

顔と人体画像認識に生きる機械学習*

Machine Learning as a Tool for Face and Human Detection

勞世竝†

Shihong Lao

概要: 近年、顔画像認識技術が急速に実用化してきた。その代表的な応用として、広く普及されたのは顔検出機能を搭載したデジタルカメラやビデオカメラである。その他にセキュリティ分野における顔認識技術の応用や、自動マーケティングにおける性別年齢推定技術の応用なども実用化されている。これらの技術の実現には機械学習がなくてはならない重要な役割を果たしている。本稿では、主に顔検出の開発を事例として、どのように機械学習技術を応用し、そしてどのようにこれらの技術が実用されているかを紹介する。

Keywords: 顔検出, 人体検出, SGF 特徴量, AdaBoost, Real AdaBoost

1 はじめに

デジタルカメラやビデオカメラに顔検出、顔認識の機能が搭載され、身近に最先端のパターン認識技術の恩恵を受けることができるようになった。これが可能になったのは、ハードウェアの進歩による面もあるが、機械学習の理論と技術の発展によって、大量データによる学習が可能になり、性能的、速度的に実用的なアルゴリズムが実現できたお陰だと言っても過言ではない。

本稿では、顔画像処理の実応用の現状を簡単に紹介してから、主に顔検出技術の実用化に際して、実施したいいくつかの改良と機械学習を行う際の工夫について説明する。

2 顔画像処理技術の概況

近年、顔検出や顔認識などの技術がリアルタイムで実現できるようになったことで、顔画像処理の応用分野が急速に広がっている。図 1 に、主要要素技術とそれらを応用したアプリケーションの例を示す。

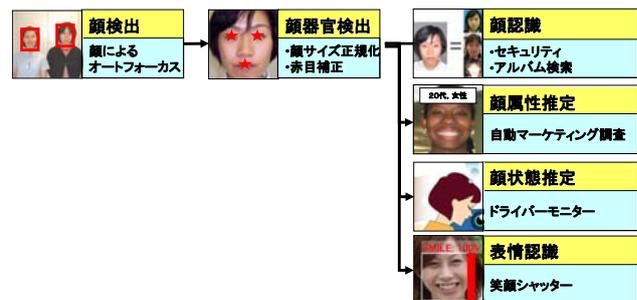


図 1 顔画像処理技術の要素技術例

顔画像処理の主な応用分野として、次のようなものがあげられる：

1) デジタル機器分野

写真やビデオの画質向上に繋がる応用を目的として、デジタルカメラ、ビデオカメラのオートフォーカスやプリンタ、写真現像機などにおける自動補正機能が実用化されている。また、PC や携帯電話などにおける個人認証や映像インデクシング、写真整理などに応用されている。

2) エンタテインメント分野

顔写真を楽しむためのサービスも多い。写真シール機の写真画質向上、顔写真による古いサービスや写真変装サービスが若年層を中心に使われている。また携帯型ゲーム機器にも顔画像処理が導入されている。

3) 車載分野

顔画像処理によってわき見や居眠りを検出するドライバーモニターが搭載された事例がある。

4) セキュリティ分野

*企画セッション 「広がる機械学習応用のフロンティア」

†オムロン株式会社, コアテクノロジーセンター

〒610-0283 木津川市木津川台 9-1, tel. 0774-74-2891,

e-mail: lao_shihong@omron.co.jp, OMRON Corp.

駅、空港、商業施設などでの不審者検出、出入国管理、施設へのアクセスコントロールなどの分野で顔認識技術が実用化されている。

3 顔検出技術

顔画像処理の中で、最も基本的で、重要な技術として顔検出技術が上げられる。デジタル写真や映像が急速に普及している中で、顔検出技術の役割がますます重要になってきた。例えば、コンピュータで顔認識を行う場合は画像の中からどこに顔があるかをまず検出する必要がある。特にデジタルカメラにおける顔オートフォーカスや顔オートアイリスにはカメラに搭載でき、リアルタイム処理できる顔検出技術が必要である。

3.1 顔検出の課題

3.1.1 顔検出の難しさ

コンピュータで顔をロバストに検出する難しさは顔の多様性に由来する。人には個人差があり、性別、年齢、人種などの違いによって、顔の見た目はかなり違う。また、同じ人の顔でも、顔の向き、照明環境と表情によっても見え方が違う。実アプリケーションの場合、顔の大きさ、回転角度も変化していることが多いので、さらに複雑になる(図2)。



図2 顔の多様性

3.1.2 従来の顔検出技術とその課題

人はまったく努力せずに、自然に顔を見つけることができる。しかしながら、コンピュータで顔を検出するには画像の中で順番に領域を切り出して、その領域が顔かどうかを判定することによって顔の位置を特定する(図3)。そのため、膨大な数の領域の判別が必要となり、高速な判別器が重要になる。

90年後半CMUのRowleyらが開発したニューラルネットワークによる顔検出技術は学習技術を使った顔検出のフレームワークを確立し、当時は最も性能

の良い顔検出技術として知られていたが、計算速度に関しては実時間処理には及ばなかった。

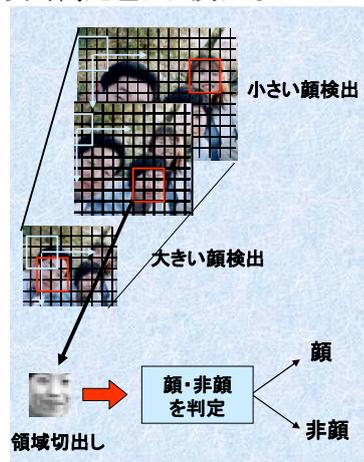


図3 領域切り出して顔かどうかを逐一判別することで顔を検出する方法

2001年ごろ、ViolaとJones[1]が高速に顔検出を行う手法を提案し、実時間での顔検出を可能にした。この手法は近年最も脚光を浴びたパターン認識手法の一つである。従来の顔識別器は入力された画像がどんなものであってもすべて同じ処理を行って顔かどうかを判別する。彼らは画像の中の多くの部分は顔ではなく、ほとんどの領域は明らかに顔に似てないことに注目し、これらの領域においてはより単純な計算で高速に顔ではないと判断できることを巧妙に利用して高速化を実現した。

彼らの方法には3つ重要な貢献がある：

1. 積分画像を用いて、単純で解像度に関係なく高速演算できる Haar-like の特徴量を導入した。
2. AdaBoost アルゴリズムを導入して上述した高速に計算できる特徴量を使った弱判別器を選定し、線形結合で強判別器を構築する方法を提案した。弱判別器は単純な閾値演算でバイナリの出力を決める。
3. 計算量の少ない判別器と計算量の多い判別器を直列に順番につなぐ構造の顔検出器を提案した。計算量の少ない識別器で非顔と判断された場合は途中で計算を打ち切ることで、画像の中で大部分の領域での計算を減らし、高速な顔検出を実現した。

ViolaとJonesが提案したHaar-likeの特徴量は矩形の中の領域の平均輝度差である(図4に白い領域と黒い領域との平均輝度差)。ViolaとJonesは4種類のHaar-likeの矩形特徴量を提案している。これらの特徴量の良いところは領域の大きさによらず、任意領域の平均輝度値は積分画像を使って3回の足し算、引き算と1回の割り算で高速に計算できることである(図5)。ここで、積分画像の画素 $S(x, y)$ と元画像の画素 $I(i, j)$ の関係は次のように定義される：

$$S(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

図4 Viola と Jones が提案した 4 種類の Haar-like の特徴量，隣接する白い矩形領域の平均輝度と黒い矩形領域の平均輝度の差で表す

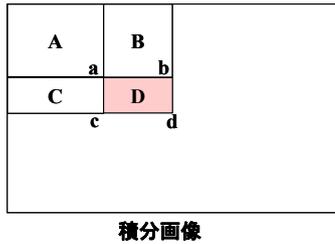


図5 元画像中のある領域の輝度合計を矩形の大きさによらず 3 回の加減算だけで求めることができる (D=d-c-b+a)

Viola らの手法は非常に高速で，画期的であったが，正面顔以外，横顔も対応する必要がある場合，性能的，速度的に不十分であった．実アプリケーションに応用すると，更なる高速化や検出性能の向上とハードウェア化しやすくするためのメモリ使用量の削減などが要求される．

3.2 実用化に向けた顔検出の改良

実アプリケーションにおいて，顔検出に対して，以下のようなニーズがある：

- ・写真印刷で逆光や白とびの顔の輝度を自動補正する場合，カメラの持ち方などによって，顔が上向きとは限らないため，回転した顔をも検出する必要がある．
- ・顔の向きは正面だけでなく，横顔の対応も必要．
- ・組み込み機器に搭載するためにはプログラムのROM と RAM の使用量を削減する必要がある．

これらのニーズに答えるために，特徴量，学習アルゴリズム，検出器の構造に対して改良を加えている．

3.2.1 特徴量の改良

筆者のグループでは，Haar-likeの特徴量よりもより高速に計算でき，しかも識別能力が高い“SGF特徴量” (Sparse Granular Feature) を提案している[2].

$$F(\boldsymbol{\pi}) = \sum_i \alpha_i g_i(\boldsymbol{\pi}; x, y, s), \quad \alpha_i \in \{-1, +1\}$$

ここで， $\boldsymbol{\pi}$ は入力画像の濃淡データ， $g_i(\boldsymbol{\pi}; x, y, s)$ はGranule “粒子” で x, y は位置， s はサイズを表すパラメータ (

図6図7) である． g はその領域の平均輝度とし，元画像をスケール $1/2^s$ に縮小した画像をあらかじめ生成しておけば，各粒子 g の値は直接各 $1/2^s$ に縮小した画像の画素値から得られるため，計算量が少ない．

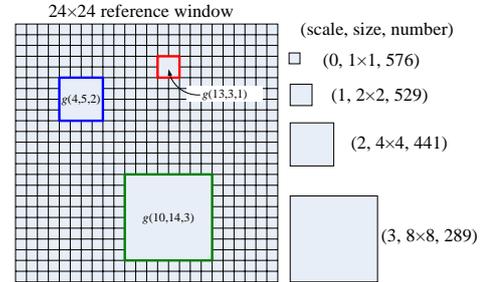


図6 SGF 特徴量(Sparse Granular Feature)の概念

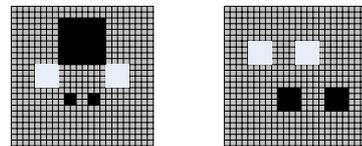


図7 学習した結合粒子特徴量の例 黒は $\alpha = -1$ ，白は $\alpha = 1$ を示す

このような SGF 特徴量は速度的にも性能的にも優れていることが実証できている．また，計算が単純なため，ハードウェアでの実現がしやすいメリットがある．

3.2.2 学習アルゴリズムの改良

ブースティングアルゴリズムは複数の判別能力の低い弱判別器 h を組合わせてより性能の高い強判別器 H を構成する方法である．AdaBoostは適応型の学習法で，弱判別器を組合わせる方法を問題に適応して決める．ViolaとJonesがAdaBoostを最初に顔検出に適用し，優れた学習能力があることを示した．但し，ViolaとJonesの顔検出アルゴリズムの場合，弱判別器の出力は閾値処理によってバイナリの値を出力するタイプのもので，顔と非顔の弱判別器における入力データの分布がそれぞれ分かれた場合には有効であるが，そうでない場合は弱判別器の識別能力が落ちる弱点がある (図8にそのイメージ図を示す)．しかしながら，実際の弱判別器の入力データの分布は複雑なものが多いため (図9)，そこに改善の余地があった．

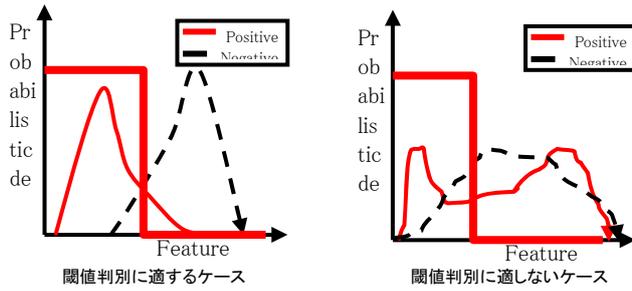


図 8 弱判別器の閾値によるバイナリな判別



図 9 弱判別器の比較例, バイナリの出力の場合には誤差が大きい

より判別能力の高い弱判別器として, 出力値を実数で表せるものを提案した. これは計算上の利便性を考え, **Parewise Function** (ロックアップテーブル) によって実装した図 10). ここで, W_{+1} と W_{-1} はそれぞれ顔と非顔の分布で, ヒストグラムで表す. ここで, **Real AdaBoost** を使って学習を行うことができる.

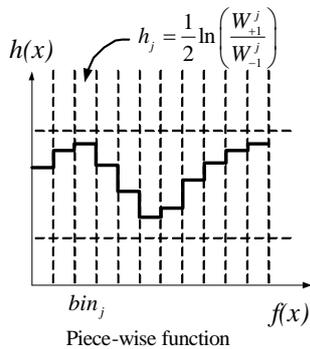


図 10 顔らしさを実数で出力する弱判別器

Real AdaBoost は **Schapire** と **Singer** によって提案されたもので, サンプル空間 X から 2 値の予測でなく, 確信度に関する実数の空間 R へ射影する判別器の学習アルゴリズムである.

以下, **Real Adaboost** を使って学習を行う時に, 計算上の利便性を考慮して, 学習した出力は等間隔に分けた入力空間における実数型のテーブルとする. 仮に特徴量 f が $[0, 1]$ に正規化したものとし, この領域を n 個のビンに分割する.

$$bin_j = [(j-1)/n, j/n], j = 1, \dots, n$$

そして, 弱判別器は次のように定義できる:

$$\text{If } f(\mathbf{x}) \in bin, \text{ then } h(\mathbf{x}) = \frac{1}{2} \ln \left(\frac{\bar{W}_{+1}^j + \varepsilon}{\bar{W}_{-1}^j + \varepsilon} \right)$$

ここで,

$$\bar{W}_l^j = P(f(\mathbf{x}) \in bin_j, y = l), l = \pm 1, j = 1, \dots, n.$$

実験の結果, **Real Adaboost** 学習アルゴリズムは離散 **Adaboost** よりすぐれた性能が得られることがわかった. 特に最初の数層のカスケードにおいて, 少ない弱判別器しか使用してないため, **Real Adaboost** のほうがより正確な確信度を出力できる.

SGF 特徴量の場合, 可能な形の特徴量の探索空間非常に広いため, ヒューリスティックな探索手法を導入する必要がある. 特徴量の探索はローカルサーチを行い, 以下の 3 つの **Expansion operator** を使う:

1. **Remove a granule**: **SGF** の中の一部を外す
2. **Add a granule**: **SGF** の中に一部を足す
3. **Refine**: **SGF** の相対位置を調整する

その他, 初期の **SGF** の選び方などに関する内容は参考文献[2]をご参照ください.

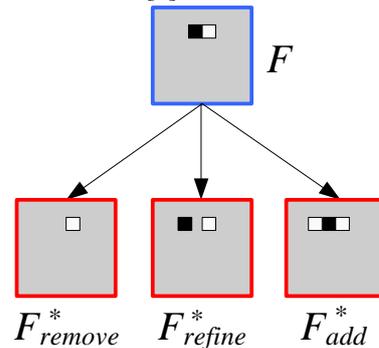


図 11 **SGF** のローカルサーチ

3.2.3 学習データの整備

高性能な識別器を構築するために, 最も性能に影響するのは学習データの質である. 我々は長年学習データの充実化を図ってきた. 例えば, 顔の向きに対応するためには同時に 80 種類の異なる顔向きのデータを採取するためのマルチカメラ顔画像データ収集装置を作成した (図 12).

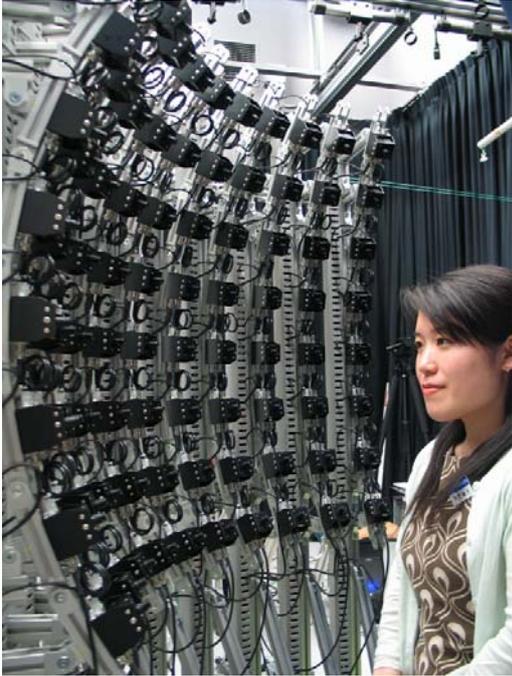


図 12 マルチアングルの顔データ採取装置

また、収集されたデータに顔や顔器官の正解点位置情報を入力するためには、コストの安い中国でのデータ入力を実施した。

3.2.4 顔の回転角、向きへの対応

全方向の顔検出を実現するために、顔の回転角、と顔の向き（正面顔、半横顔、横顔）への対応はそれぞれの検出器を用意することで対応している。顔の向きが5方向、顔の回転角が12方向を用意する必要があつて、合計60個の検出器を使っている（図13）。



図 13 全方向対応顔検出器

図 14に顔検出の検出例を示す。

3.3 ハードウェアによる実現

デジタルカメラで顔検出を行う場合、速度が非常に重要になる。そのため、リアルタイムでの検出を可能にするハードウェアでの実装が必要になる。いかに少ないゲート数で高速な検出器を実装できるかはハードウェア設計者の腕の見せ所と言える。設計において、重要なポイントは2つある：

1. 大量演算処理の並列化

データを分割して並列処理するか、識別の演算を分解して並列処理するかを最適に行う。

2. メモリアクセスの効率化

並列処理と同時に考慮する必要があるのはメモリアクセスの効率化。回路規模を減らすためには内部で使うメモリを最小限にすることが有効で、共有メモリを設けるのが良いが、そのアクセスのタイミングを上手に設計しないと他の処理のメモリアクセスとバッティングしてしまうことが起こり、計算速度の低下を招くことがある。

4 人体検出の実現に向けて

I顔検出よりも難しいのは人体の検出である。それは顔よりも人体の見え方の変化の幅がより広いからである。

現在人体検出の研究開発がまだ模索中で、顔検出のフレームワークが人体検出にも有効であることが確認できた。顔と違って、人体の場合、内部の様子が服装によって変化するため、安定した特徴抽出が難しい。そのため、人体の輪郭の特徴を如何に検出できるかがキーポイントになる。図 15に人体の輪郭情報を検出するための特徴量の例を示す。このような特徴量を機械学習のアルゴリズムを通して、人体の検出器を学習することができる。

図 16に人体検出の結果の例を示す。

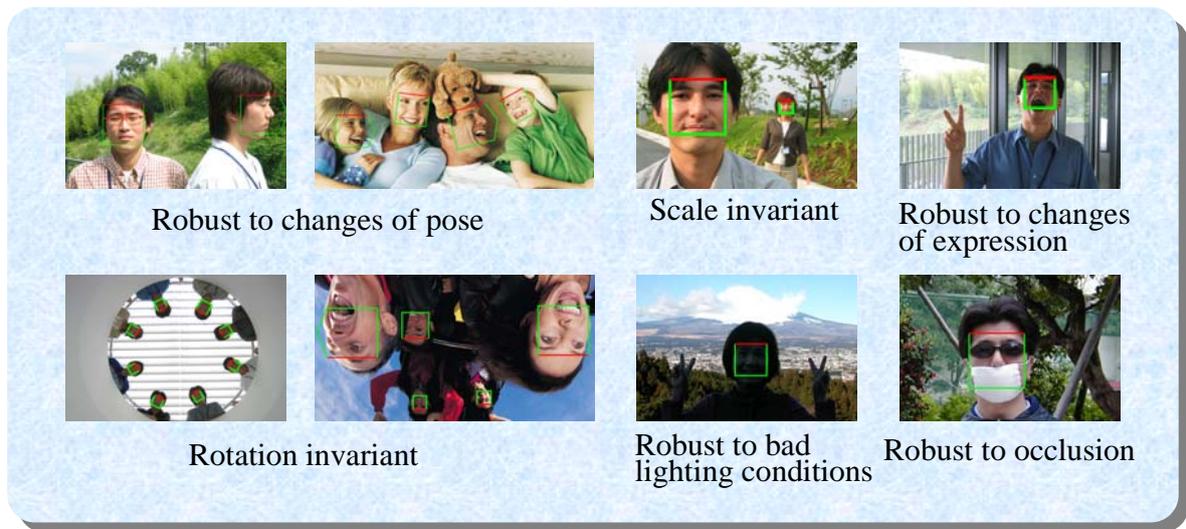


図 14 顔検出の結果例

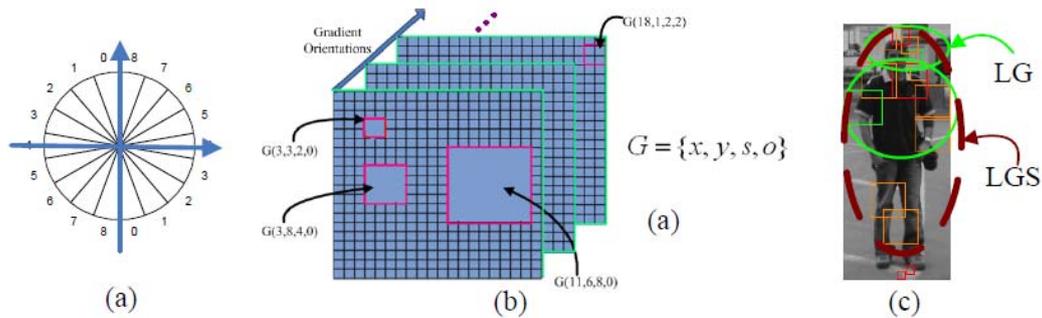


図 15 人体検出のために開発した勾配情報を扱う SGF

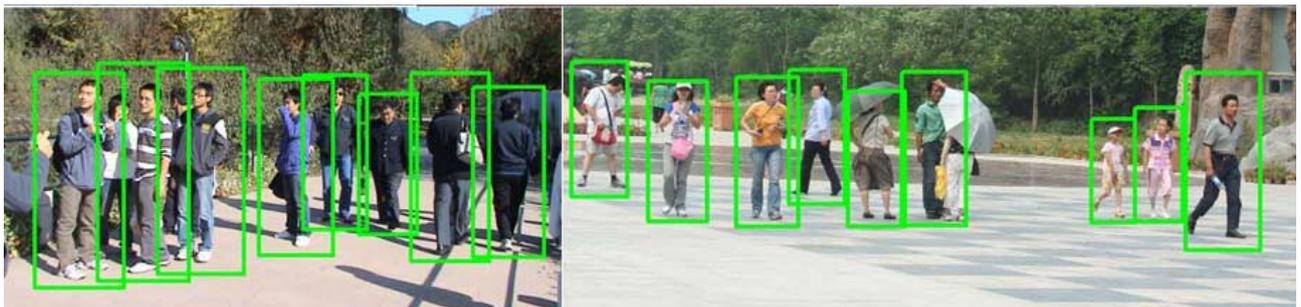


図 16 人体検出の結果例

5 まとめ

顔検出と人体検出の開発において、機械学習が極めて重要である。実用化においては速度の向上プログラムサイズ及び実行時メモリ使用量の削減も重要な課題になる。顔検出において、実応用多様な顔の向きを対応でき、高速に計算できる特徴量 **Sparse Granular Feature** が実用化の鍵となった。また、**Real AdaBoost** による学習アルゴリズムは強力な機械学習

手法で、性能向上に有効な手法である。学習データは量と質も識別器の性能を左右するため、地道なデータ収集、整理が欠かせない。人体検出の開発にも顔検出の手法が活用されているが、人体の見え方の変化の幅が広く、さらに強力な学習アルゴリズムの出現を期待する。将来的にはよりロバストな人体検出や、より多くの種類の物体検出の研究開発につなげていきたい。

参考文献

- [1] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, USA, 2001.
- [2] C. Huang, H. Ai, Y. Li, S. Lao , "High Performance Rotation Invariant Multi-View Face Detection", IEEE PAMI Vol. 29, No. 4, pp. 671-686, 2007.
- [3] W. Gao, H. Ai, S. Lao, "Adaptive Contour Features in Oriented Granular Space for Human Detection and Segmentation", CVPR2009 pp. 1786-1793, 2009.