilde{\mathsf{W}}^* \begin{bmatrix} \mathbf{z} \\ 1 \end{bmatrix} = \mathsf{W}^*\mathbf{z} + \boldsymbol{\mu}^*. \tag{6}$$

Note that maximum likelihood estimation inherently suffers from overfitting, that is, increasing the size of the filters beyond certain complexity results in increased generalization errors, although the training errors always decrease [3].

3. Bayesian Modeling of RS

According to the Bayesian framework, all parameters are treated as *random variables*, and prior distributions are put on them as follows:

$$p(\mathsf{W}|\mathsf{A}) = \prod_{d=1}^{D}\prod_{q=1}^{Q} \mathcal{N}(w_{dq}|0, \alpha_{dq}^{-1}), \tag{7}$$

$$p(\boldsymbol{\mu}|\rho) = \mathcal{N}(\boldsymbol{\mu}|\mathbf{0}, \rho^{-1}\mathsf{I}_D), \tag{8}$$

$$p(\beta) = \mathcal{G}(\beta|a_{\beta 0}, b_{\beta 0}), \tag{9}$$

where the gamma distribution is denoted by $\mathcal{G}(\tau|a,b) = \frac{1}{\Gamma(a)}b^a\tau^{a-1}e^{-b\tau}$. We further put hierarchical priors

$$p(\mathsf{A}) = \prod_{d=1}^{D}\prod_{q=1}^{Q} \mathcal{G}(\alpha_{dq}|a_{\alpha 0}, b_{\alpha 0}), \tag{10}$$

$$p(\rho) = \mathcal{G}(\rho|a_{\rho 0}, b_{\rho}). \tag{11}$$

The joint density is decomposed according to the model as

$$p(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta, \mathsf{X}|\mathsf{Z}) = p(\mathsf{A})p(\mathsf{W}|\mathsf{A})p(\rho)p(\boldsymbol{\mu}|\rho)\prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n, \mathsf{W}, \boldsymbol{\mu}, \beta). \tag{12}$$

The prior for the filtering matrix W, (7), is similar to that in a sparse Bayesian treatment called *automatic relevance determination* (ARD), which was first introduced for neural networks [4]. The parameters α_{dq} work as regularizers that pull w_{dq} toward the prior mean 0. Therefore, if α_{dq} are large, estimated values of w_{dq} become small. It is known that in this "sparse Bayesian" type of estimation [5], the elements of W irrelevant to the filtering subspace are automatically pruned because the corresponding elements of A diverge to infinity.

The filtering equation that maps a low-resolution patch $\mathbf{z}$ to a corresponding high-resolution patch $\mathbf{x}$ is given by the mean value of the predictive distribution:

$$\mathbb{E}(\mathbf{x}) = \int d\mathbf{x}\, \mathbf{x} p(\mathbf{x}|\mathbf{z}, \mathcal{D}). \tag{13}$$

The predictive distribution $p(\mathbf{x}|\mathbf{z}, \mathcal{D})$ is given by

$$p(\mathbf{x}|\mathbf{z}, \mathcal{D}) = \int d\mathsf{A} d\mathsf{W} d\boldsymbol{\mu} d\rho d\beta\; p(\mathbf{x}|\mathbf{z}, \mathsf{W}, \boldsymbol{\mu}, \beta) \times p(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta|\mathcal{D}), \tag{14}$$

where the posterior is given by the Bayes theorem as

$$p(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta|\mathcal{D}) = \frac{p(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta, \mathsf{X}|\mathsf{Z})}{\int d\mathsf{A} d\mathsf{W} d\boldsymbol{\mu} d\rho d\beta\, p(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta, \mathsf{X}|\mathsf{Z})}. \tag{15}$$

However, analytical evaluation of the true predictive distribution is intractable because it is a complex of Gaussian and gamma variables. Therefore, we adopt an efficient computation procedure based on variational estimation.

4. Variational Estimation

The posterior distribution $p(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta|\mathcal{D})$ is approximated by a trial distribution q, which is a distribution restricted to have a factorization property:

$$q(\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta) = q(\mathsf{A})q(\mathsf{W})q(\rho)q(\boldsymbol{\mu})q(\beta). \tag{16}$$

We denote the latent variables by $\boldsymbol{\tau} = \{\mathsf{A}, \mathsf{W}, \rho, \boldsymbol{\mu}, \beta\}$ for simplicity. Within the restricted distribution space, we search for the optimal trial distribution that minimizes the Kullback-Leiber (KL) divergence to the true posterior distribution:

$$q^\star(\boldsymbol{\tau}) = \operatorname*{argmin}_{q} D_{\mathrm{KL}}(q(\boldsymbol{\tau})\|p(\boldsymbol{\tau}|\mathcal{D})), \tag{17}$$

where the KL divergence is defined by

$$D_{\mathrm{KL}}(q(\boldsymbol{\tau})\|p(\boldsymbol{\tau}|\mathcal{D})) = -\int d\boldsymbol{\tau}\, q(\boldsymbol{\tau}) \ln \frac{p(\boldsymbol{\tau}|\mathcal{D})}{q(\boldsymbol{\tau})} \tag{18}$$

$$= -\Big\langle \ln \frac{p(\boldsymbol{\tau}|\mathcal{D})}{q(\boldsymbol{\tau})} \Big\rangle. \tag{19}$$

Here, $\langle\cdot\rangle$ is the expectation operator with respect to $q(\boldsymbol{\tau})$. The KL divergence is always nonnegative, $D_{\mathrm{KL}}(q\|p) \geqq 0$, for any q and p, and $D_{\mathrm{KL}}(q\|p) = 0$ if and only if q and p

are equivalent distributions. This variational optimization problem can be analytically solved if we optimize only one factor, fixing the other factors. We then iterate computing optimal factors $q^*(\mathsf{A})$, $q^*(\mathsf{W})$, $q^*(\rho)$, $q^*(\boldsymbol{\mu})$, and $q^*(\beta)$ in a sequential manner until convergence to find a minimum $q^\star$.

The optimal trial factors are found as follows:

$$q^*(\mathsf{A}) = \prod_{d=1}^{D}\prod_{q=1}^{Q} \mathcal{G}(\alpha_{dq}|a_{\alpha dq}, b_{\alpha dq}), \tag{20}$$

$$q^*(\mathsf{W}) = \prod_{d=1}^{D} \mathcal{N}(\mathbf{w}_d|\mathbf{m}_{\mathrm{w}}^{(d)}, \boldsymbol{\Sigma}_{\mathrm{w}}^{(d)}), \tag{21}$$

$$q^*(\rho) = \mathcal{G}(\rho|a_\rho, b_\rho), \tag{22}$$

$$q^*(\boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\mu}|\mathbf{m}_\mu, \boldsymbol{\Sigma}_\mu), \tag{23}$$

$$q^*(\beta) = \mathcal{G}(\beta|a_\beta, b_\beta), \tag{24}$$

where the parameters are

$$a_{\alpha dq} = a_{\alpha 0} + \frac{1}{2}, \quad b_{\alpha dq} = b_{\alpha 0} + \frac{1}{2}\langle w_{dq}^2\rangle, \tag{25}$$

$$\boldsymbol{\Sigma}_{\mathrm{w}}^{(d)} = \Big(\langle \mathrm{diag}(\alpha_{d1}, \ldots, \alpha_{dQ})\rangle + \langle\beta\rangle \sum_{n=1}^{N} \mathbf{z}_t \mathbf{z}_t^{\mathrm{T}}\Big)^{-1}, \tag{26}$$

$$\mathbf{m}_{\mathrm{w}}^{(d)} = \langle\beta\rangle \boldsymbol{\Sigma}_{\mathrm{w}}^{(d)} \sum_{n=1}^{N} (x_{dn} - \langle\mu_d\rangle)\mathbf{z}_n, \tag{27}$$

$$a_\rho = a_{\rho 0} + \frac{D}{2}, \quad b_\rho = b_{\rho 0} + \frac{1}{2}\langle \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\rangle, \tag{28}$$

$$\boldsymbol{\Sigma}_\mu = \frac{1}{\langle\rho\rangle + N\langle\beta\rangle}\mathsf{I}_D, \tag{29}$$

$$\mathbf{m}_\mu = \langle\beta\rangle \boldsymbol{\Sigma}_\mu \sum_{n=1}^{N} (\mathbf{x}_n - \langle\mathsf{W}\rangle\mathbf{z}_n), \tag{30}$$

$$a_\beta = a_{\beta 0} + \frac{ND}{2}, \tag{31}$$

$$\begin{aligned} b_\beta = b_{\beta 0} + \frac{1}{2}\sum_{n=1}^{N} \big\{ &\mathbf{x}_n^{\mathrm{T}}\mathbf{x}_n - 2\mathbf{x}_n\langle\mathsf{W}\rangle\mathbf{z} - 2\mathbf{x}_n^{\mathrm{T}}\langle\boldsymbol{\mu}\rangle \\ &+ \mathbf{z}_n^{\mathrm{T}}\langle\mathsf{W}^{\mathrm{T}}\mathsf{W}\rangle\mathbf{z}_n^{\mathrm{T}} + 2\mathbf{z}_n^{\mathrm{T}}\langle\mathsf{W}\rangle^{\mathrm{T}}\langle\boldsymbol{\mu}\rangle + \langle\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu}\rangle\big\}. \end{aligned} \tag{32}$$

The expectations remaining in the above equations can be evaluated easily using the well-known results in statistics.

We denote the mean of the joint trial distribution $q(\mathsf{W})$ by M_{W}, that is, we put $\mathsf{M}_{\mathrm{W}} = [\mathbf{m}_{\mathrm{w}}^{(1)}, \ldots, \mathbf{m}_{\mathrm{w}}^{(D)}]^{\mathrm{T}}$. The filtering equation for the variational BayesRS is obtained by substituting the true posterior distribution with the trial distribution, which results in

$$\mathbb{E}(\mathbf{x}) \approx \langle\mathbf{x}\rangle = \langle\mathsf{W}\rangle\mathbf{z} + \langle\boldsymbol{\mu}\rangle = \mathsf{M}_{\mathrm{W}}\mathbf{z} + \mathbf{m}_\mu. \tag{33}$$

As a criterion to check convergence and stop iterating (20)–(24), we monitor the relative change of the Frobenius norm of M_{W}

$$\Delta = \|\mathsf{M}'_{\mathrm{W}} - \mathsf{M}_{\mathrm{W}}\|_{\mathrm{F}} / \|\mathsf{M}'_{\mathrm{W}}\|_{\mathrm{F}}, \tag{34}$$

where M'_{W} is the matrix at the previous iteration step, and terminate the algorithm when $\Delta < 10^{-6}$. To accelerate the convergence, the expected values of α_{dq} are thresholded to infinity when they are greater than e^{20}.

Fig. 2 Images used as for training RSers (4.1.[01–08] in the USC-SIPI image database [6]).

We shall use a hyperparameter setting of the noninformative limit, $a_{\beta 0} = b_{\beta 0} = 0, a_{\rho 0} = b_{\rho 0} = 0$, for β and ρ but we use $a_{\alpha 0} = 20, b_{\alpha 0} = 0$ to facilitate the divergence of α_{dq}. Having zero hyperparameters makes the priors improper, but it is not a problem since the posteriors are well defined.

5. Experiments

We conducted experiments to see which subspace would be selected by BayesRS and to compare the performance of BayesRS with that of MLRS. The expanding factor was chosen to be $r = 2$. The training dataset was prepared by the following procedure. High-resolution patches were prepared by cutting the eight images of size 256×256 shown in Fig. 2 into non-overlapping pieces, resulting in $N = 8 \cdot 256^2/r^2 = 131,072$ patches in total. To make low-resolution patches, first the high-resolution images were shrunk by a factor of 2, and overlapping patches of size $m \times m$ were extracted to produce $131,072$ low-resolution patches. To extract patches near the boundaries, the low-resolution images were extended by replication.

To evaluate the generalization performance in expanding images, we used the Lena image shown in Fig. 7(a) (4.2.04 in the USC-SIPI image database) as the original image $\boldsymbol{\xi}$, which was not included in the training image dataset. This

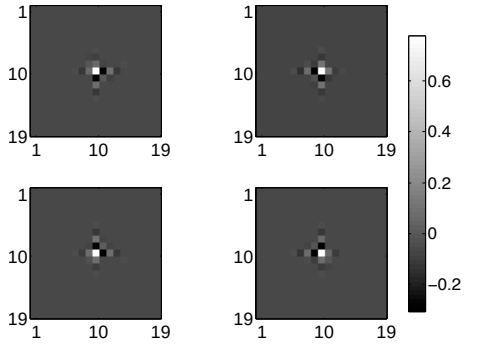


Fig. 3 Learned BayesRS filters (in log scale).

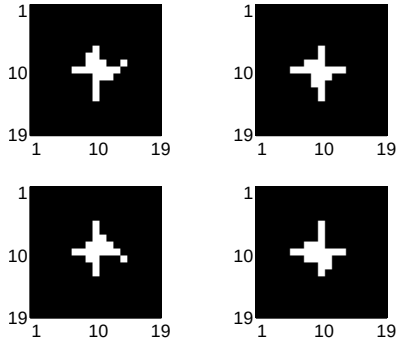


Fig. 4 Supports of BayesRS filters.

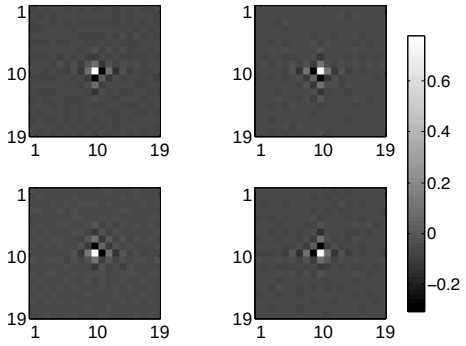


Fig. 5 Learned MLRS filters (in log scale).

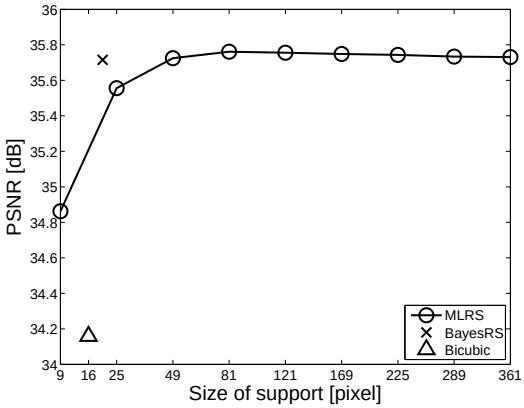


Fig. 6 Performance of RSers with effective sizes of support. The circles connected by line show the performance of the MLRSers and the cross is the one of the BayesRSer. For comparison, the performance of the bicubic interpolation method is shown by the triangle.

image was shrunk by a factor of 2 (Fig. 7(b)) and given to the trained RSers. To quantitatively assess the performance of the RSers, the peak signal-to-noise ratio (PSNR) of the expanded image $\hat{\boldsymbol{\xi}}$ was measured. PSNR is defined by

$$PSNR(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) = 10 \log_{10} \frac{\kappa^2}{\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}\|^2 / MN} \quad [\text{dB}], \tag{35}$$

where κ is the maximum pixel value and MN is the number of pixels. When displaying filters, we use a log conversion $\text{sign}(w_{dq}) \ln(1 + |w_{dq}|)$, where $\text{sign}(x) = +1/0/-1$ if x is positive, zero, or negative, respectively.

The BayesRS algorithm was executed with the size of the filters being $m \times m = 19 \times 19$. The shapes of the learned filters are shown in Fig. 3, and the supports (regions where the filters had nonzero values) are shown in Fig. 4. The sizes of the learned supports were 20, 20, 20, and 21. An interesting point is that the learned supports had asymmetric shapes. From the shapes of the learned supports, we can say that the direct horizontal and vertical pixels are highly relevant for estimating high-resolution pixels, but the diagonal pixels are of less importance. The expanded Lenna image using the learned filters is shown in Fig. 7(d) and its PSNR was 35.72 dB, which was significantly (about 1.6 dB) better than the image expanded by the bicubic method (Fig. 7(c)). The cross in Fig. 6 indicates the PSNR and its horizontal coordinate is the mean support size (20.25).

Next, we measured the performances of the MLRSers with sizes of the supports varying from $3 \times 3 = 9$ pixels to $19 \times 19 = 361$ pixels. Fig. 5 shows the shapes of the filters trained by the MLRSer when the size of the support was 19×19. There were no nonzero element in the filters. The PSNRs of the MLRSers are shown in Fig. 6 as the circles connected by the line. The maximum PSNR of 35.76 dB was attained when the support size was $9 \times 9 = 81$, and the use of larger supports only degraded the performance, showing a typical overfitting.

6. Conclusion

We showed that automatic selection of the subspace relevant to RS image expansion was successfully achieved using a sparse Bayesian methodology that incorporated the prior setting called ARD. The PSNR of BayesRS's estimation was 0.04 dB worse than that of the best MLRS, which indicates an essentially ignorable loss of performance. The mean size of BayesRS's support, 20.25, was 1/4 of that of the best MLRS, which was significantly smaller. These facts suggest BayesRS should be advantageous for future practical applications.

Acknowledgment

We thank Dr. S. Oba at Kyoto University for his insightful comments on ARD and an early version of this article.

References

[1] C. B. Atkins, Classification-Based Methods in Optimal Im-

(a) 512 × 512 original image.

(b) 256 × 256 low-resolution image.

(c) Bicubic interpolation. PSNR: 34.16 dB.

(d) Bayesian RS. PSNR: 35.72 dB.

Fig. 7 Images.

(a) Original image (close-up).

(b) Low-resolution image (close-up).

(c) Bicubic interpolation (close-up).

(d) BayesRS (close-up).

Fig. 8 Close-up images.

age Interpolation, Ph.D thesis, Purdue University, 1998.

[2] C. B. Atkins, C. A. Bouman, and J. P. Allebash, "Tree-based resolution synthesis," Proc. of PICS Conference, pp. 405–410, Cavannah, GA, Apr. 1999.

[3] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer, New York, 2001.

[4] D. J. C. MacKay, "Probable networks and plausible predictions," Network: Compt. Neural. Syst., vol. 6, no. 3, pp. 469–505, 1995.

[5] A. C. Faul, and M. E. Tipping, "Analysis of sparse Bayesian learning," Advances in NIPS 14, pp. 383–389, MIT Press, 2002.

[6] The USC-SIPI Image Database, University of Southern California, `http://sipi.usc.edu/database/`.

Generalized *N*-Dimensional Principal Component Analysis and Efficient Representation of Medical Volumes

Yen-Wei CHEN † and Rui XU †

† College of Information Science and Eng., Ritsumeikan Univ. 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan
E-mail: † chen@is.ritsumei.ac.jp

Abstract There is a growing interest in multi-dimensional image processing, such as medical volume image processing, hyperspectral image processing. In this paper, we propose a novel approach called generalized N-dimensional principal component analysis (GND-PCA) for efficient multi-dimensional data representation and modeling. In GND-PCA, the multi-dimensional data is treated as a tensor. The optimal subspaces on each mode are simultaneously calculated by minimizing the square error between the original tensor and the reconstructed tensor based on the subspace. Experiments on medical MRI dataset show that the proposed GND-PCA can represent the multi-dimensional data more efficiently compared to conventional PCA and recently proposed ND-PCA.

Keyword *N*-Dimensional Principal Component Analysis , Generalized, Tensor , High-order SVD , Efficient Representation, Medical Volume

1. Introduction

Principal component analysis (PCA) is an important technique for efficient data representation and modeling. PCA is an orthogonal linear transform that projects the data into a new coordinate system (subspace) with bases where the data varies the most. The bases are determined by the eigenvectors of the covariance matrix corresponding to the largest eigenvalues. The magnitude of the eigenvalues corresponds to the variance of the data along the eigenvector bases [1, 2]. Since the classical PCA is the method for 1-dimensional (1D) vector data, when PCA is applied to multi-dimensional data (e.g. 2D image or 3D volume), the multi-dimensional data should be initially unfolded to a long 1D vector. Such unfolding process will introduce several problems: (1) the feature vector is in high dimensional vector space resulting in huge computation cost and bad performance on generalization; (2) lost of spatial information. Yang et al proposed a new method called 2-dimensional principal component analysis (2D-PCA) to overcome the above problems [3]. This method is to calculate the bases in the column-mode subspace of the 2D image instead of finding the basis in the long unfolding vector subspace. Therefore, the 2D data can be directly used in the training without the unfolding vector preprocessing. 2D-PCA not only makes the calculation of the bases efficiently but also can accurately represent the 2D data, however its drawback is that it needs more coefficients to represent the 2D data than PCA because 2D PCA is a unilateral projection (right multiplication) scheme. Kong et al. proposed a generalized 2-dimensional PCA (G2D-PCA) [4], which is a bilateral projection scheme, to simultaneously calculate the basis of the row- and column-mode subspaces, so it can represent the 2D data not only accurately but efficiently. Recently, inspired from the work of 2D-PCA, a method called N-dimensional PCA (ND-PCA) was proposed for higher-dimensional data representation [5]. In ND-PCA, the higher-dimensional data is treated as the higher-order tensor witch is directly trained to obtain the bases on one-mode subspace by multi-linear algebra based tool called higher-order singular value decomposition (HOSVD) [6]. It was applied on the 3D facial scanning data. Since ND-PCA only compresses the data on one-mode subspace, it is also suffered from the problem that the data can not represented efficiently, similar to the problem of 2D-PCA.

Inspired from the works of G2D-PCA and ND-PCA, we proposed a new method called generalized N-dimensional principal component analysis (GND-PCA). The high–dimensional data is treated as a series of higher-order tensors and the optimal subspace on each mode are simultaneously calculated by minimizing the square error between the original tensor and the reconstructed tensor based on the subspace with an iteration algorithm. Experiments on medical MRI dataset show that the proposed GND-PCA can represent the

multi-dimensional data more efficiently compared to the conventional PCA and ND-PCA.

The paper is organized as follows: related works such as 2D-PCA, G2D-PCA and ND-PCA are briefly summarized in Sec.2; the proposed GND-PCA is presented in Sec.3; experimental results on medical MRI dataset and multi-angle view & illumilation facial dataset are presented in Sec.4; and conclusions are given in Sec.5.

2. Related Works

2.1. SVD and PCA

Suppose a series of D-dimensional vectors with zero-mean, $\mathbf{a}_i, i=1,2,\cdots,M$ are given and $\mathbf{A}=[\mathbf{a}_1,\mathbf{a}_2,\cdots,\mathbf{a}_M]$ is a $D\times M$ matrix, where M is the number of samples. PCA is based on eigenvalue decomposition of the covariance matrix $\mathbf{Cov}$, as Eq.(1).

$$\mathbf{Cov}=\mathbf{A}\mathbf{A}^T=\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T \qquad (1)$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix corresponding of eigen values and $\mathbf{W}$ is a $D\times D$ matrix, whose column vectors $\mathbf{w}_i, i=1,2,\cdots,D$ are eigenvectors of $\mathbf{Cov}$. The leading J eigenvectors, where $J\le D$ construct the subspace and the vector $\mathbf{a}\in\mathbf{R}^D$ can be represented by its coefficient vector, $\mathbf{b}=[\mathbf{w}_1,\mathbf{w}_2,\cdots,\mathbf{w}_J]^T\cdot\mathbf{a}\in\mathbf{R}^J$.

On the other hand, a matrix $\mathbf{A}\in\mathbf{R}^{D\times M}$ can be decomposed by SVD as

$$\mathbf{A}=\mathbf{U}\mathbf{S}\mathbf{V}^T \qquad (2)$$

where $\mathbf{U}\in\mathbf{R}^{D\times D}$ and $\mathbf{V}\in\mathbf{R}^{M\times M}$ are orthogonal matrices. $\mathbf{S}\in\mathbf{R}^{D\times M}$ is a diagonal matrix corresponding of singular values. From Eq.(2),

$$\mathbf{A}\mathbf{A}^T=\mathbf{U}\mathbf{S}\mathbf{V}^T\mathbf{V}\mathbf{S}^T\mathbf{U}^T=\mathbf{U}\mathbf{S}^2\mathbf{U}^T \qquad (3)$$

It is clear that the squares of the non-zero singular values of $\mathbf{A}$ are equal to the non-zero eigenvalues of $\mathbf{A}\mathbf{A}^T=\mathbf{Cov}$ and the columns of $\mathbf{U}$ (left singular vectors) are eigenvectors of $\mathbf{Cov}$. Thus we just need to apply SVD to $\mathbf{A}$ to get the principal orthogonal vectors (bases).

2.2. HO-SVD and ND-PCA

A N-th order tensor $\mathcal{A}\in\mathbf{R}^{I_1\times I_2\times\cdots\times I_N}$ is defined as a multi-array with N indices. The space of the tensor is comprised by the N mode space. The tensor $\mathcal{A}$ can be unfolded to $\mathbf{A}_{(n)}\in\mathbf{R}^{I_n\times(I_1\cdot I_2\cdots I_{n-1}\cdot I_{n+1}\cdots I_N)}$ which is called as mode-n matrix, where $0<n\le N$. Unfolding of a 3rd order tensor is shown in Fig.1. The tensor $\mathcal{A}$ can be decomposed by the higher-order SVD (HO-SVD), which is also known as Tucker decomposition [6], as

$$\mathcal{A}=\mathcal{B}\times_1\mathbf{U}^{(1)}\times_2\mathbf{U}^{(2)}\times\cdots\times_N\mathbf{U}^{(N)} \qquad (4)$$

where $\mathbf{U}^{(n)}\in\mathbf{R}^{I_n\times I_n}$ is a unitary matrix, $\mathcal{B}\in\mathbf{R}^{I_1\times I_2\times\cdots\times I_N}$ is the core tensor, and $\times_n$ is the mode-n product. As shown in Eq.(5), $\mathcal{B}\times_n\mathbf{U}^{(n)}$ can be rewritten as a matrix multiplication and the result C is also a N-th order tensor.

$$\mathcal{B}\times_n\mathbf{U}^{(n)}=\mathbf{U}^{(n)}\mathbf{B}_{(n)}=\mathbf{C}_{(n)}=C \qquad (5)$$

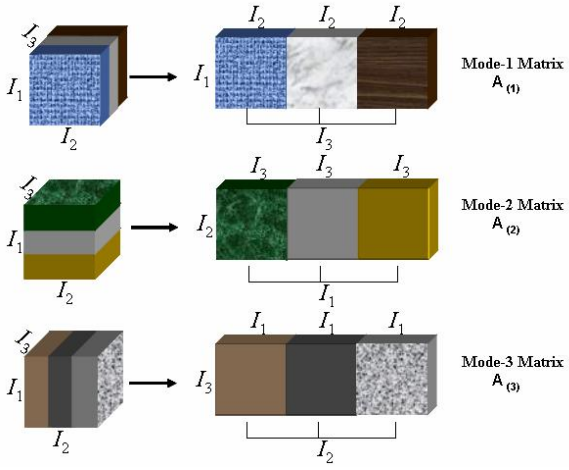


Fig.1 Example of unfolding the 3rd order tensor

HO-SVD has been applied to ND-PCA[5] with applications to 3D facial scanning data, representation of face with multiple-modes [7,8] and robust face recognition[9]. In ND-PCA, the N-dimensional dataset are directly treated as N-th order tensors $\mathcal{A}_i\in\mathbf{R}^{I_1\times I_2\times\cdots\times I_N}$, $i=1,2,\cdots,M$. In a similar manner described in Sec.2.1, instead of calculating the covariance tensor, we just need to construct a new tensor $\mathcal{X}=[\mathcal{A}_1,\mathcal{A}_2,\cdots,\mathcal{A}_M]\in\mathbf{R}^{I_1\times I_2\times\cdots\times I_N\times M}$, and apply HO-SVD on its mode-n subspace. The first leading J eigenvectors $\mathbf{u}_1^{(n)},\mathbf{u}_2^{(n)},\cdots,\mathbf{u}_J^{(n)}$, where $J<I_n$, are the bases on the mode-n subspace $\mathbf{U}^{(n)}$. The N-dimensional data $\mathcal{A}\in\mathbf{R}^{I_1\times I_2\times\cdots\times I_N}$ can be compactly represented by a tensor $\mathcal{B}=\mathcal{A}\times_n\mathbf{U}^{(n)T}\in\mathbf{R}^{I_1\times I_2\times\cdots\times I_{n-1}\times J\times I_{I+1}\times\cdots\times I_N}$, whose components are the projections (coefficients) onto the mode-n subspace.

3. Generalized ND-PCA

Though ND-PCA can make the calculation of the bases efficiently and can accurately represent the multi-dimensional data. As well as 2D-PCA, ND-PCA is also a unilateral projection scheme and only compress the data on the mode-*n* subspace. So ND-PCA needs lots of components to represent the multi-dimensional data. In this paper, we propose a generalized *N*-dimensional PCA (GND-PCA) to simultaneously calculate the basis on each mode subspace [10].

The basic idea of GND-PCA is that we want to reconstruct the original *N*-th order tensor $\mathcal{A} \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a lower rank core tensor $\mathcal{B} \in \mathbf{R}^{J_1 \times J_2 \times \cdots \times J_N}$, where $J_n < I_n$, and try to find a set of optimal matrices $\mathbf{U}^{(n)} \in \mathbf{R}^{I_n \times J_N}$, $n = 1,2,\cdots,N$ with orthogonal column for each mode. The reconstruction of *N*-th order tensor $\mathcal{A}$ can be expressed as $\hat{\mathcal{A}} = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times \cdots \times_N \mathbf{U}^{(N)}$. Illustration for 3rd order tensor reconstruction is shown in Fig.2.

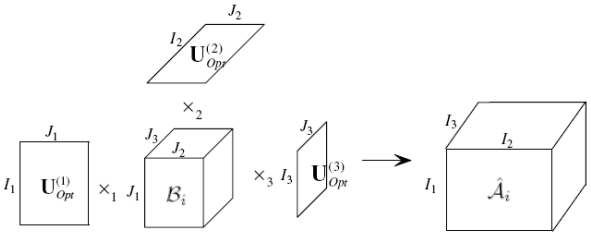


Fig.2 Illustration of reconstructing a 3rd order tensor

The optimal orthogonal matrices $\mathbf{U}^{(n)}$ can be determined by minimizing a cost function as Eq.(6).

$$S = \sum_{i=1}^{M} \left\| \mathcal{A}_i - \hat{\mathcal{A}}_i \right\|^2 = \sum_{i=1}^{M} \left\| \mathcal{A}_i - \mathcal{B}_i \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times \cdots \times_N \mathbf{U}^{(N)} \right\|^2 \quad (6)$$

In Eq.(6), only the samples $\mathcal{A}_i \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_N}$, ($i = 1,2,\cdots,M$) are known, *M* is the number of samples.

Theorem 1: *Given fixed N matrices* $\boldsymbol{U}^{(n)}$*, the tensors* $\mathcal{B}_i$ *that minimize the cost function of.(6) are given by*

$$\mathcal{B}_i = \mathcal{A}_i \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \times \cdots \times_N \mathbf{U}^{(N)^T} \quad (7)$$

Since the proof of Theory 1 is simple, it is omitted here. From Theorem 1, we can obtain Theorem 2 [10].

Theorem 2: *If the tensors* $\mathcal{B}_i$ *are given as Eq.(7), minimization of the cost function of Rq.(6) is equal to maximization of the following cost function:*

$$S' = \sum_{i=1}^{M} \left\| \mathcal{A}_i \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \times \cdots \times_N \mathbf{U}^{(N)^T} \right\|^2 \quad (8)$$

There is no close-form solution to simultaneously resolve the matrices $\mathbf{U}^{(n)}$ for the cost function S', however the explicit solution for one matrix can be obtained if the other matrices are fixed [10]. So we use an iteration algorithm to simultaneously calculate the optimal matrices $\mathbf{U}_{opt}^{(1)}, \mathbf{U}_{opt}^{(2)}, \cdots, \mathbf{U}_{opt}^{(N)}$, which are able to maximize the cost function S'. This algorithm is summarized in Algorithm 1. In Algorithm 1, we terminate the iteration when the cost of Eq.(8) is not significantly changed in two consecutive times.

Algorithm 1 Iteration Algorithm to Compute the N Matrices $\mathbf{U}_{opt}^{(1)}, \mathbf{U}_{opt}^{(2)}, \ldots, \mathbf{U}_{opt}^{(N)}$

IN: a series of N-th order tensors, $\mathcal{A}_i \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}, i = 1,2,\ldots,M$.
OUT: N Matrices $\mathbf{U}_{opt}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ $(J_n < I_n, n = 1,2,\ldots,N)$ with orthogonal column vectors.

1. Initial values: $k = 0$ and $\mathbf{U}_0^{(n)}$ whose columns are determined as the first J_n leading eigenvectors of the matrices $\sum_{i=1}^{M} (\mathbf{A}_{i(n)} \cdot \mathbf{A}_{i(n)}^T)$.
2. Iterate for k until convergence
   - Maximize $S' = \sum_{i=1}^{M} \| \mathcal{C}_i \times_1 \mathbf{U}^{(1)^T} \|^2$, $\mathcal{C}_i = \mathcal{A}_i \times_2 \mathbf{U}_k^{(2)^T} \times \ldots \times_N \mathbf{U}_k^{(N)^T}$
     Solution: $\mathbf{U}^{(1)}$ whose columns are determined as the first J_1 leading eigenvectors of $\sum_{i=1}^{M} (\mathbf{C}_{i(1)} \cdot \mathbf{C}_{i(1)}^T)$
     Set $\mathbf{U}_{k+1}^{(1)} = \mathbf{U}^{(1)}$.
   - Maximize $S' = \sum_{i=1}^{M} \| \mathcal{C}_i \times_2 \mathbf{U}^{(2)^T} \|^2$, $\mathcal{C}_i = \mathcal{A}_i \times_1 \mathbf{U}_{k+1}^{(1)^T} \times_3 \mathbf{U}_k^{(3)^T} \times \ldots \times_N \mathbf{U}_k^{(N)^T}$
     Solution: $\mathbf{U}^{(2)}$ whose columns are determined as the first J_2 leading eigenvectors of $\sum_{i=1}^{M} (\mathbf{C}_{i(2)} \cdot \mathbf{C}_{i(2)}^T)$
     Set $\mathbf{U}_{k+1}^{(2)} = \mathbf{U}^{(2)}$.
     ...
   - Maximize $S' = \sum_{i=1}^{M} \| \mathcal{C}_i \times_n \mathbf{U}^{(n)^T} \|^2$, $\mathcal{C}_i = \mathcal{A}_i \times_1 \mathbf{U}_{k+1}^{(1)^T} \times \ldots \times_{n-1} \mathbf{U}_{k+1}^{(n-1)^T} \times_{n+1} \mathbf{U}_k^{(n+1)^T} \times \ldots \times_N \mathbf{U}_k^{(N)^T}$
     Solution: $\mathbf{U}^{(n)}$ whose columns are determined as the first J_n leading eigenvectors of $\sum_{i=1}^{M} (\mathbf{C}_{i(n)} \cdot \mathbf{C}_{i(n)}^T)$
     Set $\mathbf{U}_{k+1}^{(n)} = \mathbf{U}^{(n)}$.
     ...
   - Maximize $S' = \sum_{i=1}^{M} \| \mathcal{C}_i \times_N \mathbf{U}^{(N)^T} \|^2$, $\mathcal{C}_i = \mathcal{A}_i \times_1 \mathbf{U}_{k+1}^{(1)^T} \times \ldots \times_{N-1} \mathbf{U}_{k+1}^{(N-1)^T}$
     Solution: $\mathbf{U}^{(N)}$ whose columns are determined as the first J_N leading eigenvectors of $\sum_{i=1}^{M} (\mathbf{C}_{i(N)} \cdot \mathbf{C}_{i(N)}^T)$
     Set $\mathbf{U}_{k+1}^{(N)} = \mathbf{U}^{(N)}$.

   $k = k + 1$
3. Set $\mathbf{U}_{opt}^{(1)} = \mathbf{U}_k^{(1)}$, $\mathbf{U}_{opt}^{(2)} = \mathbf{U}_k^{(2)}$, …, $\mathbf{U}_{opt}^{(N)} = \mathbf{U}_k^{(N)}$.

4. Experimental Results

The proposed GND-PCA is applied to medical MR volumes. We use eighteen MR T1-weighted 3D images (volumes) of Vanderbilt database [11]. These eighteen volumes are collected from different patients, and their dimensions are $256 \times 256 \times 26$. We choose one volume as the template and align the other seventeen volumes onto the template by

similarity-transformation based rigid registration. A 3D similarity-transformation has seven parameters, three for translations, three for rotation angles and one for scaling factor [10]. Such a registration can eliminate global difference but keep the local differences for the modeling. Three registered MR volumes are shown in Fig.3, which are used as training samples.

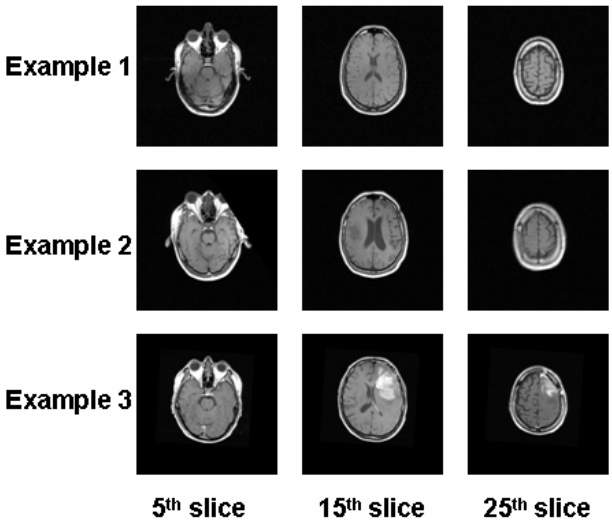


Fig.3 Examples of the registered MR volumes

The medical volumes are treated as a series of the 3rd order tensors. The leave-one-out experiment is done to test the generalization ability of GND-PCA. We use seventeen volumes as samples to learn the optimal subspaces and the left-untrained one is used as a test. In training process, the iteration is terminated when there is no dramatic change of the cost function in two consecutive times. The convergence of the training for $50\times50\times15$ mode-subspace bases is shown in Fig.4. It can be seen that the convergence is fast. Usually two times of iteration is enough.

One typical result is shown in Fig.5. The test volume is reconstructed from $50\times50\times15$ and $75\times75\times20$ mode-subspace bases, respectively. The corresponding compressing rates are 2.2% and 6.6%, respectively. It can be seen that the quality of the reconstructed images become better and better as increasing the mode-subspace bases, especially for the tumor region (the bright region in lower right). It should be noted that the training samples do not have similar tumors around that position.

In order to make a comparison, the same experiments

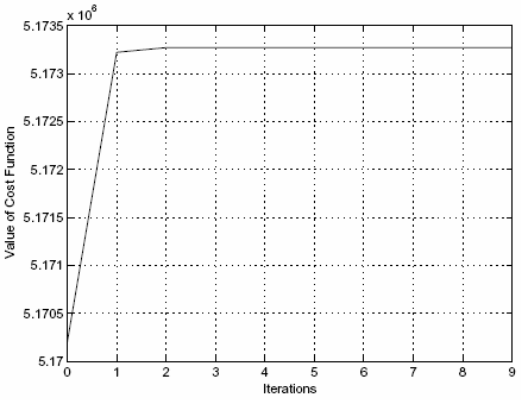


Fig.4 Convergence of GND-PCA

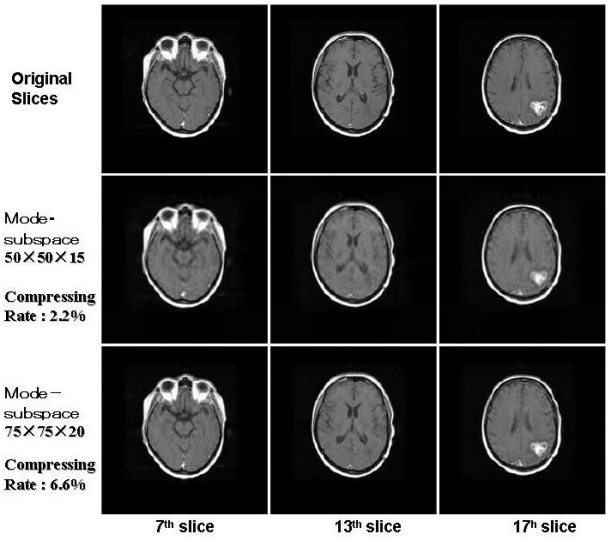


Fig.5 Reconstructed results with GND-PCA bases

are also done with classical PCA (1D-PAC) and ND-PCA. In classical 1D-PAC, the volume image should be first unfolded into a vector with a huge dimension of 1703936. So only the eigenface method [2] can be used to calculate the PCA subspace. The reconstructed result is shown in Fig.6. Since in the eigenface method, only 16 bases are available which are too few compared to the dimension of 1703936, the test volume can not be reconstructed well as shown in Fig.6. So it is clear that if the samples for training are limited, the classical PCA can not used for modeling or efficient representation of the multi-dimensional data because of its bad performance on generalization.

The reconstructed volumes by ND-PCA[5] in the leave-one-out testing experiments are shown in Fig.7. The compression rate is about 11.7%, which is corresponding to $100\times100\times20$ in GND-PCA. It can be seen that the results of ND-PCA are better than the results of classical 1D-PCA (Fig.6). But they are more blurred compared to the results of our GND-PCA (the case of $75\times75\times20$ in Fig.5).

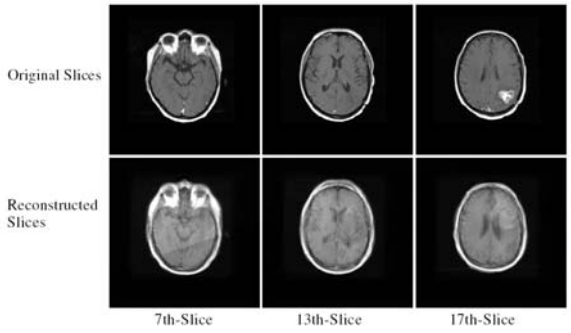


Fig.6 Reconstructed results with classical 1D-PCA bases

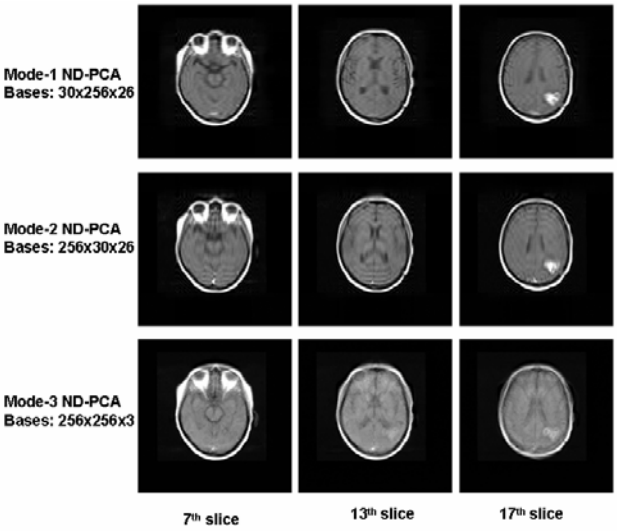


Fig.7 Reconstructed results with ND-PCA basis

In order to make a quantitative comparison, normalized correlation (*NC*) of the original volume $I(x, y, z)$ and the reconstructed volume $\hat{I}(x, y, z)$, which is defined as Eq.(9) and is used as a quantitative measure, are shown in Fig.8. The compressing rate in each method is the same (11.7%). It can be seen that the normalized correlation for GND-PCA is higher than ND-PCA and conventional PCA.

$$NC = \frac{\sum_{x,y,z} I(x, y, z) \cdot \hat{I}(x, y, z)}{\sqrt{\sum_{x,y,z} I^2(x, y, z)} \cdot \sqrt{\sum_{x,y,z} \hat{I}^2(x, y, z)}} \quad (9)$$

5. Conclusion

We proposed a novel approach called generalized *N*-dimensional principal component analysis (GND-PCA) for efficient multi-dimensional data representation and modeling. ND-PCA can be considered as a special case of GND-PCA. The effectiveness and representation ability of GND-PCA have been demonstrated by experiments on medical MR volume dataset.

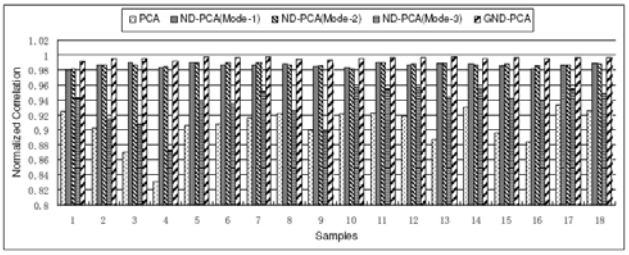


Fig.8 Comparison of results for PCA, ND-PCA, GND-PCA

This work was supported in part by the Grand-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports under the Grand No. 19500161.

References

[1] I.T. Jolliffe, Principal Component Analysis. Springer, 2002.

[2] M. Turk, A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol.3, No.1, pp. 71-86, 1991.

[3] J. Yang, et al., "To-Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition", IEEE Trans.on PAMI, Vol.26, No.1, pp. 131-137, 2004.

[4] H. Kong, et. Al., "Generalized 2D principal component analysis for face image representation and recognition ",Neural Networks, Vol.18, pp.585-594, 2005.

[5] H.C.Yu and M.Bennamoun, "1D-PCA,2D-PCA to nD-PCA", Proc. of ICPR'06,2006.

[6] L.D.Lathauwer, B.D.Moor and J.Vandewalle, "A Multilinear Singular Value Decomposition", SIAM Journal of Matrix Analysis and Application, Vol.21, No.4, pp.1253-1278, 2001.

[7] M.A.O.Vasilescu and D.Terzopoulos, "Multilinear Analysis of Image Ensembles: Tensorface", Proc. of European Conference on Computer Vision, pp.447-460,2002.

[8] M.A.O.Vasilescu and D.Terzopoulos, "Multilinear Subspace Analysis of Image Ensemble", Proc. of IEEE Conf. On Computer Vision and Pattern Recognition, Vol.2, pp.93-99, 2003.

[9] M.A.O.Vasilescu and D.Terzopoulos, "Multilinear Image Analysis for Facial Recognition", Proc. of ICPR'02, 2002

[10] R.Xu and Y.W.Chen, "Appearance models for medical volumes with few samples by generalized 3D-PCA", Lecture Notes in Computer Science, Springer, Vol.4984, pp.821-830, 2008.

[11] J. West et al., "Comparison and evaluation of retrospective intermodality brain image registration techniques," J. Comput. Assist. Tomogr., Vol.21, pp.554-566, 1997.

Dimension-Incremental Subspace Learning for High-Dimensional Data Classification

Tomoya SAKAI†

† Institute of Media and Information Technology, Chiba University
Yayoi 1–33, Inage, Chiba, 263–8522 Japan
E-mail: †tsakai@faculty.chiba-u.jp

Abstract This paper proposes novel methods for learning subspaces using dimension-incremental SVD and random sampling. The most intensive computation in the linear subspace methods is the reduction of dimensionality of the feature space by the eigen decomposition or singular value decomposition. In the present methods, the subspaces are learned by updating their orthonormal basis sets with random increment of the dimension of the feature space. The subspace learning progresses with the similarity measurement of test samples until their classification is completed. This strategy can reduce the computational expense without critical loss of recognition rate especially for the high-dimensional data, and the classification results can be assessed by observing the convergence of the similarity measures. The performance of the present methods was experimentally verified using face recognition datasets.

Key words CLAFIC, SVD, EVD, PCA, feature selection, Monte Carlo, random projection

1. Introduction

I present a substantial improvement in computation of the subspace methods by an incremental approach to the dimension of the feature space. The subspace methods [8], [12] have provided us effective techniques for applications such as optical character recognition, face recognition, and so on. The reduction in the computational costs of the subspace methods contributes to their applicable advances in the technology for large-scale and high-dimensional data.

In the linear subspace methods, the classes of given training samples are basically represented as linear subspaces spanned by the principal components of the samples in the Euclidean feature space. After learning the subspaces, test samples, i.e., queries, are classified into the classes according to (dis)similarity measures between the classes and the queries calculated with the principal components of the learned subspaces. The subspace learning by the principal component analysis is known as the reduction of dimensionality, of which computational expense is quite significant due to high dimensionality of the feature space. In particular, appearance-based vision techniques sometimes have to treat images as intolerably high-dimensional feature vectors with the pixel values in practice.

A possible solution to reduce the computational cost of the subspace learning due to the high dimensionality is incremental learning with respect to the dimension. An iterative algorithm that updates the principal components referring to the feature vectors of the training samples from low to high dimensionality can be computationally cost-effective if it can be terminated at an early iterative stage for the classification of the queries. The subspace learning with the dimensional increment of the feature space can be achieved by application of the incremental singular value decomposition (SVD) [1], [3], [4], [10], [11]. If randomly chosen dimensions are appended to the low-dimensional feature space, the (dis)similarity measures between the learned subspaces and the queries in a low-dimensional feature space are expected to approximate those between them in the high-dimensional feature space due to the same principle as the random projection [2], [5].

In this paper, I first review the traditional linear subspace methods performed in a low-dimensional feature space. Second, I propose classification methods that measure the similarity in low-dimensional feature spaces constructed by dimension increment. I call the subspace methods with random increment of the dimension the *Monte Carlo subspace methods*. Finally, I apply the Monte Carlo subspace methods to the appearance-based face recognition to show the cost effectivity.

2. Linear Subspace Methods in Low-Dimensional Feature Space

Let $\{C_l\}_{l=1}^{c}$ be a collection of classes, from each of which n_l training samples are given as d-dimensional feature vectors. A data matrix of the class C_l is defined as the matrix with the n_l feature vectors in its columns.

$$\underset{d\times n_l}{\mathbf{X}_l} := \left[\begin{array}{ccc} \mathbf{x}_1 & \dots & \mathbf{x}_{n_l} \end{array} \right], \quad l = 1, \dots, c \tag{1}$$

The linear subspace $S_l := \operatorname{span} \mathbf{X}_l \subseteqq \mathbb{E}^d$ of the class C_l is the image space of the data matrix $\mathbf{X}_l$ in the d-dimensional Euclidean feature space $\mathbb{E}^d$. A fundamental approach to determine the basis of the subspace S_l is the SVD[1)] of the data matrix as $\mathbf{X}_l = \mathbf{U}_l \mathbf{K}_l \mathbf{V}_l^\top$. Here, $\mathbf{K}_l$ is a $k_l \times k_l$ diagonal matrix with nonnegative diagonal elements, i.e., the singular values, arranged in decreasing order. $\mathbf{U}_l$ and $\mathbf{V}_l$ are respectively $d \times k_l$ and $n_l \times k_l$ matrices with orthonormal column vectors spanning the k_l-dimensional subspace S_l and a k_l-dimensional subspace $S_l^* := \operatorname{span} \mathbf{X}_l^\top \subseteqq \mathbb{E}^{n_l}$, satisfying $\mathbf{U}_l^\top \mathbf{U}_l = \mathbf{V}_l^\top \mathbf{V}_l = \mathbf{I}$.

Given the feature vector $\mathbf{q} \in \mathbb{E}^d$ of a query whose class is to be identified, most of the subspace methods geometrically evaluate the (dis)similarity between the subspaces $\{S_l\}_{l=1}^{c}$ and the query $\mathbf{q}$ in the feature space $\mathbb{E}^d$. The CLAFIC method [12], for example, measures the similarity as the squared l_2-norm of the orthogonal projection of the normalized query $\mathbf{q}/||\mathbf{q}||$ onto a subspace $\operatorname{span} \mathbf{U}$.

$$\operatorname{CLAFIC}(\mathbf{U}, \mathbf{q}) := \left|\left| \mathbf{U}^\top \frac{\mathbf{q}}{||\mathbf{q}||} \right|\right|^2 = \frac{\mathbf{q}^\top \mathbf{U}\mathbf{U}^\top \mathbf{q}}{\mathbf{q}^\top \mathbf{q}} \tag{2}$$

Here, $\mathbf{U}$ is composed of the t_l first column vectors in $\mathbf{U}_l$. One can find antecedent work, such as [9], to fix the truncated dimension t_l of the subspace S_l. The (dis)similarity measurement, however, requires the computation with the high-dimensional vectors in $\mathbb{E}^d$.

I approach this problem by choosing the basis of the feature space itself instead of the bases of the subspaces. A row of the data matrix $\mathbf{X}_l$ corresponds to a coordinate of the feature space. Let $\tilde{\mathbf{X}}_l$ be a low-dimensional data matrix consisting of r row vectors chosen from the row vectors in the data matrix $\mathbf{X}_l$, and let $\tilde{\mathbf{q}} \in \mathbb{E}^r$ be a low-dimensional feature vector of the query with r components chosen from $\mathbf{q}$ in the same manner. Then, the subspace defined as the image space of $\tilde{\mathbf{X}}_l$, i.e., $\tilde{S}_l = \operatorname{span} \tilde{\mathbf{X}}_l \subseteqq \mathbb{E}^r$, is the orthogonal projection of the subspace $S_l = \operatorname{span} \mathbf{X}_l \subseteqq \mathbb{E}^d$ onto the chosen r-dimensional feature space $\mathbb{E}^r$. The basis of $\tilde{S}_l$ is obtained by SVD of $\tilde{\mathbf{X}}_l$ as $\tilde{\mathbf{X}}_l = \tilde{\mathbf{U}}_l \tilde{\mathbf{K}}_l \tilde{\mathbf{V}}_l^\top$. Here, $\tilde{\mathbf{U}}_l$ and $\tilde{\mathbf{V}}_l$ are respectively $r \times \tilde{k}_l$ and $n_l \times \tilde{k}_l$ matrices if the dimension of $\tilde{S}_l$ is $\tilde{k}_l := \dim \tilde{S}_l = \operatorname{rank} \tilde{\mathbf{X}}_l$. The (dis)similarity between $\tilde{S}_l$ and $\tilde{\mathbf{q}}$ in the low-dimensional feature space $\mathbb{E}^r$ can be measured, for example, by the CLAFIC method as $\operatorname{CLAFIC}(\tilde{\mathbf{U}}, \tilde{\mathbf{q}})$. Here, $\tilde{\mathbf{U}}$ is composed of the $\tilde{t}_l$ first column vectors in $\tilde{\mathbf{U}}_l$. If the similarity measured in the low-dimensional feature space approximates that measured in the d-dimensional feature space, the query can be classified without referring to all features of the training samples.

1) : The compact SVD is mainly used to illustrate my method although the eigen decomposition is also available in the same way.

3. Dimension Incremental Subspace Methods

3.1 Framework

Assuming this row-incremental update algorithm, i.e., the row-incremental SVD (RiSVD), I describe a common framework of the dimension-incremental subspace methods in Algorithm 1.

Algorithm 1 Classification by dimension-incremental subspace learning

Input: the row-accessible training data matrices $\{ \underset{d\times n_l}{\mathbf{X}_l} \}_{l=1}^{c}$, and the query $\mathbf{q} \in \mathbb{E}^d$;

Output: the similarity measures $\{\tilde{g}_l\}_{l=1}^{c}$, and the learned singular value components $\{ \underset{r\times \tilde{k}_l}{\tilde{\mathbf{U}}_l} \}_{l=1}^{c}$, $\{ \underset{\tilde{k}_l\times \tilde{k}_l}{\tilde{\mathbf{K}}_l} \}_{l=1}^{c}$ and $\{ \underset{n_l\times \tilde{k}_l}{\tilde{\mathbf{V}}_l} \}_{l=1}^{c}$;

1: set $\tilde{\mathbf{q}}$ to be a zero-dimensional vector;
2: **for** all $l = 1$ to c **do**
3: set $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$ and $\tilde{\mathbf{V}}_l$ to be 0×0 matrices;
4: **end for**
5: **repeat**
6: choose the i-th dimension (disallow duplication);
7: append the i-th component of $\mathbf{q}$ to $\tilde{\mathbf{q}}$;
8: **for** all $l = 1$ to c **do**
9: set $\boldsymbol{\xi}^\top$ to be the i-th row of $\mathbf{X}_l$;
10: update $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$ and $\tilde{\mathbf{V}}_l$ by *RiSVD* using $\boldsymbol{\xi}$;
11: measure the similarity $\tilde{g}_l$ of $\tilde{\mathbf{q}}$ using $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$ and $\tilde{\mathbf{V}}_l$;
12: **end for**
13: **until** $\arg \max_{l=1,\dots,c} \tilde{g}_l$ is fixed.

We can avoid restoring the low-dimensional data matrix $\tilde{\mathbf{X}}_l$ because its SVD matrices $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$, and $\tilde{\mathbf{V}}_l$ are directly updated with the chosen row vector $\boldsymbol{\xi}^\top$ from $\mathbf{X}_l$. The low-dimensional data matrix of the class C_l is implicitly stored as $\tilde{\mathbf{X}}_l = \tilde{\mathbf{U}}_l \tilde{\mathbf{K}}_l \tilde{\mathbf{V}}_l^\top$ although the chosen row vector $\boldsymbol{\xi}^\top$ is expired after the RiSVD.

The iteration between Step 5 and Step 13 is terminated when the class with the highest similarity is settled by the similarity measurement such as the CLAFIC in the low-dimensional feature space. In case of multiple queries, the similarity measures between the classes and the queries can be calculated simultaneously in the iteration, which is terminated when the classes of all queries are identified.

The number of rows of $\tilde{\mathbf{U}}_l$, i.e., the dimension r of the low-dimensional feature space, is incremented by one after the RiSVD while the number of its columns, i.e., the dimension $\tilde{k}_l$ of the subspace $\tilde{S}_l \subsetneqq \mathbb{E}^r$, is incremented if the degeneration of $\tilde{S}_l$ is relieved by the dimension increment of the feature space. The subspace dimension $\tilde{k}_l$ can be increased up to $\operatorname{rank} \mathbf{X}_l \leqq n_l$ maintaining $\tilde{S}_l = \mathbb{E}^r$, which results in $\forall \tilde{g}_l = 1$ for any queries at the early stages. Therefore, the similarity measurement at Step 11 may be enabled after n_l iterations or after $\tilde{t}_l$ iterations in the case using the CLAFIC method.

3.2 Feature Selection

To perform effectively the subspace method in a low-dimensional feature space, we need a rule for choosing the basis of the feature space, or choosing i-th row from the data matrix $\mathbf{X}_l$ at Step 6 in Algorithm 1. The choice of the dimension is nothing more than the selection of the features. A few types of dimension-incremental subspace methods are derived by different rules of the feature selection.

a) Type-I: Monte Carlo Subspace Method by Equally Random Choice

If we do not have a priori knowledge about which rows store the important features for the similarity measures, random choice may provide us with likely measures. This strategy is well known as the Monte Carlo method. One of the advantages of the random choice is that the reliability of the classification can be tested by repeated trials using random sequences.

b) Type-II: Monte Carlo Subspace Method by Query-Dependent Random Choice

Since large components of the query contributes to the similarity measures, one would expect their faster convergence when the feature is chosen depending on the query. I design such a fast method regarding the magnitude of the query component as relative frequency of the choice. The larger query components are more likely to be chosen by this method.

c) Type-III: Query-Dependent Deterministic Choice

One can also consider the non-random choice of the dimension. For the same reason of the type-II, the dimension is chosen in decreasing order of the magnitude of the query component. Since this method does not take advantage of the random choice, the classification results cannot be assessed by repeated trials.

3.3 Row-Incremental SVD

Algorithm 2 describes the RiSVD for the dimension increment of the feature space. The RiSVD is dual to the column-incremental SVD [1], [3], [4], [10], [11] used for the data increment. Algorithm 2 ensures reconstructivity

$$\begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{K}}\tilde{\mathbf{V}}^\top \\ \boldsymbol{\xi}^\top \end{bmatrix} = \tilde{\mathbf{U}}_{\mathrm{new}}\tilde{\mathbf{K}}_{\mathrm{new}}\tilde{\mathbf{V}}_{\mathrm{new}}^\top,$$

and inductive orthonormality

$$\tilde{\mathbf{U}}_{\mathrm{new}}^\top\tilde{\mathbf{U}}_{\mathrm{new}} = \tilde{\mathbf{V}}_{\mathrm{new}}^\top\tilde{\mathbf{V}}_{\mathrm{new}} = \mathbf{I} \text{ if } \tilde{\mathbf{U}}^\top\tilde{\mathbf{U}} = \tilde{\mathbf{V}}^\top\tilde{\mathbf{V}} = \mathbf{I}.$$

3.4 Flop Count

In the usual linear subspace methods, the cost of computing the subspace basis for a class C_l by the SVD of the $d \times n_l$ training data matrix $\mathbf{X}_l$ is $O((d+n_l)\min^2(d,n_l)) \approx O(dn_l^2)$ $(d \gg n_l)$ flops [4], [7], and the similarity measurement, by the CLAFIC for example, costs $O(dt_l)$ a class. On the other hand, if Algorithm 1 requires $r_{\max}$ iterations, the SVD of $\mathbf{B}$ in Algorithm 2 costs $O(r_{\max}\tilde{k}_l^3)$, and the matrix multiplication at Step 12 or 18 and at Step 14 or 20 costs

Algorithm 2 Row-incremental SVD

Input: $\underset{r\times\tilde{k}}{\tilde{\mathbf{U}}}$, $\underset{\tilde{k}\times\tilde{k}}{\tilde{\mathbf{K}}}$, $\underset{n\times\tilde{k}}{\tilde{\mathbf{V}}}$ $(r \geqq \tilde{k})$ and $\boldsymbol{\xi} \in \mathbb{E}^n$;

Output: $\tilde{\mathbf{U}}_{\mathrm{new}}$, $\tilde{\mathbf{K}}_{\mathrm{new}}$ and $\tilde{\mathbf{V}}_{\mathrm{new}}$;

1: **if** $r = 0$ **then**

2: $\quad \underset{1\times1}{\tilde{\mathbf{U}}_{\mathrm{new}}} \leftarrow \begin{bmatrix} 1 \end{bmatrix}$;

3: $\quad \underset{1\times1}{\tilde{\mathbf{K}}_{\mathrm{new}}} \leftarrow \begin{bmatrix} ||\boldsymbol{\xi}|| \end{bmatrix}$;

4: $\quad \underset{n\times1}{\tilde{\mathbf{V}}_{\mathrm{new}}} \leftarrow \begin{bmatrix} \frac{\boldsymbol{\xi}}{||\boldsymbol{\xi}||} \end{bmatrix}$;

5: **end if**

6: $\boldsymbol{\eta} \leftarrow \tilde{\mathbf{V}}^\top\boldsymbol{\xi}$;

7: $\mathbf{p} \leftarrow \boldsymbol{\xi} - \tilde{\mathbf{V}}\boldsymbol{\eta}$;

8: $p \leftarrow ||\mathbf{p}||$;

9: **if** $p \neq 0$ **then**

10: $\quad \underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{B}} \leftarrow \begin{bmatrix} \tilde{\mathbf{K}} & \mathbf{0} \\ \boldsymbol{\eta}^\top & p \end{bmatrix}$;

11: $\quad$ do singular value decomposition of $\mathbf{B}$ to obtain $\underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{U}_{\mathrm{B}}}$, $\underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{K}_{\mathrm{B}}}$ and $\underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{V}_{\mathrm{B}}}$ such that $\mathbf{U}_{\mathrm{B}}\mathbf{K}_{\mathrm{B}}\mathbf{V}_{\mathrm{B}}^\top = \mathbf{B}$ and $\mathbf{U}_{\mathrm{B}}^\top\mathbf{U}_{\mathrm{B}} = \mathbf{V}_{\mathrm{B}}^\top\mathbf{V}_{\mathrm{B}} = \underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{I}}$;

12: $\quad \underset{(r+1)\times(\tilde{k}+1)}{\tilde{\mathbf{U}}_{\mathrm{new}}} \leftarrow \begin{bmatrix} \tilde{\mathbf{U}} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}\mathbf{U}_{\mathrm{B}}$;

13: $\quad \underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\tilde{\mathbf{K}}_{\mathrm{new}}} \leftarrow \mathbf{K}_{\mathrm{B}}$;

14: $\quad \underset{n\times(\tilde{k}+1)}{\tilde{\mathbf{V}}_{\mathrm{new}}} \leftarrow \begin{bmatrix} \tilde{\mathbf{V}} & \frac{\mathbf{p}}{p} \end{bmatrix}\mathbf{V}_{\mathrm{B}}$;

15: **else**

16: $\quad \underset{(\tilde{k}+1)\times\tilde{k}}{\mathbf{B}} \leftarrow \begin{bmatrix} \tilde{\mathbf{K}} \\ \boldsymbol{\eta}^\top \end{bmatrix}$;

17: $\quad$ do singular value decomposition of $\mathbf{B}$ to obtain $\underset{(\tilde{k}+1)\times\tilde{k}}{\mathbf{U}_{\mathrm{B}}}$, $\underset{\tilde{k}\times\tilde{k}}{\mathbf{K}_{\mathrm{B}}}$ and $\underset{\tilde{k}\times\tilde{k}}{\mathbf{V}_{\mathrm{B}}}$ such that $\mathbf{U}_{\mathrm{B}}\mathbf{K}_{\mathrm{B}}\mathbf{V}_{\mathrm{B}}^\top = \mathbf{B}$ and $\mathbf{U}_{\mathrm{B}}^\top\mathbf{U}_{\mathrm{B}} = \mathbf{V}_{\mathrm{B}}^\top\mathbf{V}_{\mathrm{B}} = \underset{\tilde{k}\times\tilde{k}}{\mathbf{I}}$;

18: $\quad \underset{(r+1)\times\tilde{k}}{\tilde{\mathbf{U}}_{\mathrm{new}}} \leftarrow \begin{bmatrix} \tilde{\mathbf{U}} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}\mathbf{U}_{\mathrm{B}}$;

19: $\quad \underset{\tilde{k}\times\tilde{k}}{\tilde{\mathbf{K}}_{\mathrm{new}}} \leftarrow \mathbf{K}_{\mathrm{B}}$;

20: $\quad \underset{n\times\tilde{k}}{\tilde{\mathbf{V}}_{\mathrm{new}}} \leftarrow \underset{n\times\tilde{k}}{\tilde{\mathbf{V}}}\,\underset{\tilde{k}\times\tilde{k}}{\mathbf{V}_{\mathrm{B}}}$;

21: **end if**

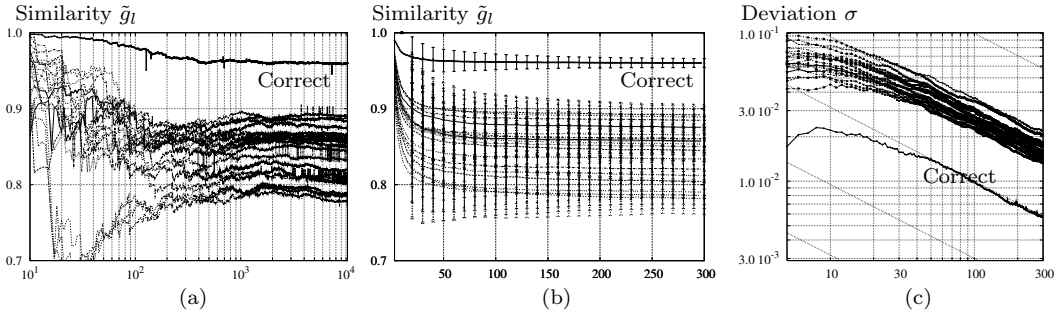


Fig. 1 Similarity measures vs dimension of feature space by the type-I Monte Carlo subspace method. In all three graphs, the horizontal axis is the iteration count r, or the reduced dimension of the feature space. (a) An example of the progress of the similarity measures. (b) Average progress of the similarity measures. The error bar indicates the standard deviation σ. (c) Evolution of σ with respect to r shows the $r^{-\frac{1}{2}}$-asymptotics (the oblique dashed lines).

$O(r_{\max}^2 \tilde{k}_l^2 + r_{\max} n_l \tilde{k}_l^2)$. The similarity measure can be done in $O(r_{\max}^2 \tilde{t}_l)$ flops. Since the total cost is approximately $O(r_{\max}^2)$, the dimension-incremental subspace methods can reduce the computation time if it completes the classification in $r_{\max} \sim O(d^{\frac{1}{2}})$ iterations. Besides, Algorithm 1 can save memory space by out-of-core computation because the input data matrices may be row-accessible.

4. Experiments

The cost-effectiveness for high-dimensional data classification is demonstrated by the appearance-based image recognition using a pre-cropped version of the UMIST face database [6]. A partial set of face images consisting of $n_l = 8$ images each of 20 individuals is used as the training dataset of a class, and the remaining images are used as the queries. The dimension of the feature space is $d = 112 \times 92 = 10304$.

Figure 1(a) shows an example of the progress of the similarity measures $\tilde{g}_l = \mathrm{CLAFIC}(\tilde{\mathbf{U}}_l, \mathbf{q})$ between the 20 classes and a query by the type-I method. The significance of the similarity between the correct class and the query becomes apparent as the dimension of the low-dimensional feature space grows. Figure 1(b) shows the average progress with over 500 trials. Remarkably, only a few dozen times of dimension increment are sufficient to clarify the correct class for the query. The computation time is reduced to 18% of the usual CLAFIC method in case of $r_{\max} = 50$ and to 56% in case of $r_{\max} = 100$. Although calculating the precise similarity requires numerous iterations because the similarity measures slowly converge with order 1/2 as shown in Fig. 1(c), the class with the highest similarity can be determined by the measurement in the low-dimensional feature spaces.

The appearance-based image recognition without any normalisation can be highly dependent on the positions of target objects in the images. It is well-known that the Fourier amplitude, or the power spectrum, is invariant under any spatial shift of the images. I have tested the dimension-incremental subspace methods for the Fourier transform of the UMIST images. As shown in Fig. 2(a), the type-I method could find a correct class in average. However, the similarity measures have large deviations, indicating low reliablity of the classification results at low dimensions. Since every similarity $\tilde{g}_l$ is greater than about 0.92, the differences between the similarity measures are relatively small. Nevertheless, the type-II method could identify the correct class at a few tens of dimensions as shown in Fig. 2(b). This implies that the query-dependent random choice is effective for improving the precision of the similarity measures at low dimensions. Among the present methods, the type-III method shows the fastest convergence as shown in Fig. 2(c).

5. Concluding Remarks

The dimension-incremental approach to the subspace learning allows us to measure the similarity between the classes and queries in low-dimensional feature spaces. The present methods achieve considerable reduction in computation time of the classification, which makes tractable the pattern recognition for large dimensional data. Another distinctive feature of the present methods is that we can observe the progress of the similarity measures with respect to the dimension, and evaluate the reliability of the classification results. The further research on the Monte Carlo scheme for the dimensionality reduction should be pursued from the viewpoint of the random projection [2], [5].

References

[1] M. Brand. Fast online svd revisions for lightweight recommender systems. In *Proc. SDM 2003*, 2003.

[2] E. Brigham and H. Maninila. Random projection in dimensionality reduction: applications to image and text data. In *ACM SIGKDD ICKDDM*, pages 245–250, 2001.

[3] J. R. Bunch and C. P. Nielsen. Updating the singular value decomposition. *Numer. Math.*, 31:111–129, 1978.

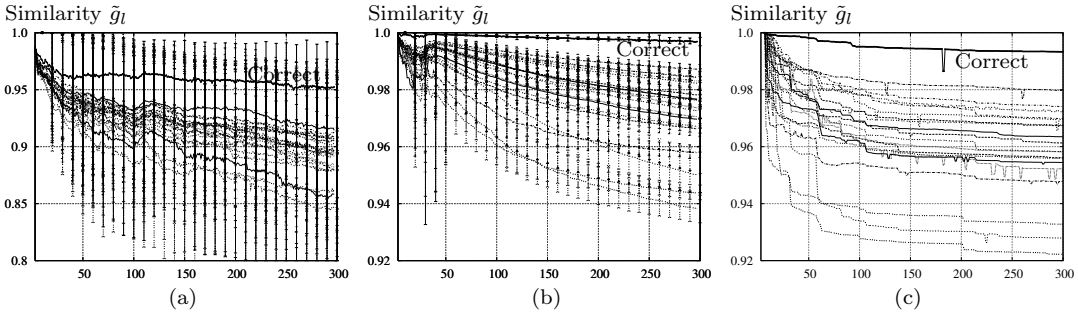


Fig. 2 Same as Fig. 1(b), but for the Fourier transformed data evaluated by (a) the type-I method, (b) the type-II method and (c) the type-III method.

[4] S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image analysis. *Graphical Models and Image Processing*, 59(5):321–332, 1997.

[5] D. Fradkin and D. Madigan. Experiments with random projections for machine learning. In *ACM SIGKDD ICKDDM*, pages 517–522, 2003.

[6] D. B. Graham and N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. In *Face Recognition: From Theory to Applications*, volume 163, pages 446–456. NATO ASI Series F, Computer and Systems Sciences, 1998.

[7] M. Gu and S. C. Eisenstat. A stable and fast algorithm for updating the singular value decomposition. In *Tech. Rep. YALEU/DCS/RR-966*. Yale University, 1994.

[8] T. Iijima, H. Genchi, and K. Mori. A theory of character recognition by pattern matching method. In *Proc. 1st IJCPR*, pages 50–56, October 1973.

[9] J. Laaksonen and E. Oja. Subspace dimension selection and averaged learning subspace method in handwritten digit classification. In *Proc. ICANN 1996*, pages 227–232, 1996.

[10] H. Murase and M. Lindenbaum. Partial eigenvalue decomposition of large images using spatial temporal adaptive method. *IEEE Trans. IP*, 4(5):620–629, May 1995.

[11] D. Skocaj and A. Leonardis. Incremental and robust learning of subspace representations. *Image and Vision Computing*, 26:27–38, 2008.

[12] S. Watanabe, P. F. Lambert, C. A. Kulikowski, J. L. Buxton, and R. Walker. Evaluation and selection of variables in pattern recognition. In *Computer and Information Sciences II*. Academic Press, New York, 1967.

抑制付きカーネル部分空間法

鷲沢　嘉一

† (独) 理化学研究所 脳科学総合研究センター　〒 351–0198 埼玉県和光市広沢２－１
E-mail: †washizawa@brain.riken.jp

あらまし　CLAFIC 法などの主な部分空間法では，対象クラスの標本のみから入力パターンとクラスの類似度を測る関数を求める．このため，識別器の独立性が高いという利点があるが，複数のクラスに同じ特徴がある場合は，識別に有用な特徴が抽出できない．そこで，他のクラスの特徴を抑制する手法がいくつか提案されている．CLAFIC 法の他の拡張として，カーネルトリックを適用したカーネル部分空間法がある．カーネルトリックは標本を高次元空間へ写像し，問題を高次元の空間から標本数次元の双対問題へと変換することにより解を導く手法である．しかしながら，他のクラスの特徴を抑制する手法にカーネルトリックを適用する場合，全クラスの標本数の行列の演算が必要であり，クラス数が大きいと計算量が非常に大きくなる問題がある．本研究では，解の作用素の空間を制限することで計算量を削減し，現実的な抑制付きカーネル部分空間法を提案する．
キーワード　CLAFIC 法，カーネル部分空間法，カーネルトリック

Kernel Subspace methods with suppression

Yoshikazu WASHIZAWA

† Brain Science Institute, RIKEN
2–1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan
E-mail: †washizawa@brain.riken.jp

Abstract　We propose efficient kernel subspace methods with supression. The experimental results demonstorate advantages of the proposed method.
Key words　CLAFIC, kernel subspace methods, kernel trick

1. はじめに

CLAFIC 法を始めとする部分空間法は，その類似度を測る類似度関数が各クラス標本のみから作成できるという特長がある．このため，文字認識や顔画像などのクラス数が非常に多い問題でも識別器の設計が容易に出来る．また，棄却判定やクラス数の増減への対応，１つのクラスを複数のサブクラスに分割するマルチテンプレート識別，１つの標本が複数のクラスに属するマルチクラス問題などにも柔軟に対応できる．

しかし，複数のクラスのパターンが同様の特徴を有する場合，類似度関数はその特徴を取り出してしまい，クラス間の有意な特徴差を取り出すことが出来ないことがある．このため，混合類似度法や相対 KL 変換法，競合部分空間を考慮した学習部分空間法 (LSM, learning subspace method) などの競合クラスの特徴を抑制しながら，対象クラスの特徴を取り出す手法がいくつか提案されている [1]〜[3]．混合類似度法は，対象クラスの類似クラスの平均パターンを用いる手法であり，相対 KL 変換法は，類似クラスの相関行列を用いる．競合クラスの特徴を抑制するための最も単純な手法として，部分空間の計算時に対象クラスの相関行列 R を固有値分解する代わりに，全クラスの相関行列 Q と正のパラメータ β を用いて，$R-\beta Q$ の固有空間を用いる手法が考えられる．この意味については次節で述べる．以降，この手法を抑制付き部分空間法と呼ぶ．

部分空間法の別の拡張として，カーネルトリックを用いたカーネル部分空間法がある [4], [5]．カーネルトリックは標本を高次元の特徴空間へ非線形写像し，問題を特徴空間の主問題から標本数空間の双対問題へ変換することにより，現実的な計算量で解を求める手法である [6]．CLAFIC 法にカーネルトリックを適用する場合は，各クラスの標本数の大きさの行列演算で解を求めることができる．しかしながら，抑制付き部分空間法では，対象クラスと競合クラス，あるいは全クラスの標本数の行列演算をする必要がある．このため，文字認識などのクラス数が非常に多い場合，計算コストが非現実的に大きくなる．例えば，アルファベットの認識で，各文字につき 1,000 文字の標本を用意した場合，カーネル部分空間法では，1,000x1,000 の行列の固有値分解と行列演算で類似度関数を設計することが出

来るが，抑制付きの場合は，26,000x26,000 の行列演算が必要となり，現実的な計算が難しい．[2], [7], [8] などは，抑制付きの部分空間法にカーネルトリックを適用したものであるが，あらかじめ，類似する標本を選んでおき，計算量を削減している．

本研究では，類似度関数の集合に制約を加えることにより，対象クラスの標本数の行列演算のみで設計できる抑制付きのカーネル部分空間法を提案する．また，手書き数字認識実験で提案手法の有効性を示す．

2. CLAFIC 法と抑制付き部分空間法

2.1 CLAFIC 法 [9], [10]

クラス数を c，入力次元数を d，入力パターンを $\boldsymbol{x}$ とすると CLAFIC 法の識別関数 $f(\boldsymbol{x})$ は，

$$f(\boldsymbol{x}) = \underset{i=1,\ldots,c}{\operatorname{argmax}}\, g_i(\boldsymbol{x}) \tag{1}$$

$$g_i(\boldsymbol{x}) = \langle \boldsymbol{x}, P_i\boldsymbol{x}\rangle = \|P_i\boldsymbol{x}\|^2 \tag{2}$$

で与えられる．ここで，$\langle\cdot,\cdot\rangle$ は内積，$\|\cdot\|$ は l_2 ノルムを表し，$g_i(\boldsymbol{x})$ はクラス i の類似度関数，P_i はクラス i の Karhunen-Loéve(KL) 部分空間への正射影である．

P_i は，以下の最適化問題の最小解として特徴付けされる．

$$\begin{array}{ll}\min\limits_{Y\in\mathbb{R}^{d\times d}} & \underset{\boldsymbol{x}\in\Omega_i}{E}\|\boldsymbol{x}-Y\boldsymbol{x}\|^2 \\ \text{subject to} & \operatorname{rank}(Y)\leqq r\end{array} \tag{3}$$

または，

$$\begin{array}{ll}\max\limits_{Y\in\mathbb{R}^{d\times d}} & \underset{\boldsymbol{x}\in\Omega_i}{E}\|Y\boldsymbol{x}\|^2 \\ \text{subject to} & \operatorname{rank}(Y)\leqq r,\ Y^\top = Y,\ YY = Y\end{array} \tag{4}$$

ここで，Ω_i はクラス i のパターンの集合である．最適化問題 (4) の 2 番目と 3 番目の制約は Y が正射影であることを表している．解は，クラス i のパターン $\boldsymbol{x}\in\Omega_i$ の KL 部分空間への正射影となり，例えば相関行列 $R_i = E_{\boldsymbol{x}\in\Omega_i}\boldsymbol{x}\boldsymbol{x}^\top$ の大きい r 個の固有値に対応する固有ベクトルなどから求めることができる．

2.2 抑制付き部分空間法

CLAFIC 法は，対象クラスの特徴のみから類似度関数を計算する．このため，複数のクラスに同じ大きい特徴が含まれている場合，その特徴のみを取り出して，有意な特徴差を取り出さないという欠点がある．ここで，対象クラスに対して，同じ特徴を持つクラスを競合クラスと呼ぶ．これより，類似度関数の設計時に，競合クラスの特徴を抑制しながら，対象クラスの特徴を取り出すよう基準で設計すれば識別性能の向上が期待できる．

式 (2) の類似度関数は，射影ベクトルのノルムであるが，P_i は正射影であるため，投影距離を用いた類似度関数

$$g_i'(\boldsymbol{x}) = -\|\boldsymbol{x}-P_i\boldsymbol{x}\|^2 \tag{5}$$

も等価である．最適化問題 (3) より，行列 P_i は対象クラスのパターン $\boldsymbol{x}$ と $P_i\boldsymbol{x}$ の距離が平均的に小さくなるような行列を求めていることになる．競合クラスの特徴を抑制するには，競合クラスのパターン $\boldsymbol{y}$ と $P_i\boldsymbol{y}$ の距離が平均的に大きくなればよい．そこで，競合クラスのパターンの集合を Ψ とおき，以下の最適化問題を考える．

$$\begin{array}{ll}\min\limits_{Y\in\mathbb{R}^{d\times d}} & \underset{\boldsymbol{x}\in\Omega_i}{E}\|\boldsymbol{x}-Y\boldsymbol{x}\|^2 - \beta\underset{\boldsymbol{y}\in\Psi}{E}\|\boldsymbol{y}-Y\boldsymbol{y}\|^2 \\ \text{subject to} & \operatorname{rank}(Y)\leqq r\end{array} \tag{6}$$

目的関数の第 2 項が競合クラスを抑制する項である．

最適化問題 (3) の目的関数を J_3 とおくと，

$$\begin{aligned} J_3 &= \underset{\boldsymbol{x}\in\Omega_i}{E}\|\boldsymbol{x}-Y\boldsymbol{x}\|^2 \\ &= \underset{\boldsymbol{x}\in\Omega_i}{E}\operatorname{Trace}[\boldsymbol{x}\boldsymbol{x}^\top - 2Y\boldsymbol{x}\boldsymbol{x}^\top + Y\boldsymbol{x}\boldsymbol{x}^\top Y^\top] \\ &= \operatorname{Trace}[R_i - 2YR_i + YR_iY^\top]\end{aligned}$$

となるのに対し，最適化問題 (6) の目的関数を J_5 とおくと同様に，

$$J_5 = \operatorname{Trace}[(R_i-\beta Q) - 2Y(R_i-\beta Q) + Y(R_i-\beta Q)Y^\top]$$

となる．ここで，Q は $\boldsymbol{y}\in\Psi$ の相関行列 $Q = \underset{\boldsymbol{x}\in\Psi}{E}\boldsymbol{y}\boldsymbol{y}^\top$ である．β は十分に小さく，$(R_i-\beta Q)$ の固有値はすべて非負であるとすると，最小解は $R_i-\beta Q$ 固有値分解して求まる固有空間への正射影となる．

2.3 正則化抑制付き部分空間法

抑制付き部分空間法を行う場合，$R-\beta Q$ が非負である必要があった．もし，$R-\beta Q$ に負の固有値がある場合には，最適化問題 (6) の目的関数が負の無限大となってしまい，類似度関数としての役割を果たさない．これを防ぐために正則化を導入することができる．部分空間法に正則化を導入した手法は，正則化 2 次識別法，ハイブリッド法と呼ばれる [11], [12].

以下の最適化問題を考える．

$$\begin{array}{ll}\min\limits_{Y\in\mathbb{R}^{d\times d}} & \underset{\boldsymbol{x}\in\Omega_i}{E}\|\boldsymbol{x}-Y\boldsymbol{x}\|^2 - \beta\underset{\boldsymbol{y}\in\Psi}{E}\|\boldsymbol{y}-Y\boldsymbol{y}\|^2 + \mu\|Y\|_F^2 \\ \text{subject to} & \operatorname{rank}(Y)\leqq r\end{array} \tag{7}$$

ここで，$\mu>0$ は正則化パラメータ，$\|\cdot\|_F$ は Frobenius ノルムを表す．

正則化を用いる場合は，ランク制約がない場合でも類似度関数を構成することができる．ランク制約がない場合 ($r=d$ の場合)，最適化問題 (7) の目的関数 J_7 は

$$\begin{aligned} J_7 &= \operatorname{Trace}[R_i - YR_i - R_iY^\top + YR_iY^\top \\ &\quad -\beta(Q - YQ - QY^\top + YQY^\top) + \mu YY^\top] \\ &= \operatorname{Trace}[Y(R_i-\beta Q+\mu I)Y^\top - Y(R_i-\beta Q) \\ &\quad -(R_i-\beta Q)Y^\top + (R_i-\beta Q)]\end{aligned} \tag{8}$$

となる．ここで I は単位行列を表す．このとき，J_7 の Y における Z 方向への Gâteaux 微分は

$$\begin{aligned} &\delta J_7(Y;Z) \\ &= \lim_{\delta\to 0}\frac{1}{\delta}\operatorname{Trace}[J_7(Y+\delta Z) - J_7(Y)] \\ &= 2\lim_{\delta\to 0}\frac{1}{\delta}\operatorname{Trace}[(\delta Z)\Big((R_i-\beta Q+\mu I)Y^\top - (R_i-\beta Q)\Big)]\end{aligned}$$

となる．$R_i - \beta Q + \mu I$ が正則となるように β, μ をとれば，任意の方向 Z に対して，$\delta J_7(Y;Z) = 0$ となるのは，

$$Y = (R_i - \beta Q)(R_i - \beta Q + \mu I)^{-1} \tag{9}$$

となるときである．$\beta = 0$ のときは，Y は正則化２次識別法の解となる．

ランク制約がある場合は，式 (8) より，

$$\begin{aligned} J_7 \;=\; & \mathrm{Trace}[\Big(Y(R_i - \beta Q + \mu I)^{1/2} \\ & -(R_i - \beta Q)(R_i - \beta Q + I)^{-1/2}\Big) \\ & \Big((R_i - \beta Q + \mu I)^{1/2} Y \\ & -(R_i - \beta Q + I)^{-1/2}(R_i - \beta Q)\Big)] \\ & -\mathrm{Trace}[(R_i - \beta Q)(R_i - \beta Q + \mu I)(R_i - \beta Q)] \\ & +\mathrm{Trace}[(R_i - \beta Q)] \\ =\; & \|Y(R_i - \beta Q + \mu I)^{\frac{1}{2}} - (R_i - \beta Q)(R_i - \beta Q + I)^{\frac{-1}{2}}\|_F^2 \\ & -\mathrm{Trace}[(R_i - \beta Q)(R_i - \beta Q + \mu I)(R_i - \beta Q)] \\ & +\mathrm{Trace}[(R_i - \beta Q)] \end{aligned} \tag{10}$$

第２項と第３項は Y に関係しないため，第１項を J_7' とおく．

$$J_7' = \|Y(R_i - \beta Q + \mu I)^{\frac{1}{2}} - (R_i - \beta Q)(R_i - \beta Q + \mu I)^{-\frac{1}{2}}\|_F^2$$

ここで，$R - \beta Q$ の固有値分解を

$$R - \beta Q = \sum_{i=1}^{d} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \tag{11}$$

とし，固有値は降順に並んでいるとする．また，μ は最小固有値よりも大きくとる．

$$(R_i - \beta Q)(R_i - \beta Q + \mu I)^{-\frac{1}{2}} = \sum_{i=1}^{d} \frac{\lambda_i}{\sqrt{\lambda_i + \mu}} \boldsymbol{u}_i \boldsymbol{u}_i^\top \tag{12}$$

であり，$R_i - \beta Q + \mu I$ が正則であるため，J_7' が最小となるときは

$$\begin{aligned} Y(R_i - \beta Q + \mu I)^{\frac{1}{2}} \;&=\; \sum_{i=1}^{r} \frac{\lambda_i}{\sqrt{\lambda_i + \mu}} \boldsymbol{u}_i \boldsymbol{u}_i^\top \\ Y \;&=\; \sum_{i=1}^{r} \frac{\lambda_i}{\lambda_i + \mu} \boldsymbol{u}_i \boldsymbol{u}_i^\top \end{aligned} \tag{13}$$

となるときである．

3. カーネル部分空間法と抑制付きカーネル部分空間法

3.1 カーネル部分空間法

カーネル部分空間法では，入力 $\boldsymbol{x}$ は非線形写像 Φ により入力次元よりも高次元の空間 $\mathcal{F}$ に写像される．

$$\Phi : \mathbb{R} \to \mathcal{F} : \boldsymbol{x} \mapsto \Phi(\boldsymbol{x}) \tag{14}$$

このとき，写像 Φ を明示的に定義するのではなく，Mercer カーネルと呼ばれる以下の条件を満たす関数を用いて計算を行う．

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle \Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2) \rangle \tag{15}$$

代表的な Mercer カーネルとして以下のものが知られている．

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle \tag{16}$$

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; (\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle + c)^d \tag{17}$$

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; \exp(-c\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2) \tag{18}$$

ここで，実数 c と自然数 d はパラメータである．

L_i をクラス i の標本数とし，クラス i の標本を $\boldsymbol{x}_1^i, \ldots, \boldsymbol{x}_{L_i}^i$ と表す．高次元特徴空間 $\mathcal{F}$ に写像されたパターンの標本相関行列 R_i^Φ は

$$R_i^\Phi = \frac{1}{L_i} \sum_{j=1}^{L_i} \Phi(\boldsymbol{x}_j^i) \Phi(\boldsymbol{x}_j^i)^\top \tag{19}$$

で与えられる．カーネル関数に式 (18) の Gauss カーネルを用いた場合は，$\mathcal{F}$ が無限次元関数空間となり，$\Phi(\boldsymbol{x}_j^i)$ が関数となるため，転置を Neumann-Schatten 積 $\Phi(\boldsymbol{x}_j^i) \otimes \overline{\Phi(\boldsymbol{x}_j^i)}$ あるいはケットブラ表記 $|\Phi(\boldsymbol{x}_j^i)\rangle\langle\Phi(\boldsymbol{x}_j^i)|$ に読み替えればよい．これ以降，簡便のため，クラスを表す添え字 i は省略する．

CLAFIC 法では，相関行列 R を固有値分解して射影行列を求めたが，R^Φ は，非常に高次元であるため，現実的に固有値分解を計算することは難しい．そこで，Φ で写像された標本 $\Phi(\boldsymbol{x}_1), \ldots \Phi(\boldsymbol{x}_L)$ を並べた行列あるいは作用素 S を考える．

$$S = [\Phi(\boldsymbol{x}_1), \ldots, \Phi(\boldsymbol{x}_L)] = \sum_{j=1}^{L} \Phi(\boldsymbol{x}_j) \boldsymbol{e}_j^\top \in \mathcal{B}(\mathbb{R}^L, \mathcal{F}) \tag{20}$$

ここで，$\boldsymbol{e}_j$ は，第 j 要素が 1 でそれ以外が零の L 次元ベクトルを表し，$\mathcal{B}(\mathbb{R}^L, \mathcal{F})$ は，$\mathbb{R}^L$ から $\mathcal{F}$ への線形作用素全体を表す．S を用いると R^Φ は，

$$R^\Phi = \frac{1}{L} S S^\top \tag{21}$$

と表される．先ほどと同様に $\mathcal{F}$ が無限次元関数空間となる場合は，転置を随伴作用素 S^* に置き換えればよい．S の特異値分解を

$$S = \sum_{k=1}^{L'} \lambda_k \boldsymbol{u}_k \boldsymbol{v}_k^\top = U \Lambda V^\top \tag{22}$$

とおく．ここで，L' は S のランクであり，$U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{L'}]$, $V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{L'}]$ である．$SS^\top = U\Lambda^2 U^\top$ より，R^Φ の固有ベクトルは $\boldsymbol{u}_i$, $i = 1, \ldots, L'$ である．式 (22) の両辺の右から $V\Lambda^{-1}$ を掛けると

$$U \;=\; S V \Lambda^{-1} \tag{23}$$

$$\boldsymbol{u}_k \;=\; \frac{1}{\lambda_k} S \boldsymbol{v}_k \tag{24}$$

の関係が得られる．$\boldsymbol{v}_k$ は，カーネル Gram 行列 $S^\top S \in \mathbb{R}^{L \times L}$ の固有ベクトルであるため，標本数の大きさの行列を固有値分解すれば式 (24) の関係から，R^Φ の固有ベクトルが求められる．

カーネル Gram 行列の i,j 要素は $k(\boldsymbol{x}_i,\boldsymbol{x}_j)$ で計算ができる．

類似度関数は，

$$g(\boldsymbol{x}) = \|P\Phi(\boldsymbol{x})\|^2 \tag{25}$$

$$P = \sum_{k=1}^{r} \boldsymbol{u}_k\boldsymbol{u}_k^\top \tag{26}$$

であるため，式 (24) を代入すると

$$g(\boldsymbol{x}) = \langle \boldsymbol{h}(\boldsymbol{x}), \left(\sum_{k=1}^{r}\frac{1}{\lambda_k^2}\boldsymbol{v}_k\boldsymbol{v}_k^\top\right)\boldsymbol{h}(\boldsymbol{x})\rangle \tag{27}$$

$$= \sum_{k=1}^{r}\langle \boldsymbol{h}(\boldsymbol{x}), \frac{1}{\lambda_k}\boldsymbol{v}_k\rangle^2 \tag{28}$$

で与えられる．

ここで，$\boldsymbol{h}(\boldsymbol{x}) = S^\top\Phi(\boldsymbol{x}) = [k(\boldsymbol{x}_1,\boldsymbol{x}),\ldots,k(\boldsymbol{x}_L,\boldsymbol{x})]^\top \in \mathbb{R}^L$ は経験カーネル写像と呼ばれるベクトルである [6]．

射影作用素 P は，以下の最適化問題で特徴付けをすることができる．

$$\begin{aligned}\min_{Y\in\mathcal{B}(\mathcal{F})} \quad & \frac{1}{L}\sum_{j=1}^{L}\|\Phi(\boldsymbol{x}_j) - Y\Phi(\boldsymbol{x}_j)\|^2 \\ \text{subject to} \quad & \mathrm{rank}(Y)\leqq r,\ \mathcal{N}(Y)\supset\mathcal{R}(S)^\perp\end{aligned} \tag{29}$$

ここで，$\mathcal{N}(\cdot)$ は核空間，$\mathcal{R}(\cdot)$ は値域を表す．

3.2 抑制付きカーネル部分空間法

抑制付きカーネル部分空間法は，最適化問題 (7) に非線形写像 $\Phi(\cdot)$ を適用した以下の最適化問題

$$\begin{aligned}\min_{Y\in\mathcal{B}(\mathcal{F})} \quad & \frac{1}{L}\sum_{j=1}^{L}\|\Phi(\boldsymbol{x}_j) - Y\Phi(\boldsymbol{x}_j)\|^2 \\ & -\beta\frac{1}{M}\sum_{j=1}^{M}\|\Phi(\boldsymbol{y}_j) - Y\Phi(\boldsymbol{y}_j)\|^2 \\ & +\mu\|Y\|_F^2 \\ \text{subject to} \quad & \mathrm{rank}(Y)\leqq r, \mathcal{N}(Y)\supset\mathcal{R}(T)^\perp\end{aligned} \tag{30}$$

の解で与えられる．ここで，

$$\begin{aligned}T &= [\Phi(\boldsymbol{x}_1),\ldots,\Phi(\boldsymbol{x}_L)\Phi(\boldsymbol{y}_1),\ldots,\Phi(\boldsymbol{y}_M)] \\ &= \sum_{j=1}^{L}\Phi(\boldsymbol{x}_j)\boldsymbol{e}_j^\top + \sum_{j=1+L}^{L+M}\Phi(\boldsymbol{y}_{j-L})\boldsymbol{e}_j^\top,\end{aligned} \tag{31}$$

$\boldsymbol{e}_j \in \mathbb{R}^{L+M}$ である．

これをカーネル部分空間法と同様に計算すると，高次元特徴空間の主問題は，対象クラスの標本数 L と競合クラスの標本数 M の重複を省いた和，N 次元の双対問題となる．前述の通り，クラス数が非常に大きい (数十〜数百) 場合は，L が現実的な大きさ ($\sim 10{,}000$) でも N がその数十倍から数百倍となり，計算ができない．このため，解の範囲を Y の値域の直交補空間 $\mathcal{R}(Y)$ が対象クラスの標本が高次元特徴空間 $\mathcal{F}$ で張る空間 $\mathcal{R}(S)$ の直交補空間を含むような作用素であり，かつ，Y の値域が $\mathcal{R}(S)$ に含まれるような作用素に限り，最適化問題を解く．

$$\begin{aligned}\min_{Y\in\mathcal{B}(\mathcal{F})} \quad & \frac{1}{L}\sum_{j=1}^{L}\|\Phi(\boldsymbol{x}_j) - Y\Phi(\boldsymbol{x}_j)\|^2 \\ & -\beta\frac{1}{M}\sum_{j=1}^{M}\|\Phi(\boldsymbol{y}_j) - Y\Phi(\boldsymbol{y}_j)\|^2 \\ & +\mu\|Y\|_F^2 \\ \text{subject to} \quad & \mathrm{rank}(Y)\leqq r,\ \mathcal{N}(Y)\supset\mathcal{R}(S)^\perp,\ \mathcal{R}(Y)\subset\mathcal{R}(S)\end{aligned} \tag{32}$$

ここで，S は式 (20) で定義される対象クラスのみの標本を並べた作用素である．抑制の目的は，Y が抽出する対象クラスの特徴の中に対象クラス以外にも含まれる特徴があるとき，その特徴を抑制することであるため，解の空間を限っても抑制の効果が期待できる．また，非線形写像の次元が小さい場合や，原空間と高次元特徴空間の直積空間への写像

$$\Phi' : \mathbb{R}^d \to \mathcal{F}\times\mathbb{R}^d,\ \boldsymbol{x}\mapsto[\Phi(x)^\top\boldsymbol{x}^\top]^\top$$

$$k'(\boldsymbol{x}_1,\boldsymbol{x}_2) = \langle\Phi'(\boldsymbol{x}_1),\Phi'(\boldsymbol{x}_2)\rangle = k(\boldsymbol{x}_1,\boldsymbol{x}_2) + \langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle$$

を用いれば，対象クラスの標本が張る空間と標本全体が張る空間の重なりが大きくなり，抑制の効果が大きくなると考えられる．

最適化問題 (32) の解を求める．

［補題 1］ 制約条件 $\mathcal{N}(Y)\subset\mathcal{R}(S)^\perp$ は，Y がある作用素 $A\in\mathcal{B}(\mathbb{R}^L,\mathcal{F})$ を用いて，$Y = AS^\top$ と表されることと同値である．

(証明) (i)『任意の作用素 $A\in\mathcal{B}(\mathbb{R}^L,\mathcal{F})$，$Y\in\mathcal{B}(\mathcal{F})$，$S\in\mathcal{B}(\mathbb{R}^L,\mathcal{F})$ に対して，$Y = AS^\top$ ならば $\mathcal{N}(Y)\supset\mathcal{R}(S)^\perp$ である．』は，$\mathcal{R}(S)^\perp = \mathcal{N}(S^\top)$ より，自明である．

(ii)『任意の作用素 $Y\in\mathcal{B}(\mathcal{F})$，$S\in\mathcal{B}(\mathbb{R}^L,\mathcal{F})$ に対して，$\mathcal{N}(Y)\supset\mathcal{R}(S)^\perp$ ならば，$Y = AS^\top$ を満たす作用素 $A\in\mathcal{B}(\mathbb{R}^L,\mathcal{F})$ が存在する．』を証明する．X の Moore-Penrose 一般逆作用素を $X^\dagger$ で表す．$A = Y(S^\top)^\dagger$ と置けば $AS^\top = Y(S^\top)^\dagger S^\top = YP_{\mathcal{R}(S)} = Y(I - P_{\mathcal{R}(S)^\perp})$ である．ここで，$P_{\mathcal{R}(S)}$, $P_{\mathcal{R}(S)^\perp}$ はそれぞれ，$\mathcal{R}(S)$, $\mathcal{R}(S)^\perp$ への正射影を表す．$\mathcal{N}(Y)\supset\mathcal{R}(S)^\perp$ より，$YP_{\mathcal{R}(S)^\perp} = 0$ であるため，$AS^\top = Y$ である．

(i), (ii) より，補題 1 が証明できる．(証明終)

［補題 2］ 制約条件 $\mathcal{R}(Y)\subset\mathcal{R}(S)$ は，Y がある作用素 $B\in\mathcal{B}(\mathcal{F},\mathbb{R}^L)$ を用いて，$Y = SB$ と表されることと同値である．

証明は補題 1 とほぼ同様であるため，省略する．

補題 1, 2 より，Y は，行列 $X\in\mathbb{R}^{L\times L}$ を用いて $Y = SXS^\top$ と表すことができる．最適化問題 (32) の目的関数を J と置くと

$$
\begin{aligned}
J \;=\; & \frac{1}{L}\sum_{j=1}^{L}\Big(k(\boldsymbol{x}_j,\boldsymbol{x}_j) - \langle S^\top\Phi(\boldsymbol{x}_j), XS^\top\Phi(\boldsymbol{x}_j)\rangle \\
& -\langle XS^\top\Phi(\boldsymbol{x}_j), S^\top\Phi(\boldsymbol{x}_j)\rangle \\
& +\langle XS^\top\Phi(\boldsymbol{x}_j), S^\top SXS^\top\Phi(\boldsymbol{x}_j)\rangle\Big) \\
& \frac{\beta}{M}\sum_{j=1}^{M}\Big(k(\boldsymbol{x}_j,\boldsymbol{x}_j) - \langle S^\top\Phi(\boldsymbol{y}_j), XS^\top\Phi(\boldsymbol{y}_j)\rangle \\
& -\langle XS^\top\Phi(\boldsymbol{y}_j), S^\top\Phi(\boldsymbol{y}_j)\rangle \\
& +\langle XS^\top\Phi(\boldsymbol{y}_j), S^\top SXS^\top\Phi(\boldsymbol{y}_j)\rangle\Big) \\
& +\mu\|SXS^\top\|_F^2
\end{aligned}
$$

$K_x = S^\top S \in \mathbb{R}^{L\times L}$, $S_y = [\Phi(\boldsymbol{y}_1),\ldots,\Phi(\boldsymbol{y}_M)] \in \mathcal{B}(\mathbb{R}^L,\mathcal{F})$, $K_y = S^\top S_y \in \mathbb{R}^{L\times M}$ とおくと

$$
\begin{aligned}
J \;=\; & \frac{1}{L}\mathrm{Trace}[K_x - K_xXK_xK_xX^\top K_x - K_xXK_xK_xX^\top K_x \\
& +K_x^\top X^\top K_xXK_x] - \frac{\beta}{M}\mathrm{Trace}[S_y^\top S_y - K_y^\top XK_y \\
& -K_y^\top X^\top K_y + K_y^\top X^\top K_xXK_y] + \mathrm{Trace}[XK_xX^\top K_x^\top] \\
=\; & \mathrm{Trace}[K_x^{\frac{1}{2}}X(\frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top + \mu K_x)X^\top K_x^{\frac{1}{2}} \\
& -X(\frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top) \\
& -(\frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top)X^\top + \frac{1}{L}K_x - \frac{\beta}{M}S_y^\top S_y]
\end{aligned}
\tag{33}
$$

ここで，$K_{xy} = \frac{1}{L}K_xK_x^\top - \frac{\beta}{M}K_yK_y^\top$ とおく．

a） ランク制約がない場合，すなわち $r = L$ である場合

J の X における Z 方向の Gâteaux 微分 $\delta J(X;Z)$ は

$$
\begin{aligned}
\delta J(X;Z) \;=\; & \lim_{\delta\to 0}\frac{1}{\delta}\mathrm{Trace}[K_x^{\frac{1}{2}}\delta Z(K_{xy}+\mu K_x)X^\top K_x^{\frac{1}{2}} \\
& +K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)\delta Z^\top K_x^{\frac{1}{2}} - \delta ZK_{xy} \\
& -K_{xy}\delta Z^\top] \\
=\; & \lim_{\delta\to 0}\frac{2}{\delta}\mathrm{Trace}[\big(K_xX(K_{xy}+\mu K_x) - K_{xy}\big)\delta Z^\top]
\end{aligned}
$$

任意の Z に対して，$\delta J(X;Z) = 0$ となるのは，

$$
K_xX(K_{xy}+\mu K_x) = K_{xy} \tag{34}
$$

が成り立つときである．$\mathcal{R}(K_x) \subset \mathcal{R}(K_y)$ より，$\mathcal{R}(K_x) = \mathcal{R}(K_y)$ である．また，$\mathcal{N}(K_{xy}+\mu K_x) \subset \mathcal{N}(K_{xy})$ であるため，式 (34) は解を持ち，その解の 1 つは

$$
X = K_x^\dagger K_{xy}(K_{xy}+\mu K_x)^\dagger \tag{35}
$$

で与えられる．特に，K_x, $K_{xy}+\mu K_x$ が正則となる場合 Moore-Penrose 一般逆行列は逆行列に置き換わる．

b） ランク制約がある場合

Mercer カーネルを用いた場合，同一の標本がない限り，K_x は正定値である．以降，$(K_{xy}+\mu K_x)$ が正定値となるように β, μ が選ばれているものとする．

式 (33) で J の X に関係のない項を除いたものを J' とすると

$$
\begin{aligned}
J' \;=\; & \mathrm{Trace}[\big(K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}} \\
& -K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}}\big) \\
& \big((K_{xy}+\mu K_x)^{\frac{1}{2}}X^\top K_x^{\frac{1}{2}} - (K_{xy}+\mu K_x)^{-\frac{1}{2}}K_{xy}K_x^{-\frac{1}{2}}\big)] \\
=\; & \|K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}} - K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}}\|_F^2
\end{aligned}
\tag{36}
$$

式 (36) より，問題は行列の最良近似問題となる．ここで，$K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}}$ の特異値分解を

$$
K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-\frac{1}{2}} = \sum_{i=1}^{L}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top \tag{37}
$$

とおくと $K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}}$ のランクは r 以下であるため

$$
K_x^{\frac{1}{2}}X(K_{xy}+\mu K_x)^{\frac{1}{2}} = \sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top \tag{38}
$$

を満たす X が存在するときは，上式を満たすとき J' が最小となる．

$$
\mathcal{R}(K_x^{\frac{1}{2}}) \supset \mathcal{R}\Big(\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top\Big) \tag{39}
$$

$$
\mathcal{N}\Big((K_{xy}+\mu K_x)^{\frac{1}{2}}\Big) \subset \mathcal{N}\Big(\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top\Big) \tag{40}
$$

が満たされるため，行列方程式 (38) は解を持ち，その解の 1 つは

$$
X = K_x^{-\frac{1}{2}}\Big(\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^\top\Big)(K_{xy}+\mu K_x)^{-\frac{1}{2}} \tag{41}
$$

と与えられる．

特異値分解の性質を用いると

$$
\boldsymbol{v}_i \;=\; \frac{1}{\lambda_i}(K_{xy}+\mu K_x)^{-\frac{1}{2}}K_{xy}K_x^{-\frac{1}{2}}\boldsymbol{u}_i \tag{42}
$$

であり，これを に代入すると

$$
X = K_x^{-\frac{1}{2}}\Big(\sum_{i=1}^{r}\boldsymbol{u}_i\boldsymbol{u}_i^\top\Big)K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-1} \tag{43}
$$

で与えられる．ここで，$\boldsymbol{u}_i$ は，$K_x^{-\frac{1}{2}}K_{xy}(K_{xy}+\mu K_x)^{-1}K_{xy}K_x^{-\frac{1}{2}}$ の降順に並べた固有値に対応する固有ベクトルである．

［系 1］ $r = L$ かつ，K_x, $K_{xy}+\mu K_x$ が正則である場合，式 (35) と式 (b)) は等価である．

$Y = SXS^\top$ より，類似度関数は

$$
\begin{aligned}
g(\boldsymbol{x}) \;=\; & -\|\Phi(\boldsymbol{x}) - Y\Phi(\boldsymbol{x})\|^2 \\
=\; & k(\boldsymbol{x},\boldsymbol{x}) - 2\langle\boldsymbol{h}(\boldsymbol{x}), X\boldsymbol{h}(\boldsymbol{x})\rangle \\
& +\langle\boldsymbol{h}(\boldsymbol{x}), X^\top K_xX\boldsymbol{h}(\boldsymbol{x})\rangle
\end{aligned}
\tag{44}
$$

となる．

X の計算は，$K_yK_y^\top$ の計算の $L\times M$ 行列の積の計算の他は，すべて L 次の正方行列の積，逆，固有値分解で計算ができる．

表 1　手書き数字識別実験結果

手法	パラメータ	誤識別率 ± 標準偏差 [%]
CLAFIC	$r = 26$	5.03 ± 0.10
SPCA	$r = 31$, $\mu = 10^{-2.8}$, $\beta = 10^{-3}$	4.88 ± 0.10
KPCA	$r = 155$	3.57 ± 0.12
SKPCA	$\beta = 10^{-1.8}$, $\mu = 8 \times 10^{-4}$, $r = 260$	3.51 ± 0.11

4. 実　　験

手書き数字のデータベース MNIST を用いて性能の比較実験を行った．MNIST の学習用標本 60,000 サンプル中から無作為に 5,000 サンプルを学習用に抽出し，残りの 55,000 サンプルを検定用に使用した．このランダムサンプリングを 20 回行い，誤識別率の平均と標準偏差を比較した．実験では，CLAFIC 法，正則化抑制付き部分空間法 (SPCA, 式 (13))，カーネル部分空間法 (KPCA)，抑制付きカーネル部分空間法 (SKPCA) の 4 つの手法を比較した．カーネル関数は，式 (18) の Gaussian カーネル，$c = 1$ を用いた．競合クラスの標本 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M$ には対象クラス以外のすべての標本を用いた．表 1 に結果を示す．パラメータはいくつかの候補から誤識別率が最小となるものを選んだ．KPCA と提案法の SKPCA では，Student の片側 t 検定において，5.2%の有意水準で提案法の優位性が示された．

5. ま と め

抑制付きカーネル部分空間法に対して，解の存在範囲を限定することで，対象クラスの標本数の大きさの行列演算で設計ができる手法を提案した．提案法は，行列計算が対象クラスの標本数の大きさであるためクラス数が非常に大きな場合でも現実的に設計を行うことができるという利点がある．手書き数字識別実験結果は，提案法が通常のカーネル部分空間法よりも低い誤識別率を示した．これは解の空間を限っても，抑制の効果があるためであると考えられる．

今後の課題として，パラメータ選択に関する議論，棄却やクラス数の増減への対応などが挙げられる．

謝　　辞

本研究は，日本学術振興会科学研究費補助金 No. 19700153(若手 B) の補助を受けた．

文　　献

[1] 飯島泰蔵：“パターン認識理論”, 森北出版 (1989).
[2] 鷲沢, 疋田, 田中, 山下：“カーネル相対主成分分析による多クラスパターン認識”, 第二回 FIT（情報科学技術フォーラム）情報レターズ, pp. 207–208 (2003).
[3] 池野, 山下, 小川：“相対 KL 変換法によるパターン認識”, 信学論 (D-II), **J80-D-II**, 2, pp. 541–547 (1997).
[4] 前田, 村瀬：“カーネル非線形部分空間法によるパターン認識”, 信学論 (D-II), **J82-D-II**, 4, pp. 600–612 (1999).
[5] 津田：“ヒルベルト空間における部分空間法”, 信学論 (D-II), **J82-D-II**, 4, pp. 592–599 (1999).
[6] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch and A. Smola: “Input space vs. feature space in kernel-based methods”, IEEE Transactions on Neural Networks, **10**, 5, pp. 1000–1017 (1999).
[7] Y. Washizawa, K. Hikida, T. Tanaka and Y. Yamashita: “Kernel relative principal component analysis for pattern recognition”, Proc. of Joint IAPR Iinternational Workshops on Syntactical and Structural Pattern Recognition and Statistical Pattern Recognition (SSPR/SPR 2004), pp. 1105–1113 (2004).
[8] Y. Washizawa and Y. Yamashita: “Kernel projection classifiers with suppressing features of other classes”, Neural Computation, **18**, 8, pp. 1932–1950 (2006).
[9] S. Watanabe and N. Pakvasa: “Subspace method in pattern recognition”, Proc. 1st Int. J. Conf on Pattern Recognition, Washington DC, pp. 25–32 (1973).
[10] E. Oja, 小川, 佐藤：“パターン認識と部分空間法”, 産業図書 (1986).
[11] 鷲沢：“正則化を用いた２次識別器 -部分空間法との比較-”, 第 10 回　画像の認識・理解シンポジウム予稿集, pp. 510–515 (2007).
[12] Y. Washizawa:　“Regularization vs. rank reduction in quadratic classifiers”, Proc. of ACCV 2007 workshop subspace 2007, pp. 108–115 (2007).

線型多様体間距離に基づくパターン識別と学習

堀田　政二†

† 東京農工大学　〒 184–8588 東京都小金井市中町 2–24–16
E-mail: †s-hotta@cc.tuat.ac.jp

あらまし　本稿では，線型多様体間距離を利用したパターン識別と学習について考える．はじめに，線型多様体間距離が未知データとの二乗誤差を最小にする訓練データの線型結合を利用した識別から容易に導かれることを示す．次に，距離に基づく学習の一種である一般化学習ベクトル量子化 (Generalized Learning Vector Quantization, GLVQ) を線型多様体間距離に基づくパターン識別に適用した場合の学習則を導く．手書き数字を用いた実験により，多様体間距離を利用することで低いエラー率を達成でき，学習によって識別精度が改善されることを示す.
キーワード　パターン識別，線型多様体，一般化学習ベクトル量子化

Patten Classification and Learning Based on Distance Between Linear Manifolds

Seiji HOTTA†

† Tokyo Univ. of Agri. and Tech., 2–24–16 Naka-cho, Koganei-shi, Tokyo, 184–8588 Japan
E-mail: †s-hotta@cc.tuat.ac.jp

Abstract　This report describes pattern classification and learning based on a distance between linear manifolds. First, a distance between linear manifolds is derived by expanding the concept of pattern classification that uses the linear combination of training samples as the most similar prototype for an input sample. Next, generalized learning vector quantization is applied to manifold-based classification to reduce error rates and memory requirement. Experimental results on a handwritten digit dataset show that manifold-based classification and learning can improve accuracy with small memory requirement.
Key words　pattern classification, linear manifold, generalized learning vector quantization

1. はじめに

部分空間法は CLAFIC で代表される類似度型と，投影距離法のような距離型に大別できる [1]．このうち類似度型は，未知データも部分空間で表現できる場合に利用できる相互部分空間法 [2] へと拡張され，その後さらに理論拡張が進められ顔認識や物体認識において有効性が確認されている [3]．一方距離型では，投影距離法 [4] が提案されてから相互部分空間法のように線型多様体 (アフィン部分空間) 同士の距離としての拡張は行われなかったようである．しかし，文字認識において変形不変性を実現するために提案された接距離 (Tangent Distance, TD) [5] で線型多様体間の距離が利用されたのをきっかけに，joint manifold distance による動画像中の顔のクラスタリング [6] や線型多様体距離に基づく顔認識 [7] が提案された．しかし，これら以外に線型多様体間の距離を利用した研究を見つけることは難しく，研究人口も少ないようである．

これにはさまざまな理由が考えられるが，一つは投影距離法があまり知られていないことが考えられる．投影距離法は，各クラスで線型多様体を構築し，未知データの線型多様体への垂線の長さ (残差長) が最小となるクラスへ未知データを分類する簡便な識別法である．しかし，局所部分空間法 (Local Subspace Classifier, LSC) [8] や k-nearest feature line [9] の識別則は (訓練データの選択を行う点が異なるが) 投影距離法そのものであるにも関わらず，投影距離法に関する言及はない．Oja の本 [10] でも線型多様体については述べられているが，内積や射影長に基づく識別が主題であり，距離に基づく識別については触れられていない．しかし距離に基づく識別を考えることは，これまでに提案されてきた距離関数や学習を識別に取り入れられる可能性があるため有用といえる．そこで本稿では，線型多様体間距離を利用したパターン識別と学習について考える．はじめに，線型多様体間距離が未知データとの二乗誤差を最小にする訓練データの線型結合を利用した識別から容易に導かれることを示す．次に，距離に基づく学習の一種である一般化学習ベクトル量子化 (Generalized Learning Vector Quantization, GLVQ) [11] を

線型多様体間距離に基づくパターン識別に適用した場合の学習則を導く．手書き数字を用いた実験により，多様体間距離を利用した方が低いエラー率を達成でき，学習によってエラー率を低減できることを示す．

2. 二乗誤差最小パターンとの距離に基づく識別

線型多様体間距離の導出を容易にするために，はじめに未知データとの二乗誤差が最小となるような訓練データの線型結合を各クラスで作成し，最も誤差が小さくなるクラスに未知データを分類するという問題を考える (図 1 参照)．

2.1 定 式 化

d 次元の未知データを $\boldsymbol{q} = (q_1 \cdots q_d)^\top \in \mathbb{R}^d$，クラス j $(j = 1, ..., C)$ に属する n_j 個 $(n_j \leqq d)$ の訓練データのうち，第 i 訓練データを $\boldsymbol{x}_i^j = (x_{i1}^j \cdots x_{id}^j)^\top \in \mathbb{R}^d$ $(i = 1, ..., n_j)$ とする．また，$\boldsymbol{x}_i^j$ を並べた行列を $\mathbf{X}_j = (\boldsymbol{x}_1^j | \boldsymbol{x}_2^j | \cdots | \boldsymbol{x}_{n_j}^j) \in \mathbb{R}^{d \times n_j}$ とする．これらの訓練データは互いに独立であれば直交していなくても良い．なお本節では，記号の煩雑さを避けるために，クラスの添え字 j を適宜省略することにする．

まず，クラス j で未知データとの二乗誤差が最小となるような訓練データの線型結合を以下の最適化問題により求めることにする：

$$\begin{aligned} \min_{\boldsymbol{b}} \quad & \|\boldsymbol{q} - \mathbf{X}\boldsymbol{b}\|^2 = \left\| \boldsymbol{q} - \sum_{i=1}^{n} b_i \boldsymbol{x}_i \right\|^2 \\ \text{s.t.} \quad & \boldsymbol{b}^\top \mathbf{1}_n = \sum_{i=1}^{n} b_i = 1 \end{aligned} \tag{1}$$

ここで $\mathbf{1}_n = (1 \cdots 1)^\top \in \mathbb{R}^n$ は要素が全て 1 のベクトル，$\boldsymbol{b} = (b_1 \cdots b_n)^\top \in \mathbb{R}^n$ は訓練データに対する係数ベクトルである．制約条件の $\boldsymbol{b}^\top \mathbf{1}_n = 1$ は，線型結合データが訓練データの張る線型多様体上に乗ることを保障するために必要である．これは制約条件を $b_1 = 1 - \sum_{i=2}^{n} b_i$ と変形して式 (1) の目的関数に代入して制約条件を消去すれば，式 (1) の最適化問題が

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{q} - (\boldsymbol{x}_1 + \mathbf{V}\boldsymbol{\beta})\|^2 = \left\| \boldsymbol{q} - (\boldsymbol{x}_1 + \sum_{i=2}^{n} \beta_i (\boldsymbol{x}_i - \boldsymbol{x}_1)) \right\|^2 \tag{2}$$

となることからも確認できる．ここで $\mathbf{V}$ は $\boldsymbol{x}_i - \boldsymbol{x}_1$ $(i = 2, ..., n)$ を並べた $d \times (n-1)$ の行列 $\mathbf{V} = (\boldsymbol{x}_2 - \boldsymbol{x}_1 | \boldsymbol{x}_3 - \boldsymbol{x}_1 | \cdots | \boldsymbol{x}_n - \boldsymbol{x}_1)$ であり $\boldsymbol{\beta} = (\beta_1 \cdots \beta_{n-1})^\top \in \mathbb{R}^{n-1}$ は $\mathbf{V}$ の各列に対応する係数ベクトルである．すなわち式 (2) は $\boldsymbol{q}$ と線型多様体 $\mathcal{M} = \boldsymbol{x}_1 + \mathbf{V}\boldsymbol{\beta}$ との二乗誤差を最小にする問題となっている．

式 (1) の解は，$\boldsymbol{q}$ を n 個並べた行列を $\mathbf{Q} = (\boldsymbol{q} | \cdots | \boldsymbol{q}) \in \mathbb{R}^{d \times n}$，相関行列を $\mathbf{C} = (\mathbf{Q} - \mathbf{X})^\top (\mathbf{Q} - \mathbf{X}) \in \mathbb{R}^{n \times n}$ とすれば，ラグランジュ乗数法から

$$\boldsymbol{b} = \frac{\mathbf{C}^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \mathbf{C}^{-1}\mathbf{1}_n} \tag{3}$$

として閉じた形で求めることができる．もし $n > d$ の場合や，過学習を避けたい場合には，式 (1) の目的関数に罰則項を $\|\boldsymbol{q} - \mathbf{X}\boldsymbol{b}\| + \lambda \|\boldsymbol{b}\|^2$ $(\lambda > 0)$ のように付け加えて正則化を施せば良い．その場合にも閉じた形で解が得られ，$\mathbf{I}_n \in \mathbb{R}^{n \times n}$ を単位行列として $\mathbf{C}$ の逆行列を計算する前に $\mathbf{C} + \lambda \mathbf{I}_n$ とすれば良い．識別は，$\boldsymbol{b}$ を用いてクラス j との距離を $d(\boldsymbol{q}, \mathcal{M}) = \|\boldsymbol{q} - \mathbf{X}\boldsymbol{b}\|^2$ と定義して，$d(\boldsymbol{q}, \mathcal{M})$ が最小となるクラスへ $\boldsymbol{q}$ を分類することで実現できる．なお，$\mathbf{X}$ として $\boldsymbol{q}$ の k 近傍訓練データを用いて識別を行うものは Local Subspace Classifier (LSC) [8] と呼ばれる．

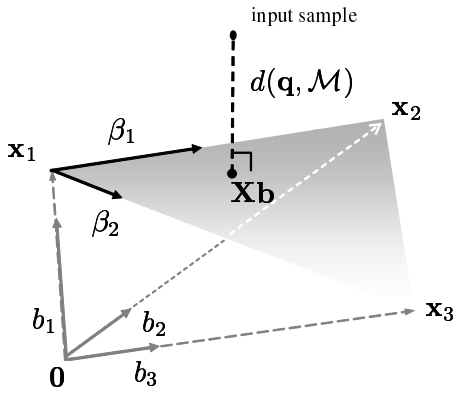


図 1 訓練データが三つの場合の未知データ $\boldsymbol{q}$ と線型多様体との関係

一方，式 (2) の解は，目的関数を $\boldsymbol{\beta}$ で偏微分して $\mathbf{0}$ と置くことにより

$$\boldsymbol{\beta} = (\mathbf{V}^\top \mathbf{V})^{-1} \mathbf{V}^\top (\boldsymbol{q} - \boldsymbol{x}_1) \tag{4}$$

として求めることができる．識別は，$\boldsymbol{\beta}$ を用いてクラス j との距離を $d(\boldsymbol{q}, \mathcal{M}) = \|\boldsymbol{q} - (\boldsymbol{x}_1 + \mathbf{V}\boldsymbol{\beta})\|^2$ と定義すれば，$d(\boldsymbol{q}, \mathcal{M})$ が最小となるクラスへ $\boldsymbol{q}$ を分類することで実現できる．

なお，$\mathbf{V}$ を $\mathbf{V}^\top \mathbf{V}$ の値の大きい固有値に対応した $n-1$ 本以下の固有ベクトルを並べた行列とすれば，$d(\boldsymbol{q}, \mathcal{M})$ は $d(\boldsymbol{q}, \mathcal{M}) = \|\boldsymbol{q} - \boldsymbol{x}_1\|^2 - \|\mathbf{V}^\top (\boldsymbol{q} - \boldsymbol{x}_1)\|^2$ と書け，文献 [4] の投影距離法 (Projection Distance Method, PDM) と同じになる．すなわち，最小の $d(q, \mathcal{M})$ は $\boldsymbol{q}$ から $\mathcal{M}$ までの垂線の長さ (残差長) で与えられることが確かめられる．また，$\boldsymbol{x}_1$ を画像パターン，$\mathbf{V}$ を x_1 の変形を表す接ベクトル (Tangent Vecotr, TV) [5] とした場合には，式 (2) は片側接距離 (one sided tangent distance, 1S TD) [5] と呼ばれている．

2.2 カーネル化

二乗誤差最小パターンとの距離に基づくパターン識別は，カーネルトリック $\Phi(\boldsymbol{x})^\top \Phi(\boldsymbol{y}) = K(\boldsymbol{x}, \boldsymbol{y})$ を利用して容易にカーネル化できる ($\Phi(\cdot)$ は高次元空間への非線型写像)．カーネル化した場合，式 (1) の最適化問題は，$\Phi(\boldsymbol{q})$ を n 個並べた行列を $\mathbf{Q}_\phi$，$\Phi(\boldsymbol{x}_i)$ $(i = 1, ..., n)$ を並べた行列を $\mathbf{X}_\phi$ とすれば

$$\begin{aligned} \min_{\boldsymbol{b}} \quad & \|(\mathbf{Q}_\Phi - \mathbf{X}_\Phi)\boldsymbol{b}\|^2 \\ \text{s.t.} \quad & \boldsymbol{b}^\top \mathbf{1}_n = 1 \end{aligned} \tag{5}$$

と書ける．目的関数を展開し，ラグランジュ乗数法とカーネルトリックを用いれば式 (5) の解は

$$\boldsymbol{b} = \frac{K_{\mathbf{C}}^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top K_{\mathbf{C}}^{-1}\mathbf{1}_n} \tag{6}$$

となる．ここで $K_{\mathbf{C}}$ は $n \times n$ の行列であり，その i,j 要素は

$$(K_{\mathbf{C}})_{ij} = K(\boldsymbol{q},\boldsymbol{q}) - K(\boldsymbol{q},\boldsymbol{x}_j) - K(\boldsymbol{x}_i,\boldsymbol{q}) + K(\boldsymbol{x}_i,\boldsymbol{x}_j) \quad (7)$$

で与えられる．正則化は $K_{\mathbf{C}} + \lambda \mathbf{I}_n$ とすれば良い．カーネル投影距離 [12] と異なり，カーネル主成分分析 [13] を用いない識別ができる．

3. 線型多様体間距離に基づくパターン識別

次に，未知データも線型多様体で与えられる場合の線型多様体間の距離について考える (図 2 参照)．未知データが m 個 ($m \leqq d$) の d 次元ベクトルの組で与えられたとする．そのうちの第 i ベクトルを $\boldsymbol{q}_i = (q_{i1} \cdots q_{id})^\top$ $(i = 1, ..., m)$ で表し，それらを並べた $d \times m$ の行列を $\mathbf{Q} = (\boldsymbol{q}_1 | \boldsymbol{q}_2 | \cdots | \boldsymbol{q}_m)$ とする．なお，これらは互いに独立であれば直交していなくても構わない．

3.1 定 式 化

$\mathbf{Q}$ の各列に対する $m \times 1$ の係数ベクトルを $\boldsymbol{a} = (a_1 \cdots a_m)^\top$ とすれば，クラス j の線型多様体までの最短距離は

$$\begin{aligned} \min_{\boldsymbol{a},\boldsymbol{b}} \quad & \|\mathbf{Q}\boldsymbol{a} - \mathbf{X}\boldsymbol{b}\|^2 \\ \text{s.t.} \quad & \boldsymbol{a}^\top \mathbf{1}_m = 1,\ \boldsymbol{b}^\top \mathbf{1}_n = 1 \end{aligned} \quad (8)$$

を解くことによって求めることができる．この問題もラグランジュ乗数法により閉じた形で解が求められるが，解が非常に複雑な式となってしまう．そこで二乗誤差最小パターンとの距離に基づく識別と同じように制約条件を消去することによって解を簡単に表すことを考える．

まず，一つの制約条件を消去した場合について考える．制約条件を $a_1 = 1 - \sum_{i=2}^{m} a_i$ として式 (8) に代入すると，式 (8) の最適化問題は

$$\begin{aligned} \min_{\boldsymbol{\alpha},\boldsymbol{b}} \quad & \|\boldsymbol{q}_1 + \mathbf{U}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{b}\|^2 \\ \text{s.t.} \quad & \boldsymbol{b}^\top \mathbf{1}_n = 1 \end{aligned} \quad (9)$$

と書き換えることができる．ここで $\mathbf{U}$ は $\boldsymbol{q}_i - \boldsymbol{q}_1$ $(i = 2, ..., m)$ を並べた行列 $\mathbf{U} = (\boldsymbol{q}_2 - \boldsymbol{q}_1 | \boldsymbol{q}_3 - \boldsymbol{q}_1 | \cdots | \boldsymbol{q}_m - \boldsymbol{q}_1) \in \mathbb{R}^{d \times (m-1)}$ であり $\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_{m-1})^\top \in \mathbb{R}^{m-1}$ は $\mathbf{U}$ の各列に対応する係数ベクトルである．ここで $\mathbf{A} \in \mathbb{R}^{d \times d}$ と $\gamma \in \mathbb{R}$ を

$$\mathbf{A} = \mathbf{U}(\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \quad (10)$$

$$\gamma = \frac{1 - \mathbf{1}_n^\top (\mathbf{X}^\top (\mathbf{I}_d - \mathbf{A})\mathbf{X})^{-1} \mathbf{X}(\mathbf{I}_d - \mathbf{A})\boldsymbol{q}_1}{\mathbf{1}_n^\top (\mathbf{X}^\top (\mathbf{I}_d - \mathbf{A})\mathbf{X})^{-1} \mathbf{1}_n} \quad (11)$$

とおくと，式 (9) を最小にする $\boldsymbol{\alpha}$ と $\boldsymbol{b}$ はラグランジュ乗数法から

$$\boldsymbol{b} = (\mathbf{X}^\top (\mathbf{I}_d - \mathbf{A})\mathbf{X})^{-1} \{\mathbf{X}^\top (\mathbf{I}_d - \mathbf{A})\boldsymbol{q}_1 + \gamma \mathbf{1}_n\} \quad (12)$$

$$\boldsymbol{\alpha} = (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top (\mathbf{X}\boldsymbol{b} - \boldsymbol{q}_1) \quad (13)$$

として求めることができる．正則化が必要な場合には式 (10) と式 (13) における $\mathbf{U}^\top \mathbf{U}$ を $\mathbf{U}^\top \mathbf{U} + \lambda \mathbf{I}_{m-1}$ と変更すれば良い．この結果は後の線型多様体間距離に基づく学習で利用する．

次に二つの制約条件を消去した場合を考える．制約条件を

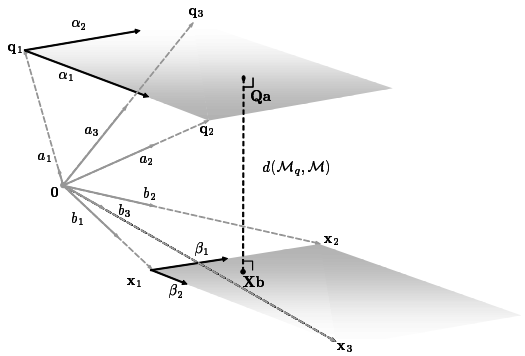


図 2　高次元空間における線型多様体間距離の概念図

$b_1 = 1 - \sum_{i=2}^{n} b_i$ として式 (9) に代入すると，式 (9) の最適化問題は

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad \|(\boldsymbol{q}_1 + \mathbf{U}\boldsymbol{\alpha}) - (\boldsymbol{x}_1 + \mathbf{V}\boldsymbol{\beta})\|^2 \quad (14)$$

と書き換えることができる．式 (14) の最適解は，式 (14) をそれぞれ $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ で偏微分して $\mathbf{0}$ とおいて式をまとめることにより求められる．すなわち

$$\mathbf{U}_1 = \mathbf{U}^\top \mathbf{U},\ \mathbf{V}_1 = \mathbf{V}^\top \mathbf{V} \quad (15)$$

$$\mathbf{U}_2 = \mathbf{U}^\top \mathbf{V},\ \mathbf{V}_2 = \mathbf{V}^\top \mathbf{U} \quad (16)$$

とおけば，$\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ は

$$\boldsymbol{\alpha} = (\mathbf{U}_1 - \mathbf{U}_2 \mathbf{V}_1^{-1} \mathbf{V}_2)^{-1} (\mathbf{U}_2 \mathbf{V}_1^{-1} \mathbf{V}^\top - \mathbf{U}^\top)(\boldsymbol{q}_1 - \boldsymbol{x}_1) \quad (17)$$

$$\boldsymbol{\beta} = (\mathbf{V}_1 - \mathbf{V}_2 \mathbf{U}_1^{-1} \mathbf{U}_2)^{-1} (\mathbf{V}^\top - \mathbf{V}_2 \mathbf{U}_1^{-1} \mathbf{U}^\top)(\boldsymbol{q}_1 - \boldsymbol{x}_1) \quad (18)$$

によって与えられる．正則化が必要な場合には式 (15) において $\mathbf{U}_1$ と $\mathbf{V}_1$ をそれぞれ $\mathbf{U}_1 + \lambda_1 \mathbf{I}_{m-1}$ と $\mathbf{V}_1 + \lambda_2 \mathbf{I}_{n-1}$ ($\lambda_1 > 0$, $\lambda_2 > 0$) に変更すれば良い．識別は，式 (9) や式 (14) の目的関数の値を線型多様体間距離 $d(\mathcal{M}_q, \mathcal{M})$ と定義し，この値が最小となるクラスを未知データのクラスとして出力することで実現できる．なお，式 (14) において未知データと訓練データの線型多様体を主成分分析で求めた場合には Inter-Subspace Distance [7], $\boldsymbol{q}_1$ と $\boldsymbol{x}_1$ を画像パターン，$\mathbf{U}$ と $\mathbf{V}$ を $\boldsymbol{q}_1$ と $\boldsymbol{x}_1$ の接ベクトルとした場合，式 (14) は両側接距離 (two sided tangent distance, 2STD) [5] と呼ばれている．

3.2 カーネル化

線型多様体間距離は二乗誤差最小パターンとの距離に基づくパターン識別と同様，容易にカーネル化できる．ここでは式 (14) を用いてカーネル化することを考える．カーネル化した場合，式 (14) の最適化問題は，$\Phi(\boldsymbol{q}_i) - \Phi(\boldsymbol{q}_1)$ $(i = 2, ..., m)$ を並べた行列を $\mathbf{U}_\phi$，$\Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_1)$ $(i = 2, ..., n)$ を並べた行列を $\mathbf{V}_\phi$ とすれば

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \|\Phi(\boldsymbol{q}_1) + \mathbf{U}_\Phi \boldsymbol{\alpha} - (\Phi(\boldsymbol{x}_1) + \mathbf{V}_\Phi \boldsymbol{\beta})\|^2 \quad (19)$$

と書ける．目的関数を展開し，$\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ でそれぞれ偏微分して $\mathbf{0}$ とおき，カーネルトリックを用いれば式 (19) の解は

$$\boldsymbol{\alpha} = (K_{\mathbf{UU}} - K_{\mathbf{UV}}K_{\mathbf{VV}}^{-1}K_{\mathbf{VU}})^{-1}(K_{\mathbf{UV}}K_{\mathbf{VV}}^{-1}k_{\boldsymbol{x}} - k_{\boldsymbol{q}}) \tag{20}$$

$$\boldsymbol{\beta} = (K_{\mathbf{VU}}K_{\mathbf{UU}}^{-1}K_{\mathbf{UV}} - K_{\mathbf{VV}})^{-1}(K_{\mathbf{VU}}K_{\mathbf{UU}}^{-1}k_{\boldsymbol{q}} - k_{\boldsymbol{x}}) \tag{21}$$

となる．ここで $K_{\mathbf{UU}} \in \mathbb{R}^{m-1\times m-1}$，$K_{\mathbf{VV}} \in \mathbb{R}^{n-1\times n-1}$，$K_{\mathbf{UV}} \in \mathbb{R}^{m-1\times n-1}$，$K_{\mathbf{VU}} \in \mathbb{R}^{n-1\times m-1}$，$k_{\boldsymbol{q}} \in \mathbb{R}^{m-1}$，$k_{\boldsymbol{x}} \in \mathbb{R}^{n-1}$ であり，その i,j 要素と第 i 行目の要素はそれぞれ

$$\begin{aligned}
(K_{\mathbf{UU}})_{ij} =& K(\boldsymbol{q}_{i+1},\boldsymbol{q}_{j+1}) - K(\boldsymbol{q}_{i+1},\boldsymbol{q}_1) \\
& - K(\boldsymbol{q}_1,\boldsymbol{q}_{j+1}) + K(\boldsymbol{q}_1,\boldsymbol{q}_1) && (22)\\
(K_{\mathbf{VV}})_{ij} =& K(\boldsymbol{x}_{i+1},\boldsymbol{x}_{j+1}) - K(\boldsymbol{x}_{i+1},\boldsymbol{x}_1) \\
& - K(\boldsymbol{x}_1,\boldsymbol{x}_{j+1}) + K(\boldsymbol{x}_1,\boldsymbol{x}_1) && (23)\\
(K_{\mathbf{UV}})_{ij} =& K(\boldsymbol{q}_{i+1},\boldsymbol{x}_{j+1}) - K(\boldsymbol{q}_{i+1},\boldsymbol{x}_1) \\
& - K(\boldsymbol{q}_1,\boldsymbol{x}_{j+1}) + K(\boldsymbol{q}_1,\boldsymbol{x}_1) && (24)\\
(K_{\mathbf{VU}})_{ij} =& K(\boldsymbol{x}_{i+1},\boldsymbol{q}_{j+1}) - K(\boldsymbol{x}_{i+1},\boldsymbol{q}_1) \\
& - K(\boldsymbol{x}_1,\boldsymbol{q}_{j+1}) + K(\boldsymbol{x}_1,\boldsymbol{q}_1) && (25)\\
(k_{\boldsymbol{q}})_i =& K(\boldsymbol{q}_1,\boldsymbol{q}_{i+1}) - K(\boldsymbol{q}_1,\boldsymbol{q}_1) \\
& - K(\boldsymbol{x}_1,\boldsymbol{q}_{i+1}) + K(\boldsymbol{x}_1,\boldsymbol{q}_1) && (26)\\
(k_{\boldsymbol{x}})_i =& K(\boldsymbol{q}_1,\boldsymbol{x}_{i+1}) - K(\boldsymbol{q}_1,\boldsymbol{x}_1) \\
& - K(\boldsymbol{x}_1,\boldsymbol{x}_{i+1}) + K(\boldsymbol{x}_1,\boldsymbol{x}_1) && (27)
\end{aligned}$$

で与えられる．二つのパターン $\Phi(\boldsymbol{q}_1)$，$\Phi(\boldsymbol{x}_1)$ の高次元空間でのユークリッド距離の二乗は $d_{qx} = \|\Phi(\boldsymbol{q}_1) - \Phi(\boldsymbol{x}_1)\|^2 = K(\boldsymbol{q}_1,\boldsymbol{q}_1) - 2K(\boldsymbol{q}_1,\boldsymbol{x}_1) + K(\boldsymbol{x}_1,\boldsymbol{x}_1)$ で与えられるから，結局，高次元空間での未知データとクラス j の多様体間距離 $d(\mathcal{M}_q,\mathcal{M})$ は

$$\begin{aligned}
& d(\mathcal{M}_q,\mathcal{M}) = \\
& d_{qx} + 2k_{\boldsymbol{q}}^\top\boldsymbol{\alpha} - 2k_{\boldsymbol{x}}^\top\boldsymbol{\beta} + \boldsymbol{\alpha}^\top K_{\mathbf{UU}}\boldsymbol{\alpha} \\
& - \boldsymbol{\alpha}^\top K_{\mathbf{UV}}\boldsymbol{\beta} - \boldsymbol{\beta}^\top K_{\mathbf{VU}}\boldsymbol{\alpha} + \boldsymbol{\beta}^\top K_{\mathbf{VV}}\boldsymbol{\beta} && (28)
\end{aligned}$$

で与えられる．正則化は $K_{\mathbf{UU}} + \lambda_1\mathbf{I}_{m-1}$，$K_{\mathbf{VV}} + \lambda_2\mathbf{I}_{n-1}$ とすれば良い．カーネル相互部分空間法[14] と比べると複雑な計算が必要である．

4. 線型多様体間距離に基づく学習

ここでは線型多様体間距離に基づくパターン分類において，誤分類や記憶容量を減らすための学習について述べる．距離に基づく識別器のための学習法としては，辞書パターンを更新するベクトル量子化[15] と，距離関数を学習するメトリック学習 [16] がよく用いられるが，ここではベクトル量子化に基づく手法を述べる．ベクトル量子化に基づく手法としては，一般化学習量子化 (GLVQ, Generalized Learning Vector Quantization) [11] がよく知られている．そこで，はじめに前節と同様に二乗誤差最小パターンとの距離を利用した GLVQ に基づく学習則を導出し，その後線型多様体間の距離に基づく学習を導く．なお，本節ではクラスの添え字をつけて説明をするので注意されたい．

4.1 二乗誤差最小パターンとの距離を利用した GLVQ に基づく学習則

学習に利用するクラスラベルの分かっている学習データを $\boldsymbol{x}$ とする．$\boldsymbol{x}$ からクラス j の線型多様体までの距離を $d_j = \|\boldsymbol{x} - \mathbf{X}_j\boldsymbol{b}_j\|^2$ とする．ただし $\boldsymbol{b}_j$ は式 (3) で得られるものとする．$\mathbf{X}_1$ を $\boldsymbol{x}$ と同じクラスに属する最近傍線型多様体を張る訓練データ，$\mathbf{X}_2$ を $\boldsymbol{x}$ と異なるクラスに属する最近傍線型多様体を張る訓練データとし，それらを用いて計算された距離をそれぞれ d_1，d_2 とする．GLVQ と同様に，$\boldsymbol{x}$ と $\boldsymbol{x}$ の属するべきクラスとの近さ μ を以下のように定義する:

$$\mu(\boldsymbol{x}) = \frac{d_1 - d_2}{d_1 + d_2} \tag{29}$$

この μ は全ての $\boldsymbol{x}$ に対して $-1 < \mu(\boldsymbol{x}) < 1$ を満たしており，$\boldsymbol{x}$ が正しく分類された場合には値が負となり，誤分類の場合には値が正となる．学習の目的は誤分類を減らすこと，すなわち全ての $\boldsymbol{x}$ について μ が減少すればよい．したがって，学習の目的を次式の評価関数 S の最小化として定式化できる [11]:

$$S = \sum_{i=1}^{N} f(\mu(\boldsymbol{x}_i)) \tag{30}$$

ここで，N は学習に用いる学習データの総数，$f(\mu)$ は μ に対する単調増加関数でありシグモイド関数 $f(\mu,t) = 1/(1-e^{-\mu t})$ が用いられる (t は学習ステップ) [11].

上記の S を最小化するために $\mathbf{X}_j$ $(j = 1,2)$ を最急降下法に基づいて修正することを考える．

$$\mathbf{X}_j \leftarrow \mathbf{X}_j - \epsilon\frac{\partial S}{\partial \mathbf{X}_j},\ j = 1,2. \tag{31}$$

ここで ϵ は微小な正の実数 $0 < \epsilon < 1$ である．$\partial S/\partial\mathbf{X}_j$ は

$$\begin{aligned}
\frac{\partial S}{\partial \mathbf{X}_j} &= \frac{\partial S}{\partial \mu}\frac{\partial \mu}{\partial d_j}\frac{\partial d_j}{\partial \mathbf{X}_j} \\
&= (-1)^j\frac{\partial f}{\partial \mu}\frac{4d_{3-j}}{(d_1+d_2)^2}(\boldsymbol{x} - \mathbf{X}_j\boldsymbol{b}_j)\boldsymbol{b}_j^\top\ (j = 1,2) && (32)
\end{aligned}$$

として与えられるので，二乗誤差最小パターンとの距離を利用した学習は $\delta_j = (-1)^j\epsilon\frac{\partial f}{\partial \mu}\frac{d_{3-j}}{(d_1+d_2)}$ $(j = 1,2)$ とおけば

$$\mathbf{X}_j \leftarrow \mathbf{X}_j - \delta_j(\boldsymbol{x} - \mathbf{X}_j\boldsymbol{b}_j)\boldsymbol{b}_j^\top \tag{33}$$

と書ける．したがって，$\mathbf{X}_j$ をなんらかの方法で初期化し，学習データ $\boldsymbol{x}$ に対する誤分類が少なくなるまで式 (33) の更新を行えば良い．なお，式 (32) から式 (33) にかけて $d_j/(d_1+d_2)$ へと変更がなされているが，この変更により ϵ の決定が容易となる（収束には影響しない) [11]. また，$\mathbf{X}_j$ の更新を $\boldsymbol{x}$ の k 近傍に限定した場合には LSC のための学習則となる．なお，式 (2) を未知データと線型多様体との距離とした場合の更新式は

$$\begin{aligned}
\boldsymbol{x}_1^j &\leftarrow \boldsymbol{x}_1^j - \delta_j(\boldsymbol{x} - (\boldsymbol{x}_1^j + \mathbf{V}_j\boldsymbol{\beta}_j)) \\
\mathbf{V}_j &\leftarrow \mathbf{V}_j - \delta_j(\boldsymbol{x} - (\boldsymbol{x}_1^j + \mathbf{V}_j\boldsymbol{\beta}_j))\boldsymbol{\beta}_j^\top && (34)
\end{aligned}$$

となり，GLVQ を利用した学習部分空間法 [17] となる．

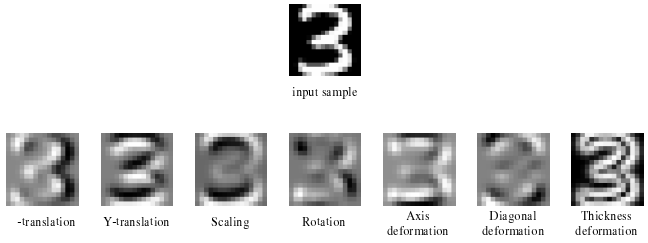


図 3　上段が未知データ $\boldsymbol{q}$，下段が 7 種類の変形を表す接ベクトル $\mathbf{U}$

4.2　線型多様体間距離に基づく学習則

次に，未知データも線型多様体で与えられる場合の学習則について考える．ラベルが既知の学習データが m 個のベクトルの組で与えられたとし，それらによって張られた線型多様体を $\mathcal{M}_x = \boldsymbol{x}_1 + \mathbf{U}\boldsymbol{\alpha}$ で表す．$\mathcal{M}_x$ からクラス j の線型多様体までの距離を $d_j = \|\boldsymbol{x}_1 + \mathbf{U}\boldsymbol{\alpha} - \mathbf{X}_j\boldsymbol{b}_j\|^2$ とする．ただし $\boldsymbol{\alpha}$ や $\boldsymbol{b}_j$ は式 (12) や式 (13) で得られるものとする．$\mathbf{X}_1$ を $\mathcal{M}_x$ と同じクラスに属する最近傍線型多様体を張る訓練データ，$\mathbf{X}_2$ を $\mathcal{M}_x$ と異なるクラスに属する最近傍線型多様体を張る訓練データとし，それらを用いて計算された距離をそれぞれ d_1，d_2 とする．二乗誤差最小パターンとの距離を利用した学習と同様にして線型多様体間距離に基づく学習を導けば，結局それは以下の更新によって実現できる：

$$\mathbf{X}_j \leftarrow \mathbf{X}_j - \delta_j(\boldsymbol{x}_1 + \mathbf{U}\boldsymbol{\alpha}_j - \mathbf{X}_j\boldsymbol{b}_j)\boldsymbol{b}_j^\top \qquad (35)$$

なお，式 (14) を線型多様体間距離とした場合には，更新式は

$$\begin{aligned}\boldsymbol{x}_1^j &\leftarrow \boldsymbol{x}_1^j - \delta_j(\boldsymbol{x}_1 + \mathbf{U}\boldsymbol{\alpha} - (\boldsymbol{x}_1^j + \mathbf{V}_j\boldsymbol{\beta}_j))\\ \mathbf{V}_j &\leftarrow \mathbf{V}_j - \delta_j(\boldsymbol{x}_1 + \mathbf{U}\boldsymbol{\alpha} - (\boldsymbol{x}_1^j + \mathbf{V}_j\boldsymbol{\beta}_j))\boldsymbol{\beta}_j^\top \end{aligned} \qquad (36)$$

となる．上記の更新では，学習データが与えられる毎に $\boldsymbol{\alpha}$ や $\boldsymbol{b}_j$ を最適化しなければならないため，学習に非常に時間がかかるという難点がある．

5.　実　　験

ここでは，公開手書き数字データ USPS [18] を用いた実験結果を示す．USPS は 7291 の訓練データと 2007 の未知データが 16×16 ピクセルのモノクロ画像で与えられている．本実験では位置合わせや正規化等の前処理を一切行っていない．実験では各手法を CPU1.86GHz，メモリ 2GB の標準 PC 上で MATLAB を用いて実装した．パラメータはランダムに選んだ 2000 個の検証データを用いて決定した. 線型多様体距離を利用した実験では，図 3 に示すように未知データを 7 種類の変形を表す接ベクトル [5] を利用して線型多様体で表現した．なお，本稿ではカーネル化した識別による実験は行っていない.

5.1　線型多様体を利用した識別法のエラー率

はじめに，未知データに対するエラー率と 1 パターンあたりの識別時間をさまざまな識別法で調べた．表 1 に各手法のエラー率と識別時間を示す．比較した手法は PDM, 線型多様体間距離を利用した識別 (Manifold Distance Classifier, MDC と略す), LSC (k は近傍数)，1S TD と 2S TD を利用した最近傍決定則である．1S TD と 2S TD では計算量削減のために，予めテストデータの 100 近傍をユークリッド距離で選択した後，選択したパターンとの接距離を計算して最近傍決定則を行った．なお，PDM と MDC では各クラスで訓練データから主成分分析により線型多様体を求め，式 (2) と式 (14) で正則化無しの分類を行った．表 1 の n_j は各クラスの線型多様体の次元数を表す.

表 1　未知データに対するエラー率と識別時間

method	test error [%]	time [s]
PDM ($n_j = 30$)	5.1	0.001
MDC ($n_j = 20$)	4.2	0.01
LSC (k=11)	3.9	0.05
1S TD	3.3	0.1
2S TD	2.4	0.13
kNN ($k = 3$)	5.3	0.2
SVM (# SVs= 3220)	4.6	0.005

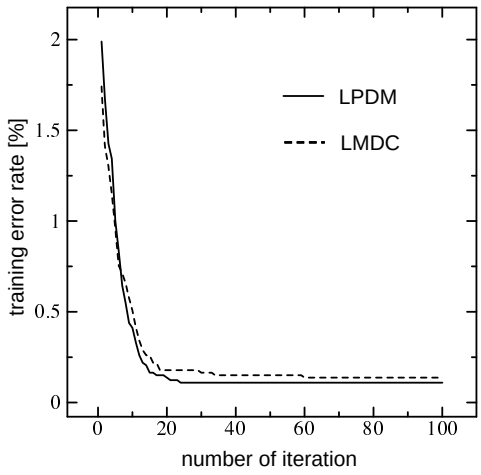


図 4　学習データに対するエラー率と学習回数との関係

比較のためユークリッド距離による kNN と RBF カーネルを用いた Support Vector Machine (SVM) の結果も示す．SVM は SVMLIB [19] を利用した．表 1 より PDM が最も高速であるが，エラー率は SVM よりも高い．一方，MDC の識別時間は SVM よりも 2 倍遅いがエラー率は低く，使用するメモリ量 (辞書サイズが 200) も SVM (サポートベクトルの数が 3220) よりも少ない．LSC や 1S TD，2S TD のエラー率は低いが，全訓練データを利用しているため識別に時間がかかり，必要なメモリ量も大きい.

5.2　学習の効果

次に，**4** で述べた学習の効果について実験した．実験では式 (34) と式 (36) を利用した学習 PDM (Learning PDM, LPDM) と学習 MDC (LMDC) について検証を行った．線型多様体の初期値は各クラスで主成分分析によって得られた重心と直交基底を用いた．次元数は表 1 で示したものと同じにしたので識別時間の変化は無い．学習係数は $\epsilon = 10^{-7}$ とし，正則化は行わなかった．学習では全ての訓練データを用いてバッチ型で線型多様体の更新を行った．本稿では一回の更新を一回の学習と呼ぶ.

図 4 に訓練データのエラー率と学習回数との関係を示す．実線が LPDM，点線が LMDC を表している．図から約 30 回でエラー率に変化が見られなくなったので以後の実験では学習回数

表 2　学習後の未知データに対するエラー率と学習時間

method	test error [%]	learning time [s]
LPDM ($n_j = 30$)	4.9	2789
LMDC ($n_j = 20$)	3.7	3437
GLVQ	8.7	37
SVM	4.6	34.9

を 30 回とした．表 2 には LPDM と LMDC の学習後のテストデータに対するエラー率と学習に要した時間を示す．比較のため，ユークリッド距離に基づく GLVQ と SVM の結果も示す．表 2 から，学習により PDM や MDC においてエラー率を低減できたことがわかる．一方，SVM と比較すると学習に要する時間が非常に長く，高速化が望まれる．

6. ま と め

本稿では，線型多様体間距離が未知データとの二乗誤差を最小にする訓練データの線型結合を利用した識別から容易に導かれることを示した．また，一般化学習ベクトル量子化を線型多様体間距離に基づくパターン識別に適用した場合の学習則を導いた．手書き数字を用いた実験により，線型多様体間距離を利用することで低いエラー率を達成でき，学習によって識別精度が改善されることを示した．

本稿では，相互部分空間法との比較を行っていないが，識別精度に関しては (CLAFIC と投影距離法の場合と同様に) あまり差は無いと思われる．実際，文献 [7] では顔認識で比較を行っているが，ほとんど差が無いことが報告されている．線型多様体間距離を用いることの利点は，線型多様体の基底が直交してなくても計算が可能なため，距離に基づく学習の適用やカーネル化などが容易なことが挙げられる．相互部分空間法はこれまでにさまざまな拡張がなされているので，線型多様体間距離に基づく識別も拡張を行うことが今後の課題である．

文　献

[1] 黒沢由明, “Subspace2006 部分空間法入門,” 部分空間法研究会 Subspace2006, pp. 136–143, 2006.

[2] 前田賢一，渡辺貞一，“局所構造を導入したパターン・マッチング法”, 信学論, vol. J68-D, no. 3, pp. 345–352 1985.

[3] 福井和広, 山口修，“部分空間法の理論拡張と物体認識への応用,” 情処論, vol. 46, no. DIG15, pp. 21–34, 2005.

[4] 池田正幸, 田中英彦, 元岡 達, “手書き文字認識における投影距離法,” 情処学論, vol. 24, no. 1, pp. 106–112, 1983.

[5] P.Y. Simard, Y. LeCun, and J.S. Denker, “Efficient pattern recognition using a new transformation distance,” Proc. of NIPS, vol. 5, pp. 50–58, 1993.

[6] A. Fitzgibbon and A. Zisserman, “Joint manifold distance: A new approach to appearance based clustering,” IEEE Int. Conf. on CVPR, 2003.

[7] J.H. Chen, S.L. Yeh, and C.S. Chen, “Inter-subspace distance: A new method for face recognition with multiple samples,” Proc. of ICPR04, vol. 3, pp. 140–143, 2004.

[8] J. Laaksonen, “Subspace classifiers in recognition of handwritten digits,” PhD thesis, Helsinki University of Technology, 1997.

[9] J.-T. Chien and C.-C. Wu, “Discriminant waveletfaces and nearest feature classifiers for face recognition,” IEEE Trans. Pattern Anal. Machine Intell., vol. 24, no. 12, pp. 1644–1649, 2002.

[10] E. Oja, “Subspace methods of pattern recognition,” Research Studies Press, 1983. (小川英光，佐藤 誠訳，パターン認識と部分空間法，産業図書，1986)

[11] A. Sato and K. Yamada, “Generalized learning vector quantization,” *Prop. of NIPS*, 7:423–429, 1995.

[12] 前田 英作, 村瀬 洋, “カーネル非線形部分空間法によるパターン認識,” 信学論, vol. J82-D-II, no. 4, pp. 600–612, Apr. 1999.

[13] B. Schölkopf, A.J. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” Neural Computation, vol. 10, pp. 1299–1319, 1998.

[14] 坂野 鋭, 武川直樹，中村太一, “核非線型相互部分空間法による物体認識,” 信学論, vol. J84-D-II, no. 8, pp. 1549–1556, 2001.

[15] T. Kohonen, “Self-organizing map,” Springer-Verlag, 1995.

[16] L. Yang and R. Jin, “Distance metric learning: A comprehensive survey,” Technical report, Department of Computer Science and Engineering, Michigan State University, 2006.

[17] 佐藤 敦, 山田敬嗣, “一般学習ベクトル量子化に基づく学習部分空間法,” 信学全大, p. 236, 1997.

[18] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel, “Backpropagation applied to handwritten zip code recognition,” Neural Computation, vol.1, no.4, pp.541–551, 1989.

[19] C.C. Chang and C.J. Lin, “LIBSVM: A library for support vector machines,” 2001.
`http://www.csie.ntu.edu.tw/~cjlin/libsvm`

重み付き相関行列による局所部分空間法

山下　幸彦†

† 東京工業大学大学院理工学研究科　〒 152–8552 東京都目黒区大岡山 2–12–1–S6–19
E-mail: †yamasita@ide.titech.ac.jp

あらまし　飯島，Watanabe によって独立に提案された部分空間法は，考え方が分かりやすいこと，性能が高いこと，実現が容易であることなどの理由により，文字認識，顔画像認識など極めて広い範囲で応用されている．本稿では，筆者らが提案してきた部分空間法に関係するパターン認識法として，相対 Karhunen-Loève 変換法，カーネル Wiener フィルタ法を紹介する．また，高次相関から計算される重み付き相関行列による局所部分空間法を提案する．そして，その有効性を計算機実験によって確認する．

キーワード　局所部分空間法，重み付き相関行列，部分空間法，相対 Karhunen-Loève 変換法，カーネル Wiener フィルタ法

Local subpsace method with weighted correlation matrix

Yukihiko YAMASHITA†

† Graduate School of Engineering and Science, Tokyo Institute of Technology　O-okayama 2–12–1–S6–19, Meguro-ku, Tokyo, 152–8552 Japan
E-mail: †yamasita@ide.titech.ac.jp

Abstract　The subspace method originally proposed by Iijima and Watanabe is used for many areas in pattern recognition since its concept is simple, its performance is high, and its implementation is not complex. In this paper, we introduce the relative Karhunen-Loève method and the kernel Wiener method that are classification methods related to the subspace method, and are proposed by the authors. We propose a new local subspace method based on a weighted correlation matrix that can be calculated from the third or higher order correlations. We show its advantages by experimental results.

Key words　Local subpsace method, weighted correlation matrix, subspace method, relative Karhunen-Loève transform method, kernel Wiener filter method

1. はじめに

飯島 [1], [2]，Watanabe [3] によって独立に提案された部分空間法は，考え方が分かりやすいこと，性能が高いこと，実現が容易であることなどの理由により，文字認識，顔画像認識など極めて広い範囲で応用されている．また，この部分空間法を改良するために，学習部分空間法，相互部分空間法，局所部分空間法，非線型部分空間法などが提案されている．

基本的な部分空間法では，認識に用いられる部分空間は主成分分析 (PCA)，Karhunen-Loève 変換 (KLT)，または，モード関数展開によって求められている．PCA の計算のためには，固有値問題を解くことが必要となるが，パターンベクトルの次元が高い場合は計算が困難になる．また，直接的にカテゴリーの分離に適した部分空間を求めているのではないという問題も存在する．学習部分空間法は，カテゴリーを分離し，認識に適する部分空間を繰り返し法によって逐次的に求める手法である．そして，単純な計算の繰り返しであるため，パターンベクトルの次元が高い空間でも計算が可能であるという特徴を持っている．

PCA ではカテゴリーの平均的な特徴を捉えている．従って，平均的でないパターンに対して認識率が劣化する．局所部分空間法は，クラスタリング等により，カテゴリーをサブカテゴリーに分解し，そのサブカテゴリーを独立のカテゴリーとして部分空間法を適応するものである [4]〜[6]．平均的でない類似のパターンを集め，ひとつのカテゴリーとすることによって認識率を向上させることができる．しかしながら，このような手法には，クラスタリングアルゴリズムやその初期値によって，結果が変わってしまうという問題が存在する．また，適応局所部分空間法 [7] は，入力パターンに近い学習パターンで相関行列を構成するものである．この手法では，認識時に学習パター

ン数オーダーの計算が必要になるという問題がある．

相互部分空間法は，ひとつの対象に対して複数のパターンが得られる場合に，部分空間法を拡張したものである [8]〜[10]．入力パターンベクトルに対して部分空間を構成し，カテゴリーを表す部分空間とのなす角 (複数) を評価することによって，パターン認識を行うものである．実際に，顔画像認識などに応用され，その有効性が確認されている．

サポートベクトルマシンのために導入されたカーネル法は，パターンベクトルを非線型写像によって Hilbert 空間に写像し，その Hilbert 空間で線形あるいは 2 次識別を行う手法である [11], [12]．そして，パターンベクトルの写像先の Hilbert 空間における内積をカーネル関数で与えることができれば，認識器を構成するための計算が，ヒルベルト空間でなく，学習パターン数次元のベクトル空間で実行できる [13]．このことによって，計算量を大幅に削減することができる．非線形部分空間法は，部分空間法にカーネル法を導入したものであり，部分空間はカーネル PCA(KPCA) によって計算される．非線型写像によって，入力パターンベクトルの空間で複雑な境界を表現することが可能で，認識率を向上させることができる．しかしながら，学習パターン数が非常に多くなることがあり，そのような場合は，計算量が大きくなりすぎ，実質的に実現不可能になる．

本稿では，筆者らが提案してきた部分空間法に関係するパターン認識法として，相対 Karhunen-Loève 変換法 (RKL 変換法) [14]〜[16]，カーネル Wiener フィルタ法 (KWF 法) [17], [18] を紹介すると共に，重み付き相関行列による局所部分空間法を提案する．

PCA により得られる正射影行列を，そのカテゴリーの特徴を抽出するものと考える．このとき，RKL 変換法により得られる行列は，自カテゴリーの特徴を抽出しながら，他カテゴリーの特徴は抽出しないようにしたものである．そのため，カテゴリーの分離に適した特徴を抽出できると考えられる．

KWF 法は，信号復元に使われる Wiener フィルタを，カーネル法を使って拡張し，パターン認識に応用したものである．カテゴリーごとにそのカテゴリーを代表する原信号を選び，入力パターンベクトルを観測信号と考える．識別は，その観測信号からのカテゴリーを表す原信号への復元と考えることによって実現する．さらに，KWF に元のパターンベクトルの空間における距離に従った正則化法を導入している．

さらに，本稿では，入力パターンベクトルと学習パターンベクトルとの内積で，その学習パターンに重みを付けた相関行列による部分空間法を提案する．本手法では，この重み付き相関行列を使って PCA を計算し部分空間を求める．相関行列は，この重みによって入力パターン周辺の学習パターンの性質を強くを表していることになる．従って，本手法は局所部分空間法の一つと考えることができる．また，非線型な識別法ではあるが，クラスタリングのような問題が生じることはなく，安定した認識結果が得られる．そして，高次相関を計算しておけば，認識時に学習パターン数のオーダの計算を不要にすることができる．最後に，計算機実験によって本手法の有効性を示す．

2. 相対 Karhunen-Loève 変換法 (RKL 変換法)

R^N を，パターンベクトルが存在する N 次元ベクトル空間とする．R^N のベクトルは列ベクトルで表されているものする．R^N のベクトル $\boldsymbol{f}$ のノルムと転置したものを，それぞれ，$\|\boldsymbol{f}\|$ and $\boldsymbol{f}^{\mathrm{T}}$ で表す．R^N の 2 つのベクトル $\boldsymbol{f}$，$\boldsymbol{g}$ の内積を，$\langle \boldsymbol{f}, \boldsymbol{g} \rangle (= \boldsymbol{f}^{\mathrm{T}} \boldsymbol{g})$ によって表す．

部分空間法では，各カテゴリー $\Omega^{(c)}$ $(c = 1, 2, \ldots, C)$ に対して部分空間 $S^{(c)}$ を用意する．ここで，C はカテゴリー数である．$P^{(c)}$ を $S^{(c)}$ への正射影行列とする．未知パターンベクトル $\boldsymbol{x}$ の識別結果となるカテゴリーの番号を $D(\boldsymbol{x})$ で表せば，

$$D(\boldsymbol{x}) = \underset{c=1,2,\ldots,C}{\operatorname{argmax}} \|P^{(c)} \boldsymbol{x}\| \tag{1}$$

となる．

カテゴリー $\Omega^{(c)}$ に属しているパターンベクトル $\boldsymbol{f}$ に対する，母集団期待値を $E_{f \in \Omega^{(c)}}$ で表す．PCA によって得られる正射影行列 $P^{(c)}$ は，そのランクが一定であるという条件の元で，次の評価基準を最大にするものとして特徴づけることができる．

$$E_{f \in \Omega^{(c)}} \|P^{(c)} \boldsymbol{f}\|^2 \tag{2}$$

$\Omega^{(c)}$ の相関行列 $R_{ff}^{(c)}$ は，

$$R_{ff}^{(c)} = E_{f \in \Omega^{(c)}} \boldsymbol{f} \boldsymbol{f}^{\mathrm{T}} \tag{3}$$

によって定義される．$R_{ff}^{(c)}$ の固有ベクトルを $\lambda_1^{(c)} \geqq \lambda_2^{(c)} \geqq \cdots \geqq \lambda_N^{(c)}$ で表し，それに対応する固有ベクトルを $\phi_1^{(c)}, \phi_2^{(c)}, \ldots, \phi_N^{(c)}$ で表す．ここで，一般性を失うことなく，固有ベクトルは正規直交基底をなしているものと仮定することができる．n を N 以下の自然数とする．このとき，PCA によって与えられる部分空間法のための n 次元部分空間 $S^{(c)}$ は，$\phi_1^{(c)}, \phi_2^{(c)}, \ldots, \phi_n^{(c)}$ によって張られる．

式 (2) を最大にする $P^{(c)}$ は，ランク一定の条件のもとで，

$$E_{f \in \Omega^{(c)}} \|X^{(c)} \boldsymbol{f} - \boldsymbol{f}\|^2 \tag{4}$$

を最小にする $X^{(c)}$ と本質的には同じものになる．すなわち，ランクが一定という条件の元で，平均 2 乗的にカテゴリのパターンを最良に近似できるものとなり，カテゴリーの特徴を抽出していると考えることができる．

しかしながら，式 (4) には他カテゴリーのことが考慮されていないため，この $P^{(c)}$ は他のカテゴリーの特徴も抽出する可能性がある．従って，他のカテゴリのパターンベクトル $\boldsymbol{g}$ に対しても，$\|P^{(c)} \boldsymbol{g}\|$ の値が大きくなる可能性がある．この問題を解決するために，RKL 変換法が提案された．RKL 変換は，ランク一定の元で，

$$E_{f \in \Omega^{(c)}} \|X^{(c)} \boldsymbol{f} - \boldsymbol{f}\|^2 + \alpha E_{l \neq c} E_{g \in \Omega^{(l)}} \|X^{(c)} \boldsymbol{g}\|^2 \tag{5}$$

を最小にする行列 $X^{(c)}$ として定義される．ここで，$E_{l \neq c}$ は，$\Omega^{(c)}$ 以外のカテゴリーの出現確率による期待値であり，α は正の定数 (パラメータ) である．

式 (5) の第 1 項は，$X^{(c)}$ が抽出した $\Omega^{(c)}$ の特徴による近似誤差の 2 乗平均を表している．式 (5) の第 2 項は，$X^{(c)}$ が抽出した $\Omega^{(c)}$ 以外のカテゴリーの特徴量の 2 乗平均を表している．$X^{(c)}$ が，自分のカテゴリーの特徴は正確に抽出し，他カテゴリーの特徴を抽出しないようにするためには，この 2 つの項を小さくすれば良い．従って，式 (5) では，その 2 項のどちらにどれだけ重点をおくか決めるためのパラメータ α を第 2 項にかけて和をとったものとした．

RKL 変換と式が類似したものに，ランク低減 WF が提案されている [19]．この場合，式 (5) の $\boldsymbol{f}$ と $\boldsymbol{g}$ は，それぞれ，同じ観測信号の信号成分と雑音成分を表すが，RKL 変換の場合は，異なるカテゴリーの全く別のパターンベクトルを表している．

RKL 変換法によって，他カテゴリーとの境界付近のパターンに対する認識率の向上が期待されたが，向上の幅はあまり大きくなかった．平均的な値，例えば，$E_{f\in\Omega^{(c)}}\|X^{(c)}\boldsymbol{f}\|^2$ と $E_{l\neq c}E_{g\in\Omega^{(l)}}\|X^{(c)}\boldsymbol{g}\|^2$ の比などは向上している．しなしながら，誤認識するパターンは，カテゴリの平均を使った評価では捉えることが難しいため，認識率があまり改善しなかったものと考えられる．

3. カーネル Wiener フィルタ法 (KWF 法)

カーネル法では，非線型写像 Φ を使って，もとのパターンや信号が存在する空間から，特徴空間と呼ばれるヒルベルト空間に写像し，その空間で処理を行う手法である．このとき，

$$k(\boldsymbol{x},\boldsymbol{y}) = \langle \Phi(\boldsymbol{x})\Phi(\boldsymbol{y})\rangle \tag{6}$$

を満たす，カーネル関数 $k(\boldsymbol{x},\boldsymbol{y})$ が与えられていれば，ヒルベルト空間での処理をカーネル関数を使った処理に書き換えることができる．PCA にカーネル法を適用した KPCA を使った部分空間法が提案され，その有効性が確認されている．

本稿では，カーネル Wiener フィルタを用いたパターン認識に関して説明する．Wiener フィルターは，劣化などを受けた観測信号から，原信号を推定するための線形復元フィルタである．原信号空間および観測画像空間から，それぞれの特徴空間への写像を，Φ_S および Φ_O とおく．学習信号として，L 個の原信号と観測信号の組 $\{(\boldsymbol{x}_l,\boldsymbol{y}_l)\}_{l=1}^{L}$ が与えられているとすれば，KWF は，

$$\frac{1}{L}\sum_{i=1}^{L}\|X\Phi_O(\boldsymbol{y}_l)-\Phi_S(\boldsymbol{x}_l)\|^2 \tag{7}$$

を最小にする線形作用素 X として与えられる．

これをパターン認識に応用する．カテゴリー数を C とし，各カテゴリー $\Omega^{(c)}$ $(c=1,2,\ldots,C)$ に対して，標準ベクトル $\boldsymbol{s}_c$ を定める．L 個の学習パターンベクトル $\boldsymbol{f}_l$ $(l=1,2,\ldots,L)$ と，そのカテゴリ $\Omega_{c(l)}$ が与えられているときに，

$$\frac{1}{L}\sum_{i=1}^{L}\|X\Phi_O(\boldsymbol{f}_l)-\Phi_S(\boldsymbol{s}_{c(l)})\|^2 \tag{8}$$

を最小にする作用素 X を使ってパターン認識を行う．未知パターンベクトル $\boldsymbol{x}$ に対する認識結果は，

$$D(\boldsymbol{x}) = \underset{c=1,2,\ldots,C}{\operatorname{argmin}}\|X\boldsymbol{x}-\boldsymbol{s}_c\| \tag{9}$$

によって与えられる．

正則化は，パターンの微小な変動に対して結果が大きく変化しないようにし，オーバーフィッティングを防ぐための有効な手段である．この正則化を特徴空間ではなく，元の空間の距離に従って行うことを考える．そのため，学習パターンベクトル $\boldsymbol{f}_l$ に仮想的な雑音 $\boldsymbol{e}_l$ が加算されるものと考える．この $\boldsymbol{e}_l$ は $\boldsymbol{f}_l$ と独立で，$\boldsymbol{e}_l$ は無相関であるものとする．このとき，

$$\frac{1}{L}\sum_{i=1}^{L}E_{\boldsymbol{e}_l}\|X\Phi_O(\boldsymbol{f}_l+\boldsymbol{e}_l)-\Phi_S(\boldsymbol{s}_{c(l)})\|^2 \tag{10}$$

を最小にする線形作用素 X を考える．ここで，$E_{\boldsymbol{e}_l}$ は，$\boldsymbol{e}_l$ に関する期待値を表している．

通常のカーネル法では，$\boldsymbol{e}_l$ に対しても標本平均を計算する必要がある．しかしながら，その方法では計算量が膨大になるため，$\boldsymbol{e}_l$ の大きさが小さいと考え，Φ_O をその 1 次の Taylor 級数による近似に置き換え，式 (10) を評価する．Φ_O を微分したものの内積は，カーネル関数を微分して得ることができる．このとき，Φ_O の 1 次微分したものどうしの内積は，カーネル関数の 2 次微分になる．この項を導入すると，逆行列を計算する行列の次数が正則化しない KWF のものよりも大幅に増加するため，省略して解を求めている．

表 1 に，13 個の 2 クラス標準データセット [20] に対する，KWF 法の認識実験の結果を示す．KWF 法には，元空間の距離に従った正則化法を導入している．パラメータは，5 fold のクロスバリデーションによって求めている．表の中で太文字は，SVM，AdaBoost Regression，KWF の中で誤認識率が最も低いもの示している．また，KWF における*は，優位水準 5%の t 検定で，SVM に対して優位にあるものを示している．表 1 より，KWF 法の有効性が確認できる．

4. 重み付き相関行列による局所部分空間法

はじめに，重み付き相関行列による局所部分空間法の最も簡単なものについて説明する．$\boldsymbol{x}$ を入力パターンベクトルとする．カテゴリー $\Omega^{(c)}$ の重き付き相関行列 $R^{(c)}(\boldsymbol{x})$ を，

$$R^{(c)}(\boldsymbol{x}) = E_{f\in\Omega^{(c)}}\langle \boldsymbol{x},\boldsymbol{f}\rangle \boldsymbol{f}\boldsymbol{f}^{\mathrm{T}} \tag{11}$$

で定義する．パターンベクトルがノルムで正規化されているとすれば，重み $\langle \boldsymbol{x},\boldsymbol{f}\rangle$ は，$\boldsymbol{f}$ と $\boldsymbol{x}$ の距離が近い方が大きくなる．従って，式 (11) の中で，平均される学習パターンに対する重みは，その学習パターンが入力パターンに近いほど大きくなる．カテゴリー $\Omega^{(c)}$ を表す部分空間 $S^{(c)}(\boldsymbol{x})$ は，$R^{(c)}(\boldsymbol{x})$ から PCA によって得られた固有ベクトルで張られる空間とする．$P^{(c)}(\boldsymbol{x})$ を部分空間 $S^{(c)}(\boldsymbol{x})$ への正射影行列とする．識別結果は，

$$D(\boldsymbol{x}) = \underset{c=1,2,\ldots,C}{\operatorname{argmax}}\|P^{(c)}(\boldsymbol{x})\boldsymbol{x}\| \tag{12}$$

によって与えられる．$P^{(c)}(\boldsymbol{x})$ は $\boldsymbol{x}$ に依存するため，全体としては非線型な識別法となる．

ベクトル $\boldsymbol{f}$，行列 R，3 階のテンソル T，4 階のテンソル U

表 1　Error rate of KWF-1 and KWF-2'

	Data set Name	error rate SVM	error rate AdaBoost Reg	error rate KWF'
1	banana	11.53 ± 0.66	10.85 ± 0.42	$\mathbf{10.29 \pm 0.39}$ *
2	breast cancer	26.04 ± 4.47	26.51 ± 4.47	$\mathbf{24.45 \pm 4.29}$ *
3	diabetis	23.53 ± 1.73	23.79 ± 1.80	$\mathbf{22.97 \pm 1.83}$ *
4	flare solar	$\mathbf{32.43 \pm 1.82}$	34.20 ± 2.18	33.18 ± 1.75
5	german	$\mathbf{23.61 \pm 2.07}$	24.34 ± 2.08	23.63 ± 2.24
6	heart	15.95 ± 3.26	16.47 ± 3.51	$\mathbf{15.46 \pm 3.30}$
7	image	2.96 ± 0.60	$\mathbf{2.67 \pm 0.61}$	2.69 ± 0.53
8	ringnorm	1.66 ± 0.12	$\mathbf{1.58 \pm 0.12}$	1.62 ± 0.10 *
9	splice	10.88 ± 0.66	$\mathbf{9.50 \pm 0.65}$	10.83 ± 0.67
10	thyroid	4.80 ± 2.19	4.55 ± 2.19	$\mathbf{3.61 \pm 2.05}$ *
11	titanic	22.42 ± 1.02	22.64 ± 1.20	$\mathbf{22.00 \pm 0.75}$ *
12	twonorm	2.96 ± 0.23	2.70 ± 0.24	$\mathbf{2.34 \pm 0.11}$ *
13	waveform	9.88 ± 0.43	9.79 ± 0.81	$\mathbf{9.50 \pm 0.35}$ *

* : superior to SVM by t-test of significance level 5%

に対して，第 i, (i,j), (i,j,k), and (i,j,k,l) 成分を，それぞれ，f_i，R_{ij}，T_{ijk}，U_{ijkl} で表す．

認識時における $R^{(c)}(\boldsymbol{x})$ の計算を高速化するために，3 次相関テンソル $R^{(c)}_{fff}$ を，

$$(R^{(c)}_{fff})_{ijk} = E_{f\in\Omega^{(c)}} f_i f_j f_k \tag{13}$$

によって与える．このとき，$R^{(c)}(\boldsymbol{x})$ は

$$R^{(c)}(\boldsymbol{x})_{ij} = \sum_{k=1}^{N} (R^{(c)}_{fff})_{ijk} x_k \tag{14}$$

によって計算することができる．式 (14) を使えば，認識時に $R^{(c)}(\boldsymbol{x})$ を求めるために，学習パターンごとに計算する必要がなくなる．

計算量を削減したり，効率的な重みを構成するために，部分空間を張るためのパターンベクトルと，重みのためのパターンベクトルを分けることができる．また，部分空間法を行うために，前処理等を行うことも多い．それを一般的に表すために，2 つの変換 $\boldsymbol{p}(\,\cdot\,)$ および $\boldsymbol{q}(\,\cdot\,)$ を考える．前者を部分空間を構成するベクトルを生成する変換，後者を相関行列の重みのためのベクトルを生成するための変換とする．N および M を，それぞれ，$\boldsymbol{p}(\,\cdot\,)$ および $\boldsymbol{q}(\,\cdot\,)$ の像の次元とする．今回，扱う相関行列は，$n = 1, 2$ に対して，

$$(R^{(c)}(\boldsymbol{x}) = E_{f\in\Omega^{(c)}} (\langle \boldsymbol{q}(\boldsymbol{x}), \boldsymbol{q}(\boldsymbol{f})\rangle + a)^n \boldsymbol{p}(\boldsymbol{f})\boldsymbol{p}(\boldsymbol{f})^{\mathrm{T}} \tag{15}$$

である．

$$\left(R^{(c)}_{p(f)p(f)}\right)_{ij} = E_{f\in\Omega^{(c)}} p(\boldsymbol{f})_i p(\boldsymbol{f})_j \tag{16}$$

$$\left(R^{(c)}_{p(f)p(f)q(f)}\right)_{ijk} = E_{f\in\Omega^{(c)}} p(\boldsymbol{f})_i p(\boldsymbol{f})_j q(\boldsymbol{f})_k \tag{17}$$

$$\left(R^{(c)}_{p(f)p(f)q(f)q(f)}\right)_{ijkl} = E_{f\in\Omega^{(c)}} p(\boldsymbol{f})_i p(\boldsymbol{f})_j q(\boldsymbol{f})_k q(\boldsymbol{f})_l \tag{18}$$

とおく．式 (15) で $n = 1$ のときは，

$$R^{(c)}(\boldsymbol{x})_{ij} = \sum_{k=1}^{M} \left(R^{(c)}_{p(f)p(f)q(f)}\right)_{ijk} q(\boldsymbol{x})_k + a\left(R^{(c)}_{p(f)p(f)}\right)_{ij} \tag{19}$$

によって計算できる．同様に，式 (15) で $n = 2$ のときは，

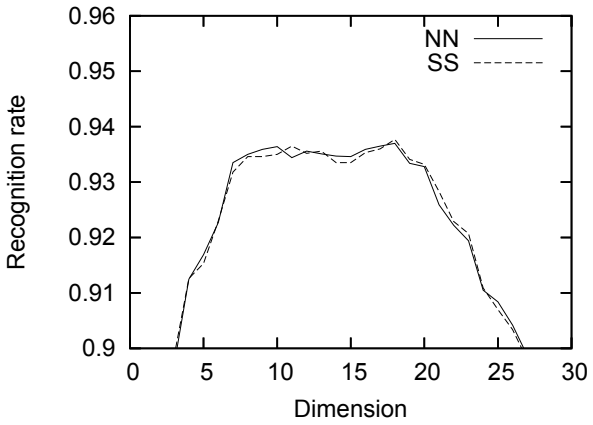
$$\begin{aligned} &R^{(c)}(\boldsymbol{x})_{ij} \\ &= \sum_{k=1}^{M}\sum_{l=1}^{M} \left(R^{(c)}_{p(f)p(f)q(f)q(f)}\right)_{ijkl} q(\boldsymbol{x})_k q(\boldsymbol{x})_l \\ &\quad + 2a\sum_{k=1}^{M} \left(R^{(c)}_{p(f)p(f)q(f)}\right)_{ijk} g(\boldsymbol{x})_k + a^2 \left(R^{(c)}_{p(f)p(f)}\right)_{ij} \end{aligned} \tag{20}$$

によって計算できる．$\boldsymbol{q}(\,\cdot\,)$ に，全カテゴリーの PCA などを適用し，写像して得られるベクトルの次元 M を小さくすれば，計算量を削減することができる．また，全カテゴリーに対するマハラノビス距離を使うなどすれば，パターンの分布を均一化させ，重みの作用が効果的になると考えられる．

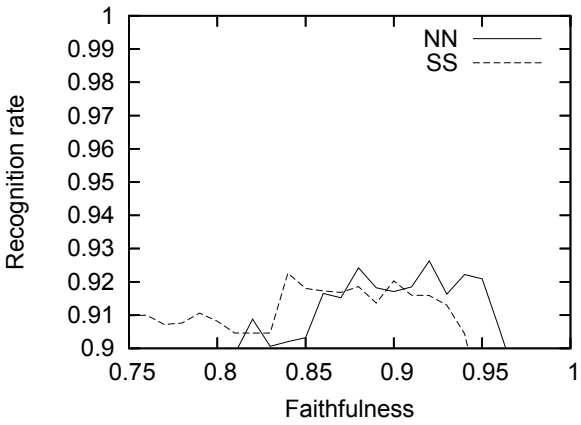
計算量に関して検討する．計算量のオーダが L のとき，$O(L)$ と表す．提案手法で，認識時の処理において計算量が大きい部分は，重み付き相関行列の計算と固有ベクトルの計算である．重み付き相関行列のための計算量は，$n = 1, 2$ のとき，それぞれ，$O(MN^2)$，$O(M^2N^2)$ である．また，固有ベクトルの計算量は $O(N^3)$ である．従って，$n = 1$ の場合は，$M = N$ の場合でも，固有ベクトルの計算量と同程度であり，固有ベクトル計算を行う識別法ならば本手法が適用可能である．$n = 2$ の場合は，パターンベクトルの次元が特に低い場合を除いて，次元圧縮などによって $M << N$ とする必要があると考えられる．

5. 計算機実験

通常の部分空間法と重み付き相関行列による局所部分空間法を比較する計算機実験を行う．元となるデータとして，手書き数字のデータセットである MNIST を用いる [21]．データ数は，学習用 60,000，およびテスト用 10,000 で，それぞれ，(高次) 相関，および認識率を求めるために用いる．今回は，両識別法の比較が目的であるため，エッジ特徴抽出などの特別な特徴抽出法は用いていない．1 つのパターンは 784 次元 (28×28 画素) のベクトルからなる．ここでは計算量削減のため，上下，

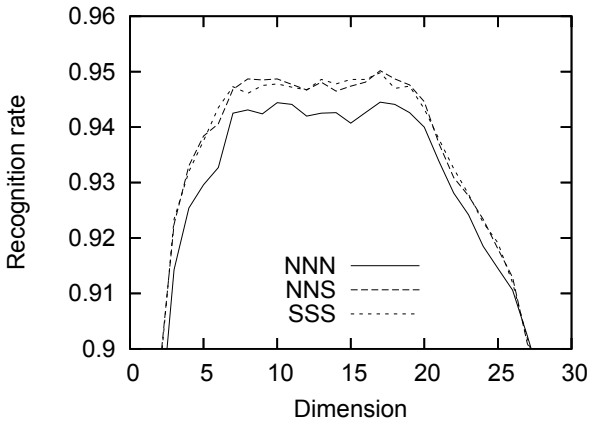


(a) Dimension

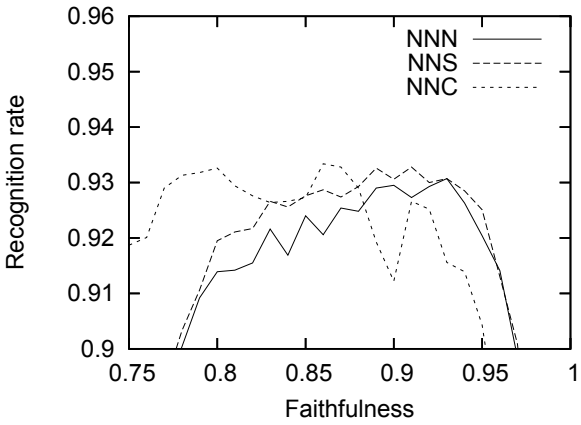


(b) Faithfulness

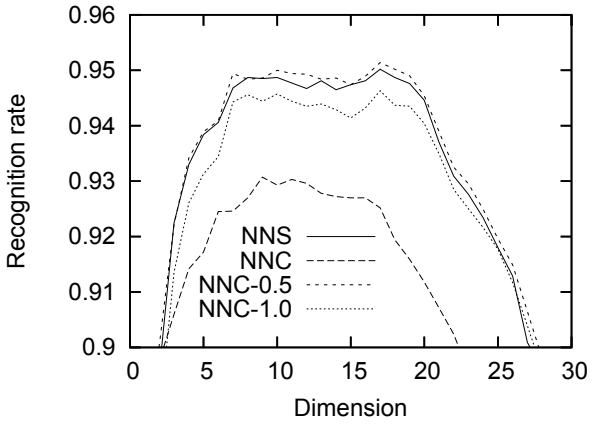
図 1　Subspace method



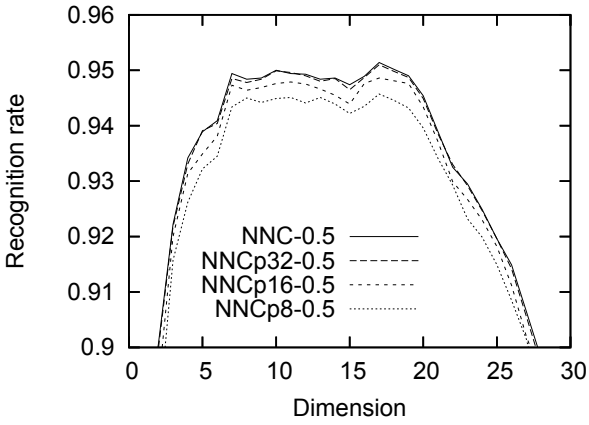
(a) Dimension



(b) Faithfulness

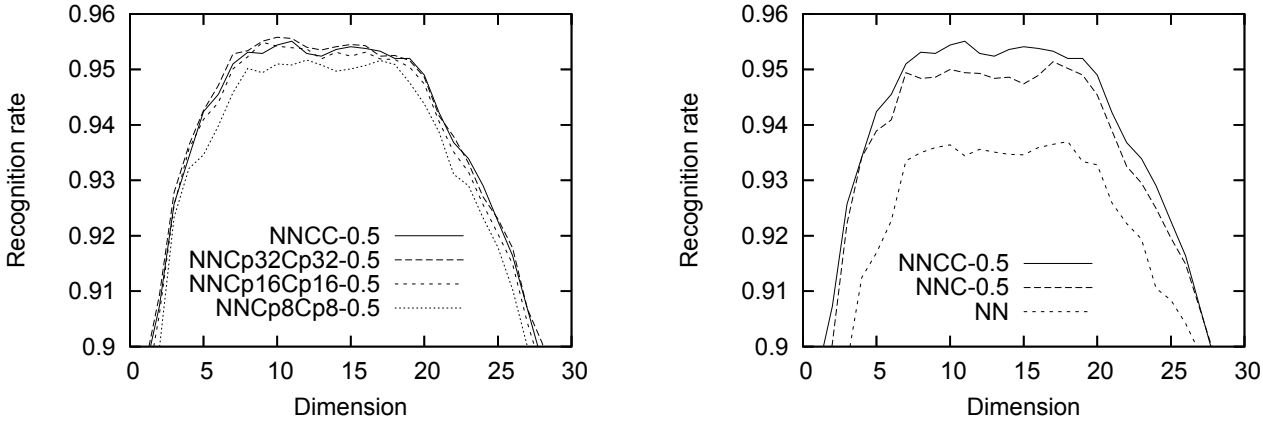


(c) Parameter



(d) Dimension of $g(\boldsymbol{f})$

図 2　Local subspace method with third order correlations



(a) With fourth order correlations (b) Summary

図 3 Local subspace method with fourth order correlations and summary

表 2 Data type and parameters in experiments

Symbol	$\boldsymbol{p}(\boldsymbol{x})$	n	$\boldsymbol{q}(\boldsymbol{x})$	a
'NN'	'N'	0	N.A.	N.A.
'SS'	'S'	0	N.A.	N.A.
'NNN'	'N'	1	'N'	0
'NNS'	'N'	1	'S'	0
'NNC'	'N'	1	'C'	0
'NNC-0.5'	'N'	1	'C'	0.5
'NNC-1.0'	'N'	1	'C'	1.0
'NNC8-0.5'	'N'	1	'C8'	0.5
'NNC16-0.5'	'N'	1	'C16'	0.5
'NNC32-0.5'	'N'	1	'C32'	0.5
'NNCC-0.5'	'N'	2	'C'	0.5
'NNC8C8-0.5'	'N'	2	'C8'	0.5
'NNC8C16-0.5'	'N'	2	'C16'	0.5
'NNC32C32-0.5'	'N'	2	'C32'	0.5

左右の 2 つのラインを切り取り，3×3 画素の値の単純平均を取ることによって，64 次元のベクトルに変換している．

上記のようにして 64 次元に変換したパターンベクトルをノルムで正規化したデータの種類を'N' で表す．種類'N' のデータのそれぞれのベクトルの成分の平均を計算し，全ての要素からその平均を引き，ノルムで正規化したデータの種類を'S' で表す．種類'S' のデータに対して，全てのカテゴリーの全てのパターンの平均ベクトルを計算し，種類'S' のそれぞれのパターンベクトルからその平均ベクトルを引き，ノルムで正規化したデータの種類を'C' で表す．また，種類'C' の全てのカテゴリーの全てのパターンに対する PCA を計算し，ベクトルを k 次元に次元圧縮したデータの種類を'Ck' で表す．

実験結果の種類を示す記号と，使用したデータの種類とパラメータをまとめたものを表 2 に示す．この表で，例えば，実験結果 NNCp16Cp16-0.5 は，部分空間を構成するデータとして種類'N' を，式 (15) の重みの次数 $n = 2$ を，重みを計算するデータとして種類'C16' を，パラメータ a として 0.5 を使用していることを示している．

図 1 は，通常の部分空間法の実験結果である．図 1(a) および (b) は，それぞれ，全てのカテゴリーで部分空間の次元を一定にしたもの，および，全てのカテゴリーで寄与率が一定になるように各カテゴリーの部分空間の次元を定めたものである．これらの図の結果の種類'NN' は，データの種類'N' に対するものであり，'SS' は，データの種類'S' に対するものである．最良の結果 93.64%は，次元を 11 に固定した結果の種類'SS' から得られている．

図 2 は，式 (15) で $n = 1$ のときの提案手法の結果である．2(a) と (b) は，それぞれ，次元一定と寄与率一定の場合の結果である．また，結果の'NNN' および'SSS' は，それぞれ，$\boldsymbol{p}(\boldsymbol{x})$ と $\boldsymbol{q}(\boldsymbol{x})$ の両方にデータの種類'N' および'S' を使ったものである．'NNS' は，$\boldsymbol{p}(\boldsymbol{x})$ および $\boldsymbol{q}(\boldsymbol{x})$ に，それぞれ，データの種類'N' および'S' を使ったものである．最良の結果 95.02%は，次元を 17 に固定した結果の種類'NNS' から得られている．

次元を固定した方が良い結果が得られているため，以下の実験では次元を固定した場合だけについて行うことにする．また，$\boldsymbol{p}(\boldsymbol{x})$ を変更してもあまり結果が変わらないため，$\boldsymbol{p}(\boldsymbol{x})$ には，データの種類'N' だけを使うことにする．

図 2(c) は，$\boldsymbol{q}(\boldsymbol{x})$ にデータの種類'C' を使ったものである．式 (15) の重みの値が負になることがあるため，正の定数 a を加算する．結果の種類'NNC'，'NNC-0.5'，および，'NNC-1.0' は，それぞれ，$a = 0.0, 0.5, 1.0$ のときの結果を表している．最良の結果 95.15%は，次元を 17 に固定した結果の種類'NNC-0.5' から得られている．以下では，$a = 0.5$ に固定して実験を行う．

図 2(d) は，次に計算量を削減するために，重みのためのベクトル $\boldsymbol{q}(\boldsymbol{x})$ に，PCA によって次元削減を行った場合の結果を示す．結果の種類'NNC-0.5'，'NNC32-0.5'，'NNC16-0.5'，および，'NNC8-0.5' は，それぞれ，次元を圧縮しない場合，32，16，8 次元に圧縮した場合を表している．実験結果より，次元を圧縮すると共に認識率は低下するが，32 次元，すなわち，1/2 程度に次元を圧縮しても認識率が大きく変化しないことがわかる．

図 3(a) は，式 (15) で $n = 2$ のときの提案手法の結果である．この場合，4 次相関が必要となり計算量が非常に大きくなる．そのため，重みのためのベクトル $\boldsymbol{q}(\boldsymbol{x})$ に次元圧縮を行った場合の結果も示す．結果の種類 'NNCC-0.5'，'NNC32C32-

0.5’，’NNC16C16-0.5’，および，’NNC8C8-0.5’ は，それぞれ，次元を圧縮しない場合，32，16，8 次元に圧縮した場合を表している．実験結果より，次元を圧縮すると認識率は低下するが，16 次元，すなわち，1/4 程度に次元を圧縮しても認識率が大きく変化しないことがわかる．

図 3(b) は，実験結果をまとめたものである．この図より，重み付き相関行列によって，部分空間法の認識率を向上させることができることがわかる．

6. ま と め

本稿では，筆者らが提案してきた部分空間法に関係するパターン認識法として，RKL 変換法，KWF 法を紹介した．さらに，重み付き相関行列による局所部分空間法を提案し，その有効性を計算機実験によって確認した．本手法は，クラスタリングのような複雑な処理を行うことなく，部分空間法の認識率を向上させることができる．従って，パターンベクトルの次元があまり高くないときに，部分空間法の認識率を手軽に向上させたいときに適していると考えられる．今後の課題として，重みを計算するための次元を削減法，および，クラスタリングしたサブカテゴリーに対する本手法の有効性に関して研究する必要がある．

文　献

[1] 飯島：“パターン認識”, 日刊工業新聞社, 東京 (1969).

[2] 飯島：“視覚情報の基礎理論–パターン認識問題の源流–”, コロナ社, 東京 (1999).

[3] S. Watanabe and N. Pakvasa: “Subspace method in pattern recognition”, Proc. 1st International Joint Conference on Pattern Recognition, Washington DC, pp. 25–32 (1973).

[4] J. Laaksonen: “Local subspace classifier”, Proceedings of the 7th International Conference on Artificial Neural Networks, pp. 637–642 (1997).

[5] 堀田, 喜安, 宮原：“局所部分空間法と変形不変性を用いた画像パターン分類”, 部分空間法研究会 Subspace2006, pp. 111–118 (2006).

[6] H. Cevikalp, D. Larlus, M. Douze and F. Jurie: “Local subspace classifier: Linear and nonlinear approaches”, 2007 IEEE Workshop on Machine Learning for Signal Processing, pp. 57–62 (2007).

[7] J. Laaksonen, M. Aksela, E. Oja and J. Kangas: “Adaptive local subspace classifier in on-line recognition of handwritten character”, Proceedings of International Joint Conference on Neural Networks, Vol. 4, pp. 2812–2815 (1999).

[8] 福井, 山口, 鈴木, 前田：“制約相互部分空間法を用いた環境変動にロバストな顔画像認識：照明変動の影響を抑える制約部分空間の学習”, 電子情報通信学会論文誌 (D-II), **J82-D-II**, 4, pp. 613–620 (1999).

[9] 坂野, 武川, 中村：“核非線形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 (D-II), **J84-D-II**, 8, pp. 1549–1556 (2001).

[10] 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 (D-II), **J88-D-II**, 8, pp. 1349–1356 (2005).

[11] 前田, 村瀬：“カーネル非線形部分空間法によるパターン認識”, 信学論 (D-II), **J82-D-II**, 4, pp. 600–612 (1999).

[12] 津田：“ヒルベルト空間における部分空間法”, 信学論 (D-II), **J82-D-II**, 4, pp. 592–599 (1999).

[13] V. N. Vapnik: “Statistical Learning Theory”, Wiley, New-York (1998).

[14] Y. Yamashita and H. Ogawa: “Relative Karhunen-Loève transform”, IEEE Trans. on Signal Processing, **44**, 2, pp. 371–378 (1996).

[15] 池野, 山下, 小川：“相対 KL 変換法によるパターン認識”, 信学論 (D-II), **J80-D-II**, 2, pp. 541–547 (1997).

[16] 鷲沢, 疋田, 田中, 山下：“パターン認識のための相対 KL 変換法の高精度化”, 情報科学技術フォーラム 2002 一般講演論文集,I-48, pp. 95–96 (2002).

[17] Y. Washizawa and Y. Yamashita: “‘non-linear Wiener filter in reproducing kernel Hilbert space”, Proceedings of 18th International Conference on Pattern Recognition, Vol. 1, pp. 967–970 (2006).

[18] H. Yoshino and Y. Yamashita: “Pattern recognition by kernel Wiener filter”, Proceedings of Signal Processing, Pattern Recognition, and Applications, pp. 7–12 (2008).

[19] L. Scharf: “The SVD and reduced rank signal processing”, Signal Processing, **25**, 2, pp. 113–133 (1991).

[20] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda and B. Schölkopf: “An introduction to kernel-based learning algorithms”, IEEE Transactions on Neural Networks, **12**, 2, pp. 181–201 (2001).

[21] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, **86**, 11, pp. 2278–2324 (1998).

球面最小二乗法による球面上の曲線あてはめ

藤木　淳†　赤穂昭太郎†

† 産業技術総合研究所　〒 305–8568 茨城県つくば市梅園 1–1–1 つくば中央第 2
E-mail: †{jun-fujiki,s.akaho}@aist.go.jp

あらまし　高次元ベクトルデータの類似性は，ユークリッド距離よりもむしろ相関係数によって測られることが多い．このとき，高次元ベクトルデータは正規化することにより単位超球面上の点とみなすことができ，2 つのベクトルの相関係数は単位超球面上の測地線に沿う長さとなる．つまり高次元ベクトルデータを超球面データとみなす場合，測地線に沿う長さを基準として色々な推定が行われる必要がある．そこで本稿では単位超球面上の点列に超曲面をあてはめるための手始めとして，まずは二次元球面上のデータに対して測地線に沿う距離の二乗を最小化することによって一次元曲線あてはめる手法を提案する．そしてこの提案手法が計量のユークリッド化を利用した曲線のあてはめ問題と関係があることを特殊な場合について証明する．

キーワード　球面最小二乗法，曲線あてはめ，測地線，ユークリッド化

Curve fitting by Spherical Least Squares on two-dimensional sphere

Jun FUJIKI† and Shotaro AKAHO†

† National Institute of Advanced Industrial Science and Technology, Tsukuba-Central 2, 1–1–1 Umezono, Tsukuba-shi, Ibaraki 305–8568 Japan
E-mail: †{jun-fujiki,s.akaho}@aist.go.jp

Abstract　To measure the similarity between two high dimensional vector data, correlation coefficient is often used instead of Euclidean distance. For this purpose, the high dimensional vectors are mapped into hyperspherical points by normalization and the distance is measured as the length along geodesic on the hypersphere. Then estimations from high dimensional vector data should be resolved as minimizing appropriate energy function of the length along geodesic when high dimensional vector data are regarded as hyperspherical data. In this paper, for the first step of hyper surface fitting to hyperspehrical data, the method of curve fitting to two-dimensional spherical data by Spherical Least Squares is proposed. It is also shown that the proposed method is closely related to the curve fitting by Euclidenization of the metric is special case.

Key words　Spherical Least Squares, curve fitting, geodesic, Euclideanization

1. はじめに

近年超球面上のデータ解析が重要な地位を占めつつある．

コンピュータビジョンにおいて反射屈折カメラや魚眼カメラなどの全方位カメラが広く用いられ，ロボットナビゲーションや監視システム等に利用されている．そして全方位カメラを統一的に理解するために球面カメラが定義されている [2], [4], [8], [11], [14], [17], [20], [21]．ここで球面カメラにおいて，3 次元空間の点や直線は 2 次元球面 S^2 上の点や大円に射影されるので，2 次元球面 S^2 上のデータ解析はコンピュータビジョンにおいて重要である．またコンピュータビジョンにおけるカメラ運動は，カメラの正面方向にのみ着目すると，その軌跡は球面上の点列と対応づけることができる [6], [7], [16], [18]（詳細は [18] 参照）．このときカメラ運動の平滑化は球面上の点列に小円や大円 [7], [16]，そしてそれらをつなげた連接小円 [18] をあてはめることによって実現できる．

金融工学では，2 つの超球面データの内積を並べてできる相関行列が投資モデルにおいて重要な役割りを果たす [3]．シミュレーションにおいて計算コストを削減するには相関行列の次元削減を行うことが重要であり [13]，これは単位超球面データに大超球面をあてはめることに相当する．そしてバイオインフォマティクスにおける遺伝子表現プロファイルやデータマイニングにおける文章解析においても，データのノルムを無視し，超球面上のデータと等価な方向データとみなすことが行なわれている [19]．以上の例からもわかるように超球面データ解析は近年重要な意味を持っている．

超球面データ解析における一つの主題は高次元データの次元圧縮である．その際，超球面の部分空間として大超球面や小超球面をあてはめるのが簡明である．というのは大超球面や小超球面は超球面とユークリッド空間との交わりとしてとらえられるからである．しかし，このような線型に近い構造のあてはめではなく，非線型な構造をあてはめることを考えるとき，超球面の部分空間として大超球面や小超球面以外の一般の超曲面をあてはめる必要が生じるだろう．

本稿では，そのあてはめ手法として球面上の点列に助変数で表現される曲線をあてはめることを考え，手始めに二次元球面上の点列に緯度経度を座標変数とする曲線をあてはめる問題に対する解法を提案する．この際，曲線とデータ点のずれを測地線に沿う距離によって測る，つまり**球面最小二乗基準（Spherical Least Squares；SLS）**により曲線をあてはめる．

また，球面を等方向射影によって平面に射影し，球面上の測地線に沿う距離に関する最小二乗法を平面上でのユークリッド距離に関する重みつき最小二乗法で近似するユークリッド化 [9], [10] が提案されているが，本稿では，航程線のあてはめ問題における提案手法と，メルカトル図法によって平面に射影場合のユークリッド化を用いた推定とが本質的に等価であることを示す．

2. 問 題 設 定

ノイズを含むデータに対して曲線をあてはめることはデータ解析において基本的かつ重要な過程である．そこで本稿では，二次元球面 S^2 上において，ノイズを含むデータに対して緯度と経度の関係式で表現された助変数を含む曲線をあてはめる手法を提案する．

2.1 球面最小二乗基準

観測されたノイズを含むと考えられるデータ点を $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^m$ とする．ここで $||\boldsymbol{x}^p|| = 1$（$p = 1, \ldots, m$）である．

球面上において $\boldsymbol{x}^p$ から曲線に下した垂線の足[(注1)]を $\widehat{\boldsymbol{x}}^p$ とするとき

$$r^p = \cos^{-1}\left\{\left(\widehat{\boldsymbol{x}}^p\right)^\top \boldsymbol{x}^p\right\} \tag{1}$$

として，

$$\sum_{p=1}^{m} (r^p)^2 \tag{2}$$

を最小とするような曲線（を記述する助変数）を求める．

2.2 球面の緯度経度

二次元球面 S^2 上の点は，緯度と経度で表現することができる．本稿では，緯度及び経度が非負の値となるように修正した**補緯度（colatitude）**と**補経度（colongitude）**[(注2)]を座標変数として二次元球面を表現する．

補緯度とは，北極が 0，南極が π となるように緯度を修正したもので，補緯度 ϕ（$0 \leqq \phi \leqq \pi$）と地球を表現する緯度 ϕ^{Earth}（北緯を $+$，南緯を $-$ とする）との間には

$$\phi + \phi^{\mathrm{Earth}} = \frac{\pi}{2} \tag{3}$$

という関係があり，補経度 ψ（$0 \leqq \psi < 2\pi$）と地球を表現する経度 ψ^{Earth}（西経を $-$，東経を $+$ とする）との間には

$$\psi = \begin{cases} \psi^{\mathrm{Earth}} + 2\pi & (-\pi < \psi^{\mathrm{Earth}} < 0) \\ \psi^{\mathrm{Earth}} & (0 \leqq \psi^{\mathrm{Earth}} \leqq \pi) \end{cases} \tag{4}$$

という関係がある．

補緯度と補経度を用いると，二次元球面上の点の三次元座標は

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin\phi\cos\psi \\ \sin\phi\sin\psi \\ \cos\phi \end{pmatrix} \tag{5}$$

と表現される．この三次元座標から補緯度，補経度を求めるには

$$\phi = \cos^{-1} z, \quad \psi = \arg(x + yi) \tag{6}$$

（i は虚数単位）とすれば良い（ここで極の補経度は 0 であるとする）．

2.3 微小距離〜弧と弦〜

補緯度と補経度の組を $\boldsymbol{\phi} = \begin{pmatrix} \phi \\ \psi \end{pmatrix}$ とすると，それに対応する三次元座標は

$$\boldsymbol{x}(\boldsymbol{\phi}) = \begin{pmatrix} \sin\phi\cos\psi \\ \sin\phi\sin\psi \\ \cos\phi \end{pmatrix} \tag{7}$$

となる．今，$\boldsymbol{x}(\boldsymbol{\phi})$ が微小変化して

$$\boldsymbol{x}(\boldsymbol{\phi} + \delta\boldsymbol{\phi}) = \boldsymbol{x}(\boldsymbol{\phi}) + \delta\boldsymbol{x} \tag{8}$$

になったとする．このとき，$\delta\boldsymbol{x}$ は 2 点 $\boldsymbol{x}(\boldsymbol{\phi})$ と $\boldsymbol{x}(\boldsymbol{\phi} + \delta\boldsymbol{\phi})$ を結ぶ弦の長さであり，幾何学的に，弦の長さと弧の長さ r との間には

$$\frac{||\delta\boldsymbol{x}||}{2} = \sin\frac{r}{2} \tag{9}$$

という関係が成立する．よって

$$r = 2\sin^{-1}\frac{||\delta\boldsymbol{x}||}{2} \tag{10}$$

が成立する．なお式 (10) を一次近似すると

$$r = ||\delta\boldsymbol{x}|| \tag{11}$$

となり，弦の長さと弧の長さの差は 2 次以上の微小量となることがわかる．つまり誤差が小さいと考えて良い場合，弧を弦で近似しても良いということになる．

(注1)：ある点から曲線に対して測地線に沿う距離が最小となるような測地線を**垂線**と呼び，測地線に沿う距離が最小となるような点を**垂線の足**と呼ぶ．

(注2)：緯度と補緯度は明確に区別されているが，経度と補経度は 2π を法として合同なため，特に区別しないことが多く，文献 [15]p.160 においても，補緯度のみしか定義されていない．

2.4 計　　量

二次元球面上の点の三次元座標は補緯度と補経度の組 $\boldsymbol{\phi}$ から三次元座標 $\boldsymbol{x}$ への写像と考えることができ，その写像を

$$\boldsymbol{x}:\begin{pmatrix}\phi\\ \psi\end{pmatrix}\mapsto\begin{pmatrix}x\\ y\\ z\end{pmatrix} \tag{12}$$

と表現すると，その**ヤコビ行列（Jacobian matrix）** $J_{\boldsymbol{x}}$ は

$$J_{\boldsymbol{x}}=\frac{\partial\boldsymbol{x}}{\partial\boldsymbol{\phi}}=\begin{pmatrix}\cos\phi\cos\psi & -\sin\phi\sin\psi\\ \cos\phi\sin\psi & \sin\phi\cos\psi\\ -\sin\phi & 0\end{pmatrix} \tag{13}$$

となり，計量テンソル $G_{\boldsymbol{x}}$ は

$$G_{\boldsymbol{x}}=J_{\boldsymbol{x}}^{\top}J_{\boldsymbol{x}}=\begin{pmatrix}1 & 0\\ 0 & \sin^2\phi\end{pmatrix} \tag{14}$$

となる．このとき，

$$\delta\boldsymbol{x}(\boldsymbol{\phi})=J_{\boldsymbol{x}}\delta\boldsymbol{\phi} \tag{15}$$

であるから，

$$||\delta\boldsymbol{x}||^2 \;=\; \delta\boldsymbol{\phi}^{\top}J_{\boldsymbol{x}}^{\top}J_{\boldsymbol{x}}\delta\boldsymbol{\phi}=\delta\boldsymbol{\phi}^{\top}G_{\boldsymbol{x}}\delta\boldsymbol{\phi} \tag{16}$$

$$\;=\;\delta\phi^2+\sin^2\phi\,\delta\psi^2 \tag{17}$$

が成立する．これは微小変化を 1 次近似したものであるから，1 次近似の範囲内では，測地線に沿う誤差 r は

$$r^2=\delta\phi^2+\sin^2\phi\,\delta\psi^2 \tag{18}$$

と近似されることがわかる．

3. 緯度経度を変数とする曲線

緯度経度を変数とする曲線として主なものは**大圏航路（great circle cource）**と**航程線（rhumb line または loxodrome）**である．大圏航路は測地線に沿う距離であり，二次元球面 S^2 と原点を通る平面との交線としてとらえた方が簡単ではあるが，本稿では助変数を利用してとらえることにする．また，航程線は大航海時代に活躍した航行法であり，曲線と経線とのなす角度が常に一定となる曲線である．このとは**メルカトル図法（Mercator projection）** において航程線が直線で表現されることを意味する．

本節では，まずこれら 2 つの曲線について考察し，次に一般的な曲線について考察することとする．

3.1 大 圏 航 路

大圏航路は大円であり，S^2 と原点を通る平面との交わりとしてとらえることができるので，大圏航路の助変数表示は，三次元空間における原点を通る平面の助変数表示 $ax+by+cz=0$ を利用して

$$a\sin\phi\cos\psi+b\sin\phi\sin\psi+c\cos\phi=0 \tag{19}$$

となる．

なお，2 点 (ϕ_1,ψ_1)，(ϕ_2,ψ_2)（$0\leqq\psi_1\leqq\psi_2<2\pi$）を通る大圏航路は

$$\cot\phi=\cot\phi_1\frac{\sin(\psi_2-\psi)}{\sin(\psi_2-\psi_1)}+\cot\phi_2\frac{\sin(\psi_1-\psi)}{\sin(\psi_1-\psi_2)} \tag{20}$$

となる．

3.2 航　程　線

航程線はメルカトル図法

$$\alpha=\sinh^{-1}(\cot\phi))=\log\left|\cot\frac{\phi}{2}\right|,\quad \beta=\psi \tag{21}$$

において直線をあらわすので，航程線の助変数表示は，直線の助変数表示 $a\alpha+b\beta+c=0$ を利用して

$$a\sinh^{-1}(\cot\phi))+b\psi+c=0 \tag{22}$$

となる．

なお，2 点 (ϕ_1,ψ_1)，(ϕ_2,ψ_2)（$0\leqq\psi_1\leqq\psi_2<2\pi$）を通る航程線の方程式は，

$$\boldsymbol{n}=\begin{pmatrix}\psi_1-\psi_2\\ \sinh^{-1}(\cot\phi_2))-\sinh^{-1}(\cot\phi_1))\\ \psi_2\sinh^{-1}(\cot\phi_1)-\psi_1\sinh^{-1}(\cot\phi_2)\end{pmatrix} \tag{23}$$

として

$$\boldsymbol{n}^{\top}\begin{pmatrix}\sinh^{-1}(\cot\phi))\\ \psi\\ 1\end{pmatrix}=0 \tag{24}$$

となる．

3.3 緯度経度を変数とする曲線

補緯度 ϕ，補経度 ψ を変数とし，助変数 $\boldsymbol{a}$ によって記述される曲線群を

$$f(\phi,\psi;\boldsymbol{a})=\boldsymbol{a}^{\top}F(\phi,\psi)=0 \tag{25}$$

とする．

観測されたデータ点を $\begin{pmatrix}\phi^p\\ \psi^p\end{pmatrix}$（$p=1,\ldots,m$）とし，それらデータ点の真の値，つまり曲線 $f(\phi,\psi;\boldsymbol{a})=0$ へ下した垂線の足を $\begin{pmatrix}\widehat{\phi}^p\\ \widehat{\psi}^p\end{pmatrix}$（$p=1,\ldots,m$）とする．

このとき，

$$\begin{pmatrix}\widehat{\phi}^p\\ \widehat{\psi}^p\end{pmatrix}-\begin{pmatrix}\phi^p\\ \psi^p\end{pmatrix}=\begin{pmatrix}\delta\phi^p\\ \delta\psi^p\end{pmatrix} \tag{26}$$

とすると，

$$f(\widehat{\phi}^p,\widehat{\psi}^p;\boldsymbol{a})\;\approx\;f(\phi^p,\psi^p;\boldsymbol{a}) \tag{27}$$

$$+\,\delta\phi^p\frac{\partial f}{\partial\phi}(\phi^p,\psi^p;\boldsymbol{a})$$

$$+\,\delta\psi^p\frac{\partial f}{\partial\psi}(\phi^p,\psi^p;\boldsymbol{a})=0 \tag{28}$$

が成立する．ここで

$$\frac{\partial f(\phi^p,\psi^p;\boldsymbol{a})}{\partial\phi}\;=\;\boldsymbol{a}^{\top}\frac{\partial F}{\partial\phi}(\phi^p,\psi^p), \tag{29}$$

$$\frac{\partial f(\phi^p,\psi^p;\boldsymbol{a})}{\partial\psi}\;=\;\boldsymbol{a}^{\top}\frac{\partial F}{\partial\psi}(\phi^p,\psi^p) \tag{30}$$

により，

$$\partial\phi^p\;=\;\frac{\partial F}{\partial\phi}(\phi^p,\psi^p), \tag{31}$$

$$\partial\psi^p\;=\;\frac{\partial F}{\partial\psi}(\phi^p,\psi^p) \tag{32}$$

とおくと，

$$\delta\phi^p(\boldsymbol{a}^\top\partial\phi^p) + (\sin\phi^p\delta\psi^p)\cdot\frac{1}{\sin\phi^p}(\boldsymbol{a}^\top\partial\psi^p) \tag{33}$$

$$= -\boldsymbol{a}^\top F(\widehat{\phi}^p,\widehat{\psi}^p) \tag{34}$$

だから，$(r^p)^2 = (\delta\phi^p)^2 + \sin^2\phi^p\,(\delta\psi^p)^2$ の最小値は

$$(r^p)^2 = \frac{\boldsymbol{a}^\top\left[F(\widehat{\phi}^p,\widehat{\psi}^p)F(\widehat{\phi}^p,\widehat{\psi}^p)^\top\right]\boldsymbol{a}}{\boldsymbol{a}^\top\left[(\partial\phi^p)(\partial\phi^p)^\top + \dfrac{(\partial\psi^p)(\partial\psi^p)^\top}{\sin^2\phi^p}\right]\boldsymbol{a}} \tag{35}$$

である．ここで，

$$V_{\boldsymbol{x}^p}[\boldsymbol{\phi}^p] \;=\; (\partial\phi^p)(\partial\phi^p)^\top + \frac{(\partial\psi^p)(\partial\psi^p)^\top}{\sin^2\phi^p} \tag{36}$$

とおくと，

$$(r^p)^2 = \frac{\boldsymbol{a}^\top\left[F(\widehat{\phi}^p,\widehat{\psi}^p)F(\widehat{\phi}^p,\widehat{\psi}^p)^\top\right]\boldsymbol{a}}{\boldsymbol{a}^\top V_{\boldsymbol{x}^p}[\boldsymbol{\phi}^p]\,\boldsymbol{a}} \tag{37}$$

となり，球面最小二乗法による曲線のあてはめ問題は

$$\mathcal{E}(\boldsymbol{a}) \;=\; \sum_{p=1}^{m}(r^p)^2 \tag{38}$$

$$= \sum_{p=1}^{m}\frac{\boldsymbol{a}^\top\left[F(\widehat{\phi}^p,\widehat{\psi}^p)F(\widehat{\phi}^p,\widehat{\psi}^p)^\top\right]\boldsymbol{a}}{\boldsymbol{a}^\top V_{\boldsymbol{x}^p}[\boldsymbol{\phi}^p]\,\boldsymbol{a}} \tag{39}$$

を最小にする $\boldsymbol{a}$ を求めれば良い．

3.4 一 般 化

n 次元空間 $\boldsymbol{R}^n$ のベクトル $\boldsymbol{x}$ の集合である k 次元曲面を記述する k 個の座標変数を並べてできるベクトルを $\underset{(k\times 1)}{\boldsymbol{\phi}}$ とし，l 個の助変数を並べてできるベクトル $\underset{(l\times 1)}{\boldsymbol{a}}$ によって記述される曲線群を

$$f(\boldsymbol{\phi};\boldsymbol{a}) = \underset{(1\times l)}{\boldsymbol{a}^\top}\;\underset{(l\times 1)}{F(\boldsymbol{\phi})} = 0 \tag{40}$$

とする．

観測されたデータ点を $\boldsymbol{\phi}^p$ （$p=1,\ldots,m$）とし，それらデータ点の真の値，つまり曲線 $f(\boldsymbol{\phi};\boldsymbol{a})=0$ へ下した垂線の足を $\widehat{\boldsymbol{\phi}}^p$ （$p=1,\ldots,m$）とする．

このとき，

$$\widehat{\boldsymbol{\phi}}^p - \boldsymbol{\phi}^p = \delta\boldsymbol{\phi}^p \tag{41}$$

とすると，

$$f(\widehat{\boldsymbol{\phi}}^p;\boldsymbol{a}) \approx f(\boldsymbol{\phi};\boldsymbol{a}) + \underset{(1\times k)}{\frac{\partial f}{\partial\boldsymbol{\phi}}}(\boldsymbol{\phi}^p;\boldsymbol{a})\;\delta\boldsymbol{\phi}^p = 0 \tag{42}$$

が成立する．

$$\underset{(1\times k)}{\frac{\partial f}{\partial\boldsymbol{\phi}}}(\boldsymbol{\phi}^p;\boldsymbol{a}) \;=\; \underset{(1\times l)}{\boldsymbol{a}^\top}\;\underset{(l\times k)}{\frac{\partial F}{\partial\boldsymbol{\phi}}}(\boldsymbol{\phi}) \tag{43}$$

により，写像 F に対するヤコビ行列を J_F を

$$J_F = \underset{(l\times k)}{J_F(\boldsymbol{\phi}^p)} \;=\; \underset{(l\times k)}{\frac{\partial F}{\partial\boldsymbol{\phi}}}(\boldsymbol{\phi}^p) \tag{44}$$

とおくと，

$$\underset{(1\times k)}{\frac{\partial f}{\partial\boldsymbol{\phi}}}(\boldsymbol{\phi}^p;\boldsymbol{a})\;\delta\boldsymbol{\phi}^p = \boldsymbol{a}^\top J_F(\boldsymbol{\phi}^p)\;\delta\boldsymbol{\phi}^p = -f(\widehat{\boldsymbol{\phi}}^p;\boldsymbol{a}) \tag{45}$$

が成立するので，写像

$$\boldsymbol{x}:\boldsymbol{\phi}\mapsto\boldsymbol{x}=\boldsymbol{x}(\boldsymbol{\phi}) \tag{46}$$

に対するヤコビ行列 $J_{\boldsymbol{x}}$，計量テンソル $G_{\boldsymbol{x}}$ を

$$J_{\boldsymbol{x}} \;=\; \underset{(n\times k)}{J_{\boldsymbol{x}}(\boldsymbol{\phi})} = \frac{\partial\boldsymbol{x}}{\partial\boldsymbol{\phi}}, \tag{47}$$

$$G_{\boldsymbol{x}} \;=\; \underset{(k\times k)}{G_{\boldsymbol{x}}(\boldsymbol{\phi})} = J_{\boldsymbol{x}}^\top J_{\boldsymbol{x}} \tag{48}$$

と定義すると，

$$(r^p)^2 = ||\delta\boldsymbol{x}^p||^2 = \underset{(1\times k)}{\delta\boldsymbol{\phi}^{p\top}}\;\underset{(k\times k)}{G_{\boldsymbol{x}}}\;\underset{(k\times 1)}{\delta\boldsymbol{\phi}^p} \tag{49}$$

であるから，条件

$$\boldsymbol{a}^\top\left[J_F\,\delta\boldsymbol{\phi}^p\,\delta\boldsymbol{\phi}^{p\top}\,J_F^\top\right]\boldsymbol{a} = f(\widehat{\boldsymbol{\phi}}^p;\boldsymbol{a})^2 \tag{50}$$

における

$$(r^p)^2 = \underset{(1\times k)}{\delta\boldsymbol{\phi}^{p\top}}\;\underset{(k\times k)}{G_{\boldsymbol{x}}}\;\underset{(k\times 1)}{\delta\boldsymbol{\phi}^p} \tag{51}$$

の最小値を求めれば良く，コーシー・シュワルツの不等式（**附録 A**）により

$$(r^p)^2 \;=\; \frac{f(\widehat{\boldsymbol{\phi}}^p;\boldsymbol{a})^2}{\boldsymbol{a}^\top\left[J_F\,G_{\boldsymbol{x}}^{-1}\,J_F^\top\right]\boldsymbol{a}} \tag{52}$$

$$= \frac{\boldsymbol{a}^\top\left[(\widehat{F}^p)(\widehat{F}^p)^\top\right]\boldsymbol{a}}{\boldsymbol{a}^\top\left[J_F\,G_{\boldsymbol{x}}^{-1}\,J_F^\top\right]\boldsymbol{a}} \tag{53}$$

となる．ここで，

$$\widehat{F}^p = F(\widehat{\boldsymbol{\phi}}^p) \tag{54}$$

である．

さて，式 (53) において

$$\widehat{X}^p \;=\; (\widehat{F}^p)(\widehat{F}^p)^\top,\quad D^p = J_F\,G_{\boldsymbol{x}}^{-1}\,J_F^\top \tag{55}$$

とおくと，球面最小二乗法による曲線のあてはめ問題は

$$\mathcal{E}(\boldsymbol{a}) = \sum_{p=1}^{m}\frac{\boldsymbol{a}^\top\widehat{X}^p\boldsymbol{a}}{\boldsymbol{a}^\top D^p\boldsymbol{a}} \tag{56}$$

を最小にする $\boldsymbol{a}$ を求めることとなる．

4. 球面上の曲線のあてはめ問題の解法

本節では，Akaho [1] による式 (56) の最小化アルゴリズムを紹介する．ここで $f(\phi,\psi;\boldsymbol{a})$ は助変数 $\boldsymbol{a}$ に対して十分に滑らかであるとし，$\boldsymbol{a}$ が少し変化した場合の $\partial f/\partial\phi$，$\partial f/\partial\psi$ の変化は十分に小さいものとする．

さて，助変数 $\boldsymbol{a}$ の近似値 $\boldsymbol{a}_0$ が得られたとき，

$$\mu^p = \boldsymbol{a}_0^\top D^p\boldsymbol{a}_0 \tag{57}$$

とするとエネルギー関数 $\mathcal{E}(\boldsymbol{a})$ は

$$\mathcal{E}(\boldsymbol{a}) \approx \widehat{\mathcal{E}}(\boldsymbol{a}) = \sum_{p=1}^{m} \boldsymbol{a}^\top \left(\frac{1}{\mu^p} X^p\right) \boldsymbol{a} = \boldsymbol{a}^\top \left(\sum_{p=1}^{m} \frac{1}{\mu^p} X^p\right) \boldsymbol{a} \tag{58}$$

と近似でき，これを最小にする $\boldsymbol{a}$ は行列

$$\sum_{p=1}^{m} \frac{1}{\mu^p} X^p \tag{59}$$

の最小固有値に対応する固有ベクトルとなる．

4.1 アルゴリズム

(1) $\mu^p := 1$ for $p = 1, \ldots, m$

(2) (a)，(b) を繰り返す

(a) $\sum_{p=1}^{m} \frac{1}{\mu^p} X^p$ の最小固有値に対応する単位固有ベクトル $\widehat{\boldsymbol{a}}$ を求める．

(b) $\mu^p := \widehat{\boldsymbol{a}}^\top D^p \widehat{\boldsymbol{a}}$ for $p = 1, \ldots, m$

4.2 大圏航路のあてはめ

大圏航路の表現は

$$f(\phi, \psi; \boldsymbol{a}) = \boldsymbol{a}^\top \begin{pmatrix} \sin\phi\cos\psi \\ \sin\phi\sin\psi \\ \cos\phi \end{pmatrix} = 0 \tag{60}$$

であるから，

$$J_F = \begin{pmatrix} \cos\phi\cos\psi & -\sin\phi\sin\psi \\ \cos\phi\sin\psi & \sin\phi\cos\psi \\ -\sin\phi & 0 \end{pmatrix} \tag{61}$$

となる．また

$$G_{\boldsymbol{x}} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\phi \end{pmatrix} \tag{62}$$

より

$$G_{\boldsymbol{x}}^{-1} = \frac{1}{\sin^2\phi} \begin{pmatrix} \sin^2\phi & 0 \\ 0 & 1 \end{pmatrix} \tag{63}$$

となるので $\sin\phi = s_\phi$ などと書くと

$$X^p = \begin{pmatrix} s_\phi^2 c_\psi^2 & s_\phi^2 c_\psi s_\psi & s_\phi c_\psi c_\phi \\ s_\phi^2 c_\psi s_\psi & s_\phi^2 s_\psi^2 & s_\phi s_\psi c_\phi \\ s_\phi c_\psi c_\phi & c_\phi s_\phi s_\psi & c_\phi^2 \end{pmatrix}, \tag{64}$$

$$D^p = \begin{pmatrix} c_\phi^2 c_\psi^2 + s_\phi^2 s_\psi^2 & -s_\phi^2 c_\psi s_\psi & -s_\phi c_\phi c_\psi \\ -s_\phi^2 c_\psi s_\psi & c_\phi^2 s_\psi^2 + s_\phi^2 c_\psi^2 & -s_\phi c_\phi s_\psi \\ -s_\phi c_\phi c_\psi & -s_\phi c_\phi s_\psi & s_\phi^2 \end{pmatrix} \tag{65}$$

が成立する．

4.3 航程線のあてはめ

航程線の表現は

$$f(\phi, \psi; \boldsymbol{a}) = \boldsymbol{a}^\top \begin{pmatrix} \sinh^{-1}(\cot\phi) \\ \psi \\ 1 \end{pmatrix} = 0 \tag{66}$$

であるから，

$$J_F = \begin{pmatrix} -\frac{1}{\sin\phi} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{67}$$

となる．また

$$G_{\boldsymbol{x}} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\phi \end{pmatrix} \tag{68}$$

より

$$G_{\boldsymbol{x}}^{-1} = \frac{1}{\sin^2\phi} \begin{pmatrix} \sin^2\phi & 0 \\ 0 & 1 \end{pmatrix} \tag{69}$$

となるので，$s^p = \sinh^{-1}(\cot\phi^p))$ とおくと

$$X^p = \begin{pmatrix} (s^p)^2 & s^p\psi^p & s^p \\ s^p\psi^p & (\psi^p)^2 & \psi^p \\ s^p & \psi^p & 1 \end{pmatrix}, \tag{70}$$

$$D^p = \frac{1}{\sin^2\phi^p} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{71}$$

が成立する．

5. 計量のユークリッド化

球面最小二乗法を近似するために超球面上の**計量のユークリッド化（Euclideanization of Metric）**が提案された[9], [10]．計量のユークリッド化とは，空間を変換する際に，もとの空間の計量をなるべく保存するために導入される計量の変換規則のことである．

計量のユークリッド化は藤木・赤穂[9], [10] によって**等方向射影（equi-directional projection）**の場合に提案された．

k 次元ユークリッド空間の正規直交座標から k 次元空間への写像

$$\boldsymbol{\alpha} : \boldsymbol{\phi} \mapsto \boldsymbol{\alpha} = \boldsymbol{\alpha}(\boldsymbol{\phi}) \tag{72}$$

において，ヤコビ行列 $J_{\boldsymbol{\alpha}}$ の行列式である**ヤコビアン（Jacobian determinant）**の絶対値は

$$|\det(J_{\boldsymbol{\alpha}})| \quad \text{where} \quad J_{\boldsymbol{\alpha}} = \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\phi}} \tag{73}$$

となる．このヤコビアンを用いると，k 次元体積要素は $|\det(J_{\boldsymbol{\alpha}})|$ 倍されるので，

k 次元超球面 $\boldsymbol{x}(\boldsymbol{\phi})$ から k 次元空間への写像

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\phi}) \mapsto \boldsymbol{\alpha} = \boldsymbol{\alpha}(\boldsymbol{\phi}) \tag{74}$$

において，k 次元体積要素は

$$J = \frac{|\det(J_{\boldsymbol{\alpha}})|}{|\det(J_{\boldsymbol{x}})|} \quad \text{where} \quad J_{\boldsymbol{x}} = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\phi}} \tag{75}$$

倍される．よって線分長は平均して $J^{\frac{1}{k}}$ 倍拡大されると期待できる．

そこで写像先の空間で，もとの空間の最小二乗法を近似的に表現するためには，写像先の空間において線分長に $J^{-\frac{1}{k}}$ の重みを加えた重みつき最小二乗法を適用すれば良いというのが

ユークリッド化の考え方である．

この考え方の下では，ユークリッド化は球面から超平面への等方向射影の場合だけでなく，より一般的な写像についても考えることができる．そこで，航程線が直線として表現されるメルカトル図法において計量のユークリッド化を考え，提案手法と比較する．

5.1 メルカトル図法のユークリッド化

球面をメルカトル図法に射影したとき，補緯度 ϕ，補経度 ψ 近辺の面積要素がどの程度拡大されるかを考えてみる．

写像 $\boldsymbol{\alpha}$ を

$$\boldsymbol{\alpha}: \begin{pmatrix} \phi \\ \psi \end{pmatrix} \mapsto \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sinh^{-1}(\cot\phi) \\ \psi \end{pmatrix} \tag{76}$$

とおくと，$\boldsymbol{\alpha}$ のヤコビ行列は

$$J_{\boldsymbol{\alpha}} = \begin{pmatrix} -\dfrac{1}{\sin\phi} & 0 \\ 0 & 1 \end{pmatrix} \tag{77}$$

であるから，$\boldsymbol{\alpha}$ のヤコビアンの絶対値は

$$|\det(J_{\boldsymbol{\alpha}})| = \left| \det \begin{pmatrix} -\dfrac{1}{\sin\phi} & 0 \\ 0 & 1 \end{pmatrix} \right| = \frac{1}{\sin\phi} \tag{78}$$

となる．また，2 次元球面上での補緯度 ϕ，補経度 ψ における面積要素は

$$\sin\phi \, d\phi \, d\psi \tag{79}$$

であるから，

$$J_{\boldsymbol{x}} = \sin\phi \tag{80}$$

が成立する．よって

$$J = \frac{1}{\sin^2\phi} \tag{81}$$

が成立し，補緯度 ϕ，補経度 ψ 近辺において線分は，その平方根である

$$\sqrt{J} = \frac{1}{|\sin\phi|} \tag{82}$$

倍拡大されることが期待できる．

よって，球面上における測地線に沿う誤差 r^p が小さいとき，$(r^p)^2$ はメルカトル図法上での誤差 $(\delta\alpha^p)^2 + (\delta\beta^p)^2$ に重みをつけた

$$\sin^2\phi^p\{(\delta\alpha^p)^2 + (\delta\beta^p)^2\} \tag{83}$$

で近似できるという考え方がユークリッド化であり，球面上で球面最小二乗基準による航程線のあてはめ問題は，ユークリッド化により，メルカトル図法上での重みつき最小二乗基準による直線あてはめ問題で近似をすることができる．

さて，このメルカトル図法上での重みつき最小二乗基準による直線あてはめ問題は

$$\alpha^p = \sinh^{-1}(\cot\phi^p), \quad \beta^p = \psi^p \tag{84}$$

と座標変換した後，直線

$$\boldsymbol{a}^\top \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} = \boldsymbol{a}^\top \boldsymbol{\alpha} = 0 \tag{85}$$

をあてはめる問題となる．ここで座標変換されたデータ点と直線の距離は条件

$$\boldsymbol{a}^\top \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} = \boldsymbol{a}^\top \boldsymbol{\alpha} = 0 \tag{86}$$

における

$$(r^p)^2 = \sin^2\phi^p\{(\delta\alpha^p)^2 + (\delta\beta^p)^2\} \tag{87}$$

の最小値であり，

$$(r^p)^2 = \frac{\boldsymbol{a}^\top[(\widehat{\boldsymbol{\alpha}}^p)(\widehat{\boldsymbol{\alpha}}^p)^\top]\boldsymbol{a}}{\boldsymbol{a}^\top \left[\dfrac{1}{\sin^2\phi^p} J_F J_F^\top\right] \boldsymbol{a}} \tag{88}$$

となる．ここで $F(\boldsymbol{\alpha}) = \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$ だから

$$J_F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{89}$$

である．すると，

$$\mathcal{E}^{\mathrm{Euclid}}(\boldsymbol{a}) = \sum_{p=1}^{m} \frac{\boldsymbol{a}^\top[(\widehat{\boldsymbol{\alpha}}^p)(\widehat{\boldsymbol{\alpha}}^p)^\top]\boldsymbol{a}}{\boldsymbol{a}^\top \left[\dfrac{1}{\sin^2\phi^p} J_F J_F^\top\right] \boldsymbol{a}} \tag{90}$$

を最小化する $\boldsymbol{a}$ を求める問題に帰着されることがわかる．

ここで

$$\boldsymbol{\alpha}^p = \begin{pmatrix} \sinh^{-1}(\cot\phi^p) \\ \psi^p \\ 1 \end{pmatrix}, \tag{91}$$

$$J_F J_F^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{92}$$

であるから，$s^p = \sinh^{-1}(\cot\phi^p))$ とおくと

$$(\boldsymbol{\alpha}^p)(\boldsymbol{\alpha}^p)^\top = \begin{pmatrix} (s^p)^2 & s^p\psi^p & s^p \\ s^p\psi^p & (\psi^p)^2 & \psi^p \\ s^p & \psi^p & 1 \end{pmatrix} = X^p, \tag{93}$$

$$\frac{1}{\sin^2\phi^p} J_F J_F^\top = D^p \tag{94}$$

となり，

$$\mathcal{E}^{\mathrm{Euclid}}(\boldsymbol{a}) = \sum_{p=1}^{m} \frac{\boldsymbol{a}^\top \widehat{X}^p \boldsymbol{a}}{\boldsymbol{a}^\top D^p \boldsymbol{a}} \tag{95}$$

は式 (56) と等しくなる．つまり，航程線のあてはめにおいて，球面最小二乗法による曲線のあてはめが，球面をメルカトル図法に射影したときのユークリッド化により実現できていることがわかる．

6. おわりに

本稿では，二次元球面上の点列に緯度経度を座標変数にもつ曲線をあてはめる手法を提案した．また，航程線のあてはめにおいて，提案手法がメルカトル図法におけるユークリッド化と一致することを証明した．今後は他の曲線のあてはめ問題とユークリッド化の関係，多次元球面上のあてはめへの拡張などを探って行きたい．

文　献

[1] S. Akaho, “Curve fitting that minimizes the mean square of perpendicular distances from sample points,” In Proc. of SPIE93, Vision Geometry II, Vol.2060, (1993).

[2] S. Baker and S. K. Nayar, “A theory of catadioptric image formation,” In Proc. of IEEE International Conference on Computer Vision (ICCV98), pp.35–42, Jan. 1998.

[3] A. Brace, D. Gątarek, M. Musiela, “The market model of interest rate dynamics,” Mathematical Finance, Vol.7, No.2, pp.127-155, (1997).

[4] K. Daniilidis, A. Makadia and T. Bulow, “Image processing in catadioptric planes: spatiotemporal derivatives and optical flow computation,” *OMNIVIS02*, pp.3-10, 2002.

[5] N. I. Fisher, T. Lewis and B. J. J. Embleton, “Statistical anaylsis of spherical data,” Cambridge Univ. Press, 1987.

[6] J. Fujiki, S. Akaho and N. Murata, “Nonlinear PCA/ICA for the structure from motion problem,” In Proc. of ICA04(LNCS 3195), pp. 750-757, Granada, Sep. 2004.

[7] J. Fujiki and S. Akaho, “Small ciecle fitting to the sequence of spherical points -Towards smoothing of camera motions-,” Technical Report of IEICE, PRMU2004-149, pp.91-96, 2004 , (In Japanese).

[8] J. Fujiki and S. Akaho, “Epipolar geometry for spherical camera and its calculations,” Technical Report of IEICE, PRMU2005-41, pp.41-46, 2005 , (In Japanese).

[9] J. Fujiki and S. Akaho, “Small hypersphere fitting by Spherical Least Square,” In Proc. of ICONIP05, pp.439-444, 2005.

[10] J. Fujiki and S. Akaho, “Spherical PCA with Euclideanization,” In Proc. of Workshop on ACCV’07, Subspace 2007.

[11] C. Geyer and K. Daniilidis, “Catadioptric Projective Geometry,” IJCV, vol.45, no.3, pp.223-243, Dec. 2001.

[12] N. H. Gray, P. A. Geiser and J. R. Geiser, “On the least-squares fit of small and great circles to spherically projected orientation data,” Mathematical Geology, vol.12, no.3, pp.173–184, 1980.

[13] I. Grubišić and R. Pietersz, “Efficient rank reduction of correlation matrices,” Utrecht Univ., preprint, 2005.

[14] A. Makadia and K. Daniilidis, “Direct 3D-rotation estimation from spherical images via a generalized shift theorem,” In proc. of CVPR03, pp.II: 217-224, 2003.

[15] K. V. Mardia and P. E. Jupp, “Directional Statistics,” John Wiley & Sons Ltd., 2000.

[16] J. Mimura, N. Murata and J. Fujiki, “Smoothing of camera motions via small circle fitting for the sequence of spherical points,” Technical Report of IEICE, PRMU, 2005, (In Japanese).

[17] K. Miyamoto, “Fish eye lens,” Journal of Optical Society of America, vol.54, no.8, pp.1060-1061, Aug. 1964.

[18] 野田容士, 藤木淳, 村田 昇, “球面上の点列に対する連接小円回帰を用いたカメラ運動の平滑化,” 信学論 D-II, vol.J91-D, no.5, pp.1336-1348, 2008

[19] S. Oba and S. Ishii, “Kernel density estimation on hypersphere,” 2004 Workshop on Information-Based Induction Sciences(IBIS2004), pp.197-202, (2004), (In Japanese).

[20] T. Svoboda and T. Pajdla, “Epipolar Geometry for Central Catadioptric Cameras,” IJCV, vol.49, no.1, pp.23-37, 2002.

[21] A. Torii and A. Imiya, “Analysis of Central Camera Systems for Computer Vision,” CVIM 154-30, 2006.

附録 A. コーシー・シュワルツの不等式

n 次元実ベクトル $\boldsymbol{x}$，$\boldsymbol{y} \in \boldsymbol{R}^n$ に対して

$$(\boldsymbol{x}^\top \boldsymbol{x})(\boldsymbol{y}^\top \boldsymbol{y}) \geqq (\boldsymbol{x}^\top \boldsymbol{y})^2$$

（等号は $\boldsymbol{x}$ と $\boldsymbol{y}$ が平行なとき）が成立する．これを**コーシー・シュワルツの不等式（Cauchy-Schwarz’s inequality）**と呼ぶ．

証明は任意の λ に対して

$$(\boldsymbol{x} - \lambda \boldsymbol{y})^\top (\boldsymbol{x} - \lambda \boldsymbol{y}) \geqq 0$$

だから，λ について判別式をとれば良い．

今，正値対称行列 G があるとする．そして $G = LL^\top$ とコレスキー分解されるものとする．このとき，

$$\boldsymbol{z} = L^\top \boldsymbol{x}, \quad \boldsymbol{w} = L^{-\top} \boldsymbol{y}$$

とおくと $\boldsymbol{z}$，$\boldsymbol{w}$ に対するコーシー・シュワルツの不等式から

$$(\boldsymbol{z}^\top \boldsymbol{z})(\boldsymbol{w}^\top \boldsymbol{w}) \geqq (\boldsymbol{z}^\top \boldsymbol{w})^2$$

であるから，

$$(\boldsymbol{x}^\top G \boldsymbol{x})(\boldsymbol{y}^\top G^{-1} \boldsymbol{y}) \geqq (\boldsymbol{x}^\top \boldsymbol{y})^2$$

が成立する．

よって

$$\boldsymbol{x}^\top G \boldsymbol{x} \geqq \frac{(\boldsymbol{x}^\top \boldsymbol{y})^2}{\boldsymbol{y}^\top G^{-1} \boldsymbol{y}}$$

が成立する．

Computational Complexities of Dimensionality Reduction Schemes for Dissimilarity-Based Classification

Sang-Woon KIM† and Jian GAO†

† Dept. of Computer Engineering, Myongji University Yongin, 449–728 Korea
E-mail: †{kimsw,marsgao}@mju.ac.kr

Abstract Dissimilarity-Based Classifiers (DBCs) are a way of defining classifiers based on a suitable dissimilarity measure between individual patterns. In the attempt to find the most appropriate dimensionality reduction method for DBCs, in this paper, we report a comparison between the computational complexities of prototype selection methods (PSM) and dimensionality reduction schemes (DRS). This is done by theoretically and experimentally demonstrating the strength in terms of processing time and classification accuracy.

Key words Dissimilarity-Based Classifiers (DBCs), Prototype Selection Methods (PSM), Dimensionality Reduction Schemes (DRS)

1. Introduction

One of the most recent and novel developments in pattern classification is the concept of dissimilarity-based classifiers (DBCs) proposed by Duin and his co-authors [1], [2]. DBCs are a way of defining classifiers between the classes, which are not based on the feature measurements of the individual patterns, but rather on a suitable *dissimilarity measure* between them [3].

The problem with this strategy, however, is that we need to select a representative set of data that is both compact and capable of representing the entire data set. In DBCs, a good selection of prototypes seems to be crucial to succeed with the classification algorithm in the dissimilarity space. The prototypes should avoid redundancies in terms of selection of similar samples, and prototypes should include as much information as possible [1], [2], [3]. However, it is difficult to find the optimal number of prototypes, and there is also a possibility that we lose some useful information for discrimination when selecting the prototypes.

Recently, to avoid these problems, researchers [4], [5] proposed an alternative approach where they used *all* available samples from the training set as prototypes, and subsequently apply dimensionality reduction schemes [5]. In the attempt to find the most appropriate reduction method, in this paper, we report a comparison between the computational complexities of prototype selection methods (PSM) and dimensionality reduction schemes (DRS) for dissimilarity-based classification. This is done by theoretically and experimentally demonstrating its strength in terms of processing time and classification accuracy.

2. Dissimilarity-Based Classification

Foundations of DBCs : A dissimilarity representation of a set of samples, $T = \{\mathbf{x}_i\}_{i=1}^n \in \Re^d$, is based on pairwise comparisons and is expressed as an $n \times m$ dissimilarity matrix $D_{T,Y}[\cdot,\cdot]$, where $Y = \{\mathbf{y}_1, \cdots, \mathbf{y}_m\}$, a prototype set, is extracted from T, and the subscripts of D represent the set of elements, on which the dissimilarities are evaluated. Thus, each entry $D_{T,Y}[i,j]$ corresponds to the dissimilarity between the pairs of objects $\langle \mathbf{x}_i, \mathbf{y}_j \rangle$, where $\mathbf{x}_i \in T$ and $\mathbf{y}_j \in Y$. Consequently, an object $\mathbf{x}_i$ is represented as a column vector as follows:

$$[d(\mathbf{x}_i, \mathbf{y}_1), d(\mathbf{x}_i, \mathbf{y}_2), \cdots, d(\mathbf{x}_i, \mathbf{y}_m)]^T, 1 \leqq i \leqq n. \qquad (1)$$

Here, the dissimilarity matrix $D_{T,Y}[\cdot,\cdot]$ is defined as a *dissimilarity space*, on which the d-dimensional object, $\mathbf{x}$, given in the feature space, is represented as an m-dimensional vector $\delta_Y(\mathbf{x})$.

For a training set $\{\mathbf{x}_i\}_{i=1}^n$ and an evaluation sample $\mathbf{z}$, the modified training set and sample now become $\{\delta_Y(\mathbf{x}_i)\}_{i=1}^n$ and $\delta_Y(\mathbf{z})$, respectively. From this perspective, we can see that the dissimilarity representation can be considered as a *mapping*, by which any arbitrary $\mathbf{x}$ is translated into $\delta_Y(\mathbf{x})$, and thus, if m is selected sufficiently small (i.e., $m << p$), we are essentially working in a space with much smaller dimensions. The literature [1] reports the use of many traditional decision classifiers, including k-NN rule and the lin-

ear/quadratic normal-density-based classifiers, to the task of classifying $\mathbf{z}$ using $\delta_Y(\mathbf{z})$ in the dissimilarity space.

Prototype Selection Methods (PSM) : To select the representative set that is compact and capable of simultaneously representing the entire data set, Duin and colleagues [1] discussed the followings : *Random*, *RandomC*, *KCentres*, *ModeSeek*, *LinProg*, *FeatSeal*, *KCentres-LP*, and *EdiCon*.

In the interest of completeness, we briefly explain below the methods that are pertinent to our present study. Here, we assume c classes, a training set T, and the training subset T_i of the class ω_i. Each method selects m objects for the prototype set Y. If the algorithm is applied to each class separately, then m_i objects are chosen, such that $m = \sum_{i=1}^{c} m_i$.

(1) *Random* : This method involves a *random* selection of m samples from the training data set T.

(2) *RandomC*: This method involves a random selection of m_i samples per class, ω_i, from T_i.

(3) *KCentres* : This method consists of a procedure that is applied to each class separately. For each class ω_i, the algorithm is invoked so as to choose m_i samples which are "evenly" distributed with respect to the dissimilarity matrix $D_{T_i,T_i}[\cdot,\cdot]$. The algorithm can be summarized as follows: (a) Select an initial set $Y_i = \{\mathbf{y}_1, \cdots, \mathbf{y}_{m_i}\}$ consisting of m_i objects, e.g., randomly chosen from T_i. (b) For each $\mathbf{x} \in T_i$, find its nearest neighbor in Y_i. Let $N_j, j = 1, \cdots, m_i$, be a subset of T_i consisting of objects that yield the same nearest neighbor $\mathbf{y}_j$ in Y_i. This means that $T_i = \cup_{j=1}^{m_i} N_j$. (c) For each N_j, find its center $\mathbf{c}_j$, which is the object for which the maximum distance to all other objects in N_j is minimum (this value is called the radius of N_j). (d) For each center $\mathbf{c}_j$, if $\mathbf{c}_j \neq \mathbf{y}_j$, then replace $\mathbf{y}_j$ by $\mathbf{c}_j$ in Y_i. If any replacement is done, then return to Step (b). Otherwise exit. (e) Return the final representation set Y which consists of all the final sets Y_i.

(4) *ModeSeek* : This method focuses on the modes in the dissimilarity data in the specified neighborhood size s. For each class ω_i, the algorithm proceeds as follows: (a) Set a relative neighborhood size as an integer $s > 1$. (b) For each $\mathbf{x} \in T_i$, find the dissimilarity $d_{s-NN}(\mathbf{x})$ to its s-th neighbor. (c) Find a set Y_i consisting of all $\mathbf{x}_j \in T_i$ for which $d_{s-NN}(\mathbf{x}_j)$ is minimum within its set of s neighbors.

From the experimental results of [2], the authors seem to have deliberated that systematic approaches lead to better results than those that rely on random selection, especially when the number of prototypes is small. Furthermore, although there is no single winner (inasmuch as the results depend on the characteristics of the data), they indicate that, in general, the KCentres works well. The details of the other methods, such as *FeatSel*, *LinProg*, *KCentres-LP*, and *EdiCon*, are omitted here in the interest of compactness, but can be found in the existing literature, including [1] and [3].

Dimensionality Reduction Schemes (DRS) : Various strategies have been used to tackle the "dimensionality reduction" problem (some of them are [6], [8], [9], [10], and [11]). To optimize DBCs, we can use a strategy of reducing the dimensionality after computing the dissimilarity matrix with the entire training samples. With regard to reducing the dimensionality of the dissimilarity matrix, we make use of the well-known dimensionality reduction schemes (DRSs) proposed in the literature. In the interest of completeness, we briefly explain below the methods that are pertinent to our present study (注1).

(1) *PCA* : We first compute the covariance matrix of the training set T after normalizing T. Next, we determine the eigenvectors $\mathbf{e}_i$ corresponding to the nonzero eigenvalues λ_i of the covariance matrix, where $\lambda_1 \geqq \cdots \geqq \lambda_d \geqq 0$. Then we can reduce the dimensionality of an object by representing it in a new coordinate system defined by the eigenvectors corresponding to the $m(< d)$ highest eigenvalues.

(2) *LDA* : This method uses the concept of a within-class scatter matrix, S_w, and a between-class scatter matrix, S_b, to maximize a separation criterion, such as $J = tr(S_W^{-1} S_B)$. We can obtain the solution by solving an eigenvalues problem on the matrix $S_w^{-1} S_b$, if S_w^{-1} is nonsingular, or on $S_b^{-1} S_w$ if S_b^{-1} is nonsingular. There are at most $c-1$ eigenvectors corresponding to nonzero eigenvalues since the rank of the matrix S_b is bounded by $c-1$. Therefore, the reduced dimension is at most $c-1$.

(3) *PCA-plus-LDA* : In this two-stage algorithm, the discriminant stage is preceded by a dimension reduction stage using PCA. However, its computation is expensive, and the PCA stage may also potentially lose some useful information for discrimination.

(4) *DCV* : This approach extracts the common properties of the training samples T_i of ω_i. The common vectors are then used for recognition. A common vector x_{com}^i is obtained by removing all the features that are in the direction of the eigenvectors $\mathbf{e}$'s corresponding to the nonzero eigenvalues of the scatter matrix of ω_i: set $Q = [\mathbf{e}_1, \cdots, \mathbf{e}_r]$; then $x_{com}^i = x_j^i - QQ^T x_j^i$, $j = 1, \cdots, n_i$, $i = 1, \cdots, c$, where r is the rank of the scatter matrix; finally obtain $Y = Q^T X$.

The details of other methods are omitted here in the interest of compactness, but can be found in the existing literature, including [9], [10], and [11].

3. The Computational Complexity

First, the time complexity of PSM is analyzed. Following

(注1) : Our overview is necessarily brief, but additional details can be found in [8], [9], [10], and [11].

this, the time complexity of DRS is given.

The Time Complexity of PSM : First of all, the time required for the *Random* algorithm is $t_{Rand} = t_{rand}(m)$, where m is the number of prototypes and $t_{rand}(m)$ is the time for generating m random numbers. From this analysis, the reader can observe that the time complexity of the algorithm is $O(m)$.

Next, the time complexity of the *RandomC* algorithm can be analyzed as follows: First, let the computation times for the operations of addition (or subtraction), substitution (or comparison), and multiplication (or division) be t_a, t_s, and t_m, respectively. The time required for initializing the algorithm is $t_1 = t_{rand}(m - m_i c) + (m - m_i t_s + 1)(t_m + t_a + t_s)$. Then, the time required for iterating a sub-step c times is $t_2 = c \times ((2 + m_i)t_a + t_{rand}(m_i) + 2m_i t_s + (m_i + 1)t_m)$. Thus, the total time required for the entire procedure to process many kinds (classes) of images under the condition $t_1 << t_2$ is $t_{RandC} = t_1 + t_2 \simeq t_2 = c \times (m_i(t_a + 2t_s + t_m) + t_{rand}(m_i))$. From the above analysis, the reader can observe that the time complexity of the algorithm is $O(m_i c) \simeq O(m)$, and the required time primarily depends on the parameter of m.

Third, the time complexity of the *KCentres* algorithm can be analyzed as follows: First, the time required for initializing the algorithm is $t_1 = t_{min}(n) + t_{min}(d) + t_s$. Then, the time required for the other steps is a sum of times to iterate the followings η times: $t_{(21)} = 3kt_{min}(d) + 3kt_s + t_{rand}(n)$; $t_{(22)} = k(d+1)t_a + k(n_i+1)t_{min}(n_i)$. Thus, the total time required for the entire procedure under the condition $t_1 << t_2$ is $t_{KCent} \simeq \eta(t_{(21)} + t_{(22)}) = \eta(3kt_{min}(d) + t_{rand}(n) + 3kt_s + dt_{min}(k) + kn_i t_{min}(n_i) + kdt_a)$, where k is the k in k-NN strategy, $t_{min}(n)$ is the time required to search for the minimum number from n numbers, and η, an experimental constant, is the number of trials to achieve the selection. From the above analysis, the reader can observe that the time complexity of the algorithm is $O(kn_i^2 + kd + n + k)$, and the required time primarily depends on the parameters of n, d, and k.

Finally, the time complexity of the *ModeSeek* algorithm can be analyzed as follows: First, the time required for step (a) is $t_{(a)} = (n + dn + nk)t_s$. Next, the time required for step (b) is $t_{(b)} = 2(n-2)dt_a + (d+1)(n-2)t_m + 2(n-2)t_s$. Then, the time needed for step (c) is a sum of times to compute the three sub-steps of (c1), (c2), and (c3) for all the n samples: $t_{(c1)} = (dn + 2n + 3)t_s + t_{min}(n)$; $t_{(c2)} = (k-1)(t_{min}(n) + 2t_s)$; $t_{(c3)} = n((2d+1)t_s + t_a)$. Thus, the total time required for repeating the three steps c times is $t_{ModeS} = c \times (t_{(a)} + t_{(b)} + nt_{(c)}) = c \times ((2dn + 5n + 2k - 6 + nk)t_s + (3n-4)t_a + (d+1)(n-2)t_m + t_{min}(n))$. From the above analysis, the reader can observe that the time complexity of the algorithm under the condition $k << d$ is $O(ncd + nck) \simeq O(ncd)$, and the required time primarily depends on the parameters of n, c, and d.

表 1 A comparison of the time complexities of the PSM and DRS techniques. The details of the table are discussed in the text.

Reduction Methods	Big-oh Computation
Random	$O(m)$
RandomC	$O(m)$
KCentres	$O(kn_i^2 + kd + n + k)$
ModeSeek	$O(ncd + nck) \simeq O(ncd)$
PCA	$O(n^2 d)$
LDA	$O(n^2 d)$
PCALDA	$O(n^2 d)$
DCV	$O(n^2 d + c^2 d)$

The Time Complexity of DRS : In [7], it is reported that the time complexities of LDA based methods, such as PCA, PCA+LDA, LDA/GSVD, and RLDA, respectively, are $O(n^2 d)$, $O(n^2 d)$, $O((n+c)^2 d)$, and $O(n^2 d)$, and their space complexities are all the same as $O(nd)$. The details of the above analysis are omitted here in the interest of compactness. Also, in [11], it is reported that DCV requires approximately $(2d(n-c)^2 + 4dnc)$ flops. From this report, the reader can observe that the time complexity of the algorithm is $O(n^2 d + c^2 d)$, and the required time primarily depends on the parameters of n, d, and c.

In summary, to examine the rationality of employing the dimensionality reduction schemes for DBCs, the time complexity required to reduce the dimensionality has been investigated. Table 1 shows a comparison between the time complexities of PSM and DRS techniques.

4. Experimental Results

The PSM and DRS based techniques have been tested and compared. This was done by performing experiments on the well-known benchmark database, namely, "UMIST" face database(注2).

We reduced the dimensionality of the dissimilarity matrix with a DRS, such as PCA [8], LDA [10], PCA-plus-LDA [9], or DCV [11]. In the DRS based techniques, we reduced the dimension $n-1$ to $c-1$, where n is the total number of training samples and c is the number of classes. In the PSM based techniques of *Random*, *RandomC*, *KCentres*, and *ModeSeek*, on the other hand, we selected $c-1$, c, c, and c samples from the training data set as the prototypes of DBCs.

We experimented different classifiers, such as the k-nearest neighbor classifiers (1-NN, 3-NN, 5-NN, 7-NN), the nearest mean classifiers (NMC), the support vector classifier (SVC), and the regularized normal density-based linear/quadratic classifiers (RLDC, RQDC). The classifiers were implemented

(注2) : http://images.ee.umist.ac.uk/danny/database.html

表 2 A comparison of the classification accuracy rates (%) of DBCs for the UMIST database between the PSM and DRS based techniques.

Experimental	1NN	3NN	5NN	7NN
Methods	NMC	RLDC	QLDC	SVC
Random	98.00	98.67	98.67	97.33
	95.33	95.33	99.33	98.67
RandomC	99.33	99.33	98.00	97.33
	97.33	98.00	99.33	100
KCentres	98.67	98.67	98.67	98.00
	96.00	88.00	93.00	98.67
ModeSeek	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	98.67
PCA	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	99.33
LDA	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	99.33
PCALDA	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	100
DCV	99.33	99.33	99.33	99.33
	99.33	99.33	98.67	100

with PRTools (注3).

In comparing the PSM and DRS based techniques, first of all, we measured the classification accuracies (%) of the DBCs for the real benchmark database. Table 2 shows a comparison of the classification accuracy rates (%) (for the process of prototype selection or dimensionality reduction) of DBCs for UMIST.

From Table 2, the reader can observe that the classification performances of the classifiers are improved with the DRS based techniques.

In comparing the PSM and DRS based techniques, we also measured the processing CPU-times (seconds) of the DBCs for the face database. Table 3 shows a comparison of the averaged processing CPU-times (for the process of prototype selection or dimensionality reduction) of DBCs for UMIST. Here, to measure the dissimilarities, we used Euclidean distance (ED), Hamming distance (HD), regional distance (RD) [12], or spatially weighted gray-level Hausdorff distance (WGHD) [13].

From Table 3, it is clearly observed that the processing CPU-times (seconds) increases when the DRS technique is applied.

5. Conclusion

In the attempt to reduce the dimensionality of dissimilarity representation, we can use dimensionality reduction schemes (DRS) as well as prototype selection methods (PSM). In this paper, we considered a comparison between the computational complexities of PSM and DRS techniques. This has been done by theoretically and experimentally analyzing their computational complexities. Our experimental results for the well-known benchmark facial images demonstrated the possibility that DRS could be used efficiently for dissimilarity-based classifiers (DBCs). However, it was also observed that the processing CPU-times (seconds) increased when the DRS technique was applied. The research concerning the reduction of the processing CPU-times is a future aim of the authors.

(注3)：http://www.prtools.org/

表 3 A comparison of the averaged processing CPU-times (seconds) of DBCs for the UMIST database. Each number of the table is obtained by averaging the results of five iterations on a Windows platform (CPU: 2.40 GHz, RAM: 2GB).

Experimental	UMIST			
Methods	ED	HD	RD	WGHD
Random	0.15	0.10	0.10	0.10
RandomC	0.27	0.25	0.26	0.27
KCentres	24.10	23.27	24.44	22.06
ModeSeek	5.91	5.95	5.85	5.85
PCA	52.76	46.32	44.91	43.00
LDA	1.94	1.32	1.35	1.33
PCALDA	42.500	42.50	38.98	41.29
DCV	13.88	13.63	13.751	13.86

Acknowledgments This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD- KRF-2007-313-D00714).

文　献

[1] E. Pekalska and R. P. W. Duin, *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, World Scientific Publishing, Singapore, 2005.

[2] E. Pekalska, R. P. W. Duin, and P. Paclik, "Prototype selection for dissimilarity-based classifiers", *Patt. Recogn.*, vol. 39, pp. 189 - 208, 2006.

[3] S. -W. Kim and B. J. Oommen, "On using prototype reduction schemes to optimize dissimilarity-based classification", *Patt. Recogn.*, vol. 40, pp. 2946–2957, 2007.

[4] K. Riesen, V. Kilchherr, and H. Bunke, "Reducing the dimensionality of vector space embeddings of graphs", In *Proceedings of 5th Int. Conf. on Machine Learning and Data Mining*, vol. LNAI-4571, pp. 563–573, 2007.

[5] S. -W. Kim and J. Gao, "On Using Dimensionality Reduction Schemes to Optimize Dissimilarity-Based Classifiers", In *Proceedings of CIARP 2008*, Havana, Cuba, vol. LNCS-5197, pp. 302–309, 2008.

[6] M. Loog and R. P. W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The ChernoCriterion", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-26, no. 6, pp. 732–739, 2004.

[7] J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 6, pp. 929 - 941, Jun. 2005.

[8] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-22, no. 1, pp. 4–37, 2000.

[9] P. N. Belhumeour, J. P. Hespanha, and D. J. Kriegman,

"Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-19, no. 7, pp. 711–720, 1997.

[10] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data - with application to face recognition", *Patt. Recogn.*, vol. 34, pp. 2067–2070, 2001.

[11] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition" , *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-27, no. 1, pp. 4–13, 2005.

[12] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The problem of compensating for changes in illumination direction", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PAMI-19, no. 7, pp. 721 - 732, 1997.

[13] S. -W. Kim, "Optimizing dissimilarity-based classifiers using a newly modified Hausdorff distance", In *Proceedings of 2006 Pacific Knowledge Acquisition Workshop*, Guilin, China, vol. LNAI-4303, pp. 177–186, 2006.

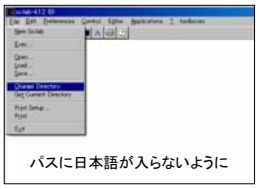
使ってみよう部分空間法
― 部分空間法体験実習 ―

部分空間法研究会
2008年7月28日

堀田政二，天野敏之，玉木徹

はじめに

準備としてこの演習で使用するScilab, Octave, MATLAB
を起動してダウンロードしたプログラムのあるフォルダや
ディレクトリに移動しておいてください



パスに日本語が入らないように

☆ PCを持っていない等，作業できない方は
スライドを見ているだけでも大丈夫（なはず）

実習の目的

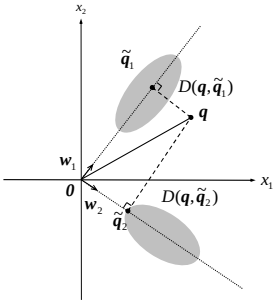
実際に部分空間法のプログラムを動かしてみて
使いやすさやプログラミングの容易さを実感する

実習の内容

- Step1 自己相関行列の固有値と固有ベクトルを見よう
- Step2 画像パターンの直交展開してみよう
- Step3 部分空間法で手書き数字を認識してみよう

部分空間法のCLAFICとは

クラスらしさを部分空間で表現し
識別にも直接部分空間を利用



テストパターンを良く近似できる
部分空間の属すクラスを出力

識別規則

$$\omega = \arg \max_{j=1,\cdots,C} \|\mathbf{W}_j^\top \mathbf{q}\|^2$$

部分空間は原点共通

クラス毎の部分空間の求め方

$$\max_{\mathbf{w}} \sum_{i \in C_j} (\mathbf{w}^\top \mathbf{x}_i)^2 \quad \text{s.t. } \|\mathbf{w}\|^2 = 1$$

ラグランジュ関数は

$$L(\mathbf{w}, \lambda) = \mathbf{w}^\top (\sum_{i \in C_j} \mathbf{x}_i \mathbf{x}_i^\top) \mathbf{w} - \lambda(\mathbf{w}^\top \mathbf{w} - 1)$$

$\partial L / \partial \mathbf{w} = \mathbf{0}$ として実際に解くと

$$(\textstyle\sum_{i \in C_j} \mathbf{x}_i \mathbf{x}_i^\top)\mathbf{w} = \lambda \mathbf{w}$$

各クラスの正規直交基底は自己相関行列の固有ベクトル

部分空間法の識別

クラス j の自己相関行列の固有値の大きい
上位 r 本の固有ベクトルを並べた行列：$\mathbf{W}_j$

$$\omega = \max_{j=1,\cdots,C} \|\mathbf{W}_j^\top \mathbf{q}\|^2$$

パラメータ r (一つだけ！) は累積寄与率や
パラメータ推定法によって決定（実験例は後述）

メモリ容量や計算時間を削減したい場合にはrで
調節可能な点が便利（SVMでは無理）

線型部分空間を使うと何がうれしいのか

辞書を少ないメモリ容量で保存できる

高速にパターンを多クラスに分類できる

簡単に実装できて高い識別率を達成できる

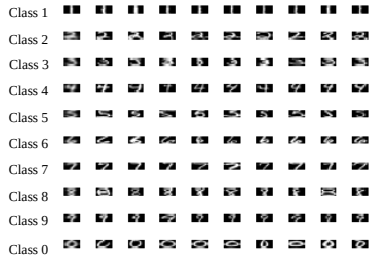
これらを実際に体感してみましょう

7

演習で用いる手書き数字データUSPSの詳細

訓練パターン数 7291，未知パターン数 2007

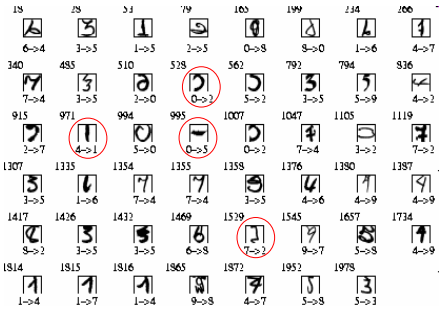
クラスは0から9までの10クラス 16×16pixelの256階調のモノクロ画像



訓練データの一例：左端はクラスの重心

8

テストデータの例

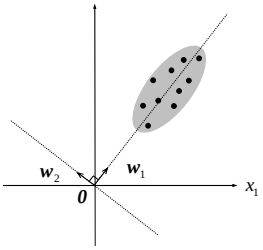


ミスラベルや意味不明のパターンが含まれている

9

主成分分析でデータの分布を観察する

データ分布全体を最も良く近似する部分空間を求める



データの分布を楕円で近似して長軸と短軸を求める
上の図の場合，長軸に正射影すれば元のデータ分布を良く近似できる

10

主成分分析

$$\max_{\mathbf{w}} \ \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \qquad \text{s.t. } \|\mathbf{w}\|^2 = 1$$

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^\top$$ (標本) 共分散行列

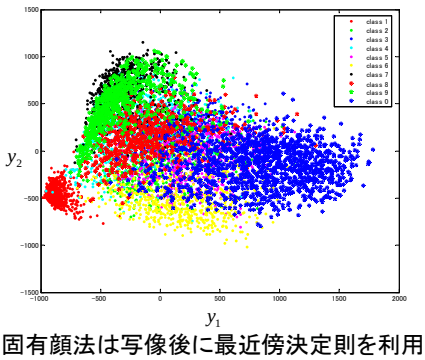
ラグランジュ乗数法から解は $\boldsymbol{\Sigma}\mathbf{w} = \lambda\mathbf{w}$

共分散行列の大きい固有値（分散）に
対応する上位r本の固有ベクトルでデータを写像

$$\mathbf{y} = \mathbf{W}^\top(\mathbf{x} - \mathbf{m})$$

11

主成分分析による二次元空間への写像

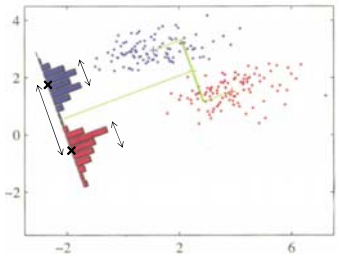


固有顔法は写像後に最近傍決定則を利用

12

線形判別分析でデータの分布を観察する

異なるクラス同士をなるべく離すような
線型写像を求めて識別に利用する



13

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}}$$

クラス間共分散行列

$$\mathbf{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^\top$$

クラス内共分散行列

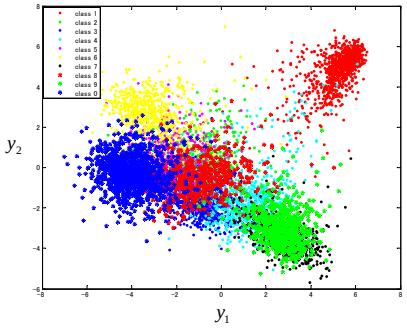
$$\mathbf{S}_W = \sum_{j=1,2} \sum_{i \in C_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^\top$$

極大点は微分して0となるから　$\partial J(\mathbf{w}) / \partial \mathbf{w} = \mathbf{0}$

結局求めたいWは　$\mathbf{w} \propto \mathbf{S}_W^{-1}(\mathbf{m}_2 - \mathbf{m}_1)$

14

線形判別分析による二次元空間への写像



識別は閾値処理，最小距離法等を利用

15

代表的な手法のテストデータに対する誤識別率

識別法	誤識別率 (%)
最近傍決定則	5.5
Tangent Distance [1]	2.6
P2DHMDM with 3-nearest neighbor [2]	1.9
Human error rate [3]	2.5
Human error rate [4]	1.5

[1] Simard et al., NIPS, pp.50-58, 1993
[2] Keysers et al., ICPR, vol.4, pp.511–514, 2004.
[3] Bromley and Sackinger, Tech. Rep., 11359--910819-16TM, AT&T, 1991
[4] Dong, Report. CENPARMI, Concordia Univ., 2001.

16

演習　Step1
部分空間の直交基底をクラス毎に見てみる

17

演習で用いるデータファイルの内訳

画像は画素値を左上から順に縦に並べてベクトル化

test_data.txt　　テストデータを格納した256x2007 の行列

test_label.txt　　テストデータのラベルを格納した
2007x1 のベクトル

trai_data.txt　　訓練データを格納した256x7291 の行列

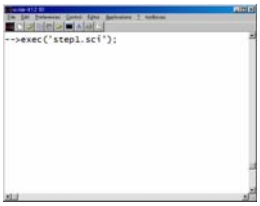
trai_label.txt　　訓練データのラベルを格納した
7291x1 のベクトル

18

Step1のプログラムを実行する

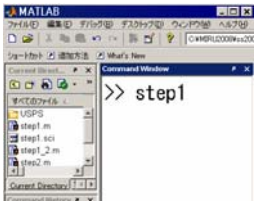
Scilab

以下のように入力



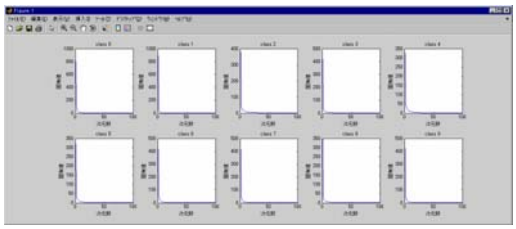
Octave, MATLAB

プロンプトでstep1と入力



19

実行画面



自己相関行列の固有値の変化
縦軸が固有値，横軸が固有値の番号 (降順ソート)

少ない基底数でクラスの分布が近似できる

20

各クラスの直交基底

各クラスの基底を使えばパターン変動が表現できる

変動の大きいパターンに有効

21

プログラムの中身（Octave，MATLABの例）

```
%% 各クラスで部分空間を求める
W=zeros(Dim,100,10); % 各クラスの直交基底 次元数×基底数×クラス数
figure(1), clf, axis square;
for j = 0 : 9
  X=D(:,find(trai_label==j)); % ラベルがjの訓練パターンをXに格納
  C=XXX';                     % 自己相関行列を計算
  [eig_vec, eig_val]=eig(C);  % 固有ベクトル，固有値を計算
  [value index]=sort(-diag(eig_val)); % 固有値を降順に並べ替え
  W(:,:,j+1)=eig_vec(:,index(1:100)); % 固有値の大きい上位100本に対応する固有ベクトルをWに格納
  figure(1),subplot(2,5,j+1), plot(-value(1:100)); % 固有値をプロット
  s=sprintf('class %d', j);, title(s);
  xlabel('次元数');,ylabel('固有値');
  fprintf('class %d ... OKYn',j);
end
```

各クラスの直交基底を求めるプログラムは
5，6行で書けてしまう！

22

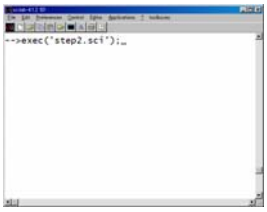
演習　Step2
画像パターンを直交展開してみる

23

Step2のプログラムを実行する

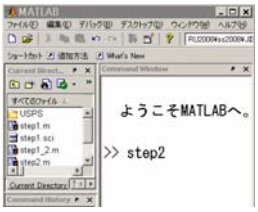
Scilab

以下のように入力



Octave, MATLAB

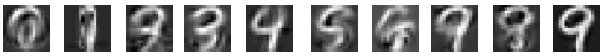
プロンプトでstep2と入力



24

実行画面

テストパターン



0 1 2 3 4 5 6 7 8 9

各クラスで30本の正規直交基底の線型結合によって
近似されたテストパターン

🙂 プログラム中のrの値を変えると近似の
程度が変わるので変えて実行してみよう

テストパターンを変えたいときはプログラムの先頭付近の
xの部分を1から2007の値で指定してください

プログラムの中身（Octave, MATLABの例）

どのように近似パターンを作成しているか

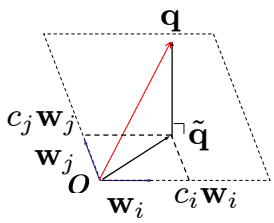
```
for j = 0 : 9
 c=W(:,1:r,j+1)'*Q(:,x); % 係数を計算
 QA=W(:,1:r,j+1)*c;        % 線型近似パターンの作成
 for i = 1 : 16, IMG(i,:)=QA((i-1)*16+1:i*16,1)';,end
 IMG=IMG-min(min(IMG));, IMG=IMG./max(max(IMG));
 figure(1),subplot(2,10,11+j),imshow(IMG);,
 s=sprintf('class %d',j);, title(s);
end
```

なぜこれで近似パターンを作成できる？

正規直交基底による信号の展開

元信号 ← 近似信号 ← 正規直交基底

$$\mathbf{q} \simeq \tilde{\mathbf{q}} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_r\mathbf{w}_r$$



係数は元信号と正規直交基底との
内積値

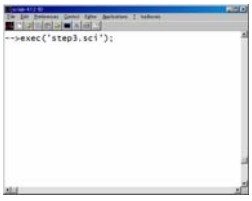
$$c_i = \mathbf{w}_i^\top \mathbf{q}$$

演習　Step3
部分空間法で手書き数字認識

Step3のプログラムを実行する

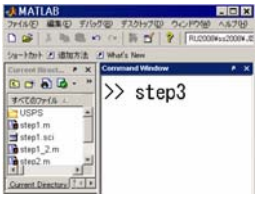
Scilab

以下のように入力

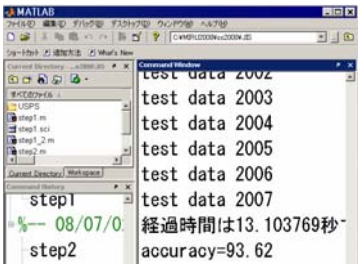


Octave, MATLAB

プロンプトでstep3と入力



実行画面



識別に要した時間と識別精度が表示される

※ 環境によって遅いかもしれませんがATLASで高速化できます

CONFと入力

```
>> CONF

CONF =

   351     3     3     0     1     0     0     0     1     0
     0   257     0     0     3     1     3     0     0     0
     7     0   180     2     3     1     0     1     4     0
     2     0     3   147     0    10     0     0     4     0
     0     3     2     0   182     1     0     1     2     9
     4     1     0     3     1   148     0     0     1     2
     0     0     1     0     2     5   161     0     1     0
     0     1     1     0     3     0     0   138     1     3
     4     5     3     4     0     2     2     0   146     0
     0     3     0     0     0     1     0     3     1   169
```

混同行列（confusion matrix）が表示される

☺プログラム中のrの値を変えて識別精度や識別時間がどう変化するか観察しましょう

31

プログラムの中身

```
for i = 1 : test_num
  for j = 0 : 9, S(j+1)=sum((W(:,1:r,j+1)'*Q(:,i)).^2);, end
  [value index]=max(S);
  CONF(test_label(i)+1,index)=CONF(test_label(i)+1,index)+1;
  fprintf('test data %d¥n',i);
end
```

識別部分は一行で書ける！

32

パラメータの決め方

累積寄与率：r次元の部分空間を張る基底に対する固有値の総和を全固有値の総和で割ったもの

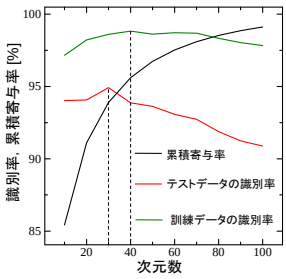
$$\sum_{i=1}^{r} \lambda_i / \sum_{j=1}^{d} \lambda_j$$

分割学習法：交差確認法などのパラメータ推定法

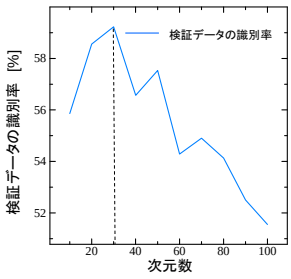
訓練データをテストデータと識別部を設計するためのデータに分けて識別率を算出しパラメータを推定

33

実験例



累積寄与率



分割学習法

パラメータ推定法を用いた方が良い

34

比較実験

CPU 1.86GHz，メモリ 2GbのPC上のMATLABを使用

識別法	
最小距離法	各クラスの重心との距離で識別
最近傍決定則	最近傍パターンのラベルを出力
判別分析(次元数9)	写像先の空間で最小距離法
固有数字(次元数40)	写像先の空間で最近傍決定則
部分空間法(次元数30)	CLAFIC
SVM (RBF Kernel)	SVMLIB※ を利用

※ http://www.csie.ntu.edu.tw/~cjlin/libsvm

35

各手法の比較

テストデータに対する誤識別率，1パターンあたりの識別時間，辞書サイズ

識別法	誤識別率 (%)	識別時間	辞書サイズ
最小距離法	18.6	3.8×10^{-5}	256×10
最近傍決定則	5.5	0.02	256×7291
判別分析(次元数9)	11.5	3.9×10^{-5}	9×10
固有数字(次元数40)	4.9	0.03	40 ×7291
部分空間法(次元数30)	5.1	0.001	256×300
SVM (RBF Kernel)	4.6	0.005	256×3220

36

部分空間法の特長 (1)

☆ クラスを表現する直交基底を
そのまま識別に利用できる

固有顔や固有空間法では識別則は別途必要

☆ 2クラス専門の識別器でないので
多クラス化が容易

SVMは多クラス化は面倒

学習により他クラスを考慮できる

37

部分空間法の特長 (2)

☆ 辞書サイズを部分空間の
次元数で調節できる

SVMでは辞書サイズの調節は困難

☆ 大量の訓練データを扱える

カーネル行列を使う手法では訓練データ数×訓練データ数の行列が必要なため，訓練データ数が多いと学習できない

☆ 理論的な拡張が容易

相互部分空間法や相対KL展開

38

さらに学びたい人のために

39

参考文献

- E. Oja, ``Subspace methods of pattern recognition,'' Research Studies Press, 1983. (小川英光，佐藤 誠訳，パターン認識と部分空間法，産業図書，1986) 絶版
- 黒沢由明, `` Subspace2006 部分空間法入門," 部分空間法研究会 Subspace2006, pp.136-143, 2006.
- 石井ら，わかりやすいパターン認識，オーム社，1998.
- 黒沢由明, ``部分空間法の今昔（上）：歴史と技術的俯瞰：誕生から競合学習との出会いまで ,'' 情報学誌，vol.49, no.5, pp.76-82, 2008.
- 福井和広 , ``部分空間法の今昔（下）：最近の技術動向：相互部分空間法への拡張とその応用事例 ,'' 情報学誌，vol.49, no.6, pp.82-87, 2008.

40

WhatとHow：部分空間法の歴史から学ぶもの

小川　英光†

† 東京福祉大学教育学部　〒372-0831 群馬県伊勢崎市山王町 2020-1
E-mail: †hidemitsu-ogawa@kuramae.ne.jp

あらまし　我々が研究を進めていく上で最も重要なことは，問題をどのようにして解くか（How）ではなく，どのような問題を論じるか（What）ということである．問題を定式化する際には，問題の解き方や計算の仕方を紛れ込ませることなく，何をしたいかだけを明確にして事に当たることが重要である．部分空間法の歴史に例をとりながら，What と How を切り分けることが如何に困難であるか，また，両者を混同するとどのような問題が生じるか，両者を切り分けるためにはどうすればよいかという問題について論じる．

キーワード　部分空間法，K-L 展開，K-L 部分空間，K-L 固有元，what と how

What and How: Message from a History of Subspace Methods

Hidemitsu OGAWA†

† School of Education, Tokyo University and Graduate School of Social Welfare
Sanno-cho 2020-1, Isesaki, Gunma, 372-0831 Japan
E-mail: †hidemitsu-ogawa@kuramae.ne.jp

Abstract　The key to doing good research often lies not in the method to solve a problem, i.e., *how* of the problem but in making clear the exact problem to be solved, i.e., the *what* of the problem; and the ability to clearly distinguish between these two components during problem formulation. By using examples from a history of research on subspace methods, we discuss how difficult it is to keep *what* and *how* separate, illustrate the consequences of confusing them, and provide tips to clearly distinguish between these two fundamental components of problem solving.

Key words　Subspace method, K-L expansion, K-L subspace, K-L eigenelement, what and how

1. はじめに

我々が研究を進めていく際に最も重要なことは，問題をどのようにして解くか（How）ではなく，どのような問題を論じるか（What）ということである．

「問題の定式化と問題の解き方は同時に決まる」と言われることがあるが，それは間違いである．「解き方」に対す意識が少しでも心の中にあれば，問題を定式化する際に，その心の中の「解き方」に無意識のうちに引きずられてしまい，表面的な定式化，複雑な定式化，あるいは間違った定式化を行ってしまう．

逆に，正しい問いかけをすれば，解き方も自然に湧き出てきて，意外に容易に正しい答えに到達できる場合もあるのである．そのような例は，歴史に残る重要な研究でも多く見ることができる(注1)．幸か不幸か，部分空間法の歴史においても，同様な例をみることができる．そこで本論文では，部分空間法の問題を例にとりながら，What と How を切り分けることが如何に重要であるかという問題を論じることにする．

Karhunen-Loève 展開（K-L 展開）は，パターン認識や画像処理の分野に定着している概念である．ほとんどの標準的書籍 [1]–[14] にその記述が見られる．しかし，それらの書籍 [2]–[14] 及び論文 [15]–[17] を注意深く読んでみると，意外にも，不完全な記述しかなされていないことに驚かされる．

不完全という意味には二つある．第一は，K-L 展開の意味づけの問題である．ほとんどの文献 [2]–[17] では，直接・間接に

(注1)：一例を示す．ハイゼンベルグは，1925 年に量子力学の理論を作った後，霧箱の中の電子の軌跡をどう記述すればよいかという問題で 2 年間も苦しんだ．ボーアとの激しい議論が続いた後，疲れ果てて星空の下の公園を散策しているとき，アインシュタインに初めて会ったときに言われた言葉，すなわち，「我々が自然界を認識することは，理論という枠組みをとおして初めて可能になるのだ．」という言葉をふと思い出した．そして，「霧箱の中の電子の軌跡をどう記述するか」と問うのではなく，「量子力学の理論で何が分かるのか」という問いかけをした．するとたちまち，不確定性関係が導かれ，あの不確定性原理が生まれたのである [26] ．

「パターン集合を最良に近似できるものは K-L 展開だけであり，それ以外の直交系では無理である」という表現がとられている．これを本論文では,「K-L 展開の必要十分性」と呼ぶことにする．

ところで，よく知られているように，K-L 展開を N 項で打ち切った場合，この n 個の固有元（これを以下では K-L 固有元と呼ぶ）の線形和で表現できるパターンは，その固有元が張る部分空間の中の任意の正規直交基底を使って完全に表現できる．しかも，それらの直交基底は必ずしも K-L 固有元になっていないのである．したがって，K-L 展開は十分であるが，必要ではないのである．本質的な意味を持っているものは，K-L 固有元そのものではなく，K-L 固有元によって張られる部分空間である．

ところが，ほとんどの書籍では，K-L 展開は，特徴抽出に関する章の中で論じられている．そこでは，次のような表現がとられている．すなわち,「ある直交系でパターンを展開したとき，その展開係数からできるベクトルを特徴ベクトルと呼び，そのとき使った直交系を特徴軸と呼ぶ．そして，あるパターン集合を最良に近似する特徴軸が，その集合に対する K-L 固有元である.」という表現がとられている．そこでは，軸そのものが重要な意味を持っている．しかし，すぐ上で述べたように，パターンの近似を論じるときには，軸そのものは何も本質的な意味を持っておらず，その軸によって張られる部分空間が重要である．したがって，特徴軸という概念そのものが，この文脈では意味をなさなくなるのである．

K-L 展開に関する第二の問題は，証明の不完全さである．これらの文献の中には，あたかも K-L 展開が必要十分であるかのごとき表現をとりながら，十分性しか証明していないもの，しかも，その十分性の証明が不完全なもの，必要性まで証明してしまったもの等々がある．前述のごとく，パターン集合の近似という立場からみたとき，K-L 展開は十分条件になっているが，必要条件にはなっていない．したがって，もし必要条件を証明できたとしたら，その証明は間違っていることになる．

本論文では，部分空間法における基本定理ともいうべき問題を整理し，その証明に挑んだいくつかの事例をとおして，なぜ従来多くの研究者が間違いを繰り返してきたのか，なぜその間違いに気づかなかったのか，間違いを起さないようにするためにはどうすればよいか，という問題について論じることにする．

2. 問題の定式化

議論の対象になるパターンが属する空間を H で表す．H は複素ヒルベルト空間になっているものとする(注2)．したがって，H の元はベクトルになることもあるし，関数になることもある．さらに，音声のように１変数関数のこともあるし，文字や画像のように２変数関数になることもある．それらを一般的に f で表すことにする．

認識対象になるカテゴリを任意に１つ固定し，Ω で表す．パターン集合 Ω (注3) を最良に近似する部分空間を求めることが，部分空間法における最も基本的な課題である．

Ω の相関作用素を R で表す．R は，ヒルベルト空間の言葉を使って，次のように定義される．

$$R = E(f \otimes \overline{f}) \tag{1}$$

ここで，E は $f \in \Omega$ に関する平均である．$f \otimes \overline{f}$ は，ノイマン・シャッテン積と呼ばれるものであり，一般の $f, g \in H$ と $h \in H$ に対して，

$$(f \otimes \overline{g})h = \langle h, g \rangle f \tag{2}$$

によって定義される [22] ．式 (2) の右辺の $\langle h, g \rangle$ は，H における内積である．なお，式 (2) の g の上についている横棒は，複素共役の意味ではなく，ノイマン・シャッテン積の記号の一部である．

相関作用素 R は半正値自己共役作用素であり，任意の $\phi \in H$ に対して，次の関係が成立している．

$$E|\langle f, \phi \rangle|^2 = \langle R\phi, \phi \rangle \tag{3}$$

R の固有値，及び，大きさを 1 に正規化した固有元を λ_n, φ_n で表す：

$$R\varphi_n = \lambda_n \varphi_n \qquad : \lambda_1 \geqq \lambda_2 \geqq \cdots \geqq 0 \tag{4}$$

$\{\varphi_n\}_{n=1}^{\infty}$ に関する f の展開は，Karhunen-Loève 展開，あるいは K-L 展開と呼ばれる．そこで，$\{\varphi_n\}_{n=1}^{\infty}$ を K-L 固有元と呼ぶことにする．

カテゴリ Ω を近似する部分空間として，２種類の重要な部分空間を導入する．第１はカテゴリ Ω を最良に近似する M 次元部分空間であり，S_0 で表す．第２は，R の固有値の大きな方から選んだ M 個の K-L 固有元 $\{\varphi_n\}_{n=1}^{M}$ で張られる部分空間であり，S_{KL} で表す．この空間を K-L 固有空間と呼ぶことにする．

部分空間 S_0 として S_{KL} がよく使われるが，その理論的根拠は，次の定理による．

［**定理** 1］ (部分空間法における基本定理) カテゴリ Ω を最良に近似する M 次元部分空間は S_{KL} であり，S_{KL} に限る．すなわち，$S_0 = S_{\mathrm{KL}}$ である．

本論文は，この定理にまつわる出来事から得られる貴重な教訓について論じたものである．

3. 従来の証明法とその問題点

定理 1 の従来から行われている証明を２種類紹介し，その問題点を明らかにする．

部分空間は，その空間を張るような正規直交系によって表すことができる．したがって，

$$『S_0を張るような正規直交系が \{\varphi_n\}_{n=1}^{M} になる』 \tag{5}$$

ことを示すことが，従来とられてきた多くの証明の基本方針である．これから紹介する証明も，この方針に従っている．まず準備として，レイリーの原理を示す．

(注2)：実ヒルベルト空間の場合も，ほとんど平行した議論ができる．

(注3)：Ω はカテゴリを表す記号であり，f はヒルベルト空間 H の元であるから，本来別の概念である．しかし，混乱の恐れがないことと，記述を簡単にするために，Ω に属するパターンの全体も同じ記号 Ω で表し,「パターン集合 Ω」といったり，$f \in \Omega$ と表記することにする．

3.1 レイリーの原理

2次形式の最大値問題に関するレイリーの原理を述べる(注4).

［**補題** 1］ (レイリーの原理) A を H 上の半正値自己共役作用素とする．A の固有値を大きさの順に並べて λ_n で表し，対応するノルムを1に正規化した固有元を u_n で表す：

$$Au_n = \lambda_n u_n \qquad : \lambda_1 \geqq \lambda_2 \geqq \cdots \geqq 0 \tag{6}$$

A の2次形式

$$J_1[\phi] = \frac{\langle A\phi, \phi \rangle}{\| \phi \|^2}$$

に対して，次の関係が成立する．ここで，$\| \phi \|$ は H におけるノルムである．

(i) 任意の $\phi \neq 0$ に対して，

$$J_1[\phi] \leqq J_1[u_1] = \lambda_1$$

となる．

(ii) $\{u_n\}_{n=1}^{m-1}$ に直交している $\phi \neq 0$ に対して，

$$J_1[\phi] \leqq J_1[u_m] = \lambda_m$$

となる．

3.2 逐次法（レイリーの原理による方法）

パターン認識関係の多くの書籍では，定理1が次のように証明されている．一般に，H の元 $\{f_n\}_{n=1}^{M}$ で張られる部分空間を $\mathcal{L}(\{f_n\}_{n=1}^{M})$ で表すことにする．例えば $S_{\mathrm{KL}} = \mathcal{L}(\{\varphi_n\}_{n=1}^{M})$ である．

(i) まず $M = 1$ の場合を考える．H に属するノルム1の元 ϕ で張られる1次元部分空間が Ω を最良に近似するように，すなわち，$\mathcal{L}(\{\ \phi\ \}) = S_0$ となるように ϕ を決定する．$f \in H$ の $\mathcal{L}(\{\ \phi\ \})$ への正射影成分は，ϕ のノルムが1であるから，$\langle f, \phi \rangle$ で与えられる．そこで，Ω に属するすべての f に関する2乗平均

$$J_2[\phi] = E|\langle f, \phi \rangle|^2 = < R\phi, \phi > \tag{7}$$

を考える．この式の第2の等号は，式 (3) によるものである．式 (7) を最大にする ϕ が，S_0 を張る正規直交系である．したがってレイリーの原理により，式 (7) は $\phi = \varphi_1$ に対して最大値 λ_1 をとる．よって，$S_0 = \mathcal{L}(\{\varphi_1\})$ である．

(ii) $M = 2$ の場合，レイリー の原理により，

$$\langle \phi, \varphi_1 \rangle = 0 \tag{8}$$

なる条件のもとで，式 (7) は $\phi = \varphi_2$ に対して最大値 λ_2 をとる．よって，$S_0 = \mathcal{L}(\{\varphi_1, \varphi_2\})$ である．

(iii) 以下同様にして，一般の M に対して定理1が成立する．■

この証明法において (i) は正しい．しかし，(ii) の最後の部分は，これだけでは正しい論理の進め方になっていない．すなわち，(ii) で主張していることは，

『$\phi = \varphi_2$ が，条件 (8) のもとで式 (7) を最大にする』

ということであって，

『Ω を最良に近似する2次元部分空間が $\mathcal{L}(\{\varphi_1, \varphi_2\})$ である』

ということは，論理的には何も示されていない．証明を完成させるためには，更に論理の詰めが必要である．

次節で述べる証明法は，この問題に直接答えようとしたものである．

3.3 直 接 法

本節で紹介する証明法は，Ω を最良に近似する M 次元部分空間 S_0 を張るような M 個の元からなる正規直交系を直接求めようとする試みである．$\{\phi_n\}_{n=1}^{M}$ を H の正規直交系とする．すなわち，

$$\langle \phi_m, \phi_n \rangle = \delta_{m,n} \qquad : 1 \leqq m, n \leqq M \tag{9}$$

とする．ここで $\delta_{m,n}$ はクロネッカーのデルタと呼ばれ，

$$\delta_{m,n} = \begin{cases} 1 & : \quad m = n \\ 0 & : \quad m \neq n \end{cases}$$

によって定義される．

この正規直交系 $\{\phi_n\}_{n=1}^{M}$ を使って Ω を最良に近似するためには，汎関数

$$J_3[\{\phi_n\}] = E\sum_{n=1}^{M} |\langle f, \phi_n \rangle|^2 \tag{10}$$

を最大にする $\{\phi_n\}_{n=1}^{M}$ を求めればよい．式 (3) より，式 (10) は

$$J_3[\{\phi_n\}] = \sum_{n=1}^{M} \langle R\phi_n, \phi_n \rangle \tag{11}$$

と表すことができる．したがって問題は，条件 (9) のもとで式 (11) を最大にする $\{\phi_n\}_{n=1}^{M}$ を求める問題になる．

ここで，式 (9) の条件を少し緩めて，

$$\| \phi_n \|^2 = 1 \qquad : 1 \leqq n \leqq M \tag{12}$$

だけを要請してみる．すなわち，式 (12) の条件のもとで式 (11) を最大にする $\{\phi_n\}_{n=1}^{M}$ を求める問題をまず解くことにする．

ラグランジュ未定乗数の組を $\{\lambda_n\}_{n=1}^{M}$ とすれば，式 (11) は，$\{\phi_n\}_{n=1}^{M}$ と $\{\lambda_n\}_{n=1}^{M}$ に関する条件なしの変分問題

$$J_3[\{\phi_n, \lambda_n\}] = \sum_{n=1}^{M} \langle R\phi_n, \phi_n \rangle - \sum_{n=1}^{M} \lambda_n (\| \phi_n \|^2 - 1) \tag{13}$$

に変換できる．

式 (13) を各 λ_n について微分し零とおけば，式 (12) を得る．式 (13) を各 ϕ_n について変分し零とおけば，R が自己共役であることから，

$$R\phi_n = \lambda_n \phi_n \qquad : 1 \leqq n \leqq M \tag{14}$$

となる．すなわち，λ_n 及び ϕ_n は，相関作用素 R の固有値と固有元になる．

そこで，何番目の固有元を採用すればよいかを調べる．式 (14) を式 (11) に代入すれば，式 (12) より，

(注4)：固有値を求めるための数値計算法の1つであるベキ乗法は，このレイリーの原理を使ったものである [24] .

$$J_3[\{\phi_n\}] = \sum_{n=1}^{M} \lambda_n \tag{15}$$

となる．よって，式 (15) の J_3 を最大にするためには，R の固有値の大きな方から M 個とればよい．すなわち，

$$\phi_n = \varphi_n \qquad : 1 \leqq n \leqq M \tag{16}$$

とすればよい．

式 (12) の条件のもとで式 (11) を最大にする問題を解いたところ，結果は自動的に式 (9) の条件を満たしていたのである．よって，式 (9) の条件のもとで式 (11) を最大にする ϕ_n は，式 (16) で与えられることになる． ■

こうして，$\{\phi_n\}_{n=1}^{M}$ が K-L 固有元でなければいけないこと，すなわち，K-L 展開の必要性が導かれた．しかし，式 (14) が導出されたこと自体が間違いなのである．実際，$\{\varphi_n\}_{n=1}^{M}$ を空間 S_{KL} の中で回転したものを $\{\phi_n\}_{n=1}^{M}$ として採用した場合，固有値 $\{\lambda_n\}_{n=1}^{M}$ がすべて縮退していない限り，$\{\phi_n\}_{n=1}^{M}$ はもはや R の固有元にはならない．しかし，空間そのものは $\mathcal{L}(\{\phi_n\}_{n=1}^{M}) = S_{\mathrm{KL}}$ と変化しないし，J_3 も，$J_3[\{\phi_n\}] = J_3[\{\varphi_n\}]$ と，同じ値をとるのである．

4. 厳密な証明

文献 [18] に従って，定理 1 の厳密な証明を与える．そのための準備として，まず，作用素に関するシュミットの内積，及び，相関作用素 R の不変部分空間についてまとめておく．

4.1 シュミットノルムとシュミットの内積

ヒルベルト空間 H 上の有界線形作用素 A を考える．$\{u_n\}_{n=1}^{\infty}$ を H の正規直交基底とする．

$$\sum_{n=1}^{\infty} \| Au_n \|^2$$

が有限な値をとるとき，その値は，正規直交基底 $\{u_n\}_{n=1}^{\infty}$ の取り方によらず一定になる．その値の平方根を A のシュミットノルムといい，$\| A \|_2$ で表す [22] [23] (注5)：

$$\| A \|_2 = \left(\sum_{n=1}^{\infty} \| Au_n \|^2 \right)^{1/2} \tag{17}$$

シュミットノルムが有限になるような有界線形作用素の全体を，シュミットクラスの完全連続作用素といい，$\sigma\mathrm{c}$ で表す [22]．

例えば，値域が有限次元になるような作用素は $\sigma\mathrm{c}$ に属す．しかし，恒等作用素は $\sigma\mathrm{c}$ に属さない．また，パターン認識の分野で重要な働きをするボケの変換は $\sigma\mathrm{c}$ に属すけれども，平行移動を行う作用素は $\sigma\mathrm{c}$ に属さない．このように，$\sigma\mathrm{c}$ は有界線形作用素の全体がなす空間の部分空間になっている．

$\sigma\mathrm{c}$ に属す作用素 A, B に対して，

$$\langle A, B \rangle = \sum_{n=1}^{\infty} \langle Au_n, Bu_n \rangle \tag{18}$$

を定義することができる．右辺の内積は，ヒルベルト空間 H における内積である．右辺の総和の値は，正規直交基底 $\{u_n\}_{n=1}^{\infty}$ の取り方によらず一定になる．そこで，$\langle A, B \rangle$ をシュミットの内積という [22] [23] (注6)．

作用素 A, B, X，及び，H の元 f に対して，次の公式が成立する．ただし，X^* は X の共役作用素である．

$$\langle AX, B \rangle = \langle A, BX^* \rangle \tag{19}$$

$$\langle XA, B \rangle = \langle A, X^*B \rangle \tag{20}$$

$$\langle A, B \rangle = \mathrm{tr}(AB^*) \tag{21}$$

$$\langle A, f \otimes \overline{f} \rangle = \langle Af, f \rangle \tag{22}$$

$$\| f \|^2 = \mathrm{tr}(f \otimes \overline{f}) \tag{23}$$

ここで，$\mathrm{tr}(A)$ は作用素 A のトレースであり，H の任意の正規直交基底 $\{u_n\}_{n=1}^{\infty}$ を用いて，

$$\mathrm{tr}(A) = \sum_{n=1}^{\infty} \langle Au_n, u_n \rangle \tag{24}$$

により定義される(注7)．この式の右辺が有限な値になれば，その値は $\{u_n\}_{n=1}^{\infty}$ の取り方によらず一定になる．なお，$A \in \sigma\mathrm{c}$ であっても，$\mathrm{tr}(A)$ が存在しないことがある．しかし，A の値域が有限次元の場合，$\mathrm{tr}(A)$ が必ず存在するので，以下の議論では特に問題は生じない．

4.2 相関作用素 R の不変部分空間

S を H の閉部分空間とする．H 上の作用素 A に対して，

$$AS \subseteqq S \tag{25}$$

が成立するとき，すなわち，S の元を A で変換しても再び S に含まれるとき，S を A の不変部分空間という．相関作用素 R は自己共役であるから，次の補題が成立する．

(注5)：シュミットノルムは，行列に対するフロベニウスノルム [25]，すなわち，行列 $A = (a_{m,n})$ に対して

$$\| A \|_2 = \left(\sum_{m=1}^{N} \sum_{n=1}^{N} |a_{m,n}|^2 \right)^{1/2}$$

で定義されるノルムの拡張概念である．それは次のようにして分かる．$\mathbf{C}^N$ の標準基底を $\{\boldsymbol{e}_n\}_{n=1}^{N}$ で表す．すなわち，$\boldsymbol{e}_n$ は第 n 成分が 1 で，それ以外の成分が 0 となる N 次元ベクトルである．式 (17) で $H = \mathbf{C}^N$, $u_n = \boldsymbol{e}_n$ と置けば，フロベニウスノルムになる．

(注6)：シュミットの内積は，行列 $A = (a_{m,n})$, $B = (b_{m,n})$ に対して

$$\langle A, B \rangle = \sum_{m=1}^{N} \sum_{n=1}^{N} a_{m,n} \overline{b_{m,n}}$$

で定義される内積の拡張概念である．すなわち，この式は，式 (18) で $H = \mathbf{C}^N$, $u_n = \boldsymbol{e}_n$ と置いたものになっている．

(注7)：作用素のトレースは，行列 $A = (a_{m,n})$ に対して

$$\mathrm{tr}(A) = \sum_{n=1}^{N} a_{n,n}$$

で定義されるトレースの拡張概念である．すなわち，この式は，式 (24) で $H = \mathbf{C}^N$, $u_n = \boldsymbol{e}_n$ と置いたものになっている．

［**補題 2**］（R の不変部分空間 1) [20] [21] H の閉部分空間 S が相関作用素 R の不変部分空間になるための必要十分条件は，

$$PR = RP \tag{26}$$

が成立することである．

［**補題 3**］（R の不変部分空間 2) [19] H の閉部分空間 S が相関作用素 R の不変部分空間になるための必要十分条件は，S が R の固有元で張られる部分空間になることである．

4.3 厳密な証明

S を H の有限 M 次元部分空間とし，S への正射影作用素を P で表す．S と P は一対一に対応している．そこで，P に関する汎関数

$$J_4[P] = E \parallel Pf \parallel^2 \tag{27}$$

を考えることにする．定理 1 を証明するためには，式 (27) の $J_4[P]$ が最大になるための必要十分条件が $S = S_{\rm KL}$ であることを示せばよい．

まず，式 (27) をシュミットの内積を用いて表現する．ノイマン・シャッテン積に関して，

$$(Af) \otimes \overline{(Bg)} = A(f \otimes \overline{g})B^*$$

なる関係が成立している．よって，式 (27), (23), (1), (21) より，

$$\begin{aligned} J_4[P] &= E \parallel Pf \parallel^2 \\ &= E{\rm tr}[(Pf) \otimes \overline{(Pf)}] \\ &= E{\rm tr}[P(f \otimes \overline{f})P^*] \\ &= {\rm tr}[PE(f \otimes \overline{f})P^*] \\ &= {\rm tr}(PRP^*) \\ &= \langle PR, P\rangle \end{aligned}$$

となり，

$$J_4[P] = \langle PR, P\rangle \tag{28}$$

となる．

ところで，P が正射影作用素であること，すなわち，

$$P^2 = P,\ P^* = P \tag{29}$$

が成立することと，

$$P^*P = P \tag{30}$$

が成立することとは同値である．また，部分空間 S の次元が M であるということと，

$$\langle P, P\rangle = M \tag{31}$$

とは同値である．よって問題は，

『式 (30) , (31) を満たす P の中で式 (28) を最大にするものを求めよ.』

ということになる．

この条件付変分問題は，C と λ をそれぞれラグランジュ未定作用素，及び，ラグランジュ未定乗数とするとき，

$$\begin{aligned} J_4[P, C, \lambda] &= \langle PR, P\rangle + 2\langle C, P^*P - P\rangle \\ &\quad +\lambda(\langle P, P\rangle - M) \qquad (32) \\ &= \langle PR, P\rangle + 2(\langle PC, P\rangle - \langle C, P\rangle) \\ &\quad +\lambda(\langle P, P\rangle - M) \qquad (33) \end{aligned}$$

を最大にするような P と C と λ を求める条件なしの変分問題と等価になる．

そこで，式 (32) , (33) の変分問題を解くことにする．まず，式 (32) を C について変分し零とおけば，式 (30) が成立し，式 (29) が成立する．また，式 (32) を λ について微分し零とおけば，式 (31) が成立する．

次に，式 (33) を P について変分したものを δJ とおき，P の微小変化を δP で表せば，式 (29) , (19) より，

$$\begin{aligned} \delta J &= \langle \delta PR, P\rangle + \langle PR, \delta P\rangle \\ &\quad +2\big(\langle \delta PC, P\rangle + \langle PC, \delta P\rangle - \langle C, \delta P\rangle\big) \\ &\quad +\lambda\big(\langle \delta P, P\rangle + \langle P, \delta P\rangle\big) \\ &= 2\Re(\langle P(R + C + C^* + \lambda P) - C, \delta P\rangle) \end{aligned}$$

となる．ここで $\Re(\cdot)$ は，複素数 $\cdot$ の実部を表す．よって，$\delta J = 0$ とおけば，

$$\langle P(R + C + C^* + \lambda P) - C, \delta P\rangle = 0$$

となる．この式の δP は任意の作用素であるから，

$$P(R + C + C^* + \lambda P) = C \tag{34}$$

となる．式 (34) を，この変分問題の正規方程式という．

次に，式 (34) より，

$$PR = RP \tag{35}$$

を導く．まず，式 (34) の左から P を掛けて，式 (29) を使えば，

$$PR + PC^* + \lambda P = 0 \tag{36}$$

となる．この式の両辺の共役をとれば，λ は実数であるから，式 (29) より

$$RP + CP + \lambda P = 0 \tag{37}$$

となる．一方，式 (34) の右から P を掛ければ

$$P(R + C + C^* + \lambda P)P = CP$$

となる．この式の左辺は自己共役であるから，右辺の CP も自己共役になる．よって，式 (29) を考慮すれば，

$$PC^* = CP \tag{38}$$

となる．式 (36)〜(38) より，確かに式 (35) が成立している．

式 (35) が成立したので,，補題 2，補題 3 より，S は R の K-L 部分空間，すなわち，R の固有元で張られる部分空間になる．そこで，何番目の固有元を採用すればよいかを調べる．R

の固有元を $\{\varphi_{m_n}\}_{n=1}^{M}$ とする．ここで $\{m_n\}_{n=1}^{M}$ は，M 個の相異なる自然数を表す．このとき，$\{\varphi_{m_n}\}_{n=1}^{M}$ で張られる部分空間 $\mathcal{L}(\{\varphi_{m_n}\}_{n=1}^{M})$ への正射影作用素 P は，ノイマン・シャッテン積を用いて，

$$P = \sum_{n=1}^{M} (\varphi_{m_n} \otimes \overline{\varphi_{m_n}}) \tag{39}$$

と表すことができる．こうして問題は，式 (28) を最大にするような固有元の組 $\{\varphi_{m_n}\}_{n=1}^{M}$ を求める問題に帰着された．

以下，この問題を解くことにする．式 (28), (20), (30), (39), (22), (4) より

$$\begin{aligned} J_4[P] &= \langle PR, P\rangle \\ &= \langle R, P^*P\rangle \\ &= \langle R, P\rangle \\ &= \langle R, \sum_{n=1}^{M} (\varphi_{m_n} \otimes \overline{\varphi_{m_n}})\rangle \\ &= \sum_{n=1}^{M} \langle R\varphi_{m_n}, \varphi_{m_n}\rangle \\ &= \sum_{n=1}^{M} \lambda_{m_n} \end{aligned}$$

となり，

$$J_4[P] = \sum_{n=1}^{M} \lambda_{m_n} \tag{40}$$

となる．λ_n は大きさの順に番号付けられているので，式 (40) を最大にするためには，

$$\{\lambda_{m_n}\}_{n=1}^{M} = \{\lambda_n\}_{n=1}^{M} \tag{41}$$

とおけばよい．これは $S_0 = S_{\mathrm{KL}}$ を意味している．よって，定理 1 が成立する． ■

なお，この証明の最後の部分は，「カテゴリ Ω を最良に近似する M 次元部分空間 S_0 は，R の固有元 $\{\varphi_n\}_{n=1}^{M}$ で張られる部分空間 S_{KL} と一致する」ということを主張しているのであり，「S_0 を張るような正規直交基底が $\{\varphi_n\}_{n=1}^{M}$ である」ということを主張しているのではないことは，注意を要する．

5. What と How

学問の発展という立場から見たとき，次の問題を考えてみることは大切である．

(i) なぜ間違いに気付かなかったか．

(ii) なぜ間違えたか．

(iii) 間違いを起こさないようにするためにはどうすればよいか．

まず，(i) の問題を考えてみる．上述のように，従来の証明の結果が間違っているということは容易にわかる．それならば，なぜそのような簡単なことに気付かなかったのであろうか．しかも，この証明を与えた人々の中には，パターン認識の分野では世界的な理論家として認められている人達も含まれている．ここに，研究者の心の動きが窺える．これは私の推察にすぎないが，このような証明を考えた人々の意識の底に，$\phi_n = \varphi_n$ を導こうとする気持ちが強く働いていたために，式 (14) を導出できたことで安心してしまったのではないかと思われる．このような一瞬の隙に，間違いが忍び込んでくるのである．

次に，(ii) の問題を考えてみよう．このような間違いを犯した遠因は，部分空間を求める問題を，式 (5) や式 (10) のように，その部分空間を張るような正規直交系を求める問題として定式化したところにある．確かに部分空間は，ある正規直交系を使って表現することができる．しかし，同じ部分空間を表現できる正規直交系は，無限に存在する．したがって，式 (10) の形式の変分問題を解いても，K-L 固有元という特定の正規直交系を求めることはできないのである．

(iii) 間違いを防ぐためには，その問題にふさわしい道具を使うことである．部分空間法の例でいえば，正規直交系は部分空間を構成するための手段（How）に過ぎない．今問題になっていること（What）は，最適な部分空間 S_0 がどのような空間であるかを特徴づけることである．すなわち，$S_0 = S_{\mathrm{KL}}$ を導くことである．S_0 をどのようにして構成するかということは，本来の問題とは別の問題である．How に惑わされないで，問題そのもの (What) を直接議論できるような数学的道具を使えば，従来法のような間違いが混入する余地はなくなる．前節で示した証明法でいえば，部分空間とその空間への正射影作用素とは一対一に対応している．そこで，正射影作用素 P を使えば，正規直交系を導入することなく，直接問題を定式化できるのである．式 (27) の表現がまさにその定式化になっている．How を経由しないで，問題そのもの (What) を直接議論できる数学的道具を用いて問題を定式化することが，いかに重要であるかがわかる．

繰り返しになるが，What と How の違いを，別の角度から示すことにする．パターン f の M 次元部分空間 S_0 への正射影 Pf は，

$$Pf = \sum_{n=1}^{M} \langle f, \varphi_n\rangle \varphi_n \tag{42}$$

と表すことができる．式 (42) は，2. で述べたように，f の K-L 展開と呼ばれている．一方，$\{\varphi_n\}_{n=1}^{M}$ を S_0 の中で任意に回転してできる正規直交系を $\{\phi_n\}_{n=1}^{M}$ とすれば，同じ Pf を

$$Pf = \sum_{n=1}^{M} \langle f, \phi_n\rangle \phi_n \tag{43}$$

と表すこともできる．$\{\phi_n\}_{n=1}^{M}$ はもはや，一般には R の固有元になっていないので，式 (43) を K-L 展開ということはない．

しかし，式 (42) と式 (43) は，同じ Pf の別表現にすぎない．パターンの近似という立場からみれば，式 (42) でも式 (43) でもよいのである．つまりこの文脈においては，K-L 展開は Pf を求めるための一つの計算手段（How）にすぎないのであって，本質を担っているもの（What）は，あくまでも f の S_0 における最良近似 Pf である．

6. おわりに

問題を定式化する際には，How（問題の解き方，計算の仕方）を紛れ込ませないで，What（何をしたいか）だけを明確にして事に当たることが肝要である．そのためには，How を経由することなく，また，問題の表面的な様相に惑わされることなく，問題の本質を直接議論できる数学的手法を用いることが大切である．

文　献

[1] E. Oja, Subspace Methods of Pattern Recognition. Research Studies Press, Letchworth, 1983; 小川英光，佐藤誠（訳），パターン認識と部分空間法, 産業図書, 東京, 1986.

[2] P.A. Devijver and J. Kittler, Pattern Recognition: A Statistical Approach, Prentice Hall, Englewood Cliffs, 1982.

[3] R.O. Duda, P.E.Hart, and D.G.Stork, Pattern Classification, Second Edition, John Wiley & Sons, Inc., New York, 2001.

[4] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.

[5] U. Grenander, Pattern Synthesis: Lectures in Pattern Theory, Springer-Verlag, Berlin, 1976.

[6] 飯島泰蔵, パターン認識, 日刊工業新聞社, 東京, 1969.

[7] 飯島泰蔵, パターン認識, コロナ社, 東京, 1973.

[8] 飯島泰蔵, パターン認識理論, 森北出版, 東京, 1989.

[9] 森俊二，坂倉栂子, パターン認識の基礎 II, オーム社, 東京, 1990.

[10] 長尾真, 画像認識論, コロナ社, 東京, 1983.

[11] 長尾真, パターン情報処理, コロナ社, 東京, 1983.

[12] 長尾真 (編), パターン認識と図形処理, 岩波書店, 東京, 1983.

[13] 中田和男 (編), パターン認識とその応用, コロナ社, 東京, 1978.

[14] 上坂吉則, パターン認識と学習の理論, 総合図書, 東京, 1971.

[15] S. Watanabe, Karhunen-Loève expansion and factor analysis, in J. Sklansky ed., Pattern Recognition, Introduction and Fundation. Dowden Hutchinson & Ross. Inc., pp.635–660, 1973

[16] 大津展之, パターン認識における特徴抽出に関する数理的研究, 電子技術総合研究所研究報告, 1981.

[17] J.P. Keating, J.E. Michalek, and J.T. Riley, A noise on the optimality of the Karhunen-Loève expansion, Pattern Recognition, vol.1, no.4, pp.203–204, 1983.

[18] H. Ogawa, “Karhunen-Loève subspace”, Proc. 11th ICPR, Int. Conf. on Pattern Recognition, The Hague, The Netherlands, vol.2, pp.75–78, Aug.-Sept. 1992.

[19] H. Ogawa and E. Oja, Projection filter, Wiener filter, and Karuhunen-Loève subspaces in digital image restoration, J. Math. Anal. Appl., vol.114, no.1, pp.37–51, Feb. 1986.

[20] M.A. ナイマルク（功力，井関，笠原訳），関数解析入門　，共立全書, 共立出版, 東京, 1968.

[21] S.L. Campbell and C.D. Meyer, Jr., Generalized Inverses of Linear Transformations, Dover Publications, Inc., New York, 1979.

[22] R. Schatten, Norm Ideals of Completely Continuous Operators. 2nd. Printing. Springer–Verlag, Berlin, 1970.

[23] 加藤敏夫, 位相解析–理論と応用への入門–, 共立出版, 東京, 1969.

[24] 戸川隼人, マトリクスの数値計算, オーム社, 東京, 1978.

[25] 伊理正夫, 一般線形代数, 岩波書店, 東京, 2003.

[26] W. ハイゼンベルク (山崎和夫訳), 科学における伝統, みすず書房, 東京, 1989.

部分空間法研究会 Subspace2008 委員会

実行委員会

実行委員長	坂野 鋭	(NTT CS 研)
実行委員	天野 敏之	(奈良先端大)
	大町 真一郎	(東北大)
	佐藤 敦	(NEC)
	玉木 徹	(広島大)
	福井 和広	(筑波大)
	堀田 政二	(東京農工大)
	牧 淳人	(東芝 Cambridge 研究所)
顧問	前田 賢一	(東芝)

部分空間法研究会 (Subspace2008) 予稿集

編集　部分空間法研究会　実行委員会
発行日　2008 年 7 月 22 日