

## GibbsBoostによる正面顔画像検出

事前情報を考慮する Bayes 的アプローチ

Frontal face detection method via GibbsBoost: Bayesian approach with prior information

木村 彰夫  
Akio Kimura松井 淳  
Atsushi Matsui中田 洋平  
Yohei Nakada松本 隆  
Takashi Matsumoto

## 1. まえがき

自然画像から人物の顔を検出する手法はさまざまなものが提案されているが、なかでも、Viola らの矩形特徴量と AdaBoost を用いた統計的手法[1]が近年注目を集めている。本研究では、この Viola らの手法に、逐次モンテカルロ法を用いた GibbsBoost[2]を適用し、その有効性を検証する。さらに、複数の弱学習機からなる判別器を多段接続する際、前段の信頼度の和の値を後段の学習に利用する、新たなカスケード構造を提案する。

## 2. GibbsBoost 学習アルゴリズム

$t$ 個の弱学習機(弱仮説)  $h_1, \dots, h_t$ の線形結合で構成された判別機械  $F_t$ について考える:

$$F_t(x; \Theta_t) = \sum_{i=1}^t \alpha_i h_i(x; \theta_i) \quad (1)$$

ここで、 $x$ は入力画像、 $\alpha_i$ は  $i$ 番目の弱学習機  $h_i$ の信頼度、 $\theta_i$ は  $h_i$ のパラメタである。ここで  $\theta_i := (s_i, z_i)$ とし、 $s_i$ と  $z_i$ はそれぞれ画素位置パラメタと矩形形状パラメタである。また、 $\Theta_t := (\alpha_1, \dots, \alpha_t, \theta_1, \dots, \theta_t)$ とする。 $N$ 組の学習データ  $\{y_i, x_i\}_{i=1}^N$ が与えられたとき、ブースティングと呼ばれる集団学習アルゴリズムは、 $t = 1, \dots, T$ における損失関数  $\sum_{i=1}^N L(y_i F_t(x_i; \Theta_t))$ を  $\Theta_t$ について逐次的に最適化する。

評価関数の最適化(最小化)は、局所最適解などのさまざまな問題を伴うことが多い。本稿で用いる GibbsBoost では、損失関数に対応するエネルギー関数  $L(z)$ を用いて定義したパラメタの確率分布

$$P_t(\Theta_t) \propto \pi(\Theta_t) \prod_{i=1}^N \exp\left(-\beta_i L\left(y_i \frac{F(x_i; \Theta_t)}{\sqrt{t}}\right)\right) \quad (2)$$

を用いて期待値操作を行うことで、パラメタの最適化に伴うさまざまな問題を回避する。ここで、 $\pi(\Theta_t)$ は  $\Theta_t$ のとりうる範囲に定義された確率密度関数であり、ベイズ学習の事前確率に相当する。また  $t^{1/2}$ による除算は、 $F_t$ の分散が  $t$ に比例して増えることを防ぐためである。

このとき、最終的な判別機械  $F_T$ は、パラメタの確率分布  $P_T(\Theta_T)$ についての期待値で定義される:

$$F_T(x) := \int F(x; \Theta_T) P_T(\Theta_T) d\Theta_T \quad (3)$$

## 3. 逐次モンテカルロ

式(3)は多次元連続空間上の積分であり、解析的に解を求めることが非常に困難である場合が多い。本稿では、逐次モンテカルロ法を用いて式(3)の積分計算を実行する。

逐次モンテカルロ法で得られた  $M$ 個のサンプルを  $\{\Theta_t^{(j)}\}_{j=1}^M$ とすると、式(3)は次式で近似される:

$$F_T(x) \approx \frac{1}{M} \sum_{j=1}^M F_T(x; \Theta_T^{(j)}) \quad (4)$$

## 4. 学習時間と処理時間の高速化手法

## 4.1 先見の情報にもとづく事前確率分布

顔画像をある大きさの四角い領域で切り出したとき、その画像の4隅の部分は、切り出す大きさによっては顔の輪郭または背景に当たる領域と異なるため、それら周辺部分の画像が顔を特徴づける情報を多く持っているとは考えがたい。また、Viola らの先行研究によれば、ブースティングの初期段階で選ばれた弱学習機のうちいくつかは、目や鼻などの顔の器官に対応する部分に特徴量計算の矩形を持つことが経験的にわかっている。

そこで本稿では、このような矩形特徴量の配置に関する主観的・先見の情報を考慮し、顔の中心に近い特徴量ほど確率が徐々に高くなる次のような確率分布を考える。

$$\pi(s_i) \propto \exp\left[-\frac{1}{2}(s_i - \mu) \Sigma^{-1}(s_i - \mu)\right] \quad (5)$$

$$\mu = \begin{bmatrix} 12 \\ 12 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 144 & 0 \\ 0 & 144 \end{bmatrix}$$

$$s_i \in [1, 24] \times [1, 24]$$

$z_i \in \{1, \dots, k_z(s_i)\}$  ( $k_z(s_i)$ は位置  $s_i$ で選択可能な矩形の種類数)

このように、全特徴量を探索対象とするのではなく、目や鼻などのより顔の情報を多く含んでいる特徴量を中心に探索することで、1つの弱学習機(弱仮説)を選択する計算処理を短縮でき、学習時間の高速化の効果が期待できる。

## 4.2 最大確率判別器

GibbsBoost にもとづく判別機械は、式(4)に示したように  $M$ 個のパラメタのサンプルを用いて計算した各々の出力の平均を取ることで最終的な判別結果を決定する。

$M$ の数が大きいほど逐次モンテカルロ法の近似誤差は小さくなるが、最終的な判別結果を得るまでの処理時間は  $M$ の数に比例して増大する。

そこで、本研究では式(3)で定義した期待値計算において、確率分布のピーク以外の寄与が無視できるほど小さいと仮定する。すなわち、パラメタ  $\Theta$ の確率分布についての期待値  $F_T(x)$ が、逐次モンテカルロ法で抽出した  $M$ 個のサンプルのうち、最大確率を与えるパラメタ

$$\Theta_T^* = \arg \max_{\Theta_T \in \{\Theta_T^{(j)}\}} \log P(\Theta_T) \quad (6)$$

のみを用いて計算した判別器  $F_T(x; \Theta_T^*)$  と十分に近いとみなす。このとき、最終的な判別結果は次式で与えられる:

$$H_T(x) := \text{sign}(F_T(x; \Theta_T^*)) \quad (7)$$

### 5. 非顔の顔類似度を考慮するカスケード構造

Violaらは弱仮説 $h$ を多数組み合わせさせたカスケード構造を用いて判別時間の効率化を図った。図1において、 $H_k$ は識別機を表す。

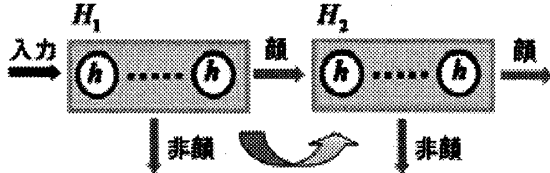


図1. カスケード構造

識別機 $H_k$ の学習には、 $H_1$ から $H_{k-1}$ をすべて通過したデータを用いる。本稿では、 $H_{k-1}$ の非顔画像データについての判別機械 $F^{k-1}$ の出力を $H_k$ の学習に利用する仕組みを提案する。具体的には、次式のように学習データ $\{y_i, x_i\}_{i=1}^N$ について学習が終了した $H_{k-1}$ の判別機械の出力 $F^{k-1}(x; \Theta_i^*)$ を、 $H_k$ の判別機械 $F^k$ のバイアス項 $F_0$ として、学習時に利用する。

$$F_0^k(x_i) = \frac{N_N F_T^{k-1}(x; \Theta_T^*)}{N_F + N_N} \quad (8)$$

$N_N$ : 非顔画像データ数  $N_F$ : 顔画像データ数

## 6. 実験

### 6.1 実験条件

検出器は識別機が2段のカスケード構造にし、各識別機は弱学習機列を50個とした。また、各識別機の学習に5000枚の正面顔画像と5000枚の非顔画像を用いた。顔画像は目とあごを基準に切り取ったものを、非顔画像は顔の映っていない画像からランダムに切り取り、それぞれ24x24の解像度に正規化したものを用いた。

実験は、まずAdaBoostとGibbsBoostの比較実験を行い、次にGibbsBoostを用いて、カスケード構造において顔の類似度を用いた場合と用いない場合との比較実験を行った。

評価にはMIT+CMU画像データベース[3]の評価用画像507枚を用いた。実験にはXeon™ 3.00GHzのPCを用いた。

### 6.2 実験結果

GibbsBoostとAdaBoostの実験結果を図2に示す。ひとつの弱仮説を学習するのにかかる時間を表1に示す。精度に差はさほど見られないが、学習時間が150分の1以下に削減している。これはAdaBoostでは、弱学習機(特徴量)を選択する際にすべての特徴量(134,736個)を探索対象としているのに対し、GibbsBoostでは、サンプル数(500個)分の処理ですむことが、影響をしていると思われる。

また、本稿で導入した先見的信息にもとづく事前確率分布により、効率的に弱学習機(特徴量)探索処理を行っている点が、学習時間の短縮に貢献していることも、著者らの予備の実験により確認されている。

またGibbsBoostで学習した2段のカスケード構造をもつ識別機に式(8)に示した提案手法を適用したときの評価を図3に示す。顔類似度を考慮しない場合と比べ、性能が少し向上している。本実験では、カスケード構造を最少構成の2段であるとしたが、さらに段数を増やした場合より大きな効果が見込める。

## 7. まとめ

正面顔画像検出問題において、事前情報を考慮したGibbsBoostによる顔画像検出手法を提案した。MIT+CMU評価画像において、AdaBoostによる検出手法と提案手法を比較したところ、検出精度にはさほど差は見られなかったが、学習時間が大幅に短縮された。今後はカスケードを多段に接続し、弱学習機列の増加による検出精度の向上を図りたい。

表1. 弱学習機1つあたりの学習時間

学習アルゴリズム	AdaBoost	GibbsBoost	
バイアス項 $F_0$	なし	なし	あり
時間(分)	308.8	1.8	1.8

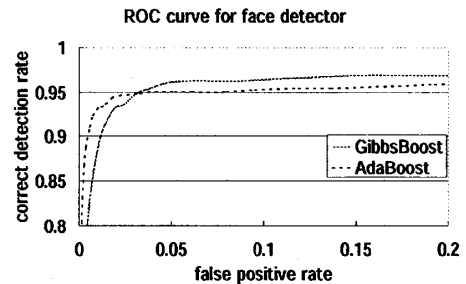


図2. AdaBoostとGibbsBoost比較

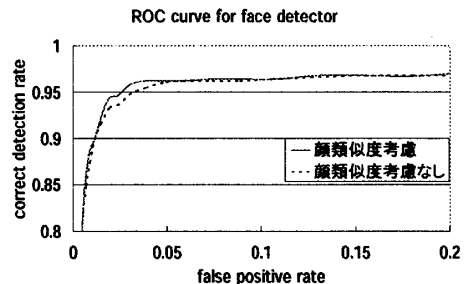


図3. 顔類似度を考慮する効果

## 8. 参考文献

- [1]Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," IEEE Conference on Computer Vision and Pattern Recognition, Vol.1, pp.511-518 (2001).
- [2]Y. Nakada, Y. Mouri, Y. Hongo and T. Matsumoto, "GibbsBoost: A Boosting Algorithm Using a Sequential Monte Carlo Approach"(accepted), IEEE Machine Learning for Signal Processing Workshop 2006 (2006).
- [3]H.A Rowley, S. Baluja, and T. Kanade. "Neural Network-Based Face Detection," IEEE Trans. on PAMI, vol.20, pp.23-38, (1998)