Vol.2013-CVIM-188 No.22
2013/9/3

部分空間法の応用と魅力

堀田　政二[1,a)]

概要：部分空間法は光学的文字認識装置や顔認識装置などのパターン認識機械で高度に実用的な方法であり，既に多くの製品が市場で実用化されている．更に部分空間法には理論的な美しい背景もあるが，その本質については余り知られていない．そこで部分空間法の魅力と応用について，理論面，アプリケーションの観点から述べる．具体的には，主成分分析と自己相関行列の固有値分解の本質の説明，および部分空間法とベイズ決定則との関係をまず述べ，部分空間法の利点と欠点を述べる．そのうえで，一般画像認識と猫顔検出への部分空間法の応用例を述べる．

1. 部分空間識別器の魅力

本稿における部分空間識別器 (subspace classifier) とは，渡辺による class-featuring information compression (CLAFIC) [1] や，飯島による複合類似度法 (multiple similarity) [2] を含む (重み付き) 部分空間法のことである．はじめに，部分空間識別器のアルゴリズムについて述べることにする (図 1 に概念図を示す)．そのために，本講演で用いる記号について約束しよう．任意の標本 (パターン) $\mathbf{x}$ は d 次元ユークリッド空間 $\mathbb{R}^d$ に属すると仮定し，$\mathbf{x}$ を d 個の実数の組である d 次元実ベクトル

$$\mathbf{x} = (x_1, \cdots, x_d)^\top \in \mathbb{R}^d \tag{1}$$

によって表現する．ここで，$\top$ は転置を表す．この $\mathbf{x}$ は適当な前処理，特徴抽出を経て得られたものと考えて良い．また，L_2 ノルムを $||\mathbf{x}|| = \sqrt{\mathbf{x}^\top \mathbf{x}}$ で表し，c 個のクラスを $\{\omega_j\}_{j=1}^c$ として，その添え字 j をクラスを表す番号とする．

部分空間識別器は以下のアルゴリズムによって，未知標本 $\mathbf{x}$ を，それが属すであろう一つのクラスへ分類する．

Step1: クラスごとに次の標本自己相関行列を計算する．

$$\mathbf{R}_j = \frac{1}{n_j} \sum_{\mathbf{x}_i \in \omega_j} \mathbf{x}_i \mathbf{x}_i^\top \tag{2}$$

ここで n_j はクラス ω_j に属する訓練標本数である．

Step2: $\mathbf{R}_j$ の固有値のうち，値の大きい r 個の固有値 $\lambda_{j1}, ..., \lambda_{jr}$ (r はクラス共通とする) に対応する固有ベクトル $\mathbf{u}_{j1}, ..., \mathbf{u}_{jr}$ を並べた $d \times r$ の行列を $\mathbf{U}_j = (\mathbf{u}_{j1}|\mathbf{u}_{j2}|\cdots|\mathbf{u}_{jr})$ とする (図 1 右上)．

1 東京農工大学
184–8588 東京都小金井市中町 2-24-16
a) s-hotta@cc.tuat.ac.jp

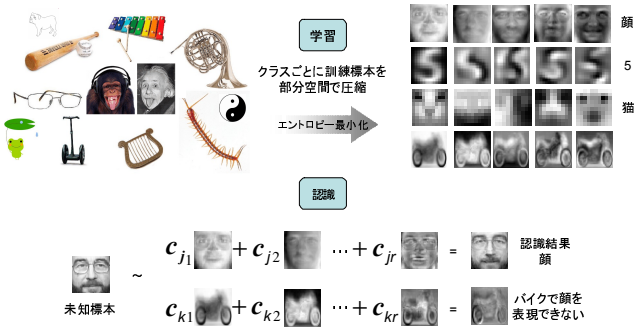

図 1 部分空間識別器の識別までの流れ．クラスごとに部分空間を張る正規直交ベクトル (自己相関行列の固有ベクトル) を求め，それらと未知パターンとの内積値の (重み付き) 総和をそのクラスとの類似度とする．

Step3: r 個の重み $w_1, ..., w_r$ (非負でクラス共通とする) を何らかの方法で決め，これと $\mathbf{U}_j$ を用いて以下の式でクラス ω_j と $\mathbf{x}$ との類似度 $S_j(\mathbf{x})$ を計算する (図 1 下段)．

$$S_j(\mathbf{x}) = \sum_{l=1}^{r} w_l (\mathbf{u}_{jl}^\top \mathbf{x})^2 = ||\mathbf{W}^{\frac{1}{2}} \mathbf{U}_j^\top \mathbf{x}||^2 \tag{3}$$

ここで $\mathbf{W}$ =diag$(w_1, ..., w_r)$ である．

Step4: $S_j(\mathbf{x})$ が最大となるクラスへ $\mathbf{x}$ を分類する．

上記のアルゴリズムにおいて $\{w_i\}_{i=1}^r$ の値を全て一定としたものが CLAFIC，クラスごとに $w_i = \lambda_{ji}/\lambda_{j1}$ としたものが複合類似度とみなせる．

上記のように部分空間識別器は，学習も認識も単純であり，特に認識に関しては高速であるという特長がある．また，標本のノルムが 1 であれば，類似度が部分空間と任意の標本 $\mathbf{x}$ との成す角度 (または垂線の足の長さ) となることから，幾何学的な解釈が容易である (図 2)．さらに，逆行列の計算が不要であるため，標本数の不足に伴う不良設定問題を回避するための正則化が必要なく，固有ベクトル

ⓒ 2013 Information Processing Society of Japan

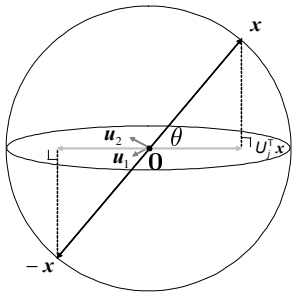

図 2 三次元空間における部分空間 (球の原点を通る断面) と任意の標本 $\mathbf{x}$ ($||\mathbf{x}|| = 1$) との類似度は $\cos^2\theta = ||\mathbf{U}_j^\top \mathbf{x}||^2$ で与えられる．$-\mathbf{x}$ も同じ類似度となることに注意．

(図 1 右上) の線型結合によって多様な (人工) 標本を表現できるため，標本数が少ない場合であっても高い認識率を達成できるという特長がある．

部分空間識別器は登場から既に 40 年以上経過しているが，いまだなお多くの拡張や改良がなされている．例えば，直交部分空間法や学習部分空間法 [3]，非線型性を吸収するためのカーネル非線型部分空間法 [4], [5] や局所部分空間識別器 [6]，部分空間同士の角度を類似度とする相互部分空間法 [7] やその拡張 [8]，および核非線型相互部分空間法 [9] など，正確にカウントすることが不可能と思われるほど存在する．ただし，それらのほとんどが認識精度の改善を目的としており，部分空間識別器の性質を明らかにしたものは少ない．

しかし，パターン認識 (クラス分類) ということを考えた場合，その理想形はベイズ決定則であるにも関わらず，部分空間識別器と (パラメトリックな学習を伴う) ベイズ決定則との関連性を指摘したものはほとんどない．これは元々部分空間識別器がエントロピー最小化 (相互情報量最大化) の枠組みで導出 [1] されてることや，テンプレートマッチングの拡張として提案 [2] されているためであり，当然といえば当然のことと言える．ただし，ベイズとの関連性を示したものは無いかというとそうではなく，黒沢による球面ガウス分布を用いた調査 [10] が行われている．しかし，球面ガウス分布を用いていることと，ベイズ決定則ではなく擬似ベイズ識別法との比較を行っているために，古典的な単峰ガウス分布に基づくベイズ決定則からマハラノビス距離や判別分析が自然と導出されることと比較すると技巧的である．ただし，この調査から部分空間識別器はガウス分布をモデルとするベイズ決定則と密接な関係があることが予想され，実際，これまでの多くの実験結果もそのような傾向を示していたといえる．

それならばと，球面ではないガウス分布に基づくベイズ決定則から部分空間識別器を導こうとすると，まず躓くのが自己相関行列 (auto correlation matrix) と呼ばれる行列が何者であるか，ということである．なぜなら，この行列は (最尤推定によって得られる) ガウス分布の式

$$\mathcal{N}(\mathbf{x}|\mu_j, \mathbf{\Sigma}_j) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathbf{\Sigma}_j|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_j)^\top \mathbf{\Sigma}_j^{-1}(\mathbf{x}-\mu_j)\right\} \quad (4)$$

に現れる分散共分散行列 $\mathbf{\Sigma}_j$ ではなく，相関行列でもないからである (μ_j は標本平均)．多くの参考書では，この自己相関行列についての説明はなく，あっても分散共分散行列と自己相関行列との間に成り立つ関係式 ($\mathbf{\Sigma} = \mathbf{R} - \mu_j\mu_j^\top$) が示される程度である．そのためか，$\mu_j = \mathbf{0}$ とおけば $\mathbf{\Sigma}_j = \mathbf{R}_j$ となるのだから，自己相関行列と分散共分散行列は同じものである，と考えている人も多いようである．しかし，式が同じだからと言って，その行列の性質 (固有値や固有ベクトルの意味) まで同じである保証はどこにもなく，実際大きく異なるのである．したがって，この自己相関行列に真っ正面から向き合わなかったことが，部分空間識別器の理論的な解明を遅らせた一因になったのではないかと思われる．

そこで筆者は標本とその原点に関する鏡像を考慮に入れた単峰ガウス分布というものを考え，自己相関行列の持つ性質について考察し，そこから部分空間識別器を (少々の仮定を用いることで) 導けることを示した [11]．また，自己相関行列の固有値の意味も明らかになったため，固有値を認識時に用いるよりも，それに代わる重みを用いた方が良いことを主張した．特に，重みとして直線重みを用いるとアンサンブル学習のような効果を持たせることができ，その結果，パラメータである部分空間の次元数に対して認識率が敏感でなくなり，また CLAFIC や複合類似度よりも認識率が高くなる傾向を持つことを実験により示した．更に，同じ鏡像を含んだガウス分布をベイズ決定則の拡張である複合決定問題に適用することにより，複数の未知標本 (ただしすべて同じクラスに属するとする) のクラスを同時に推定できる複合部分空間識別器というものを導いた [12]．本稿では，それらの導出の詳細よりも分布の定性的な意味について述べることと，部分空間識別器の画像認識・検出への応用について紹介する．

2. 鏡像を考慮した共分散行列 = 自己相関行列

ここでは文献 [11] で述べられている部分空間識別器と最大事後確率則との関係について概説する．

部分空間識別器は，クラス ω_j において，標本平均が $\mathbf{0} \in \mathbb{R}^d$，分散共分散行列が $\mathbf{R}_j \in \mathbb{R}^{d\times d}$ となる単峰ガウス分布

$$\mathcal{N}(\mathbf{x}|\mathbf{0}, \mathbf{R}_j) = \frac{1}{(2\pi)^{d/2}|\mathbf{R}_j|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{R}_j^{-1}\mathbf{x}\right) \quad (5)$$

を $p(\mathbf{x}|\omega_j)$ とした最大事後確率則から導くことができる [11]．ここで $\mathbf{R}_j = \sum_{i=1}^{n_j} \mathbf{x}_i\mathbf{x}_i^\top / n_j$ はクラス ω_j に属する n_j 個の訓練標本 $\mathcal{X}_j = \{\mathbf{x}_i\}_{i=1}^{n_j}$ から得られる標本自己相関行列であり，$|\mathbf{R}_j|$ と $\mathbf{R}_j^{-1}$ はそれぞれ $\mathbf{R}_j$ の行列式と逆

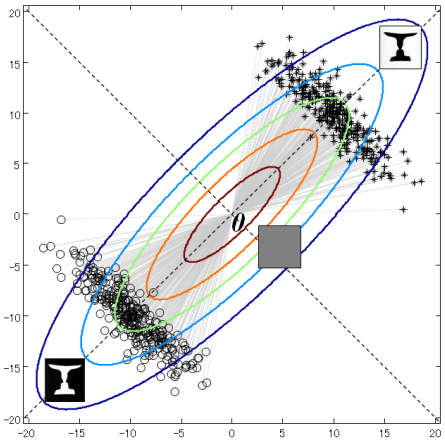

図 3 鏡像を考慮したガウス分布．画像を例にすれば，これは白黒が反転した画像にも同じ尤度を与えようという意図の分布である．原点は，光を発していない画像 (灰色) に対応する．

行列を表す．この $\mathbf{R}_j$ は，原点に関して点対称の標本 $-\mathbf{x}_i$ (鏡像) を含めた標本集合に，単峰ガウス分布を当てはめた場合の最尤推定から求められる．

ここで，正定値実対称行列 $\mathbf{R}_j$ の固有値分解を $\mathbf{R}_j = \mathbf{U}_j\mathbf{\Lambda}_j\mathbf{U}_j^\top$ とする．ただし，$\mathbf{\Lambda}_j \in \mathbb{R}^{d\times d}$ は d 個の固有値 $\{\lambda_{ji}\}_{i=1}^d$ が降順に対角要素に並んだ対角行列，$\mathbf{U}_j \in \mathbb{R}^{d\times d}$ はそれぞれの固有値に対応する固有ベクトル $\{\mathbf{u}_{ji}\}_{i=1}^d$ を並べた行列 $\mathbf{U}_j = (\mathbf{u}_{j1}|\mathbf{u}_{j2}|\cdots|\mathbf{u}_{jd})$ である．なお $\mathbf{R}_j$ の固有値は，対応する固有ベクトルに $\{\mathbf{x}_i\}_{i=1}^{n_j}$ と $\{-\mathbf{x}_i\}_{i=1}^{n_j}$ を正射影したときの標本分散に等しくなる．

この分布を用いた最大事後確率則において，固有値に代わる全クラス共通の非負で非増加な重みの導入と，重みの大きい上位 r 本 ($r < d$) の固有ベクトルを用いることによって

$$S_j(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^{r} w_i(\mathbf{u}_{ji}^\top\mathbf{x})^2 = ||\tilde{\mathbf{W}}^{\frac{1}{2}}\tilde{\mathbf{U}}_j^\top\mathbf{x}||^2 \tag{6}$$

が最大となるクラスへ $\mathbf{x}$ を分類する識別則，すなわち部分空間識別器 (linear subspace classifier, SC) が導かれる．ここで，$\tilde{\mathbf{U}}_j = (\mathbf{u}_{j1}|\mathbf{u}_{j2}|\cdots|\mathbf{u}_{jr}) \in \mathbb{R}^{d\times r}$ である．また，w_i は i 番目に大きい固有値に代わる非負で非増加な重み $w_1 \geq w_2 \geq \cdots \geq w_r > 0$ であり，それらを対角要素に持つ行列が $\tilde{\mathbf{W}}$ =diag$(w_1, ..., w_r)$ である．この部分空間識別器において $\{w_i\}_{i=1}^r$ の値を全て一定としたものが CLAFIC，クラスごとに $w_i = \lambda_{ji}/\lambda_{j1}$ としたものが複合類似度と解釈できる．一方，この重みは非増加であればなんでもよく，直線重み

$$w_i = r - i + 1 \ (i = 1, ..., r) \tag{7}$$

でもよい．この重みは単純であるが，これを用いることで CLAFIC と同等以上の認識率を達成できるようになり，さらに複合類似度のように部分空間の次元数に影響を受けにくくなるという望ましい性質を持つようになる [11]．

3. 複合決定問題と部分空間識別器

はじめに記号の意味について説明し，続いて複合決定問題 (compound decision problem) [13] と関連する決定則，および部分空間識別器 (linear subspace classifier) [11] について概説する．

本稿では，任意の標本 (パターン) $\mathbf{x}$ は d 次元ユークリッド空間 $\mathbb{R}^d$ に属すると仮定し，$\mathbf{x}$ を d 個の実数の組である d 次元実ベクトル

$$\mathbf{x} = (x_1, \cdots, x_d)^\top \in \mathbb{R}^d \tag{8}$$

によって表現する．ここで，$\top$ は転置を表す．本稿では，$\mathbf{x}$ から複数の特徴量を抽出することも考える．すなわち，$\mathbf{x}$ を特徴づける T 種類の特徴量のうち，第 f 特徴量を左下添え字を使って ${}_f\mathbf{x} \in \mathbb{R}^{d_f}$ $(f = 1, ..., T)$ で表すことにする．ここで d_f は ${}_f\mathbf{x}$ の次元数である．以後，これらのベクトルは列ベクトルで統一する．また，L_2 ノルムを $||\mathbf{x}|| = \sqrt{\mathbf{x}^\top\mathbf{x}}$ で表し，c 個のクラスを $\{\omega_j\}_{j=1}^c$ として，その添え字 j をクラスを表す番号とする．

3.1 複合決定問題とその緩和

複合決定問題とは，連続して観測された n 個の標本 $\mathbf{X} = (\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_n)$ に対応する n 個の状態 (context) $\omega = (\omega(1), ..., \omega(n))^\top$ を，連続する状態の統計的独立性を仮定しないで決定する問題である [13], [14]．ここで $\omega(i)$ は $\mathbf{x}_i$ に対する状態であり，c 個のクラス $\{\omega_j\}_{j=1}^c$ のうち一つをとるものとする．$P(\omega)$ を n 個の状態についての事前確率とし，$p(\mathbf{X}|\omega)$ を真の状態集合が ω であるときの $\mathbf{X}$ の条件付確率密度関数とする．これらを用いることで，事後確率は

$$P(\omega|\mathbf{X}) = \frac{P(\omega)p(\mathbf{X}|\omega)}{p(\mathbf{X})} = \frac{P(\omega)p(\mathbf{X}|\omega)}{\sum_\omega P(\omega)p(\mathbf{X}|\omega)} \tag{9}$$

と計算できる．もし $0-1$ 損失関数を用いれば，$P(\omega|\mathbf{X})$ が最大となる ω を選ぶこととなる [14]．しかし，ω の組み合わせは c^n もあり，$p(\mathbf{X}|\omega)$ の推定も難しいことから，計算機でこれらを推定することはほとんど不可能といえる．

しかし，この問題において，n 個の標本が全て同じクラスから由来していることが分かっている場合，事前確率を $P(\omega_j)$ とすることができ，さらに個々の $\mathbf{x}_i$ が i.i.d. に従っている場合には，$p(\mathbf{X}|\omega_j) = \prod_{i=1}^n p(\mathbf{x}_i|\omega_j)$ とできる．このような場合，式 (9) は $\mathcal{X} \stackrel{\text{def}}{=} \{\mathbf{x}_i\}_{i=1}^n$ として

$$P(\omega_j|\mathcal{X}) = \frac{P(\omega_j)\prod_{i=1}^n p(\mathbf{x}_i|\omega_j)}{\sum_{j=1}^c P(\omega_j)\prod_{i=1}^n p(\mathbf{x}_i|\omega_j)} \tag{10}$$

と書ける．したがって，最大事後確率則に基づいて $\mathcal{X}$ を分

ⓒ 2013 Information Processing Society of Japan

類する規則は以下のようになる．

$$\max_{j=1,\ldots,c}\{P(\omega_j|\mathcal{X})\} = P(\omega_k|\mathcal{X}) \Rightarrow \mathcal{X} \in \omega_k \tag{11}$$

なお，$P(\omega_j)$ や $p(\mathbf{x}_i|\omega_j)$ は予め与えられる必要があるが，$P(\omega_j)$ が全クラスで等しいと仮定したり，$p(\mathbf{x}_i|\omega_j)$ としてガウス分布を仮定したりすれば，一つの標本 $\mathbf{x}$ を分類するのと同じように，標本集合 $\mathcal{X}$ を分類することが可能である．

一方，ある標本 $\mathbf{x}$ から得られた複数の特徴量 ${}_1\mathbf{x},\ldots,{}_T\mathbf{x}$ に対しても同様の識別則を導くことができる [15]．すなわち，複数特徴量を考慮する場合の事後確率は $\mathcal{F} \stackrel{\text{def}}{=} \{{}_f\mathbf{x}\}_{f=1}^T$ として

$$P(\omega_j|\mathcal{F}) = \frac{P(\omega_j)\prod_{f=1}^{T} p_f({}_f\mathbf{x}|\omega_j)}{\sum_{j=1}^{c} P(\omega_j)\prod_{f=1}^{T} p_f({}_f\mathbf{x}|\omega_j)} \tag{12}$$

となる．ここで $p_f(\cdot)$ は第 f 特徴量に関する確率密度関数である．この場合，識別則は以下のように与えられる．

$$\max_{j=1,\ldots,c}\{P(\omega_j|\mathcal{F})\} = P(\omega_k|\mathcal{F}) \Rightarrow \mathbf{x} \in \omega_k \tag{13}$$

同様にして，同一クラスに由来する複数の未知標本 $\{\mathbf{x}_i\}_{i=1}^n$ の個々の標本から複数の特徴量 $\{{}_f\mathbf{x}_i\}_{f=1}^T$ が得られる場合の事後確率は $\bar{\mathcal{X}} = \{{}_1\mathbf{x}_1,\ldots,{}_T\mathbf{x}_1,\ldots,{}_f\mathbf{x}_n,\ldots,{}_T\mathbf{x}_n\}$ として

$$P(\omega_j|\bar{\mathcal{X}}) = \frac{P(\omega_j)\prod_{f=1}^{T}\prod_{i=1}^{n} p_f({}_f\mathbf{x}_i|\omega_j)}{\sum_{j=1}^{c} P(\omega_j)\prod_{f=1}^{T}\prod_{i=1}^{n} p_f({}_f\mathbf{x}_i|\omega_j)} \tag{14}$$

となる．したがって，識別則は以下で与えられる．

$$\max_{j=1,\ldots,c}\{P(\omega_j|\bar{\mathcal{X}})\} = P(\omega_k|\bar{\mathcal{X}}) \Rightarrow \bar{\mathcal{X}} \in \omega_k \tag{15}$$

4. 複合部分空間識別器の導出

ここでは式 (11)，式 (13)，式 (15) の識別において，式 (5) をクラス ω_j における確率密度関数 $p(\mathbf{x}|\omega_j)$ や特徴 f に関する確率密度関数 $p_f({}_f\mathbf{x}|\omega_j)$ とすることで，標本集合を分類したり，複数の特徴量を考慮した認識を行ったりすることが可能な部分空間識別器を導くことを考える．はじめに式 (11) の識別を考え，次に式 (13) の識別について考える．最後にこれらを組合わせた式 (15) の識別について述べる．

4.1 複数の標本が与えらた場合の部分空間識別器

事前確率（の推定値）を $P(\omega_j)$ とした場合，事後確率 $P(\omega_j|\mathcal{X})$ はベイズの公式から式 (10) で与えられる．この事後確率を用いて $\mathcal{X}$ を識別するには，分母がクラスに依らないことに注意すると，ω_j の識別関数を $P(\omega_j)\prod_{i=1}^n p(\mathbf{x}_i|\omega_j)$ として，これが最大となるクラスに $\mathcal{X}$ を識別すれば良いが，対数をとっても大小関係は変わらないので，本稿では

$$g_j(\mathcal{X}) \stackrel{\text{def}}{=} \ln P(\omega_j) + \sum_{i=1}^{n} \ln p(\mathbf{x}_i|\omega_j) \tag{16}$$

を識別関数と定義する．式 (16) に式 (5) を代入し，クラスに依存しない項を省くと

$$g_j(\mathcal{X}) = -\sum_{i=1}^{n}\left(\mathbf{x}_i^\top \mathbf{R}_j^{-1}\mathbf{x}_i\right) - n\ln|\mathbf{R}_j| + 2\ln P(\omega_j) \tag{17}$$

となる．ここで $\mathbf{R}_j$ の固有値と固有ベクトルを用いて式 (17) を表すことを考えると，$|\mathbf{R}_j| = \prod_{i=1}^d \lambda_{ji}$ に注意すれば

$$g_j(\mathcal{X}) = -\sum_{i=1}^{n}||\mathbf{\Lambda}_j^{-\frac{1}{2}}\mathbf{U}_j^\top\mathbf{x}_i||^2 - n\sum_{i=1}^{d}\ln\lambda_{ji} + 2\ln P(\omega_j) \tag{18}$$

と表すことができる．さらに，事前確率 $P(\omega_j)$ が全クラスで等しいと仮定すれば，式 (18) は

$$g_j(\mathcal{X}) = -\sum_{i=1}^{n}||\mathbf{\Lambda}_j^{-\frac{1}{2}}\mathbf{U}_j^\top\mathbf{x}_i||^2 - n\sum_{i=1}^{d}\ln\lambda_{ji} \tag{19}$$

となる．したがって識別では，この $g_j(\mathcal{X})$ が最大となるクラスへ $\mathcal{X}$ を分類すれば良い．

以上から，式 (5) をモデルとする最大事後確率則は，式 (19) を識別関数とすればよいことが明らかになった．しかし，このままでは全ての固有ベクトルを必要とするため次元削減を行うことができない．そこで，文献 [16] に倣って，確率的な厳密さは失われるが $\mathbf{R}_j$ の固有値の値そのものは信頼しないで，その大小関係のみが重要であると仮定しよう．すなわち，本来の固有値に代わる全クラス共通の重みを $w_1 \geq w_2 \geq \cdots \geq w_d > 0$ とし，それらを対角要素に持つ $d \times d$ の対角行列を $\mathbf{W} = \text{diag}(w_1,\ldots,w_d)$ とする．このような仮定を置けば，式 (19) は

$$g_j(\mathcal{X}) = -\sum_{i=1}^{n}||\mathbf{W}^{-\frac{1}{2}}\mathbf{U}_j^\top\mathbf{x}_i||^2 = -\sum_{i=1}^{n}\sum_{l=1}^{d}\frac{1}{w_l}(\mathbf{u}_{jl}^\top\mathbf{x}_i)^2 \tag{20}$$

と表すことができる (第二項が仮定により省略できることに注意)．この識別関数を幾何学的に解釈するために，式 (20) を，等価な次の式に書き換える．

$$g_j(\mathcal{X}) = \sum_{i=1}^{n}\sum_{l=1}^{d}\frac{1}{w_l}(||\mathbf{x}_i||^2 - (\mathbf{u}_{jl}^\top\mathbf{x})^2) - \sum_{i=1}^{n}\sum_{l=1}^{d}\frac{1}{w_l}||\mathbf{x}_i||^2 \tag{21}$$

ところが第二項は全てのクラスで共通なので省略すると

$$g_j(\mathcal{X}) = \sum_{i=1}^{n}\sum_{l=1}^{d}\frac{1}{w_l}(||\mathbf{x}_i||^2 - (\mathbf{u}_{jl}^\top\mathbf{x}_i)^2) \tag{22}$$

となる．これは $\mathbf{x}_i$ と，その $\mathbf{u}_{jl}$ への正射影 $\mathbf{u}_{jl}^\top\mathbf{x}_i$ との残差長の二乗 $||\mathbf{x}_i||^2-(\mathbf{u}_{jl}^\top\mathbf{x}_i)^2$ を，w_l の逆数で重み付けしたものの総和であると解釈できる．すなわち，w_l の小さい固有ベクトルへの残差長を重視しているため，高い認識率を得るためには固有値の小さい固有ベクトルを用いる必要があり，計算量や記憶容量の削減の観点から見ると効率が悪い．そこで固有値の大きい固有ベクトルへの残差長を重視するように識別則を変更することを考える．そのために，まず $1/w_1 \le 1/w_2 \le \cdots \le 1/w_d > 0$ なる関係を，分子と分母を反転させることで $w_1 \ge w_2 \ge \cdots \ge w_d > 0$ とする．更に，w_l の値が大きい上位 $r < d$ 本の固有ベクトルのみ識別に用いれば，計算量と記憶容量を削減できるので，w_l のうち値の大きい上位 r 個の値 $\{w_l\}_{l=1}^r$ を式 (22) 中の $1/w_l$ と入れ替えた次の識別関数を定義する．

$$S_j(\mathcal{X}) \stackrel{\text{def}}{=} \sum_{i=1}^{n}\sum_{l=1}^{d} w_l(||\mathbf{x}_i||^2-(\mathbf{u}_{jl}^\top\mathbf{x}_i)^2) \tag{23}$$

ただし，分子と分母を反転させたので，$\mathcal{X}$ を $S_j(\mathcal{X})$ が最小となるクラスへ分類しなければならないことに注意されたい．ところで，$\sum_{i=1}^n\sum_{l=1}^r w_l||\mathbf{x}_i||^2$ は全クラス共通なので省略し，さらにマイナスを掛けることによって式 (23) は

$$\begin{aligned} S_j(\mathcal{X}) &= \sum_{i=1}^{n}\sum_{l=1}^{r} w_l(\mathbf{u}_{jl}^\top\mathbf{x}_i)^2 \\ &= \sum_{i=1}^{n}||\tilde{\mathbf{W}}^{\frac{1}{2}}\tilde{\mathbf{U}}_j^\top\mathbf{x}_i||^2 = \sum_{i=1}^{n} S_j(\mathbf{x}_i) \end{aligned} \tag{24}$$

と書ける．したがって，$S_j(\mathcal{X})$ が最大となるクラスに $\mathcal{X}$ を分類すればよい．このような複数の未知標本と部分空間との類似度の総和を識別に用いるアプローチは既に文献 [17] などで用いられおり，単体の未知標本を分類する場合よりも安定して高い認識率を達成できることが知られている．

4.2 複数の特徴量が与えられた場合の部分空間識別器

ある標本 $\mathbf{x}$ から得られた T 種類の特徴 $\mathcal{F}=\{{}_f\mathbf{x}\}_{f=1}^T$ に基づいて $\mathbf{x}$ を分類することを考える．この場合は，第 f 特徴量の張る空間において，クラス ω_j の標本平均が ${}_f\mathbf{0}\in\mathbb{R}^{d_f}$，分散共分散行列が ${}_f\mathbf{R}_j=\sum_{i=1}^{n_j}{}_f\mathbf{x}_i\,{}_f\mathbf{x}_i^\top/n_j\in\mathbb{R}^{d_f\times d_f}$ ($\{{}_f\mathbf{x}_i\}_{i=1}^{n_j}\in\omega_j$ は訓練標本であることに注意) となる単峰ガウス分布を用いて前節と同様の方法で導くことができる．すなわち

$$\mathcal{N}({}_f\mathbf{x}|{}_f\mathbf{0},{}_f\mathbf{R}_j)=\frac{1}{(2\pi)^{d_f/2}|{}_f\mathbf{R}_j|^{1/2}}\exp\left(-\frac{1}{2}\,{}_f\mathbf{x}^\top\,{}_f\mathbf{R}_j^{-1}\,{}_f\mathbf{x}\right) \tag{25}$$

を $p_f({}_f\mathbf{x}|\omega_j)$ として前節と同様にして部分空間識別器の識別関数を求めると

$$\begin{aligned} F_j(\mathcal{F}) &\stackrel{\text{def}}{=} \sum_{f=1}^{T}\sum_{l=1}^{r_f}{}_fw_l({}_f\mathbf{u}_{jl}^\top\,{}_f\mathbf{x})^2 \\ &= \sum_{f=1}^{T}||{}_f\tilde{\mathbf{W}}^{\frac{1}{2}}\,{}_f\tilde{\mathbf{U}}_j^\top\,{}_f\mathbf{x}||^2 = \sum_{f=1}^{T}{}_fS_j({}_f\mathbf{x}) \end{aligned} \tag{26}$$

となる．ここで，r_f は特徴 f の張る空間での部分空間の次元数，${}_fw_l$ は l 番目に値の大きい重み，${}_f\mathbf{u}_{jl}$ は ${}_fw_l$ に対応する ${}_f\mathbf{R}_j$ の固有ベクトル，${}_f\tilde{\mathbf{W}}$ は ${}_fw_l$ を対角要素に持つ $r_f\times r_f$ の行列，${}_f\tilde{\mathbf{U}}_j$ は ${}_f\mathbf{u}_{jl}$ を r_f 本並べた行列，${}_fS_j({}_f\mathbf{x})\stackrel{\text{def}}{=}||{}_f\tilde{\mathbf{W}}^{\frac{1}{2}}\,{}_f\tilde{\mathbf{U}}_j^\top\,{}_f\mathbf{x}||^2$ である．したがって，複数の特徴量が与えられた場合は，$F_j(\mathcal{F})$ が最大となるクラスに $\mathbf{x}$ を分類すればよい．ただし，部分空間の次元数や重みが特徴量ごとに異なる場合は適切な前処理を行うか，以下のように特徴量ごとに類似度の総和が 1 となるよう規格化した方が良い．

$${}_fS_j({}_f\mathbf{x})\leftarrow\frac{{}_fS_j({}_f\mathbf{x})}{\sum_{j=1}^{c}{}_fS_j({}_f\mathbf{x})} \tag{27}$$

4.3 複数の標本と特徴量に基づいて識別を行う場合

最後に，同一クラスから由来する未知標本が複数与えられ，さらにそれらから複数の特徴量が抽出されている場合の部分空間識別器について考える．すなわち，n 個の同一クラスに由来する未知標本 $\mathcal{X}=\{\mathbf{x}_i\}_{i=1}^n$ が与えられ，それぞれの $\mathbf{x}_i$ から T 種類の特徴 $\{{}_f\mathbf{x}_i\}_{f=1}^T$ が抽出されているとする．この場合もこれまでの議論と同様にすれば，以下の識別関数が最大となるクラスへ $\bar{\mathcal{X}}$ を分類すればよいことがわかる．

$$C_j(\bar{\mathcal{X}})\stackrel{\text{def}}{=}\sum_{f=1}^{T}\sum_{i=1}^{n}{}_fS_j({}_f\mathbf{x}_i) \tag{28}$$

本稿では，これまでに求めた式 (24)，式 (26)，式 (28) が最大となるクラスへ未知標本を分類する識別法を複合部分空間識別器 (compound subspace classifier, CSC) と呼ぶことにする．

5. 実験

ここでは，公開画像データベース Caltech-101 [18] と猫顔検出による実験結果を示す．Caltech-101 を用いた実験では，式 (26) を識別関数とした場合の識別を述べる．猫顔検出では，通常の部分空間識別器における原点の影響について述べる．

5.1 Caltech-101 を用いた複数特徴量に基づく認識

まず，複数の特徴量に基づいた認識について，Caltech-101 を用いた実験結果を示す．Caltech-101 は 101 種類のクラスの画像を Google Image Search を用いて人手で収集した 9144 枚の画像から構成されるデータセットである．ただし，

図 4 Caltech-101 における各クラスの平均画像 ([18] より転載)

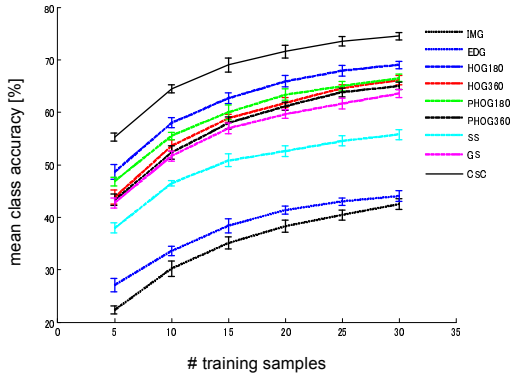

図 5 特徴量ごとのクラス平均認識率

画像の枚数はクラス毎に異なり，最小は 31 枚 (inline-skate)，最大で 800 枚 (airplanes) である．多くの物体は大きさと向きがクラス毎にほぼ揃えられており，アノテーションを用いれば背景をある程度除去することが可能である．そのため，図 4 に示すような各クラスごとの平均画像から，それが何であるかを正しく認識できるクラスも多く存在する．

実験には，アノテーションで切り出した画像から抽出した以下の 8 種類の特徴量を用いた．なお `car_side` クラスの画像が全てモノクロのため，色に基づく特徴量は使用していない．また，背景クラスは含めておらず，`faces` と `faces_easy` クラスは統合していない．画像の拡大縮小は bicubic 法を用いた．

- `IMG`: 32×32 ピクセルのモノクロ画像の輝度値を成分に持つ 1024 次元の特徴量.
- `EDG`: 32×32 ピクセルのモノクロ画像に対して Prewitt フィルタによりエッジを抽出し，そのエッジ強度を成分に持つ 1024 次元の特徴量.
- `HOG180`: 64×64 ピクセルのモノクロ画像から，cell サイズを 6×6 ピクセル，ブロックサイズを 3×3 ピクセル，エッジ方向を 180 度としてそれを 9 分割して求めた 6561 次元のエッジヒストグラム [19].
- `HOG360`: 64×64 ピクセルのモノクロ画像から，cell サイズを 6×6 ピクセル，ブロックサイズを 3×3 ピクセル，エッジ方向を 360 度としてそれを 8 分割して求めた 5832 次元のエッジヒストグラム [19].
- `PHOG180`: 128×128 ピクセルのモノクロ画像から，階層レベルを 3，エッジ方向を 180 度としてそれを 9 分割して抽出した 765 次元のピラミッドエッジヒストグラム [20]．なお，カウントに基づく特徴量は変数変換によって正規性を改善できることが知られており [21]，ピラミッド型の特徴量は特にその傾向が強いことが確かめられたので，本実験では文献 [21] に倣い，特徴成分の平方根をとった．これにより Caltech-101 において 5% 程度の認識率の向上が認められた．
- `PHOG360`: 128×128 ピクセルのモノクロ画像から，階層レベルを 3，エッジ方向を 360 度としてそれを 8 分割して抽出した 680 次元のピラミッドエッジヒストグラム．この特徴に対しても成分の平方根をとる変数変換を施した．
- `SS`: 32×32 ピクセルのモノクロ画像から，パッチサイズを 3，ウィンド半径を 2，半径の量子化数を 2，角度の量子化数を 10，着目画素の間隔を 4 ピクセルとして抽出した 980 次元の self-similarity 特徴量 [22].
- `GIST`: 128×128 ピクセルのモノクロ画像から抽出した 320 次元の gist 特徴量 [23].

表 1 他手法との比較

訓練画像数	MKL [24]	LP-beta [24]	CSC
5	46.5 ± 1.1	59.5 ± 0.2	55.3 ± 0.8
10	59.2 ± 0.5	69.2 ± 0.4	64.4 ± 0.8
15	66.0 ± 0.9	74.6 ± 1.0	69.1 ± 1.4
20	70.8 ± 0.9	77.6 ± 0.3	71.6 ± 1.2
25	74.3 ± 0.8	79.6 ± 0.4	73.6 ± 0.9
30	77.7 ± 0.5	82.1 ± 0.3	74.5 ± 0.7

認識時には，特徴量に対して独立に SELFIC を適用し，次元削減を施した．また，式 (27) を用いて特徴量ごとに類似度の正規化を行った．

本実験の認識率は，クラスあたり 15 枚の未知画像を，訓練画像の枚数をクラスあたり 5 枚から 5 刻みに 30 枚まで変化させて分類したときのクラス平均認識率である．このような分類を訓練画像と未知画像の選択をランダムに 10 回繰り返して行い，クラス平均認識率の平均値と標準偏差を求めた．ただし，画像総数が訓練画像数と未知画像数の合計に満たない場合には訓練画像の枚数を削減した．CSC の部分空間の次元数は，訓練画像の枚数と同じとし，重みは直線重みを用いた．

図 5 に訓練画像の枚数 (部分空間の次元数) を 5 から 30 まで変化させた場合の，特徴量ごとのクラス平均認識率と，式 (26) を識別関数とした CSC のクラス平均認識率を示す．図から，CSC の認識率は個々の特徴量を単独で用いた場合よりも高いことがわかる．表 1 には MKL [24] と 2010 年 10 月現在で世界最高水準の認識率を達成している LP-beta [24] との比較を示す．この表から (用いている特徴

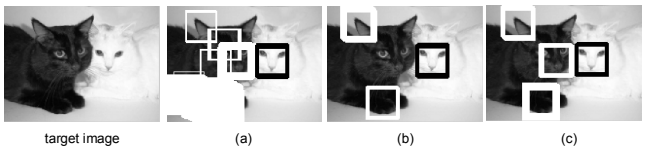

図 6 原点移動の猫顔検出への影響．左端が探索対象とした画像．(a) 原点移動なし．(b) ネガ画像に不変な原点移動．(c) 各クラスの重心の平均に原点移動．

の種類と数が異なるので厳密ではないが) 本手法は MKL と同程度の認識率を達成でき，訓練画像が少ないときは LP-beta と同程度の認識率が達成できていることがわかる．

5.2 猫顔検出

これまでの議論から，部分空間識別器は零元をピークに持つ単峰ガウス分布をモデルとした最大事後確率則から導かれることを示した．しかし，このことから部分空間識別器は零元付近の標本の認識が苦手であることも予想される．文字認識などのように，真っ黒な文字，すなわち零元付近の標本を含まない識別問題では意識する必要はないが，零元付近の標本 (真っ黒なパターン) を含む識別や検出問題には望ましくない性質である．

そこで零元付近のパターンを含む猫顔検出で検証した結果を示す．まず，32×32 ピクセルの猫顔と猫顔以外の画像をそれぞれ 7658 枚準備した後，それぞれで 50 次元の部分空間を張った．次に，猫が撮影された探索対象となる画像を準備し，32×32 ピクセルの探索窓を使って猫を以下の手順で検出した．すなわち，猫のクラスにおける識別関数の出力値を $S_c(\mathbf{x})$，猫以外のクラスの識別関数の出力値を $S_n(\mathbf{x})$ として $S(\mathbf{x}) = S_c(\mathbf{x}) - S_n(\mathbf{x})$ が閾値以上であれば猫と判定した．なお，閾値は，結果が見やすくなるよう，全猫顔標本に対する $S(\mathbf{x})$ の平均値の 1/3 とした．

図 6 に結果を示す．左端が探索対象とした画像であり，二匹の猫の顔がおよそ 32×32 ピクセルとなるようサイズ調整を施している．図 6(a) は原点を移動しないで探索を行った結果である．これから，黒猫上で誤検出が多く発生していることがわかるが，これは部分空間識別器が零元をピークに持つ単峰ガウス分布を仮定しており，黒い領域に対して非常に不安定な結果を出力するためと考えられる．そこでネガ画像に不変な原点移動を施して検出してみた (図 6(b))．これから，誤検出が低減されたが，黒猫が検出できていないことがわかる．そこで，最後に各クラスの重心の平均に原点を移動させる方法 [11] で原点移動を行った (図 6(c))．これから，誤検出を増加させることなく黒猫の顔も検出できるようになり，原点移動が有効であることが確認できた．

図 7 には得られた猫の固有顔と多くの猫の写った画像に対する検出結果を示す．固有顔から猫の模様，明暗の反転，顔の方向きが上手く抽出されており，そのため検出結果が良好になったものと考えられる．

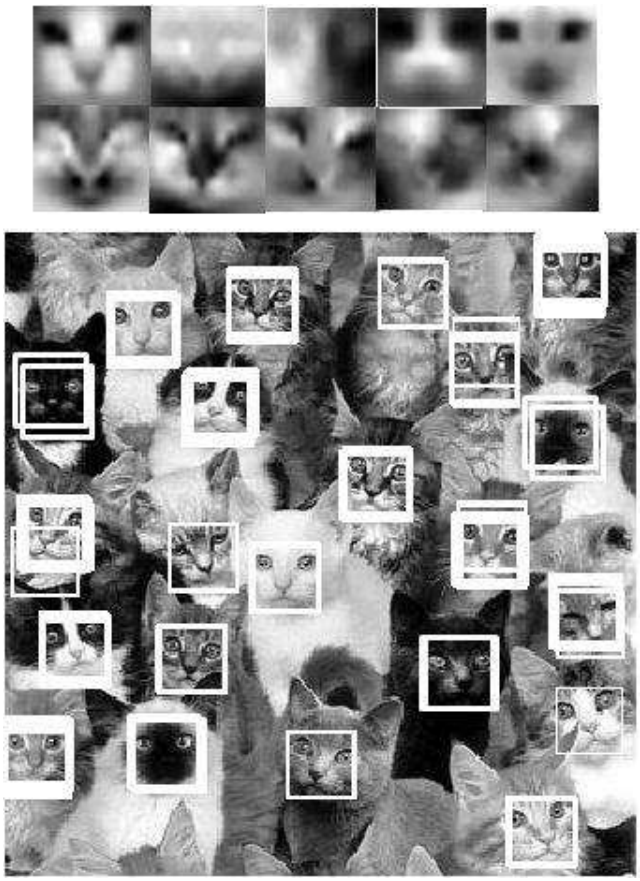

図 7 上段：猫の固有顔．下段：様々な猫の顔検出結果

6. まとめ

本稿では，同じクラスに由来していることが既知の複数の未知標本が与えられた場合，未知標本から複数の特徴量が得られる場合，さらにはそれらが組み合わさった場合の複合決定問題に対応する以下の三種類の複合部分空間識別器を導いた．

- 複数の標本が与えらた場合の複合部分空間識別器
- 複数の特徴量が与えらた場合の複合部分空間識別器
- 複数の標本・特徴が与えらた場合の複合部分空間識別器

実験では，それぞれの識別器の有効性を Caltech-101 や猫顔検出などの問題によって検証した．その結果，それぞれが非常に高い認識率を達成できることを示した．今後の課題として，未知標本数や特徴数と認識率の関係を明らかにすることや，未知標本も圧縮した場合の認識率への影響などを調査することである．

参考文献

[1] S. Watanabe, P. F. Lambert, C. A. Kulikowski, J. L. Buxton, and R. Walker, “Evaluation and selection of variables in pattern recognition,” Comp. & Info. Sciences, vol. 2, (Julius Tou, ed.)., N.Y., Academic Press, pp. 91–122, 1967.

ⓒ 2013 Information Processing Society of Japan

[2] T. Iijima, H. Genchi, and K. Mori, “A theory of character recognition by pattern matching method,” Proc. of 1st Int’l J. Conf. on Pattern Recognition, pp. 50–56, 1973.
[3] 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)
[4] 津田宏治, “ヒルベルト空間における部分空間法,” 信学論, vol. J82-D-II, no. 4, pp. 592–599, Apr. 1999.
[5] 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.
[6] J. Laaksonen, “Local subspace classifier,” Proc. of ICANN’97, pp. 637–642, 1997.
[7] 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, 1985.
[8] 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, 1999.
[9] 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, 2001.
[10] 黒沢由明, “球面ガウス分布から導出される部分空間法,” 信学論 D-II, vol. J81-D-II, no. 6, pp. 1205–1212, 1998.
[11] 堀田政二, “線型部分空間識別器と最大事後確率則の関係,” 信学技報, PRMU2010, vol.110, no.189, to appear.
[12] 堀田政二, ‘複合決定問題から導かれる複合部分空間法,” 信学技報, PRMU2010, vol.110, no.190, to appear.
[13] J.F. Hannan and H. Robbins, “Asumptotic solutions of the compound decision problem for two completely specified distributions,” Annals of Mathematical Statistcs, vol. 26, no. 1, pp. 37-51, 1955.
[14] R.O. Duda, P.E. Hart, and D.G. Stork, “Pattern classification,” John Wiley & Sons, USA, 2nd ed., p. 62, 2001.
[15] J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas, “On combining classifiers,” IEEE Trans. on PAMI, vol. 20, no. 3, pp. 226-239, 1998.
[16] S. Omachi, F. Sun, and H. Aso, “A noise-adaptive discriminant function and its application to blurred machine-printed Kanji recognition,” Trans. on PAMI, vol. 22, no. 3, pp. 314–319, 2000.
[17] 柳詰進介，高橋友和，井出一郎，目加田慶人，村瀬洋，“携帯ディジタルカメラにより撮影された動画像からの低解像度文字認識,” 信学誌，vol. J89-D, no. 2, pp. 323-331, 2006.
[18] L. Fei-Fei, R. Fergus, and P. Perona, “One-Shot learning of object categories,” Trans. on PAMI, vol. 28, no. 4, pp. 594–611, 2006. http://www.vision.caltech.edu/Image_Datasets/Caltech101/ 2013 年 7 月 29 日閲覧.
[19] N. Dalal & B. Triggs, “Histograms of oriented gradients for human detection,” Proc. of CVPR, pp. 886–893, 2005.
[20] A. Bosch, A. Zisserman, and X. Muñoz, “Representing shape with a spatial pyramid kernel,” Int’l Conf. on Image and Video Retrieval, pp. 401–408, 2007.
[21] 若林哲史, 鶴岡信治, 木村文隆, 三宅康二, “手書き文字認識における特徴量の次元数と変数変換に関する考察,” 信学論, vol. J76-D2, no. 12, pp. 2495–2503, 1993.
[22] E. Shechtman & M. Irani, “Matching local self-similarities across images and videos,” Proc. of CVPR, pp. 1–8, 2007.
[23] A. Oliva & A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” Int’l J. of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.
[24] http://www.robots.ox.ac.uk/~vgg/software/MKL/ 2013 年 7 月 29 日閲覧.

ⓒ 2013 Information Processing Society of Japan