

線形重判別分析法を利用した多クラス識別のためのブースティング学習法

谷川 昌司[†] 栗田多喜夫[†]

[†] 産業技術総合研究所 脳神経情報部門 〒 305-8568 茨城県つくば市梅園 1-1-1

E-mail: †{m-tanigawa,takio-kurita}@aist.go.jp

あらまし 本報告では、従来の線形判別空間を利用したクラス識別器の性能をブースティング法を用いて改善する方法について述べる。線形判別空間はクラス間とクラス内分散の比を大きくするような基底から構成され、多クラス分類を容易にする有効な空間を提供することが知られている。しかしながら、顔識別などのように、データが複雑な分布をしている場合には、必ずしも高い判別性能を得る空間を構成できるとは限らない。本報告で述べる方法では、識別器の正誤によって更新された重みづけデータに対する線形判別空間をブースティングの方法で統合する。それぞれのデータ重みは、誤判別したデータについては重みが大きく、正解したデータについては重みが小さくなるように各ブースティング回で変更される。判別器はこの空間を利用して構成される。それぞれの線形判別空間上の識別器の統合で、複雑なデータ分布のクラス識別に対しても高い識別性能を得ることができる。顔識別課題に適用したときの識別率は 98.28%であった。また、識別器の高速化のために基底特徴の数を減らし計算量を少なくした場合においても、10 個の Haar 特徴を用いた場合には 8 回のブースティングで 92.88%の識別性能を得た。

キーワード ブースティング、線形判別空間、矩形特徴

Boosted linear discriminant analysis for multi-class segmentation

Masashi TANIGAWA[†] and Takio KURITA[†]

[†] Faculty of Engineering, First University Yamada 1-2-3, Minato-ku, Tokyo, 105-0123 Japan

E-mail: †{m-tanigawa,takio-kurita}@aist.go.jp

Abstract We propose a new algorithm, boosted linear discriminant analysis (bLDA), for classification of a non-linear pattern distribution, and its fast algorithm for real-time classification. bLDA integrates classifiers on multiple linear discriminant spaces. Each linear discriminant spaces is spanned by eigenvectors, as traditional method, that maximize ratio of within- and between- class variances for weighted data. The weights of data are updated for each boosting step; increasing weights for malclassification, decreasing weights for correct classification. bLDA performs well to classify a complex data distribution, such as face images. Additionally, we propose a fast algorithm, for real-time application, using partial Haar features that approximates eigenvectors spanning bLDA. We obtained high performance (98.28% correct) for a small subset of face image pieces (10 Haar features) with 8 boostings.

Key words Boosting, Linear discriminant analysis, rectangle feature,

1. はじめに

本報告では、従来の線形判別空間を利用したクラス識別器の性能をブースティング法を用いて改善する方法について述べる。線形判別空間はクラス間とクラス内分散の比を大きくするような基底から構成され、クラス分類を容易にする非常に有効な空間を提供することが知られている [2]。しかしながら、顔識別などのように、データが複雑な分布をしている場合には、必ずしも十分に高い判別性能を得る空間を構成できるとは限ら

ず、改善の余地が残されている。これまでに、複雑な分布を持つ多クラス分類に対する判別空間の改善法としては、データ空間を高次特徴空間に変換するカーネル法 [1] [3] や、クラスタリング等の方法を応用した局所領域内で判別空間を構成する試みなど [5] が提案されてた。これらの方法は非線型変換を組み入れた判別空間上へデータ空間を変換し、その上で識別器を構成することによって、高性能の識別器を作ることができることを示している。

一方、識別器を設計するにあたって、複数の識別器を組み合

わせるブースティング法が提案されている [6] [7]。ブースティング法は複数の弱識別器 (識別性能がそれほど高くない識別器) を複数個用意し統合することによって、汎化性能を持ち認識性能が高い強識別器をつくる方法である。Schapire [6] の提案する AdaBoost の方法は、過学習になりにくい汎化性能の高いシステム構成できる方法として知られている。

本報告では、複数の線形判別空間上で構成された識別器をブースティング法により統合することで、複雑な分布に対応する強識別器を構成する方法 (bLDA: Boosted Linear Discriminant Analysis) を提案する。各ブースティングのステップで作られる線形判別空間は、従来と同様の方法で、クラス間分散とクラス内分散の比を大きくするような固有ベクトルの組で張られる空間で表される。ただし、データには重みがつけられており、各ブースティングのステップで重みの更新がなされる。すなわち、識別器が誤判断するデータについては重みを大きくし、識別器が正しく判断するデータについては重みを小さくする。この結果、過去のブースティング・ステップで誤りが大きいデータに対して重要視された複数の判別空間を作ることができ、それぞれの複数の判別空間上の識別器をブースティング法で統合することにより高い識別器が構成される。

本報告では bLDA の高速化の方法について述べる。リアルタイム処理への適用にあたっては、できるだけ計算量の少ない識別器を用いることが得策である。Viola と Jones [9] は、矩形特徴を積分画像の手法を用いて高速な対象探索が可能であることを示した。本報告でも同様の特徴である Haar 特徴を用いる。

先行研究 [8] [10] [11] では、Haar 特徴を選択する際に、考えることのできる全ての可能な Haar 特徴を用いたときの識別性能を調べ、その中で良いものを選択する方策をとった。ここで述べる方法は bLDA で得られる固有ベクトル基底を Haar 特徴で近似表現することを考え、少ない個数の Haar 特徴の組み合わせで統合された識別器の性能を高めることを考察する。

2. bLDA:複数の線形判別空間のブースティング統合

2.1 線形判別分析法

線形判別空間は、クラス内、クラス間の分散を最大にするような式 (1) で得られる線形変換行列 U を見付けることで構成される。元のデータ空間 x から $y = U^T x$ で線形判別空間 y へデータ点が座標変換される。

$$U = \operatorname{argmax}_U \left| \frac{U \tilde{\Sigma}_B U^T}{U \tilde{\Sigma}_W U^T} \right| \quad (1)$$

ここで、 $\tilde{\Sigma}_W$ はデータのクラス内分散を示し、 $\tilde{\Sigma}_B$ はデータのクラス間分散を表す。クラス内分散 $\tilde{\Sigma}_W$ の補空間においても、有効な判別空間が存在する場合があるので、その場合には、クラス間分散にのみ注目して、式 (2) で得られる U を求める [1]。

$$U = \operatorname{argmax}_U |U \tilde{\Sigma}_B U^T| \quad (2)$$

この判別基準を満たすような U は以下の一般化固有値問題を解くことと等価である。

$$\tilde{\Sigma}_B U = \tilde{\Sigma}_W U \Lambda \quad (3)$$

ここで U で表される固有ベクトルの基底が重判別空間を構成する。 Λ は固有値を表す。

実際に認識システムを構成する場合、対象とするデータセットによっては、まずデータ全体に対して主成分分析を行った後に寄与の高い次元のみを採用するような次元縮約の前処理を行う方が、汎化性を向上させることができる。必要に応じて、事前に PCA 処理を行う。(詳細は第 3 節を参照のこと)

2.2 識別器の構成

線形判別空間上でクラス分類をするための識別器には、K-最近傍法や、クラス間距離による方法などがよく用いられるが、本報告では、後述する高速化の方法を考慮にいて、線形判別空間の基底である固有ベクトルを利用した識別器を用いる Haar 特徴ベクトル上の周辺分布を用いた方法には [8] らのが顔認識システムがある。ここでは、その方法を拡張して識別器を構成する。

いま、線形判別空間へ変換されたクラス k のデータの分布を $p^{(k)}(y_1, \dots, y_m, \dots, y_n)$ とする。 y_i は線形判別空間上の座標を表す。また、 m は実験者が決めた識別器の個数であり、 n は、線形判別空間の次数である。

識別器の出力する、クラス k の尤もらしさ $g^{(k)}(y_1, \dots, y_m)$ は、周辺化された確率 $q_i^{(k)}(u_i), i = 1, \dots, n$ の積と、その事後確率によって計算する (式 (4-6))。

$$q_i^{(k)}(u_i) = \frac{1}{Z} \int_{u \in u_i^\perp} p^{(k)}(u_1, \dots, u_m) du^\perp \quad (4)$$

ここで、 Z は q を規格化するための変数。 u^\perp は u_i の補空間。実際に q を計算する場合には、ヒストグラムで近似する。

$$f^{(k)}(u_1, \dots, u_m) = \prod_i q_i^{(k)}(u_i) \quad (5)$$

$$g^{(k)}(u_1, \dots, u_m) = \frac{f^{(k)}(u_1, \dots, u_m)}{\sum_k f^{(k)}(u_1, \dots, u_m)} \quad (6)$$

一般に式 (5) で得られる f の形状は、データの確率分布 p の形状と一致することは希であるが、良い近似を与える場合が多い。また、近似できなかった誤差の部分については、以降で述べるブースティングの方法によって補正される。

2.3 ブースティング法

Schapire [6] は AdaBoost と呼ぶアルゴリズムを提案して弱識別器 (性能がそれほど良くない) 識別器を複数個組み合わせることで、汎化性能の損なわれない強識別器を構成できることを示した。

本報告では、Schapire の AdaBoost 法をベースにして、線形

判別空間上の弱識別器を統合する。ただし、ここでは多クラス識別問題を取り扱うが、Schapire の多クラス問題 AdaBoost.MH によるブースティングをそのまま用いず、式 (7) で表される評価量を正解らしさの指標として 2 クラス問題の Real AdaBoost 法でブースティングを行った。

$$l^{(k)}(u_1, \dots, u_m) = h^{(k)}(u_1, \dots, u_m) - \sum_{j \neq k} h^{(j)}(u_1, \dots, u_m) \quad (7)$$

図 1 は、本報告で用いた Boosting アルゴリズムを示したものである。

2.4 多クラス識別-Fisher あやめデータ

多クラス識別問題の例題としてよく用いられる Fisher のあやめデータを例題として、提案する方法 (bLDA) が有効に動作することを示す。Fisher あやめデータは、それぞれ 100 個の 3 クラスに分類される 4 次元の特徴ベクトルからなる。図 2(a) は線形判別空間のうちで、固有値の大きい 2 つのベクトルからなる空間に射影したときのあやめデータの分布を示す。クラス A は他のふたつのクラス B とクラス C と距離が離れているので容易に分類することができる。実際に、提案する方法で用いる周辺分布のヒストグラムを計算してみると図 2(b) のようになって、クラス A の分離は容易である。ブースティング法、容易に判別のできたデータについては、データの重みを小さくして、次ブースティングでの線形判別空間の構成において影響を小さくする。すなわち、図 2(c) のようなデータの重みに更新される。ここで、データプロット点の濃淡は、データの重みの程度を表している。判別の難しいデータ、すなわちクラス B とクラス C の分布が交差する部分については、大きな重みになっている。

図 2(d) は 5 回のブースティングによる重み更新が行われた後のデータに対して、線形判別空間を構成した結果である。この判別空間においてはクラス B とクラス C の分離に重点をおいた空間が構成されていることがわかる。

10-folds Cross-Validation で評価した結果の識別率は 5 回のブースティングの後で 95.63% であった。Cross-Validation 法では、Iris データをランダムに 10 個の部分集合に分けたあとで、そのうちの 1 つを評価用にして、のこりの 9 個を学習用に用いた。この操作を 10 回繰り返して、その平均認識率を算出した。実験で用いた識別器のヒストグラムのピン数は 4 個とした。

3. 画像識別の例

3.1 部分特徴で構成する

また、遮蔽が存在する場合に対するロバスト性を持たせるには、物体画像の全体をひとつの特徴として扱うのではなく、物体の画像を構成する部分画像に分解する方がより汎化性能の優れた識別器を構成できる。たとえば、Nishida ら (2005) は部分特徴を統合することによって、汎化性能の優れた歩行者識別器を実現している。

本報告では、画素に含まれるノイズや遮蔽に対する汎化性能

を向上させるために、非常に多くの訓練サンプルを用いた場合に結果として近傍画素のみが相関を持つと仮定する。訓練サンプルが十分に多いときにはこのような仮定は必要でないが、遮蔽やノイズを含む十分な訓練サンプルが少ない場合には、実験者が事前知識としてこのような仮定を設けることで性能改善を謀ることができる。仮定を考慮にいれた分散共分散行列を $\tilde{\Sigma}$ とすると $\tilde{\Sigma}_{ij} = G_{ij} \Sigma_{ij}$ で表される。 Σ は、利用できる訓練サンプルから導出される級内あるいは級間の分散共分散行列を示し、 G_{ij} は局所相関をあたえる行列として $G_{ij} = \exp(-\alpha \|x_i - x_j\|)$ で与えられる。 x_i は画素位置を表す。

ブースティング法は弱識別器を複数個統合することによって、強識別器を構成するために、弱識別器として、識別性能は良くないが計算コストの少ない識別器を利用することで十分な識別性能を持つ強識別器を構成することができる。たとえば、Viola ら [9] は、Haar 様の矩形特徴を用いて、処理の高速化を行った。しかしながら、実際に要する計算コストは、ブースティング回数の積で計算されるため、あまりに識別性能が悪い弱識別器をもちいても、全体のブースティング回数が増えることによって、全体の計算コストは少なくならない。したがって、弱識別器を選択する上で、その識別性能はある程度高い必要がある。

Viola らが提案したように、積分画像の方法を使うことで高速探索が行えるために、矩形特徴は高速化のために有益な方法である。このため、先行研究においても、計算コストを少なくするために、矩形特徴を用いる場合が多く、識別性能を高めるために、その種類を多くする方策をとることが多い [10] [11] [8]。

このような方策は、主に考えられる種類の識別器のすべてについて識別性能を調査し、尤も良い識別器の組み合わせを選ぶには非常に多くの学習時間が必要とする。

本報告の方法は、これまでに述べたように、判別空間の構成の段階で、クラス間距離が大きくなるような固有ベクトルを導出することができていることを利用する。すなわち、比較的理想的状態に近い特徴ベクトルを見出したのちに、最適ベクトルを近似する方法を見付けるという方策をとる。

本報告では重判別空間を構成する固有ベクトルを最適ベクトルとみなす。そして、このベクトルを計算コストの少ないベクトルへと分解していく。

計算コストを軽減する意味で、Viola & John がとったと同様の積分イメージを用いることのできる矩形特徴を基本要素とした。また、判別分析で得られる固有ベクトルを矩形特徴に分解するために、多重解像度を持つ Haar 分解を用いた。

この重判別空間の固有ベクトルを適当な局所特徴で分解することを考える。ここで、固有ベクトルで作られる画像の中で最も大きな絶対値を持つ位置の周辺の局所領域を選択する。式 (7) に置ける M 次元の固有ベクトル u_i のうち、 M' 次元のベクトル \tilde{u} のみを考慮する。

3.2 顔-非顔：2 クラス例題と高速化

ここでは、部分特徴の例を考察するために、2 クラスの識別問題の例題として、顔-非顔の識別について考察を行うことにする。例題に用いたサンプルデータは訓練データとして、顔

- (1) データに付随する重み係数 w_i を初期化する。 $w_i = 1/N_k, i_k = 1, 2, \dots, N_k$. ここで、 N_k はクラス番号 k のデータサンプルの数を表す。
- (2) For $T = 1$, to T_{max}
- (a) データの重みを考慮に入れた線形判別空間を構成する (式 (1-3)).
- (b) それぞれの i 番目のクラス k に属するデータ x_i^k に対して、判別空間上の座標を計算する。 $y_i^k = u^T x_i^k$
- (c) M 個のビンで作られるヒストグラムを、各クラス k のデータ x_i^k について、重みを考慮に入れて作成する。このときに行われるヒストグラムを q_j^k とする。ただし、 j はヒストグラムのビンの番号であり、 $j = 1, 2, \dots, M$ である。
- (d) 各データ x について式 (6) によってデータ x に対するクラス k の尤もらしさ $g^k(x)$ を計算する。
- (e) データ重みを更新する。 $w_i \leftarrow w_i e^{-l^k(x_i)}$. ここで、 $l^k(x_i)$ は式 (7) で得られる x_i が正しくクラス分けされているかと示す指標で、 $[-1, 1]$ の範囲をとる。
- (3) 任意の入力 x があつたとき、ブースティング T 回の弱識別器のクラス k の推定強度 $g_T^k(x)$ をの総和を計算する。その中の最大確率を与えるクラスを全体の識別器の判断とみなす。 $C = \operatorname{argmax}_k \sum_T g_T^k(x)$

図 1 線形判別空間の固有ベクトルを用いた AdaBoost のアルゴリズム

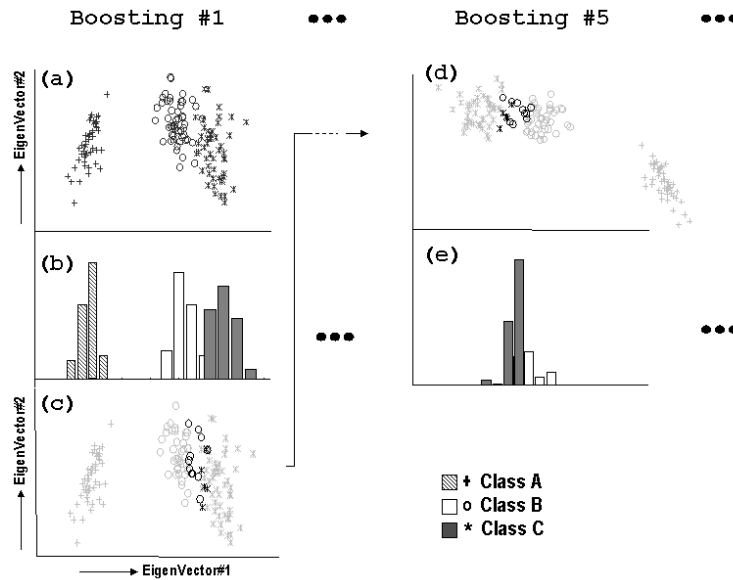


図 2 多クラス問題の例題 : Fisher あやめデータ (3 クラス、4 次元データ) (a) あやめデータを線形判別空間上にプロットしたもの。横軸は第一固有値の固有ベクトル、縦軸は第二固有値の固有ベクトルを表している。シンボルの違いはあやめの種類の違いを表す。クラス A を +、クラス B を o、クラス C を * で表す。(b) 第一固有ベクトルへ射影したときのデータ分布のヒストグラム。(c) 識別器の正誤によって更新されたデータ重み分布。濃淡の濃さが重みの強さを表す。(d)(c) の重み付けデータについて再度判別空間を構成した結果。一度目の判別空間の構成では見分けの尽きにくかったクラス B とクラス C との間が重要視した判別空間が構成される。(e) 新しい第一固有ベクトル上に投影したデータのヒストグラム。

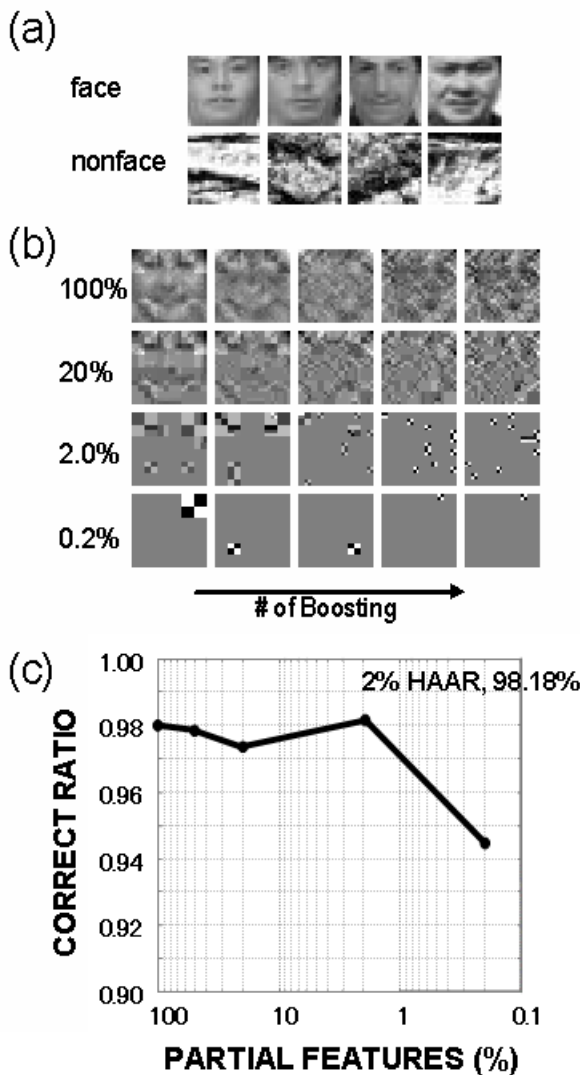


図3 (a) 顔識別問題に用いた例題パターン、(b) Haar 分解された要素のうち、大きな値を持つものを選択した。100% Haar, 20% Haar, 2% Haar, 0.2% Haar、(c) それぞれの部分特徴についての識別率の表。2% Haar の場合で尤も良い識別率を得ることができた。

300 枚、非顔 700 枚、テストパターン、顔 425 枚、非顔 1500 枚を用いた。図 3(a)

尤も良い識別性能を得たのは、部分特徴が全体の 2%(10 個の Haar 特徴) からなる場合で、140 回のブースティングの後でテストデータに対して 98.29% の認識率を得ることができた。

画像全体 (100% Haar, 576 個の Haar 特徴) を用いた場合には、認識性能は 154 回のブースティングの後で 98.13% であった。

画像全体を特徴量として用いた場合には、ブースティングの初期において良好な結果を得ることができ、一回のブースティング (単一特徴: 100% 特徴) で 92.16% の正答率を得ることができる。一方、少数の 2% の部分特徴を用いたときの結果は 75.17% であった。この結果は、全体特徴を用いた場合の方が Haar 特徴を多く含んでいるため当然の結果といえる。ここで、全体特徴 1 個を用いた場合 (92.16%) と同程度の性能を得るために必要な 2% 部分特徴のブースティング回数は 8 回で、そのとき 92.88% になる。

累積寄与率	bLDA (Test)	LDA (Train)	LDA (Test)
75%	96.31%	90.00%	91.01%
90%	98.08%	92.10%	92.16%
99%	94.75%	96.80%	86.20%
100%	70.70%	99.92%	66.23%

表 1 PCA 次元縮約の効果: bLDA は 200 回ブースティング、LDA は 1 回ブースティング。test/train は認識率を測定したデータセットの種類。

いま、計算コストがブースティング回数と部分特徴の個数の積で計算されるとするならば、部分特徴を用いた場合には、 $576/(8 \times 10) = 7.2$ 倍に高速化できることをしめしている。

3.3 男性-女性-非顔: 多クラス例題

多クラス例題として、顔-非顔のうち、顔クラスを男女のクラスに分類して、3 クラス問題として拡張した場合を計算する。

学習パターンは非顔画像は 700 枚用いた。顔画像 300 枚のうち 183 枚が男性、117 枚が女性であった。

テスト画像は、非顔画像を 1500 枚、男性 297 枚、女性 127 枚。3 クラスの識別率は 94.7% であった。識別率は識別結果が正しいクラスを示している個数を全体の個数で割算した。

本報告のように、多クラス分類を一つの学習プロセスで行うのではなく、分類を階層的に考えて、多クラス分類を 2 クラス分類を二段階に組み合わせる方策もある。すなわち、まず顔-非顔の分類をしたのち、男女の識別を考える。顔-非顔の分類は前述のように 98.13% であった。また、顔クラスに限定して、男女の識別を行った場合にはその識別率は 83.49% であった。

一方で 3 クラス識別を行った結果に対して、男性クラスと女性クラスの混同は許容して識別率を計算した結果、顔 (男性+女性) -非顔の場合には 98.8% であった。また男性-女性間の 3 クラスの識別率は 82.3% であった。

この結果は 3 クラスを同時に分類する本方法と、多クラス問題を 2 クラス分類の階層分類で識別する場合とでほぼ同程度の性能をしめしている。

3.3.1 PCA による次元縮約の効果

学習データ数が少ない場合の学習器は、その学習データに対する識別能力が高いほど、テストデータに対する汎化性能が損なわれることがある。この問題を避けるための一つの方法として前処理として主成分分析を行い、寄与率の高い次元にデータ空間を縮約する方法がある。以下に示す結果は主成分分析による次元縮約の効果、累積寄与率と認識率との関係で表にしたものである。

ブースティング後の学習データに対する認識率はどの場合についても 100% であった。PCA 縮約しない (累積寄与率 100%) 場合には、学習データに対する正答率は高いがテストデータに対する正答率は低い。従来の LDA (ブースティング回数が 1 回に相当) の場合の学習データに対する認識率は、累積寄与率が大きくなると、性能が向上し、1 つひとつの線形判別器の学習データに対する識別能力は向上する一方で、テストデータに対する許容度が低くなる。

4. 考 察

本報告では、複数の線形判別空間上に作られた識別器をブースティング法によって統合し、高性能の識別器を構成できる方法 (bLDA) について提案した。また、識別器の計算コストを削減するために、識別器のもつ特徴ベクトルを Haar 特徴で近似し、少ない Haar 特徴検出器をブースティングによって組み合わせる方法を提案した。それをアヤメデータと顔識別の例題に適用したところ良好な結果を得た。

ブースティング法は識別器の信頼性に基いた重みを決める方法を示しているにすぎないので、各ブースティング段階でどの識別器を選択するかは設計者が自由に選べる。しかしながら、ブースティング早期に識別を行おうとすると、識別性能の高い識別器を選択していく必要がある。

本報告で用いた、各ブースティング段階では、選択の指標として級内級間分散比を大きくとるものを最優先に識別器の候補として選択した。

分散比はクラス間がどの程度分離しているかを示す度合 (分離度) として従来、線形判別法で用いられてきたものである。

分離度を表現する方法にはいろいろなバリエーションが考えられる。たとえば、2 クラス問題を取り扱う場合には、サポートベクターマシンの方法で、2 クラスの分離面を基準にして決めることも可能であろう。

多クラス問題に対して如何に適切な分離度を設定し識別器を探索選択するかは課題として残されている。

文 献

- [1] J.Yang, A.F. Frangi and J.Yang and D.Zhang, KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition, IEEE Trans. PAMI, vol 27, no. 2, pp. 230–244, 2005,
- [2] P.N. Belhumeur et al. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, IEEE Trans. PAMI vol. 19, vol. 7, pp. 711–720, 1997,
- [3] T. Kurita and T. Taguchi, A kernel-based Fisher discriminant analysis for face detection, IEICE Trans. on Information and Systems, 2005
- [4] S. Schölkopf et al. Input space vs. feature space in kernel-based methods, IEEE Trans. on Neural Networks, vol. 10, no. 5, pp. 1000–1017, 1999
- [5] T. Kim and J. Kittler, Locally Linear Discriminant Analysis for Multi-modally Distributed Classes for Face Recognition with a Single Model Image, IEEE Trans. on PAMI, Vol. 27, No. 3, pp. 318–327, 2005.
- [6] R.E. Schapire and Y. Singer, Improved Boosting Using Confidence-rated Predictions, Machine Learning, Vol. 37, no. 3, 1999, pp. 297–336
- [7] J. Friedman et al. Additive logistic regression: a statistical view of boosting, 1999
- [8] B. Wu et al. Fast rotation invariant multi-view face detection based on real adaboost. Proc. IEEE int. conf. Automatic face and gesture recognition. , Korea, 2004, p.79
- [9] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, In CVPR, 2001.
- [10] R. Lienhart and J. Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002, Vol. 1, pp. 900–903, Sep. 2002.
- [11] S. Li et al., FloatBoost learning for classification, In S. Thrun S. Becker and K. Obermayer, editors, NIPS 15. MIT

Press, December 2002.

- [12] K.Nishida, T.Kurita, Pedestrian detection by boosting soft-margin SVM with local feature selection. IAPR on MVA, 2005, in press.