

グラスマン距離に基づいた部分空間学習による表情変化を含んだ顔画像列の認識

重中 亨介[†] Bisser Raytchev[†] 玉木 徹[†] 金田 和文[†]

[†] 広島大学大学院工学研究科情報工学専攻

〒739-8527 広島県東広島市鏡山 1-4-1

E-mail: †shigenaka@eml.hiroshima-u.ac.jp, ††{bisser,tamaki,kin}@hiroshima-u.ac.jp

あらまし 本論文では表情変化を含んだ顔画像列の認識問題に対して、グラスマン多様体に基づいたグラスマン距離およびグラスマンカーネルを用いることで識別能力の高い認識手法を考える。顔画像列を部分空間で表し、画像列同士の類似度に正準角を用いる相互部分空間法が既に提案されている。しかし、相互部分空間法は部分空間同士の第一正準角のみを用いるため最適ではない。そこで本論文においてグラスマン距離を相互部分空間法に適用した手法を提案する。その他にもいくつかの識別手法に対してグラスマンカーネルを適用する。これらの提案手法を評価するために、表情変化のある顔画像列データセットを用いて実験を行い、グラスマン距離に基づいた認識手法が顔画像列の認識問題に対して有効であることを確認した。

キーワード グラスマン多様体、グラスマンカーネル、正準角、正準相関、部分空間

1. はじめに

コンピュータビジョンでは多くの場合、物体の特徴を画素の集合で表現している。このとき、それぞれの画像は視点の変化や照明変化などの外部からの要因による変化や剛体・非剛体変形などの物体そのものによる変化を含んでいる。特に物体認識では、それぞれの画像間の関係よりも全体の画像集合間との関係を考慮することの方が有利であると知られている。そこで、二つの画像列を比較するために最適距離もしくは類似度を決定する必要がある。このために、Parametric Model に基づく手法や Non-Parametric Sample に基づく手法が提案されている。Parametric Model に基づく手法 [1, 2] では、画像集合はパラメータ分布関数によって表され、集合間の距離は Kullback-Leibler 情報量によって計算される。一方、Non-Parametric Sample に基づく手法 [3, 4] では、各画像集合内のそれぞれのサンプル間に対する最近傍 (Nearest-Neighbor; NN) 距離が用いられる。

近年、画像の集合を低次元線形部分空間で近似する手法が広く用いられている。また、線形部分空間法において画像列間の距離は対応する部分空間同士の距離として表現される。この部分空間同士の距離は部分空間同士の正準角で表現される。相互部分空間法 (Mutual Subspace Method; MSM) [6] では最小の正準角の余弦を 2 つの画像列同士の類似度として用いている。また、MSM の関連手法として核非線形相互部分空間法 (Kernel MSM; KMSM) [7] や多重制約相互部分空間法 (Constrained MSM) [8, 9]、ブースト多様体正準角 (Boosted Manifold Principal Angles) [10]、正準相関判別分析 (Discriminant Analysis of Canonical Correlations) [11] などが挙げら

れる。

最近の研究 [12] では、部分空間に基づく学習手法がまとめられている。これはユークリッド空間の線形部分空間の集合である、グラスマン多様体 (Grassmann manifold) [13] として問題を定式化している。グラスマン多様体の幾何学的構造を考慮し、部分空間同士の正準角を用いたさまざまな距離が提案されている [12, 14]。[12] では Projection Metric と Binet-Cauchy Metric がグラスマン多様体において有用なメトリックであると示されている。対応するカーネル関数である Projection Kernel と Binet-Cauchy Kernel を用いた Kernel Linear Discriminant Analysis は Grassmann Discriminant Analysis (GDA) と呼ばれている。また、Kernel Grassmann Discriminant Analysis (KGDA) [15] も提案されており、カーネルを用いて GDA を拡張したものである。

部分空間の類似度を近似する際、最小の正準角の余弦のみを用いる手法は幾何学的に最適ではなく、認識に対しても最適ではない。そこで本論文では、MSM の拡張手法を提案する。すなわち、グラスマン距離を画像列間の類似度の定義に用いる Grassmann Distance Mutual Subspace Method (GD-MSM) を提案する。さらにグラスマン距離に基づいて、より判別能力の高い学習関数を得るために Support Vector Machine (SVM) [16] にグラスマンカーネル [12] を適用した Grassmann Kernel Support Vector Machine (GK-SVM) も提案する。4 節で述べる実験では、2 つの異なるデータセットにおいていくつかの関連手法を用いて GD-MSM と GK-SVM の性能を比較する。データセットの一つ目は自然な会話による表情変化を含んだ 100 人分の顔画像列データセット [17] である。二つ目は 6 つの基本的な感情を含んだ 101 人分

の顔画像列データセット [18] である。また、二種類の識別辞書を用意して、その影響について調査した結果についても述べる。まず一つ目はそれぞれのクラスに属する全ての画像列から1つの識別辞書を作成する。二つ目は画像列ごとにそれぞれの識別辞書を作成する。

2. 準備

本節ではグラスマン距離とグラスマンカーネルに関連する用語について述べる。

いま、ベクトルで表した2つの画像列 $\mathbf{S}_i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_n^i\}$ および $\mathbf{S}_j = \{\mathbf{x}_1^j, \mathbf{x}_2^j, \dots, \mathbf{x}_m^j\}$ を考える。ただし、 \mathbf{x}_n^i は i 番目の画像列に含まれる n 番目の画像である。各画像列に含まれる画像の集合全体はユークリッド空間 \mathbb{R}^D 中の対応する部分空間で表せる。すなわち、 $\text{span}(Y_i) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ および $\text{span}(Y_j) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_q\}$ である。ここで、 $\text{span}(Y_i)$ は $D \times p$ 行列 $\mathbf{Y}_i = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ の列ベクトルによって張られる部分空間である。また、 $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ は正規直交基底である。 \mathbb{R}^D における全ての m 次元線形部分空間の集合はグラスマン多様体 $\mathcal{G}(m, D)$ と呼ばれ、部分空間 $\text{span}(Y_i)$ と $\text{span}(Y_j)$ は多様体 $\mathcal{G}(m, D)$ 上で2つの点として表せる。 $\mathcal{G}(m, D)$ 上の様々な距離は [14] において定義されており、その全ての距離は部分空間同士 [5] の正準角を用いて表現できる。

部分空間 $\text{span}(Y_1)$ と $\text{span}(Y_2)$ の正準角 ($0 \leq \theta_1 \leq \dots \leq \theta_m \leq \pi/2$) は以下の式で再帰的に定義できる。

$$\cos \theta_k = \max_{\mathbf{u}_k \in \text{span}(Y_1)} \max_{\mathbf{v}_k \in \text{span}(Y_2)} \mathbf{u}_k^T \mathbf{v}_k \quad (1)$$

この式は以下の制約条件を持つ。

$$\mathbf{u}_k^T \mathbf{u}_k = 1, \mathbf{v}_k^T \mathbf{v}_k = 1, \quad (2)$$

$$\mathbf{u}_k^T \mathbf{u}_i = 0, \mathbf{v}_k^T \mathbf{v}_i = 0, (i = 1, \dots, k-1) \quad (3)$$

正準角は $\mathbf{Y}_1^T \mathbf{Y}_2$ の特異値分解によって計算できる。

$$\mathbf{Y}_1^T \mathbf{Y}_2 = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_m) \quad (4)$$

ここで、直交基底行列 \mathbf{Y}_1 と \mathbf{Y}_2 は $\text{span}(Y_1)$ と $\text{span}(Y_2)$ の行列表現である。また、 $\lambda_i = \cos \theta_i$ は正準角 θ_i の余弦であり、正準相関として知られている。

本節の残りでは次のグラスマン距離 [12] について述べる。

1. Projection Metric (正準角の正弦の 2-Norm)

$$d_P(Y_1, Y_2) = \left(\sum_{i=1}^m \sin^2 \theta_i \right)^{\frac{1}{2}} = \left(m - \sum_{i=1}^m \cos^2 \theta_i \right)^{\frac{1}{2}} \quad (5)$$

2. Binet-Cauchy Metric (正準角の余弦の積)

$$d_{BC}(Y_1, Y_2) = \left(1 - \prod_i \cos^2 \theta_i \right)^{\frac{1}{2}} \quad (6)$$

3. 最大相関 (最小の正準角の正弦のみを用いており、MSM [6] と等価)

$$d_{Max}(Y_1, Y_2) = (1 - \cos^2 \theta_i)^{\frac{1}{2}} = \sin \theta_i \quad (7)$$

4. 最小相関 (最大の正準角の正弦のみを用いる)

$$d_{Min}(Y_1, Y_2) = (1 - \cos^2 \theta_m)^{\frac{1}{2}} = \sin \theta_m \quad (8)$$

5. Procrustes (Chordal) 距離 (2つの部分空間 $\text{span}(Y_1)$ および $\text{span}(Y_2)$ の異なる表現間の最小距離であり、Frobenius norm を用いる)

$$d_{CF}(Y_1, Y_2) = \min_{R_1, R_2 \in O(m)} \|\mathbf{Y}_1 R_1 - \mathbf{Y}_2 R_2\|_F \quad (9)$$

$$= 2 \left(\sum_{i=1}^m \sin^2(\theta_i/2) \right)^{\frac{1}{2}} \quad (10)$$

6. 行列 2-norm を用いた Procrustes (Chordal) 距離

$$d_{C2}(Y_1, Y_2) = \min_{R_1, R_2 \in O(m)} \|\mathbf{Y}_1 R_1 - \mathbf{Y}_2 R_2\|_2 \quad (11)$$

$$= 2 \sin(\theta_m/2) \quad (12)$$

7. Geodesic 距離 (グラスマン多様体上の2点を測地線の長さ)

$$d_G(Y_1, Y_2) = \sum_{i=1}^m \theta_i^2 \quad (13)$$

8. 平均距離

$$d_{Mean}(Y_1, Y_2) = \frac{1}{m} \sum_{i=1}^m \sin^2 \theta_i \quad (14)$$

4節では顔画像列を用いて顔認識実験と表情認識実験を行い、上記8つのグラスマン距離を比較する。

Projection Metric (5) および Binet-Cauchy Metric (6) は次の正定値グラスマンカーネルの定義に用いられている。

1. Projection Kernel

$$k_P(Y_1, Y_2) = \text{trace}[(\mathbf{Y}_1 \mathbf{Y}_1^T)(\mathbf{Y}_2 \mathbf{Y}_2^T)] \quad (15)$$

$$= \|\mathbf{Y}_1^T \mathbf{Y}_2\|_F^2 \quad (16)$$

2. Binet-Cauchy Kernel

$$k_{BC}(Y_1, Y_2) = \det(\mathbf{Y}_1^T \mathbf{Y}_2)^2 \quad (17)$$

$$= \det(\mathbf{Y}_1^T \mathbf{Y}_2 \mathbf{Y}_2^T \mathbf{Y}_1) \quad (18)$$

$$= \prod_i \cos^2 \theta_i \quad (19)$$

これらのカーネルは様々なカーネルアルゴリズム [19] と結合して用いることができる。文献 [12] では Kernel LDA に用いられている。次節では、SVM [20] にこれらのカーネルを適用することを提案する。

3. 手法

本節では、顔画像列に対して行う評価実験に用いる手法について述べる。

3.1 Grassmann Distance Mutual Subspace Method (GD-MSM)

本論文で提案する Grassmann Distance Mutual Subspace Method (GD-MSM) は前節で述べたグラスマン距離を用いて MSM を拡張した手法である。部分空間同士での最小の正準角のみを用いる代わりに全ての正準角を用いる。すなわち、式 (5) から (14) に示した 8 つの距離を用いる。

3.2 Grassmann Discriminant Analysis (GDA)

GDA [12] は判別学習におけるグラスマンカーネル $k_P(16)$ および $k_{BC}(19)$ を用いる。すなわち、グラスマンカーネルを用いた Kernel Discriminant Analysis である。GDA は LDA において判別方向 ω を求めるために用いるレイリー商 $L(\omega) = \omega^T S_b \omega / \omega^T S_w \omega$ に kernel trick を適用した手法である。ここで、 S_b および S_w はクラス間分散行列およびクラス内分散行列である。いま、 ϕ が特徴マップで $\Phi = [\phi_1 \cdots \phi_N]$ が学習サンプルの特徴行列のとき、 ω は特徴ベクトルの線形結合 $\omega = \Phi \alpha$ で表せる。レイリー商は α を用いて次式で表現できる。

$$L(\alpha) = \frac{\alpha^T \Phi^T S_B \Phi \alpha}{\alpha^T \Phi^T S_W \Phi \alpha} \quad (20)$$

$$= \frac{\alpha^T \mathbf{K} (\mathbf{V} - \mathbf{1} \mathbf{1}^T / N) \mathbf{K} \alpha}{\alpha^T (\mathbf{K} (\mathbf{I} - \mathbf{V}) \mathbf{K} + \sigma^2 \mathbf{I}) \alpha} \quad (21)$$

ここで、 \mathbf{K} は学習データにグラスマンカーネルを適用して得られるカーネル行列であり、 $\mathbf{1}$ は全ての要素が 1 の N 次元ベクトルである。また、 \mathbf{V} はブロック対角行列であり、その c 番目のブロックは $N_c \times N_c$ の全ての要素が 1 の行列を N_c で割ったものである。さらに $\sigma^2 \mathbf{I}$ は正則化項である。GDA は (21) を最大化する α を求め、 $F_{train} = \alpha^T \mathbf{K}$ と $F_{test} = \alpha^T \mathbf{K}_{test}$ 間のユークリッド距離を用いて最近傍法による識別を行う。ここで、 \mathbf{K}_{test} は学習サンプルとテストサンプルから得られるカーネル行列である。

3.3 Grassmann Kernel Support Vector Machine (GK-SVM)

Grassmann Kernel Support Vector Machine (GK-SVM) は SVM にグラスマンカーネルを用いた手法である。ここでまず、2 クラスの識別問題について考える。学習セット $S = \{(\mathbf{Y}_1, y_1), \dots, (\mathbf{Y}_N, y_N)\}$ が与えられたとする。ここで、 \mathbf{Y}_i は $span(\mathbf{Y}_i)$ の行列表現であり、 i 番目の学習サンプルに対応している。 $y_i = \{-1, 1\}$ はクラスラベルである。このとき、SVM は次の最適化問題を解く。

$$\min_{\omega, \xi, b} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N \xi_i \quad (22)$$

$$\text{subject to } y_i (\omega^T \phi_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (23)$$

また、双対表現は次のように与えられる。

$$\min_{\alpha} \frac{1}{2} \alpha^T \mathbf{Q} \alpha - \mathbf{1}^T \alpha \quad (24)$$

$$\text{subject to } \mathbf{y}^T \alpha = 0, 0 \leq \alpha_i \leq C \quad (25)$$

式 (22) - (25) において、 ξ はマージンスラックベクトルであり、 ϕ は特徴空間の変換関数である。また、 C はスラック変数とマージンのトレードオフを操作する変数であり、 $\mathbf{Q}_{ij} = y_i y_j \mathbf{K}_{ij}$ となる。決定関数は次式で書ける。

$$\text{sgn} \left(\sum_{i=1}^N y_i \alpha_i \mathbf{K}(\mathbf{Y}_i, \mathbf{Y}_T) + b \right) \quad (26)$$

ここで、 \mathbf{Y}_T は $span(\mathbf{Y}_T)$ の行列表現であり、テスト画像に対応する。本論文では多クラス問題に対応するため、one-against-one アプローチを用いる。 c はクラス数であり、 $c(c-1)/2$ 個の 2 クラス識別器が構成される。そして識別は多数決によって決定する。

3.4 相互部分空間法 (MSM)

MSM [6] は最大相関距離 (7) に対応している。

3.5 核非線形相互部分空間法 (KMSM)

KMSM [7] は入力データを非線形に扱うために、MSM に Kernel Principal Component Analysis (Kernel PCA) を適用した手法である。

3.6 CLAFIC

CLAFIC [21] は部分空間法の中でも古くから用いられている手法であり、それぞれのクラス c の部分空間 $span(\mathbf{Y}_c)$ は $\mathbf{Y}_c = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ と表せる。類似度としてテスト画像サンプル \mathbf{x}_T と学習画像の部分空間との角度が用いられる。

$$\cos^2 \theta = \frac{1}{\|\mathbf{x}_T\|_2^2} \sum_{i=1}^p (\mathbf{u}_i^T \mathbf{x}_T)^2 \quad (27)$$

さらにベースラインとして、PCA に基づいた固有顔手法 [22] および LDA に基づいたフィッシャー顔手法 [23] を用いる。これらの手法は CLAFIC と同様に、各顔画像を別々に識別する。

4. 実験

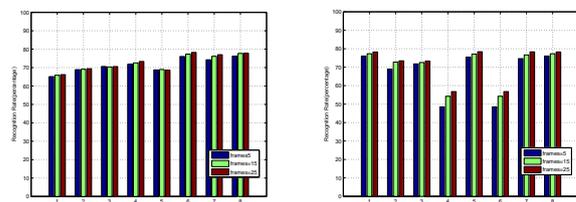
提案手法の性能を評価するために顔認識実験と表情認識実験を行った。まず、実験に用いた 2 つのデータベースについて述べる。顔認識実験には MOBIO [17] を、表情認識実験には BU-4DFE [18] を用いた。

MOBIO は 160 人の被験者が質問に答える様子を撮影した顔画像列データセットである。データは撮影環境の異なる 6 つのセッションに分かれており、各セッションごとに 21 個の画像列が入っている。また、各画像列には自然な表情変化が含まれる。本実験ではこの画像列の各

画像に対して2通りの顔検出を行った上で用いる。顔検出方法はOpenCV [24]を用いた。1つ目はViola-Jonesの顔検出器を用いた顔検出法であり、2つ目は目検出器を用いた顔検出法である。前者を用いて作成したデータをデータ1、後者を用いて作成したデータをデータ2と呼ぶこととする。ただし、検出に失敗した画像は使用しない。

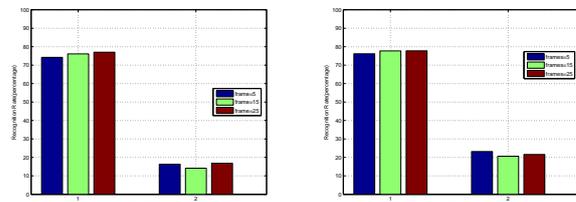
一方、BU-4DFEは101人の被験者を撮影したデータセットであり、6通りの異なる表情を含んだ画像列で構成されている。すなわち、怒り、嫌悪、恐れ、幸せ、悲しみ、驚きである。なお、データは3Dと2Dが用意されている。BU-4DFEを用いた関連研究は[18]にて行われており、2Dデータを用いた比較手法の認識率は63.72%であるのに対して3Dデータを用いた認識率は90.44%である。しかし、これは3Dデータを用いた手法の認識結果であり、また60人分のデータだけ用いているため単純には比較できない。本実験では2Dの画像列の各画像に対して先述した目検出器を用いて顔検出を行い、検出された全顔画像を用いた。

次に実験及びその結果について述べる。まず、MOBIOのデータを用いて顔認識を行った。画像列長の変化による影響を調査するために、画像列の長さを5, 15, 25枚と変化させたときの各手法の比較を行った。ただし、顔検出手法の検出率の問題から、全てのショットに対して25枚以上検出できた24クラスを用いた。また、各クラスごとに6つのセッションに分かれていることに基づきleave-one-out cross validationを用いた。なお、実験はそれぞれ2つの顔検出データに対して行い、また識別辞書の作成方法について、画像列ごとにそれぞれの識別辞書を作成する方法と全ての画像列から1つの識別辞書を作成する方法の2パターンを行った。データ1に対して画像列ごとにそれぞれの識別辞書を作成したときの認識率を図1に、データ2に対して画像列ごとにそれぞれの識別辞書を作成したときの認識率を図2に示す。一方、データ1に対して全ての画像列から1つの識別辞書を作成したときの認識率を図3に、データ2に対して全ての画像列から1つの識別辞書を作成したときの認識率を図4に示す。ただし、図の縦軸は認識率、横軸は手法または類似度である。また、(a)において横軸の1はNN (PCA)、2はNN (LDA)、3はCLAFIC法、4はMSM、5はKMSM、6はGD-MSM (平均距離)、7はGK-SVM (Projection Metric)、8はGDA (Projection Metric)を表しており、(b)において横軸の1はProjection Metric、2はBinet-Cauchy Metric、3は最大相関、4は最小相関、5はChordal距離 (F-Norm)、6はChordal距離 (2-Norm)、7はGeodesic距離、8は平均距離である。さらに、(c)と(d)において横軸の1はProjection Metric、2はBinet-Cauchy Metricを表す。これらの実験から、画像列の長さは認識率に対して大きな影響を与えない結果となった。



(a) 手法の比較

(b) GD-MSM の類似度の比較



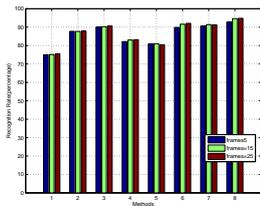
(c) GK-SVMの類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric) (d) GDAの類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)

図1 データ1 (24クラス、顔検出器使用) に対して画像列ごとにそれぞれの識別辞書を作成したときの顔認識率

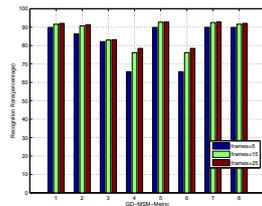
そこで次に、MOBIOのデータの100クラスを用いて各手法の比較実験を行った。ここで、実験で用いたコンピュータのスペックの問題から画像列の長さを25枚以下とした。データ1に対して画像列ごとにそれぞれの識別辞書を作成したときの認識率を図5に、データ2に対して画像列ごとにそれぞれの識別辞書を作成したときの認識率を図6に示す。一方、データ1に対して全ての画像列から1つの識別辞書を作成したときの認識率を図7に、データ2に対して全ての画像列から1つの識別辞書を作成したときの認識率を図8に示す。

最後に、BU-4DFEの100人分のデータを用いて表情認識実験を行った。なお、各表情からクラスを構成し、各クラスごとに5つのセッションを作成した。1つのセッションには20人のデータが含まれる。そして前実験と同様にleave-one-out cross validationを用いた。画像列の長さは検出された全画像枚数とした。画像列ごとにそれぞれの識別辞書を作成したときの認識率を図9に、全ての画像列から1つの識別辞書を作成したときの認識率を図10に示す。

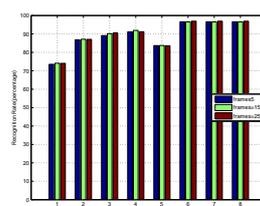
これらの実験より、一般的にGD-MSMとGK-SVM、GDAの3つの手法はその他の手法に比べて認識率が高い結果となった。また、それぞれのクラスに属する全ての画像列から1つの識別辞書を作成した場合にはGD-MSMの認識率が高い結果となった。一方で、画像列ごとにそれぞれの識別辞書を作成した場合には安定した結果が得られなかった。これは各画像列間の特徴の違いが大きいためだと思われる。したがって、全ての画像列から1つ識別辞書を作成した場合の方が高い認識率が得られやすい。さらに、GD-MSMの8つの距離の中では平均距離が最も安定して高い認識率を得た。



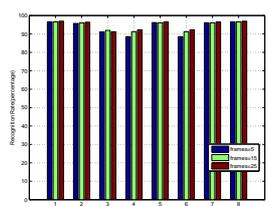
(a) 手法の比較



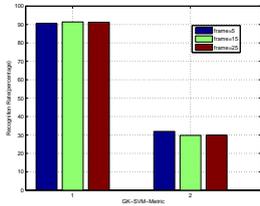
(b) GD-MSM の類似度の比較



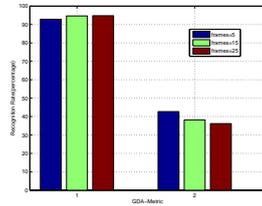
(a) 手法の比較



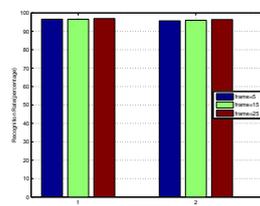
(b) GD-MSM の類似度の比較



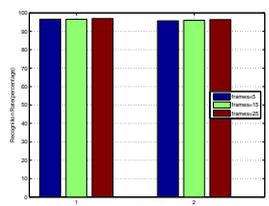
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)



(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)



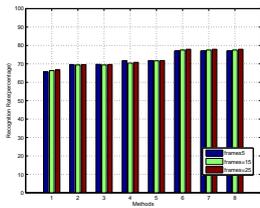
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)



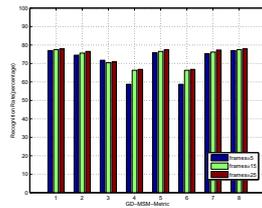
(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)

図 2 データ 2 (24 クラス、目検出器使用) に対して画像ごとにそれぞれの識別辞書を作成したときの顔認識率

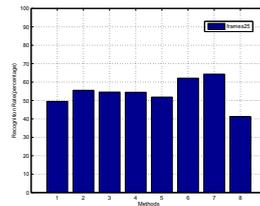
図 4 データ 2 (24 クラス、目検出器使用) に対して全ての画像列から 1つの識別辞書を作成したときの顔認識率



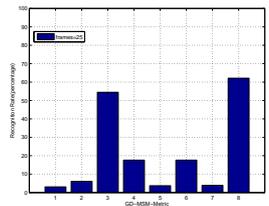
(a) 手法の比較



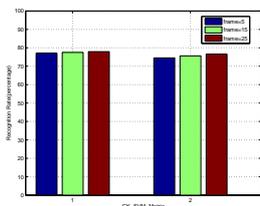
(b) GD-MSM の類似度の比較



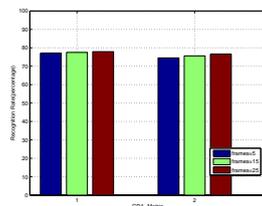
(a) 手法の比較



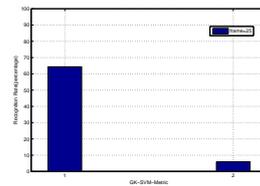
(b) GD-MSM の類似度の比較



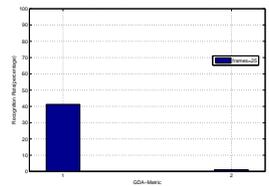
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)



(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)



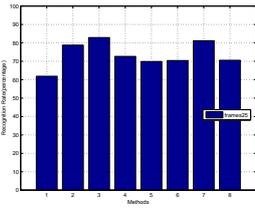
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)



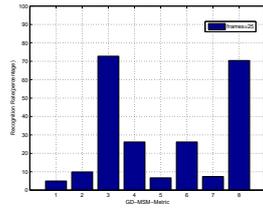
(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Cauchy Metric)

図 3 データ 1 (24 クラス、顔検出器使用) に対して全ての画像列から 1つの識別辞書を作成したときの顔認識率

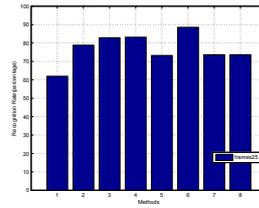
図 5 データ 1 (100 クラス、顔検出器使用) に対して画像ごとにそれぞれの識別辞書を作成したときの顔認識率



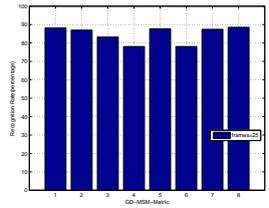
(a) 手法の比較



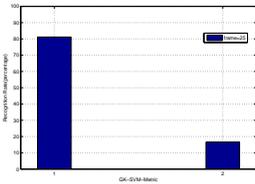
(b) GD-MSM の類似度の比較



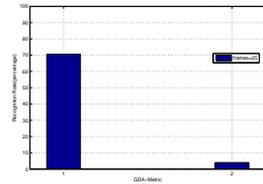
(a) 手法の比較



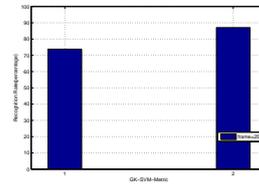
(b) GD-MSM の類似度の比較



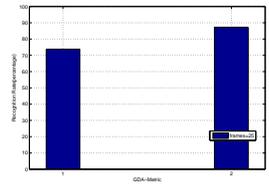
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)



(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)



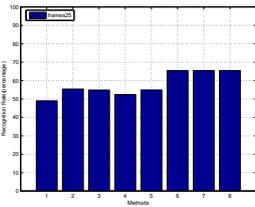
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)



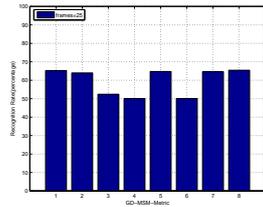
(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)

図 6 データ 2 (100 クラス、目検出器使用) に対して画像列ごとにそれぞれの識別辞書を作成したときの顔認識率

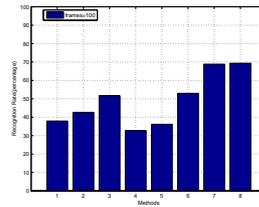
図 8 データ 2 (100 クラス、目検出器使用) に対して全ての画像列から 1 つの識別辞書を作成したときの顔認識率



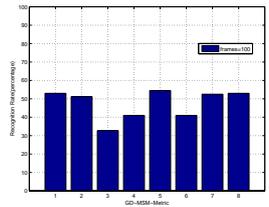
(a) 手法の比較



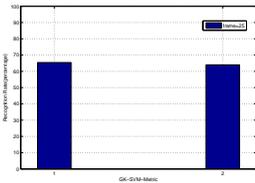
(b) GD-MSM の類似度の比較



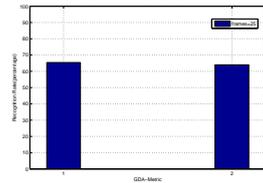
(a) 手法の比較



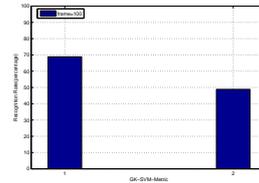
(b) GD-MSM の類似度の比較



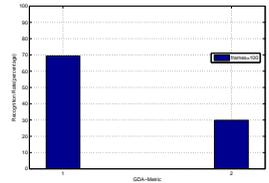
(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)



(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)



(c) GK-SVM の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)



(d) GDA の類似度の比較 (1: Projection Metric, 2: Binet-Projection Metric, 2: Binet-Cauchy Metric)

図 7 データ 1 (100 クラス、顔検出器使用) に対して全ての画像列から 1 つの識別辞書を作成したときの顔認識率

図 9 画像列ごとにそれぞれの識別辞書を作成したときの表情認識実験における認識率

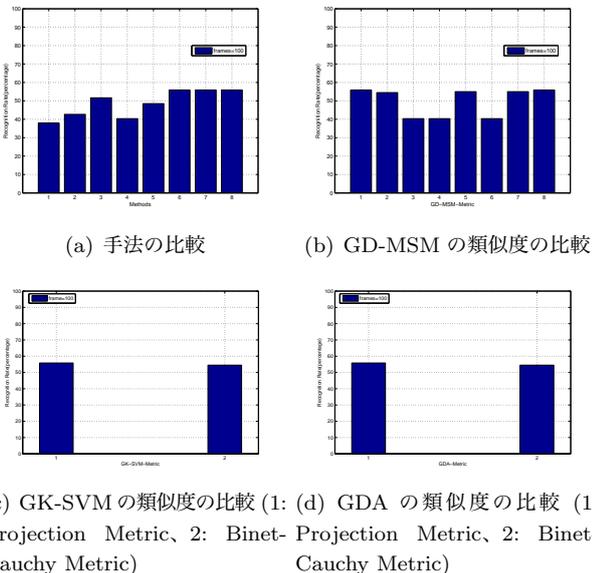


図 10 全ての画像列から 1 つの識別辞書を作成したときの表情認識実験における認識率

5. おわりに

本論文では様々な識別手法にグラスマン距離およびそれに基づいたグラスマンカーネルを適用した手法を提案し、性能の比較実験を行った。顔認識実験と表情認識実験において、データの変化による認識率の差は生じるものの、グラスマン距離に基づいた識別手法が顔画像列の認識問題に対してより有効であることを示した。

文 献

- [1] G. Shakhnarovich, J.W. Fisher, and T. Darrel, "Face Recognition from Long-Term Observations," in *Proc. European Conf. Computer Vision (ECCV)*, pp. 851-868, 2002.
- [2] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, "Face Recognition with Image Sets Using Manifold Density Divergence," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 581-588, 2005.
- [3] S. Satoh, "Comparative Evaluation of Face Sequence Matching for Content-Based Video Access," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp. 163-168, 2000.
- [4] B. Raytchev and H. Murase, "Unsupervised Recognition of Multi-View Face Sequences Based on Clustering with Attraction and Repulsion," *Computer Vision and Image Understanding*, Vol. 91, No. 1-2, pp. 22-52, 2003.
- [5] G.H. Golub and C.F. van Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd edition.
- [6] O. Yamaguchi, K. Fukui, and K. Maeda, "Face Recognition Using Temporal Image Sequence," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp. 318-323, 1998.
- [7] H. Sakano and N. Mukawa, "Kernel mutual subspace method for robust facial image recognition," in *Proc. Int. Conf. on Knowledge-Based Intell. Eng. Sys. And App. Tech.*, pp. 245-248, 2000.
- [8] K. Fukui and O. Yamaguchi, "Face Recognition Using

- Multi-Viewpoint Pattern for Robot Vision," in *Proc. Int. Symp. Robotics Research*, pp. 192-201, 2003.
- [9] M. Nishiyama, O. Yamaguchi, and K. Fukui, "Face Recognition with the Multiple Constrained Mutual Subspace Method," in *Proc. 5th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pp. 71-80, 2005.
- [10] T.-K. Kim, O. Arandjelovic, and R. Cipolla, "Learning over Sets Using Boosted Manifold Principal Angles (BoMPA)," in *Proc. British Machine Vision Conf.*, pp. 779-788, 2005.
- [11] T.-K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of images set classes using canonical correlations," in *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, 29, 1005-1018, 2007.
- [12] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: A unifying view on subspace-based learning," in *Proc. 25th Int. Conf. on Machine Learning*, pp. 376-383, 2008.
- [13] Y. C. Wong, "Differential geometry of Grassmann manifolds," in *Proc. of the Nat. Acad. of Sci.*, Vol. 57, pp. 589-594, 1967.
- [14] A. Edelman, T.A. Aris and S.T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, 20 (2), pp. 303-353, 1998.
- [15] T. Wang and P. Shi, "Kernel Grassmannian distances and discriminant analysis for face recognition from image sets," *Pattern Recognition Letters*, Vol. 30, pp. 161-165, 2009.
- [16] C.M.Bishop, *Pattern recognition and Machine Learning*, Springer-Verlag, 2006.
- [17] C. McCool and S. Marcel, "MOBIO Database for the ICPR 2010 Face and Speech Competition," *an IDIAP Research Institute Communication*, Idiap-Com-02-2009, 2009.
- [18] L. Yin, X. Chen, Y. Sun, T. Worm and M. Reale, "A High-Resolution 3D Dynamic Facial Expression Database," in *Proc. Int. Conf. Automatic Face and Recognition*, 2008.
- [19] J.S. Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [20] C. Cortes and V. Vapnik, "Support Vector Networks," *Machine Learning*, 20, pp. 273-297, 1995.
- [21] S. Watanabe and N. Pakvasa, "Subspace Method of Pattern Recognition," *1st. Int. Joint Conf. on Pattern Recognition*, pp. 25-32, 1973.
- [22] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [23] P.N.Belhumeur, J.P.Hespanha, and D.J.Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [24] OpenCV.jp, <http://opencv.jp/> (2011/03/28 accessed).